

# Bayesian Risk Management

- "Frequency Does Not Make You Smarter" -

DISSERTATION

zur Erlangung des akademischen Grades  
Doktor der Wirtschafts- und Sozialwissenschaften  
(Dr. rer. pol.)  
in der Wissenschaftsdisziplin  
Volkswirtschaftslehre

Erstgutachter: Prof. Dr. Carlo Jaeger  
Zweitgutachter: Prof. Dr. Hans Gerhard Strohe

eingereicht an der  
Fakultät für Wirtschafts- und Sozialwissenschaften  
der Universität Potsdam  
von

Markus Fucik

Potsdam, 21. Dezember 2009 (eingereicht)  
Potsdam, 20. Oktober 2010 (Verteidigung)



This work is licensed under a Creative Commons License:  
Attribution - Noncommercial - Share Alike 3.0 Germany  
To view a copy of this license visit  
<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>

Published online at the  
Institutional Repository of the University of Potsdam:  
URL <http://opus.kobv.de/ubp/volltexte/2011/5308/>  
URN <urn:nbn:de:kobv:517-opus-53089>  
<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-53089>

## Danksagung

Für die Realisierung meiner Doktorarbeit bin ich insbesondere meinem Doktorvater Prof. Dr. Carlo C. Jaeger und meinem Forschungsgruppenleiter Dr. Armin Haas dankbar. Sie eröffneten mir große Freiräume für meine Forschung, begleiteten dies gleichzeitig mit einer intensiven Betreuung. Abgesehen von wertvollen Ratschlägen im Detail hat die inhaltliche Breite dieser Arbeit gerade von ihrem Fachwissen und ihrer Lebenserfahrung profitiert. Ich bin mir darüber bewusst, dass eine Promotion innerhalb von nur drei Jahren in keinsten Weise selbstverständlich ist. Dies verdankte ich der vorausschauenden Projektplanung, die eine sich gegenseitige befruchtende Projekt- und Promotionsarbeit erlaubte.

# Contents

List of Figures VIII

List of Tables XI

Notations XIII

List of Abbreviations . . . . . XIII  
List of Statistical Distributions . . . . . XIV  
List of Mathematical Symbols . . . . . XV  
List of Indices . . . . . XV  
List of Operators & Signs . . . . . XVI  
List of Random Variables & Their Realizations . . . . . XVI  
List of Functions . . . . . XVII

Abstract XVIII

1 Intention & Goals of this Dissertation 1

2 Bayesian Way of Understanding Uncertainties 3

2.1 Hardcore- vs. Softcore-Bayesianism . . . . . 5  
2.2 Bayesian Due Diligence . . . . . 9

3 Hardcore-Bayesianism 13

4 Model Building 14

5 Frequentist Statistics 16

5.1 Parameter Estimation . . . . . 16  
5.2 Quality of Estimation Results . . . . . 19

6 Basic Idea of Bayesian Statistics 20

6.1 Bayesian Parameter Estimation . . . . . 20  
6.2 Uninformative Priors . . . . . 22  
6.3 Bayesian Posterior Inference . . . . . 23  
6.4 Bayesian Predictive Inference . . . . . 24

7 Markov Chain Monte Carlo Simulation 25

7.1 Gibbs Sampler . . . . . 27  
7.2 Griddy Gibbs Sampler . . . . . 28

7.3	Metropolis-Hastings Sampler . . . . .	29
7.3.1	Independence Metropolis-Hastings Sampler . . . . .	30
7.3.2	Random Walk Metropolis Sampler . . . . .	31
7.4	Convergence of MCMC Sampler . . . . .	32
<b>8</b>	<b>Model Weighting</b>	<b>36</b>
8.1	Posterior Model Probability/ Bayes Factor . . . . .	36
8.1.1	Sampling from the Joint Posterior . . . . .	37
8.1.2	Linking Density Approach . . . . .	39
8.2	AIC & BIC Model Weights . . . . .	43
<b>9</b>	<b>EU ETS</b>	<b>44</b>
9.1	Overview . . . . .	44
9.2	Data . . . . .	46
<b>10</b>	<b>CO2 Price Risk Assessment</b>	<b>50</b>
10.1	Models . . . . .	52
10.1.1	Geometric Brownian Motion (GBM) . . . . .	52
10.1.2	Geometric Brownian Motion with Jumps (GBMJ) . . . . .	53
10.1.3	Ornstein-Uhlenbeck Process for Log-Prices (OU) . . . . .	54
10.1.4	Cox-Ingersoll-Ross (CIR) Process . . . . .	56
10.1.5	Constant Elasticity of Variance (CEV) Process . . . . .	57
10.1.6	Student's t <i>GARCH</i> (1, 1) . . . . .	59
10.1.7	Markov Switching (MS) <i>GARCH</i> (1, 1) . . . . .	61
10.1.8	Basic Stochastic Log-Variance (BLV) . . . . .	63
10.1.9	Stochastic Log-Variance with t-Errors (tLV) . . . . .	64
10.1.10	Correlated Log-Variance (CLV) . . . . .	64
10.1.11	Correlated Heston Model . . . . .	66
10.2	Parameter Estimation Results . . . . .	67
10.2.1	Geometric Brownian Motion (GBM) . . . . .	67
10.2.2	Geometric Brownian Motion with Jumps (GBMJ) . . . . .	71
10.2.3	Ornstein-Uhlenbeck Process for Log-Prices (OU) . . . . .	74
10.2.4	Cox-Ingersoll-Ross (CIR) Process . . . . .	79
10.2.5	Constant Elasticity of Variance (CEV) Process . . . . .	81
10.2.6	Student's t <i>GARCH</i> . . . . .	84
10.2.7	Markov Switching (MS) <i>GARCH</i> (1, 1) . . . . .	91
10.2.8	Basic Stochastic Log-Variance (BLV) . . . . .	98
10.2.9	Stochastic Log-Variance with t-Errors (tLV) . . . . .	100

10.2.10	Correlated Log-Variance (CLV) . . . . .	104
10.2.11	Correlated Heston Model . . . . .	109
10.3	Model Weighting Results . . . . .	113
<b>11</b>	<b>Bayesian VaR</b>	<b>118</b>
11.1	Bayesian Ex-Post VaR . . . . .	119
11.2	Bayesian Ex-Ante VaR . . . . .	122
11.3	Bayesian Capital Charge . . . . .	123
<b>12</b>	<b>Derivative Pricing of Some Special Models</b>	<b>124</b>
12.1	Basic Idea of Option Pricing . . . . .	125
12.2	Bayesian Option Pricing . . . . .	127
12.3	Some Simulation Studies . . . . .	129
12.3.1	Geometric Brownian Motion (GBM) . . . . .	130
12.3.2	Geometric Brownian Motion with Jumps (GBMJ) . . . . .	131
12.3.3	Basic Stochastic Log-Variance (BLV) . . . . .	133
<b>13</b>	<b>Softcore-Bayesianism</b>	<b>138</b>
<b>14</b>	<b>Agents in BRM</b>	<b>138</b>
<b>15</b>	<b>Integrated Bayesian Risk Analysis (iBRA)</b>	<b>139</b>
15.1	Identification, Risk Mapping, & Model Building . . . . .	140
15.2	Specific Risk Quantification (SRQ) . . . . .	145
15.3	Integrated Risk Quantification (IRQ) . . . . .	146
15.3.1	Correlated Risk Factors . . . . .	147
15.3.2	Auto-Correlated Risk Factors . . . . .	150
15.4	Model Implementation and Analysis . . . . .	151
15.5	Diversity of Opinion Analysis . . . . .	152
15.6	Decision Management . . . . .	159
<b>16</b>	<b>Expert Elicitation</b>	<b>160</b>
16.1	Heuristics . . . . .	162
16.2	Some Theoretical Thoughts on Expert Quality & Selection . . . . .	167
16.3	Basic Elicitation Techniques . . . . .	169
16.3.1	Scoring Rules . . . . .	169
16.3.2	Promissory Notes . . . . .	172
16.3.3	Lotteries . . . . .	172
16.3.4	Lottery Insurance Market . . . . .	173

16.3.5	Fixed Interval Technique . . . . .	175
16.3.6	Variable Interval Technique . . . . .	177
16.4	Credibility Intervals . . . . .	180
16.5	Elicitation Techniques for Integrated Risks . . . . .	186
16.5.1	Visual Copula . . . . .	186
16.5.2	The Envelope Method . . . . .	190
16.5.3	Stochastic Processes . . . . .	193
16.5.4	Time-Series Copula . . . . .	196
16.6	Mathematical Expert Aggregation . . . . .	199
16.6.1	Expert Weighting . . . . .	199
16.6.2	Bayesian Expert Aggregation . . . . .	201
<b>17</b>	<b>Expert Elicitation Tools</b>	<b>207</b>
17.1	Direct Elicitation Tools . . . . .	208
17.1.1	Expert Interviews . . . . .	208
17.1.2	Expert Panels . . . . .	208
17.2	Indirect Elicitation Tools . . . . .	210
<b>18</b>	<b>Prediction Markets</b>	<b>213</b>
18.1	From Past up to Now . . . . .	216
18.2	Basic Structure of Prediction Markets . . . . .	219
18.3	Market Liquidation . . . . .	220
18.4	Real Money vs. Virtual Money Markets . . . . .	222
18.5	Traders . . . . .	223
18.6	Pricing Mechanisms . . . . .	226
18.6.1	Continuous Double Auction (CDA) . . . . .	226
18.6.2	CDA with (Non Market Maker) Artificial Traders . . . . .	226
18.6.3	CDA with Market Maker . . . . .	227
18.6.4	Pari-Mutuel Mechanism . . . . .	228
18.6.5	Wagering Markets . . . . .	229
18.6.6	Combined-Value Trading (CVT) . . . . .	230
18.6.7	Market Scoring Rule (MSR) . . . . .	230
18.6.8	Dynamic Pari-Mutuel Market (DPM) . . . . .	232
18.6.9	Comparison of Market Types . . . . .	236
18.7	Efficiency of Prediction Markets . . . . .	238
18.7.1	Illiquidity . . . . .	240
18.7.2	Foul Play . . . . .	243
18.8	Market Prices vs. Market Assessments . . . . .	246



18.8.1	Real Financial Markets . . . . .	246
18.8.2	Prediction Markets . . . . .	247
18.9	No "Objective" Liquidation . . . . .	256
18.9.1	Concept Markets . . . . .	256
18.9.2	Long-Term Prediction Markets . . . . .	257
<b>19</b>	<b>Results and Conclusions</b>	<b>265</b>
19.1	BRM Structure . . . . .	265
19.2	Improved Data Analysis . . . . .	265
19.3	Risk Management Beyond Data Analysis . . . . .	267
19.4	Non-Arbitrary Risk Management . . . . .	270
19.5	Conclusions for the Current Financial Crisis . . . . .	271
<b>A</b>	<b>Appendix</b>	<b>274</b>
A.1	Distributions . . . . .	274
A.2	Rejection & Importance Sampling . . . . .	276
A.3	Sampling-Importance-Resampling (SIR) . . . . .	277
A.4	Ito's Lemma . . . . .	278
A.5	Univariate Density Transformation . . . . .	279
A.6	Truncation of Distributions . . . . .	279
A.7	MCMC Samplers . . . . .	279
A.7.1	GBM Gibbs Sampler . . . . .	279
A.7.2	GBMJ Gibbs Sampler . . . . .	280
A.7.3	OU MCMC Sampler . . . . .	282
A.7.4	Student's t GARCH MCMC Sampler . . . . .	283
A.7.5	MSGARCH MCMC Sampler . . . . .	285
A.8	Copulas . . . . .	286
A.8.1	Basics on Copulas . . . . .	286
A.8.2	Simulation of Copulas . . . . .	291

## List of Figures

1	basic categorization of Bayesianism . . . . .	7
2	Bayesian tree . . . . .	8
3	trichotomy of Bayesian due diligence . . . . .	11
4	BRM temple . . . . .	12
5	independence Metropolis-Hastings sampler . . . . .	31
6	EU ETS market . . . . .	47
7	CO2 price data . . . . .	49
8	log-returns of synthetic spot prices . . . . .	50
9	GBM - trace, cumsum, & scatter plots . . . . .	68
10	GBM - priors/ posteriors & credible intervals . . . . .	69
11	GBM (short) - priors/ posteriors & credible intervals . . . . .	70
12	GBM - credible bounds for log-returns . . . . .	70
13	GBMJ - trace & cumsum plots . . . . .	71
14	GBMJ - posteriors & credible intervals . . . . .	72
15	GBMJ (short) - posteriors & credible intervals . . . . .	73
16	GBMJ - modelling of log-returns . . . . .	75
17	OU - trace & cumsum plots for uninformative priors . . . . .	76
18	OU - posteriors & credible intervals for uninformative priors . . . . .	76
19	OU - posteriors & credible intervals for informative priors . . . . .	77
20	OU (short) - posteriors & credible intervals for uninformative priors . . . . .	78
21	CIR - trace & cumsum plots for uninformative priors . . . . .	79
22	CIR - posteriors & credible intervals for uninformative priors . . . . .	80
23	CIR (short) - posteriors & credible intervals for uninformative priors . . . . .	81
24	CEV - trace & cumsum plots for uninformative priors . . . . .	82
25	CEV - posteriors & credible intervals for uninformative priors . . . . .	83
26	CEV (short) - posteriors & credible intervals for uninformative priors . . . . .	84
27	Student's t GARCH - cumsum and posterior mean plots . . . . .	85
28	Student's t GARCH - posteriors & credible intervals . . . . .	88
29	Student's t GARCH - stationarity & long-run unconditional volatility . . . . .	89
30	Student's t GARCH - empirical volatility . . . . .	89
31	Student's t GARCH (short) - posteriors & credible intervals . . . . .	90
32	MS GARCH - cumsum plots . . . . .	91
33	MS GARCH - mean posterior plots . . . . .	92
34	MS GARCH - posteriors & credible intervals for mean log-returns & transition probabilities . . . . .	92

35	MS GARCH - posteriors & credible intervals for GARCH parameters . . . . .	93
36	MS GARCH - stationarity & long-run unconditional volatility . . . . .	94
37	MS GARCH - empirical volatility . . . . .	95
38	MS GARCH (short) - posteriors & credible intervals for mean log-returns & transition probabilities . . . . .	95
39	MS GARCH (short) - posteriors & credible intervals for GARCH parameters	96
40	MS GARCH (short) - empirical volatility . . . . .	97
41	BLV - trace & cumsum plots . . . . .	98
42	BLV - posteriors & credible intervals . . . . .	99
43	BLV - long-run volatility . . . . .	99
44	BLV (short) - posteriors & credible intervals . . . . .	101
45	tLV - trace & cumsum plots . . . . .	101
46	tLV - posteriors & credible intervals . . . . .	103
47	tLV (short) - posteriors & credible intervals . . . . .	103
48	ASV1 - trace & cumsum plots . . . . .	104
49	ASV1 - posteriors & credible intervals . . . . .	106
50	ASV1 (short) - posteriors & credible intervals . . . . .	106
51	ASV2 - trace & cumsum plots . . . . .	107
52	ASV2 - posteriors & credible intervals . . . . .	107
53	ASV2 (short) - posteriors & credible intervals . . . . .	108
54	Heston - trace & cumsum plots . . . . .	109
55	Heston - mean posterior plots . . . . .	110
56	Heston - posteriors & credible intervals . . . . .	110
57	Heston - empirical volatility . . . . .	112
58	Heston (short) - posteriors & credible intervals . . . . .	112
59	model weighting - log-likelihoods & log-kernels . . . . .	113
60	model weighting - basic statistics . . . . .	114
61	model weighting - Bayes factors . . . . .	116
62	model weighing - AIC, BIC, & posterior model probabilities . . . . .	116
63	model weighting (short) - basic statistics . . . . .	117
64	model weighting (short) - Bayes factors . . . . .	117
65	model weighting (short) - AIC, BIC, & posterior model probabilities . . . . .	118
66	Bayesian ex-post value-at-risk . . . . .	121
67	Bayesian ex-ante value-at-risk . . . . .	123
68	Bayesian capital charge . . . . .	124
69	Bayesian option pricing - GBM . . . . .	132
70	Bayesian option pricing - GBMJ . . . . .	134

71	Bayesian option pricing - BLV . . . . .	136
72	BRM agents . . . . .	139
73	iBRA cycle . . . . .	140
74	probability impact map . . . . .	141
75	variability-impact table . . . . .	142
76	risk map . . . . .	143
77	influence matrix . . . . .	144
78	influence diagram . . . . .	144
79	consistency matrix . . . . .	148
80	diversity of opinions . . . . .	153
81	no aggregation paradigm . . . . .	155
82	reputation map . . . . .	157
83	confidence in expert's self-rating . . . . .	158
84	prospect theory . . . . .	166
85	(modified) expert histograms . . . . .	177
86	triangular distribution . . . . .	179
87	(modified) PERT distribution . . . . .	180
88	posterior credible level . . . . .	183
89	self-assessment . . . . .	185
90	independent risk factors . . . . .	187
91	Gaussian copula . . . . .	189
92	envelope method . . . . .	191
93	stochastic processes for expert elicitation . . . . .	196
94	time-series copula . . . . .	198
95	Bayesian expert aggregation (Gaussian model) . . . . .	204
96	Bayesian expert aggregation (copula approach) . . . . .	206
97	expert elicitation tools . . . . .	207
98	risk aversion . . . . .	254
99	Pittsburgh prediction market . . . . .	260
100	CIRCE prediction market . . . . .	261
101	Eberswalde prediction market . . . . .	263
102	rejection & importance sampling . . . . .	276
103	Student's t copula . . . . .	290
104	copulas and tail dependence . . . . .	292

## List of Tables

1	conjugate priors . . . . .	28
2	Bayes factors . . . . .	37
3	basic statistics of the log-returns . . . . .	48
4	results for GBM model . . . . .	69
5	parameter correlations of GBMJ model . . . . .	71
6	results of GBMJ model (full series) . . . . .	73
7	results of GBMJ model (short series) . . . . .	74
8	results of OU model (full series) . . . . .	78
9	results of OU model (short series) . . . . .	78
10	results of CIR model (full series) . . . . .	81
11	results of CIR model (short series) . . . . .	82
12	results of CEV model (full series) . . . . .	83
13	results of CEV model (short series) . . . . .	84
14	parameter correlation of Student's t GARCH model . . . . .	86
15	results of Student's t GARCH model (full series) . . . . .	87
16	results of Student's t GARCH model (short series) . . . . .	87
17	results of Markov Switching MS GARCH(1,1) model (full series) . . . . .	96
18	results of Markov Switching MS GARCH(1,1) model (short series) . . . . .	97
19	results of BLV model (full series) . . . . .	100
20	results of BLV model (short series) . . . . .	100
21	results of tLV model (full series) . . . . .	102
22	results of tLV model (short series) . . . . .	102
23	results of ASV1 model (full series) . . . . .	105
24	results of ASV1 model (short series) . . . . .	105
25	results of ASV2 model (full series) . . . . .	108
26	results of ASV2 model (short series) . . . . .	109
27	results of Heston model (full series) . . . . .	111
28	results of Heston model (short series) . . . . .	111
29	differences between the VaRs . . . . .	121
30	relative number of "hits" for different VaRs . . . . .	122
31	frequentist misspecification of the CO2 price risk . . . . .	123
32	GBM option prices . . . . .	130
33	GBMJ option prices . . . . .	133
34	BLV option prices . . . . .	137
35	expert weighting scheme . . . . .	157

36	prominent proper scoring rules (Hanson Jan. 2002, p. 4)	170
37	examples for interval weights	176
38	posteriors for credible levels (uniform priors)	184
39	posteriors for credible levels (informative priors)	184
40	self-assessment results	185
41	weighting scheme for the likelihood of joint extreme events	187
42	copulas for modelling of joint extreme events	188
43	research group assessments on future average CO2 prices	195
44	precision & bias of an expert assessment	202
45	liquidation mechanisms of prediction markets	221
46	comparison of market types	238
47	elliptical copulas	287
48	Archimedean copulas	288
49	rank correlations	290
50	lower & upper tail dependence	291

## Notations

### List of Abbreviations

AIC	Akaike information criterion
ASV1	asymmetric stochastic variance (inter-temporal dependency)
ASV2	asymmetric stochastic variance (intra-temporal dependency)
BBN	Bayesian Belief Network
BIC	Bayesian information criterion
BLV	basic stochastic log-variance
BRM	Bayesian Risk Management
CDA	continuous double auction
CDAMM	continuous double auction with market maker
cdf	cumulative distribution function
CEV	constant elasticity of variance
CIR	Cox-Ingersoll-Ross
CLV	correlated log-variance
CO2	carbon dioxide
CPI	consumer price index
CVT	combined-value trading
Dec09	ECX future on EUAs with maturity in December 2009
DGP	data generating process
DPM	dynamic pari-mutuel market
ECX	European Climate Exchange
ETS	Emissions Trading Scheme
EUA	European Union Allowance
EURIBOR	Euro interbank offered rate
GARCH	generalized auto-regressive conditional heteroskedasticity
GBM	Geometric Brownian motion
GBMJ	Geometric Brownian motion with jumps
GDP	gross domestic product
HCB	Hardcore-Bayesianism
iBRA	integrated Bayesian Risk Analysis
iM-H	independence Metropolis-Hastings

IRQ	integrated risk quantification
LMSR	logarithmic market scoring rule
LS	least squares
MCMC	Bayesian Belief Network
ML	maximum likelihood
MS	Markov Switching
MSR	market scoring rule
NRC	Nuclear Regulatory Commission
PAM	policy analysis market
PCX	Potsdam Climate Exchange
PCXquest	PCX online questionnaire platform
PCXtrade	PCX prediction market platform
pdf	probability density function
pmf	probability mass function
PERT	program evaluation and review technique
RWM	random walk Metropolis
SCB	Softcore-Bayesianism
SRQ	specific risk quantification
tLV	stochastic log-variance with t-errors
p.	page
pp.	pages

## List of Statistical Distributions

$F_Y(y)$	cdf of $Y$ at $y$
$F_Y^{-1}(q)$	inverse cdf of $Y$ at $q \in [0, 1]$
$f_Y(y)$	pdf of $Y$ at $y$
$L_y(\theta)$	likelihood function of $\theta$ given observations $y$
$P(y)$	probability for $Y = y$
$P(y_1, \dots, y_n)$	joint probability for $Y = y$
$\Phi(y \mu, \sigma)$	cdf of a Gaussian distribution with mean $\mu$ and standard deviation $\sigma$
$\Phi(y) = \Phi(y 0, 1)$	cdf of a standard Gaussian distribution
$\phi(y \mu, \sigma)$	pdf of a Gaussian distribution with mean $\mu$ and standard deviation $\sigma$



**List of Mathematical Symbols**

$B$	burn-in phase
$BF_{kl}$	Bayes factor for model $k$ given model $l$
$dt$	increment of time
$dY_t = Y_{t+dt} - Y_t$	differential of time-series variable $Y$
$H_0/ H_1$	null/ alternative hypothesis
$\mathcal{M}_k$	model $k$
$\mathbb{P}$	objective (real world) measure
$p_k$	price for bond $k$
$\mathcal{P}_k$	payoff when outcome $k$ realizes
$\mathbb{Q}$	risk neutral measure
$\mathbb{R}$	set of all real numbers
$r$	risk-free interest rate
$r_k$	report on outcome $k$
$\mathcal{T}$	maturity of a derivative
$W$	number of simulations
$w_k$	weight for outcome $k$
$\Delta$	time-step/ time to maturity
$\zeta$	strike price
$\pi(k) = \pi_k$	subjective probability for outcome $k$
$\tilde{\pi}_k$	modified expert histogram for outcome $k$
$\Sigma$	covariance matrix

**List of Indices**

$i = 1, \dots, n$	index for counting
$j = 1, \dots, J$	expert/ trader/ parameter $j$
$k = 1, \dots, K$	model $k$ / outcome $k$ of a risk factor
$t = 1, \dots, T$	time
$-i/ -j/ -t$	vector $y = (y_1, y_2, \dots)'$ without $y_i/ y_j/ y_t$
$< j/ > j$	all parameters $\theta_l$ of vector $\theta = (\theta_1, \dots, \theta_J)'$ for $l < j/ l > j$

## List of Operators & Signs

$\sum_{i=1}^n x_i = x_1 + \dots + x_n$	sum
$\prod_{i=1}^n x_i = x_1 \cdot \dots \cdot x_n$	product
$\sqrt{x}$	square root of $x$
$\int_a^b f(x) dx$	integral of function $f(x)$
$\emptyset$	empty set
$\propto$	proportional to
$\cap$	intersection of the sets
$\cup$	union of the sets
$\hat{\theta} / \hat{E} / \widehat{Var}$	estimated parameter / expected value / variance
$\cdot  _{Y \in [Y_{\min}, Y_{\max}]}$	truncated distribution

## List of Random Variables & Their Realizations

Random variables  $Y$  are written in capital letters and their realizations  $y$  in small letters. There is no differentiation for random variables in greek letters except for the random parameter (vector)  $\Theta$  and its realization  $\theta$ .

$Y = (Y_1, \dots, Y_n)'$	vector of random variable $Y$
$y = (y_1, \dots, y_n)'$	vector of realization $y$
$Y x$	$Y$ given $X = x$
$Q = q$	credible level and its realization
$X = x$	latent random variable $X$ and its realization $x$
$\Theta = \theta$	random parameter (vector) $\Theta$ and its realization $\theta$
$\theta^{(w)}$	with realization of uncertain parameter (vector) $\Theta$
$\theta^*$	"true" parameter vector of DGP
$\tilde{\Theta} = (\Theta, X)'$	augmented parameter vector

**List of Functions**

$\arg$	argument of an optimization
$Corr(X, Y)$	correlation between $X$ and $Y$ (Pearson's correlation coefficient $\rho$ )
$Cov(X, Y)$	covariance between $X$ and $Y$
$\mathcal{CR}^\beta$	credible interval given credible level $\beta$
$\mathcal{CR}(q)$	credible interval given an uncertain credible level $Q$ realized at $q$
$\dim(x)$	dimension of vector $x$
$E(Y)$	expected value of random variable $Y$
$\exp(x) = e^x$	exponential function of $x$
$\lim_{x \rightarrow c} f(x)$	limit of a function $f(x)$ for $x \rightarrow c$
$\ln x$	logarithmic function of $x$
$\max f(x)$	maximum of function $f(x)$
$mean(Y)$	arithmetic mean of random variable $Y$
$MD_k$	mean deviance of model $k$
$\min f(x)$	minimum of function $f(x)$
$mode(Y)$	mode of random variable $Y$
$s_k(r)$	scoring rule payoff for outcome $k$ on report $r$
$Std(Y)$	standard deviation of random variable $Y$
$Var(Y)$	variance of random variable $Y$
$VaR$	value-at-risk
$dW_t$	increment of a Wiener process
$\alpha(\theta^{(w-1)}, \tilde{\theta}^{(w)})$	acceptance probability for parameter proposal $\tilde{\theta}^{(w)}$ given $\theta^{(w-1)}$
$\pi(\theta)$	prior for vector $\theta$
$\pi(\theta y)$	posterior for vector $\theta$ given observations $y$
$\rho(\theta \cdot)/\rho(\theta)$	proposal distribution for vector $\theta$
$\tau(X, Y)$	Kendall's tau, rank correlation between $X$ and $Y$
$\tilde{\varphi}$	linking density (= pseudo prior)

## Abstract

Within our research group Bayesian Risk Solutions we have coined the idea of a Bayesian Risk Management (BRM). It claims (1) a more transparent and diligent data analysis as well as (2) an open-minded incorporation of human expertise in risk management. In this dissertation we formulize a framework for BRM based on the two pillars Hardcore-Bayesianism (HCB) and Softcore-Bayesianism (SCB) providing solutions for the claims above.

For data analysis we favor Bayesian statistics with its Markov Chain Monte Carlo (MCMC) simulation algorithm. It provides a full illustration of data-induced uncertainty beyond classical point-estimates. We calibrate twelve different stochastic processes to four years of CO2 price data. Besides, we calculate derived risk measures (ex ante/ post value-at-risks, capital charges, option prices) and compare them to their classical counterparts.

When statistics fails because of a lack of reliable data we propose our integrated Bayesian Risk Analysis (iBRA) concept. It is a basic guideline for an expertise-driven quantification of critical risks. We additionally review elicitation techniques and tools supporting experts to express their uncertainty.

Unfortunately, Bayesian thinking is often blamed for its arbitrariness. Therefore, we introduce the idea of a Bayesian due diligence judging expert assessments according to their information content and their inter-subjectivity.

# 1 Intention & Goals of this Dissertation

The current financial crisis since 2007/8 has shown quite plainly that risk management has encountered a myriad of risks (and opportunities) which have not adequately been considered. In this dissertation we believe in two main causes for this failure of risk management:

1. Classical risk management does not transparently disclose a rational level of uncertainty resulting from restricted historic data.
2. Classical risk management ignores relevant information beyond data analysis or accounts for it off the record as well as arbitrarily.

The first shortcoming results from the application of classical (or frequentist) statistics in risk management. It solely identifies point estimates disregarding data-induced parameter and model uncertainty. Such ignorance becomes highly critical with a low number of observations. A more prudent data analysis might unravel observed financial market anomalies like the equity premium and the capital charge puzzle.

Generally, all data analysis rests on (the assumption of) reliable data to deliver information on future structures. A risk management – solely concentrated on historical observations – can however easily miss the essentials. In a Bloomberg interview Justin Fox warns: "Frequency does not make you smarter".<sup>1</sup> This is highly relevant for singular or unprecedented regulation, reputation, and litigation risks induced by climate change.<sup>2</sup> Under such prevailing circumstances we see no alternative to an explicit consideration of human expertise in risk management.

Until now, there has been a deep-rooted averseness to such subjectivity — maybe less a personal feeling than a demand of regulations like Basel II. However, this could fundamentally change in the coming years as the financial regulators conclude (see the "79th Annual Report" of the Bank for International Settlements, BIS 2009):

"Measuring, pricing and managing risk all require modern statistical tools based largely on historical experience. [...] A long period of relative stability will lead to the perception that risk is permanently lower [...] Addressing this misperception is an enormous challenge. The major risks [...] are large, infrequent events. [...] We] need an accurate assessment of the size of the tails of the distribution

---

<sup>1</sup>At iTunes, you can download the interview: "Justin Fox Sees Need to Open OTC Derivatives Markets", Bloomberg on the Economy (September 9, 2009). See also the popular science book "The Myth of the Rational Market: History of Risk" by Fox (2009).

<sup>2</sup>see publications of the project "Mainstreaming of Climate Risks and Opportunities in the Financial Sector" at [www.climate-mainstreaming.net](http://www.climate-mainstreaming.net) and klimazwei (2009), pp. 76

of outcomes. [... The] statistical models [...] will, almost by definition, be inaccurate because of a lack of data. [...] The difficulty of assessing the tails of the distribution of outcomes is even greater for new financial instruments. With no history, their riskiness cannot be statistically measured at all."

According to the BaFin circular letter 15/2009 (BA) (BaFin 2009), the German financial supervision has corrected the minimum standards for credit and financial service institutions (MaRisk) in favor of a more expertise driven risk management. Although institutions need to understand their business activities, the supervision highlights that "understanding" does not mandatory mean "measuring risks". If an institution has no suitable approaches, a plausibility check can alternatively be applied for the risk quantification. The plausibility check itself can result from an expert appraisal.

Under such surrounding conditions, a decent risk management should make optimal but not blue-eyed *use of all reliable data* as well as of *human expertise without accepting arbitrariness*. We call such risk management accounting for these demands *Bayesian Risk Management* (BRM).

BRM has been developed in the research group *Bayesian Risk Solutions* (BRS) - headed by Carlo Jaeger and Armin Haas (both Potsdam Institute for Climate Impact Research) - while conducting research for the project *Mainstreaming of Climate Risks and Opportunities in the Financial Sector*. This project aimed at researching practicable implementations of BRM in the day-to-day business of (financial) institutions.

This dissertation aims for the presentation and discussion of some key findings. We focus to attain four fundamental goals:

(1) *BRM Structure*: We want to condense the wide scope of BRM into a clear structure based on the concepts *Hardcore-Bayesianism* (HCB) and *Softcore-Bayesianism* (SCB) - introduced by Armin Haas during the Bayesian Risk Management workshop at Carnegie Mellon University (Pittsburgh) on June 11/13, 2008.<sup>3</sup>

(2) *Improved Data Analysis*: We want to discuss alternatives to classical data analysis. We focus on the question whether Bayesian statistics is an alternative to frequentist statistics in respect of a better extraction of information contained in the data. We review our theoretical findings by an empirical analysis of European CO2 price risk where we combine a Bayesian parameter estimation (see Siliverstovs et. al. 2009) with a Bayesian model weighting (see Jaeger et. al. 2008). To understand the differences in risk appraisal, we additionally assess Bayesian and classical value-at-risks, capital charges, and option prices.

(3) *Risk Management Beyond Data Analysis*: We want to find options to transparently and rationally incorporate human expertise in risk management beyond pure data analysis.

---

<sup>3</sup>This work reflects the author's understanding of BRM. It mainly conforms to the BRS view but can slightly differ in some specific points.

Amongst others, we discuss practicable guidelines, techniques, and tools for the identification and quantification of potential risks. Since expertise is often scattered around several people, we analyze concepts to deal with competing assessments. In this context, we see need for further research on people's understanding of risk and their confidence in expert assessments. We try to contribute to a more profound understanding by two empirical surveys.

(4) *Non-Arbitrary Risk Management*: BRM explicitly propagates the incorporation of subjective appraisals into risk management. This is often opposed because of a fear of arbitrariness. Unfortunately, the common Bayesian view of "uncertainty" lacks in rules for the judgment of the usefulness of assessments while the classical understanding of "risk" and "uncertainty" totally fails to account for subjective appraisals. Therefore, we ask for a new risk classification reducing arbitrariness. This is also highly relevant for classical data analysis since it needs to be decided whether specific data is reliable for inference.

In the following section 2 we introduce the Bayesian way of understanding uncertainties and propose a clear structure for BRM with its main parts HCB and SCB. Subsequently, we discuss a new classification system for risks.

In sections 3 up to 12 we concentrate on Bayesian statistics as the most important part of HCB. In an empirical study we calibrate and weight twelve stochastic processes to four years of CO2 price data (see sections 9 and 10). We calculate Bayesian value-at-risks, capital charges, and option prices to quantify the CO2 price risk and contrast them with their classical counterparts (see sections 11 and 12).

In sections 13 up to 18 we discuss the scope of SCB in the mobilization of human expertise to quantify reliable assessments when a pure data analysis falls short of needs or fails. We start with a discussion on the agents involved in BRM (see section 14). Subsequently, we present our *integrated Bayesian Risk Analysis* (iBRA) concept offering guidelines for an expertise driven risk management beyond a pure data analysis (see section 15). As risk management often critically depends on the appraisals of in house or external experts, we overview several elicitation techniques and tools (see sections 16 and 17). Unfortunately, there is a trade-off between the quality of the collected expert appraisals and the time and resources invested in elicitation. Since we see prediction markets as a good compromise we separately dwell on them in section 18.

## 2 Bayesian Way of Understanding Uncertainties

Forecasting outcomes of future events is traditionally classified whether it is done under *certainty*, *risk*, or *uncertainty* (Knight 1921). Forecasting under certainty only exists in a moderated form. The forecast that the sun will rise again tomorrow is virtually true with a

degree of certainty that the collapse of the sun is negligible in the next days. Forecasting under risk describes stochastic randomness. The outcomes of future events depend on a probability distribution with *known* "objective" (also frequentist, mathematical, numerical) probabilities (Press 2002, p. 19). In contrast, forecasting under uncertainty is done without the knowledge of these frequentist probabilities.

There are two commonly used classical descriptions for the frequentist probabilities: (1) The *symmetry* or *exchangeability argument* defines probability as (Gelman et al. 1995, p. 12)

$$\text{probability} = \frac{\text{number of favorable cases}}{\text{number of possibilities}}.$$

(2) The more elaborate *frequency argument* describes probabilities as "the proportion of times that (an outcome of an event) occurs if we conduct a long sequence of repetitions" (O'Hagan et al. 2006, p. 11). Gelman et al. (1995), p. 12, use a more precise definition for the probability as a "relative frequency obtained in a very long sequence of (repetitions), assumed to be performed in an identical manner, physically independently of each other". Unfortunately, there is one hitch. Frequentist probabilities are no characteristic of a single event but of a sequence of "similar" events (von Mises 1981).<sup>4</sup>

From this risk and uncertainty categorization follows a subdivision of uncertainty into an epistemic and an aleatory constituent (O'Hagan et al. 2006, pp. 10, Aven 2003, pp. 16, Ötsch 2008, p. 25). The *epistemic* part describes the uncertainty arising from insufficient knowledge of the structure of a model or the world. Consequently, this uncertainty can be deleted by gathering information. In contrast, the *aleatory* part characterizes the "true" physical randomness.

This concept is better illustrated by an example: When you toss a coin, the outcome is assumed to be uncertain because of *your* (epistemic) uncertainty on the structure of the coin and the *coin's* (aleatory) physical randomness. Once you know the coin is "fair", all structural uncertainty has vanished but there is still the randomness. According to the frequentist understanding, the outcome of a fair coin is risky because you know the "true" probabilities but not the actual outcomes in advance.<sup>5</sup>

Following the concepts of frequentist probabilities and physical randomness, *classical* (or frequentist) *statistics* can only describe aleatory uncertainty by hypothesizing the existence of virtually *repeatable* events like coin tossing. However, even this prime example of a

---

<sup>4</sup>See Burdzy (2009) for a discussion on the differences between the theory on frequentist (subjective) probabilities and their practicable use in frequentist (Bayesian) statistics.

<sup>5</sup>Another example are daily returns of stocks. When there is no model for those returns, there is huge uncertainty about them. By introducing the concept of Geometric Brownian motion for the (log-)returns, the epistemic uncertainty is removed but there is still aleatory uncertainty as the (log-)returns are assumed to be Gaussian distributed.



repeatable event fails in reality. There have been several experiments showing coin tossing to be deterministic (see Diaconis et al. 2007). Indeed, the outcome of a toss solely depends on the initial conditions which can be replicated by coin tossing machines.

Because of the lack of repeatable events, there is no risk but only epistemic uncertainty. Lindley (1985), p. 8, defines uncertainty as a situation when a person does "not know what will happen on some future occasion, though sometimes the uncertainty is in the past".

When we do not want to abstain from the concept of probabilities, subjective probabilities are the only way out.<sup>6</sup> This is advisable, as in everyday life, probability statements are often used for unprecedented events or almanac questions (e.g., "Which city is located closer to the equator, Madrid or Rome?").

Consequently the *subjective probability* - the theory was developed by de Finetti (1997) and Savage (1972) - is a personal *degree of belief* in the truthfulness of a proposition (O'Hagan et al. 2006, p. 11, Press 2002, pp. 17/ 19). Moreover, the normative axioms of Savage (1954/ 1972) imply a maximization of expected utility based on the individual degree of belief. When offering persons different action alternatives, their subjective probability assessments can be observed by monitoring their decisions (O'Hagan et al. 2006, p. 21). The most sophisticated way is to quantify subjective probabilities by measuring the *willingness to bet* which is however often contaminated by a person's preferences.

All theory dealing with such an understanding of subjective probabilities we call *Bayesian theory*. The practicable application of these probabilities we term *Bayesian Risk Management* (BRM) - we discuss in this dissertation. We only present Bayesian theory to an extent that allows to correctly apply Bayesian techniques and helps to understand pitfalls.

## 2.1 Hardcore- vs. Softcore-Bayesianism

Bayesian Risk Management (BRM) focuses on a transparent risk management (1) when there is reliable data for data analysis as well as (2) when there is no reliable data, at all. Data is labeled "reliable" when the risk analyst judges a data analysis as promising with a clear conscience. Then, normative mathematical Bayesian updating discloses a rational level of uncertainty remaining after the observation of the reliable data. Unfortunately, there is often no reliable data available or no reliable statistical model can be deduced from reliable data. In such situations there is no escape to quantify subjective assessments.

Following this argumentation, BRM distinguishes two main (complementary) ways to quantify personal uncertainty on critical factors:<sup>7</sup>

---

<sup>6</sup>Indeed, Press 2002, p. 19, sees the concept of "objective" probabilities as a special case of subjective probabilities.

<sup>7</sup>We use the notion "factor" to reference to variables like oil price as well as to events like "a global emission trading scheme in 2020". Mostly, variables coincide with level variables while event is a synonym

- *Hardcore-Bayesianism* (HCB) summarizes all empirical Bayesian techniques that require reliable data to apply the formal Bayes rule for normative (mathematical) Bayesian updating.<sup>8</sup> The data can be inter-subjectively<sup>9</sup> measured observations (e.g. length, width, weight, temperature, speed, price) or subjective (probability) assessments.<sup>10</sup> HCB relies on the data's explanatory power for the inference on the critical factors.
- *Softcore-Bayesianism* (SCB) summarizes all Bayesian approaches (1) that can identify critical factors, (2) that can quantify or aggregate subjective (probability) assessments, (3) that can deal with such assessments, and (3) that can support decision makers in dealing with them. Although informal Bayesian updating of subjective assessments is generally intra-subjective,<sup>11</sup> SCB techniques intend to foster a more inter-subjective updating. A crucial task is to separate informative from uninformative or even deliberately misleading appraisals. In fact, this includes the assessment of the reliability of data and of the induced statistical models.

The BRM proposes two criteria (subjective probabilities and mathematical Bayesian updating) as a rough differentiation between HCB and SCB, although the differences are in fact much more subtle (see figure 1). Primarily, Bayesian thinking requires the concept of subjective probabilities. The famous normative mathematical Bayesian updating via the *Bayes rule*<sup>12</sup> is no premise for Bayesian theory. The quantification of uncertainty without the Bayes rule is an important, maybe, the most important part of Bayesianism.

Due to its efficiency, there have been efforts to apply the Bayes rule in empirical data analysis without informative priors. Inference purely bases on data (see section 6.2). Strictly speaking, such approaches are outside the Bayesian paradigm. Nevertheless, we label such work as Bayesianism because it is an important part of Bayesian statistics and a separation would cause a more unclear and confusing landscape.

---

for occurrence variables (see section 15.1).

<sup>8</sup>In this dissertation we use the term "updating" to describe the incorporation of additional information in an existing risk assessment.

<sup>9</sup>Inter-subjectivity: Different people interpret characteristics in the same way.

<sup>10</sup>Statements can be transformed into a numerical form like traditionally measured data. This allows to apply the mathematical Bayes rule.

<sup>11</sup>Intra-subjectivity: A person's latent, inner, and individual reconsideration and appraisal of new information.

<sup>12</sup>The Bayes rule for two events (variables)  $A$  and  $B$  follows basic statistical rules. It is  $\pi(a|b) = \pi(b|a)\pi(a)/\pi(b)$  and simply states that the conditional probability (*posterior*) of event  $A$  given the outcome  $b$  of event  $B$  can be calculated by the reverse conditional probability  $\pi(b|a)$ , by an unconditional probability (*prior*)  $\pi(a)$  for  $A = a$ , and by an unconditional probability  $\pi(b)$  for  $B = b$ . The notation  $\pi(\cdot)$  with a pi reflects the subjectivity of the probability. Indeed, we will apply the Bayes rule in the Bayesian parameter estimation (see sections 3 up to 12) where we presume a data generating process (DGP) producing measurable data (CO2 prices). Then  $\pi(b|a)$  and  $\pi(b)$  are not treated as subjective probabilities.

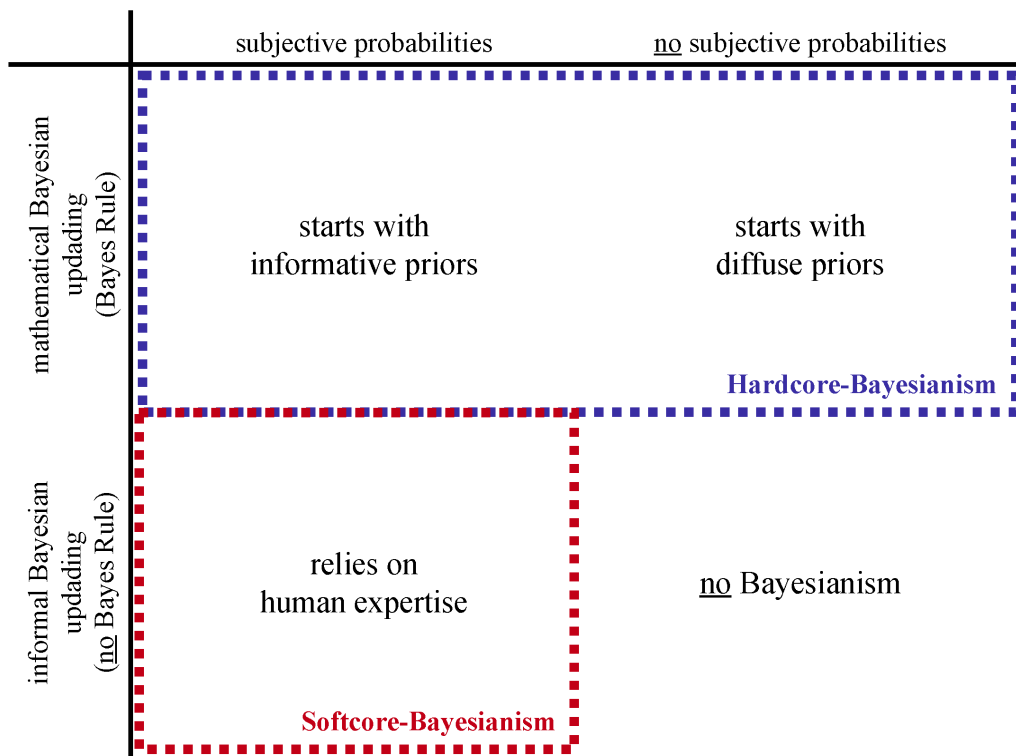


Figure 1: basic categorization of Bayesianism

In general, SCB can separately be applied while HCB cannot for two reasons (see figure 2): First, the normative mathematical Bayesian updating via the Bayes rule requires the identification of advanced information (= informative priors). The quantification of such informative priors is however part of SCB.<sup>13</sup> The controversial use of improper prior can cut this link between HCB and SCB.<sup>14</sup> Second, the pre-selection of models and the judgement of the reliability of the data and the model assumptions for inference is purely subjective even though classical testing and Bayesian model weighting can support the appraisal. The groundbreaking innovation of the BRM concept is the simultaneous and transparent handling of the normative mathematical Bayesian updating and the informal subjective updating of people in real life. Indeed, there is no clear superiority for one of both updating approaches.

The mathematical Bayesian updating guarantees a rational and non contradicting incorpo-

<sup>13</sup>In the special case of Bayesian belief networks (BBN), SCB can even deliver posteriors. Adequate priors can be calculated by means of BBN.

<sup>14</sup>The use of improper priors is a violation of statistical fundamentals which entails no practicable problems for some special cases (see section 6.2).

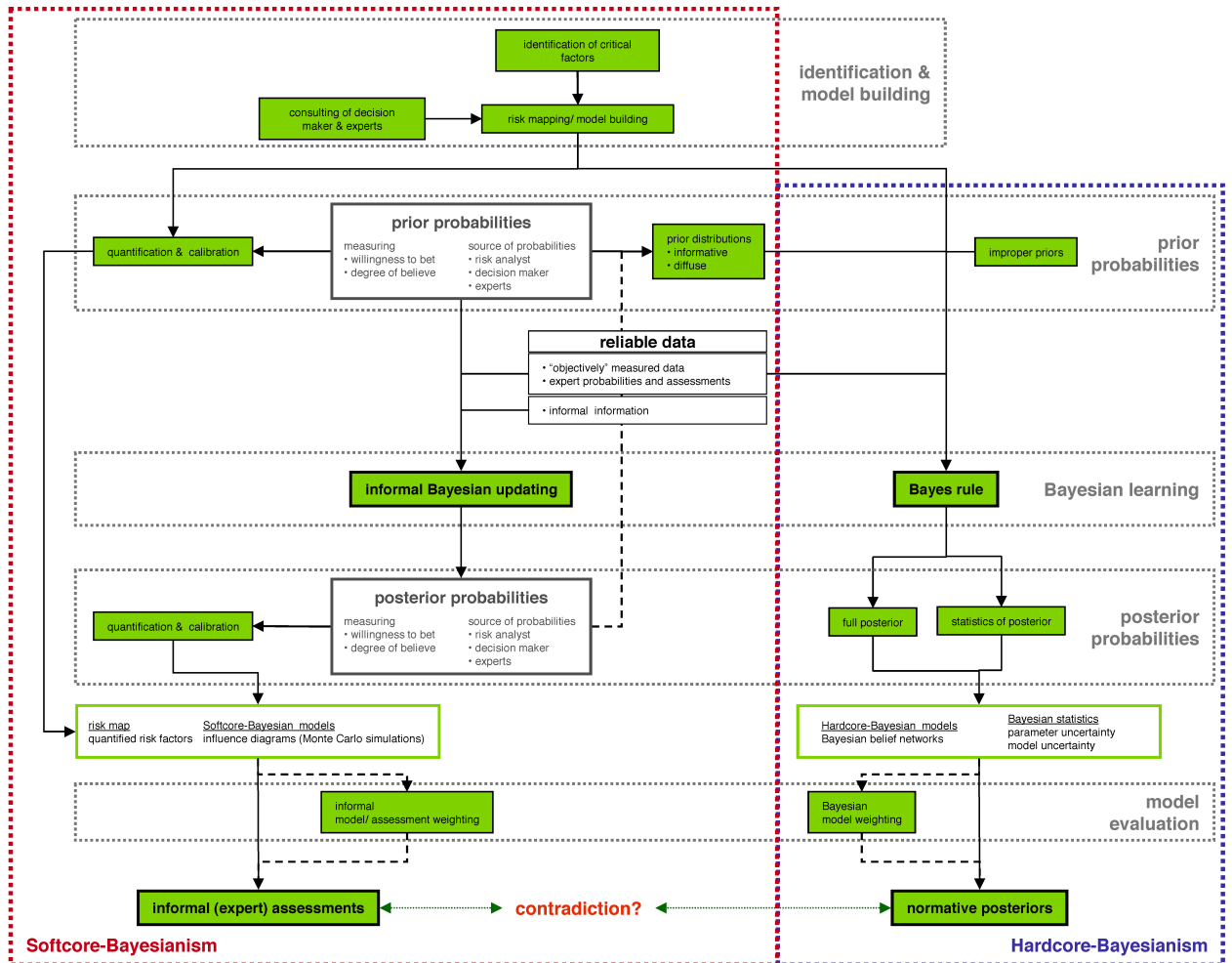


Figure 2: Bayesian tree - SCB describes an informal updating while HCB offers a mathematical updating via the Bayes rule. Since every model is formed by human expertise there is no HCB without SCB. We expect a difference between the normative HCB risk assessments (resulting from rational learning) and the observed assessments of human beings.

ration of information but requires a formal model of the world. Unfortunately, the world is too complex to find an exhaustive model and the Bayes rule can only update on clear specified data. An informal updating of individuals seems preferable when there is uncertainty on some unprecedented events or there is no reliable data, at all. Such an updating relies on the person's experience of life and often (latently) accounts for an incredible amount of information. Consequently, a prudent risk management should profit from both approaches. For instance, expertise of some experts can help to set up informative priors or to modify the result (posterior) of mathematical Bayesian updating.

## 2.2 Bayesian Due Diligence

From a theoretical point of view, Bayesian reasoning enables to quantify all imaginable uncertainties because subjective probabilities are not attached to objects (e.g. a coin) or events (e.g. global implementation of the Tobin tax) but to persons. There is no chance to verify or falsify a subjective statement irrespective the information we have access to.<sup>15</sup> Indeed, Bayesian updating is not a verification or falsification of past assessments in consideration of new information (de Finetti 1974/ 1975). Instead, an old assessment - based on old information - is replaced by a new assessment - based on old and new information (Burdzy 2009). Ideally, human beings express their uncertainty, depending on their level of information, in form of more or less flat probability distributions (Klir 1999).

In real life, a laissez-faire Bayesianism could result in a haphazard instead of a beneficial risk management. An adequate quantification and evaluation of subjective uncertainties is inevitable for a practicable implementation of BRM. In fact, we agree with Keynes (1937), pp. 213, who sees different degrees of uncertainty:

By 'uncertain' knowledge [...] I do not merely distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty; nor the prospect of a Victory bond being drawn. [...] Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence. [...] About these matters

---

<sup>15</sup> Although the Bayes rule theoretically allows for a model (= theory) falsification, i.e. probability of zero, this contradicts the confirmation holism (Goerz 2006). According to the *confirmation holism* (or *Quine-Duhem-Thesis*) a theory cannot be verified in real life because it is always tied to further theories and assumptions resulting from a person's background knowledge. It is only possible to falsify the theory and the background knowledge simultaneously. Unfortunately, there is no full access to a person's background knowledge. Consequently, there is no chance for an absolute certainty on the truth or falsity of a theory (see Goerz 2006 for more details).

there is no scientific basis on which to form any calculable probability whatever.  
We simply do not know.

A challenge is to offer principles to avoid arbitrary subjectivity within the SCB that can even translate into the HCB by a high-handed or fraudulent appraisal of the reliability of the data and the model assumptions or by a sloppy quantification of (prior) distributions.<sup>16</sup> The goal is not to reach a non-existing "objectivity" but coming as close as possible to *inter-subjectivity*.

We favor a *Bayesian due diligence* scoring the *level of information* embodied in the assessments. We have summarized some (reasonable) criteria in section 15.5 which generally account for three fundamental meta-criteria:<sup>17</sup> (1) *Substantive expertise* describes a person's level of information. (2) *Normative expertise* is the person's skill to express her beliefs in an adequate probabilistic form. (3) *Credibility* reflects our confidence whether assessments base on faithful information or misinformation.<sup>18</sup>

Regarding these meta-criteria, we demand a new risk classification for probability assessments produced by the Bayes rule as well as directly expressed by human beings. It references to the classical terms "risk" and "uncertainty" of Knight (1921):

- *risk factor*: A factor is risky when it is described by (subjective) probabilities that are judged to be *reliable* - i.e. quite inter-subjective and based on a(n economically) sufficient level of faithful information.<sup>19</sup>
- *uncertain factor*: A factor is uncertain when it is described by (subjective) probabilities that seem *unreliable* - i.e. quite intra-subjective or not based on a(n economically) sufficient level of faithful information.

We refer to risk factors and uncertain factors as *dragon kings* because they are postulated (but not necessarily correctly quantified) factors that are supposed to be relevant for risk management.<sup>20</sup> However, a prudent BRM should always bear in mind critical factors not

<sup>16</sup>However, there is no doubt about the adequacy of mathematical Bayesian updating in risk management because it is purely normative corresponding to a rational utility maximizing agent.

<sup>17</sup>The *goodness criteria* substantive and normative expertise were proposed by Winkler & Murphy (1968).

<sup>18</sup>We avoid to claim for "true" information. We believe that there is often no "objective" truth. Instead, we demand information which is not deliberately misleading but a faithful assessment of the assessor.

<sup>19</sup>Posterior distributions resulting from the mathematical Bayesian updating only describe risk factors when the data and the model assumptions are believed to be reliable and there is a general acceptance for the prior distributions within the risk management or the company. The latter demand might often result in uninformative (flat) priors.

<sup>20</sup>The original concept is different. Dragon kings correspond to meaningful outliers that coexist "with power laws in the distributions of event sizes under a broad range of conditions in a large variety of systems" (see Sornette 2009, abstract). In fact, Sornette's dragon kings can always be detected by a (sophisticated) data analysis while our dragon kings can also/ solely be identified by human reasoning.

	classical understanding	Bayesian due diligence
<b>dragon king</b>	<ul style="list-style-type: none"> <li>• known probabilities</li> </ul>	<ul style="list-style-type: none"> <li>• information based probabilities</li> </ul>
<b>uncertain factor</b>	<ul style="list-style-type: none"> <li>• unknown probabilities</li> </ul>	<ul style="list-style-type: none"> <li>• non-/ mis-information based probabilities</li> <li>• factors without probability statements</li> </ul>
<b>black swan</b>		<ul style="list-style-type: none"> <li>• some critical but ignored factors</li> </ul>

Figure 3: trichotomy of Bayesian due diligence

considered so far. Consequently, we enhance our categorization by *black swans* in the sense of Taleb (2007) - critical but personally unforeseeable factors. This results in two claims: (1) A risk management should mobilize risk cushions that can absorb unexpected adverse events of minor up to medium size. (2) Moreover, it should look for new information (sources) providing a better appraisal of possible risks and opportunities.

We hope that a Bayesian due diligence, propagating the trichotomy of risks, uncertainties, and black swans can help to reconcile conservative risk managers with the Bayesian understanding of risk management (see figure 3 for our trichotomy of risks and figure 4 for a basic structure of a prudent BRM). It curbs arbitrariness of subjective assessment and fosters a conservative risk management.

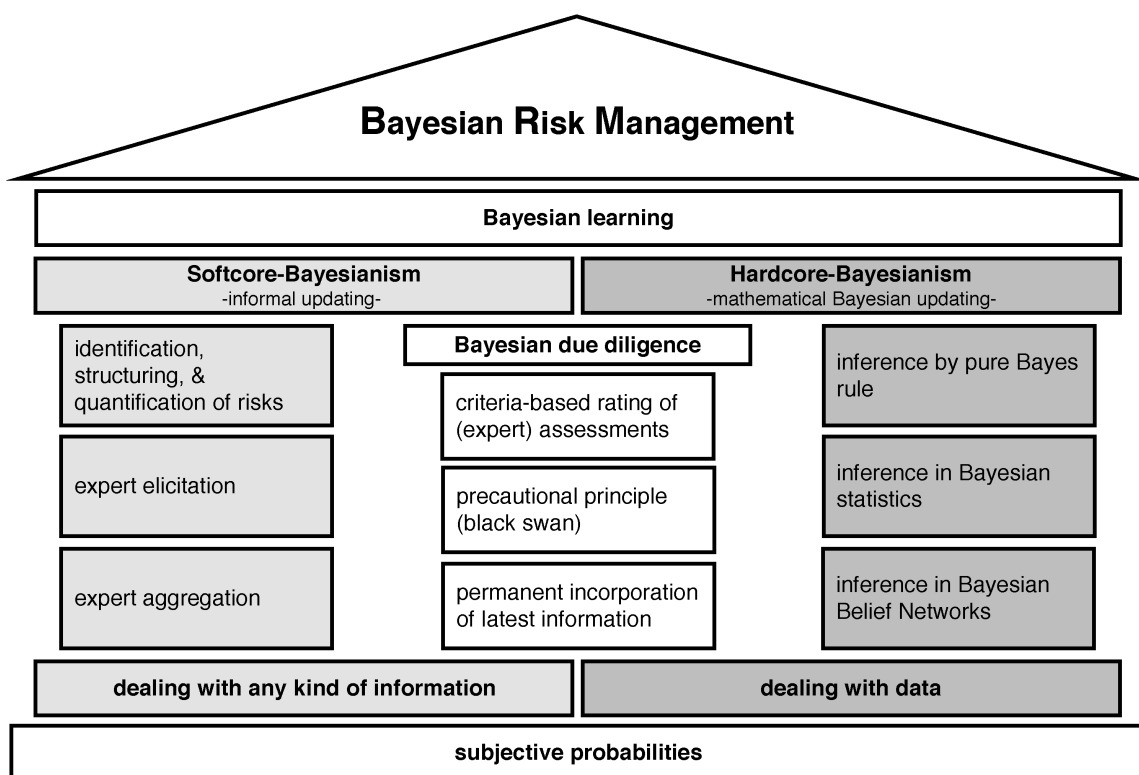


Figure 4: temple of Bayesian Risk Management



### 3 Hardcore-Bayesianism

Hardcore-Bayesianism (HCB) summarizes all empirical approaches where data-based inference relies on Bayesian updating via the Bayes rule. Just like non-Bayesian data analysis, HCB implicitly assumes data to contain relevant information for inference (= reliability).

The concept of Bayesian Belief Networks (BBN) is an exciting part of HCB. Amongst others, it is beneficial for decision making in medicine and engineering. Unfortunately, it goes beyond the scope of this work to present BBNs. Therefore, we refer to other publications like the practicable textbook Pourret et al. (2008) or Krause (2010), a BBN study on the CO2 emission reduction options for the German automotive industry.

Beside BBNs an even more important and universal area of HCB is Bayesian statistics in general and Bayesian parameter estimation in particular. In contrast to frequentist statistics, it does not restrict to the estimation of point estimates. Instead, Bayesian estimates are (multivariate) distributions reflecting a rational level of uncertainty in consideration of the available data. This has three practical consequences:

1. In the case of a low number of observations, frequentist statistics often fails to reject the null hypothesis in favor of the point estimates. Additionally, the high uncertainty of the estimators is often ignored in subsequent risk measures. In contrast, the quality of Bayesian estimates does not suffer from a low number of observations. In such a case, Bayesian estimates reflect the higher uncertainty by broader distributions.
2. It is often a drudgery in frequentist statistics to validate whether the preconditions are fulfilled to run a test, at all.<sup>21</sup> Although the convergence of the Bayesian estimates needs to be verified, there is no classical hypothesis testing. The (relative) adequacy of a model can be judged by model weights.
3. Bayesian statistics does not reject hypotheses but fully accounts for parameter as well as model uncertainty in the data. This results in a more conservative risk management and allows to explain phenomena like the equity premium puzzle or the capital charge puzzle (see section 11).

We have structured the HCB part of this dissertation in the following way: First, we describe the preliminary work of model building and compare the basic frequentist and Bayesian idea of parameter estimation. Then, we introduce the Markov Chain Monte Carlo (MCMC) simulation concept. Because of the specifics, we separately dwell on Bayesian model weighting. For a better understanding, we exemplarily run a MCMC estimation with twelve stochastic processes (models) on CO2 price observations which we subsequently compare by

---

<sup>21</sup>e.g., Gaussian distribution by Jarque-Bera or auto-correlation by Durbin-Watson

Bayesian model probabilities. To explicitly quantify the difference between the frequentist and Bayesian CO2 price risk assessment, we calculate risk premiums resulting from Bayesian value-at-risks, capital charges, and option prices.

## 4 Model Building

In the following, we explain the subjectively driven preliminary work, we call *model building*, every statistician has to do regardless whether he/ she is an "objective" frequentist or a subjective Bayesian - terms we explain later in more details. There are concepts like hypothesis testing, model selection based on information criteria (see Burnham & Anderson 2004), or probability integral transform (see Angus 1994) hoped to enable an "objective" model building. However, we are convinced that these cannot prevent subjectivity to be a critical component of model building.

To avoid confusion, in this paper we use the terms "objective" and subjective statistics to address parameter estimation, evaluation of parameter quality, and weighting of preselect models  $\mathcal{M} \in \{1, \dots, K\}$ . The (subjectively) preliminary fixing of these  $K$  models (model building) out of an infinite number of possible models is ignored for specifying the terms "objective" and subjective statistics.

Assume, we believe a time-dependent variable, e.g. log-returns on stock prices, can be described by a vector  $Y = (Y_1, \dots, Y_T)'$  of stochastic variables where<sup>22</sup>

$$Y_t \sim Dist_t(g_{tl_t}(\theta), l_t = 1, \dots, L_t) \text{ for } t = 1, \dots, T.$$

The distributions are characterized by  $L_t$  parameters, which are functions  $g_{tl_t}(\theta)$  of *basic* or *structural parameters*  $\theta = (\theta_1, \dots, \theta_J)'$ . With an assumption on the distribution of  $Y_t$ , we can construct the respective *marginal (probability) density function* (pdf)<sup>23</sup>  $f_{Y_t}(y_t|\theta)$ .

In a next step we need to presume a dependence structure for the  $T$ -dimensional (inter-temporal) joint distribution  $Y \sim Dist_T(g_l(\theta), l = 1, \dots, L)$ . Based on this dependence structure we can set up a *likelihood (joint density) function*  $L_y(\theta) = f_Y(y|\theta)$  with observations  $y = (y_1, \dots, y_T)'$ . The likelihood for stochastic dependent variables can be formulized in different ways:

---

<sup>22</sup>A more general formulation for time-series as well as cross-sectional data is  $Y_i \sim Dist_i(g_{il_i}(\theta), l_i = 1, \dots, L_i)$ , for  $i = 1, \dots, n$ . However, we only deal with time-series models in this paper. Consequently, we concentrate on this special case.

<sup>23</sup>For simplicity we do not distinguish between a *probability density function* (pdf) and a *probability mass function* (pmf) for continuous and discrete distributions. We will always use the term "pdf" to refer to a marginal distribution.

- *multivariate distributions*: When all variables  $Y_t$  originate from the same distribution  $Dist(g_{tl}(\theta), l = 1, \dots, L)$ , where distribution parameters  $g_{tl}(\theta)$  can differ for  $t = 1, \dots, T$ , the joint distribution is called multivariate distribution. Standard multivariate distributions are the multivariate Gaussian or Student's t distribution.
- *meta-distributions (via copulas)*:<sup>24</sup> The resulting joint distribution from arbitrary marginal distributions is called meta-distribution. The marginal distributions are coupled via a  $T$ -dimensional multivariate uniform distribution defined on a  $[0, 1]^T$  hypercube. The exact dependence structure originates from a copula function, which can be produced by a multivariate distribution. Consequently, there are meta-Gaussian and meta-Student's t distributions. Other prevalent meta-distributions are the meta-Clayton, meta-Gumbel, or the meta-Frank distribution.
- *Markovian-property*:<sup>25</sup> The Markovian-property can substantially ease the construction of joint distributions for time-series. A stochastic time-series  $Y = (Y_1, \dots, Y_T)'$  is a Markov chain of order  $m$  if the conditional distribution of  $Y_t$  only depends on the  $m$  preceding realizations

$$f_{Y_t}(y_t | y_{t-1}, \dots, y_m, \dots, y_1, \theta) = f_{Y_t}(y_t | y_{t-1}, \dots, y_m, \theta).$$

In the case of  $m = 1$  the likelihood function reduces to

$$L_y(\theta) = f_{Y_1}(y_1 | \theta) \cdot \prod_{t=2}^T f_{Y_t}(y_t | y_t, \theta).$$

Very often  $Y = (Y_1, \dots, Y_T)'$  are (1) assumed to be stochastically independent or (2) transformed to  $\tilde{Y}$  (e.g. prices to (log-)returns) that are assumed to be independent. Under stochastic independence, the likelihood is simply defined by a factorization of the marginal distributions, i.e.  $L_y(\theta) = f_Y(y | \theta) = \prod_{t=1}^T f_{Y_t}(y_t | \theta)$ .

In this paper we term the combination of assumptions on the marginal distributions of  $Y_t$ ,  $t = 1, \dots, T$ , and on the dependence structure of  $Y = (Y_1, \dots, Y_T)'$  as a (*stochastic*) *model*  $\mathcal{M}$ . The process of constructing  $K$  explicit models out of an infinite number of possible models we call *model building*. A model  $\mathcal{M}$  defines the domains of all model parameters  $\theta = (\theta_1, \dots, \theta_J)'$  but not their exact values. We additionally assume the observed data  $y = (y_1, \dots, y_T)'$  is generated by the *data generating process* (DGP)  $\mathcal{M} | \theta^*$ . It is described by the likelihood  $L_y(\theta^*) = f_Y(y | \theta^*)$  where  $\theta^*$  is the "true" or "objective" parameter vector.

---

<sup>24</sup>see Cech (2006), Schmidt (2007), Bouye et al. (2000), or Fermanian & Scaillet (2004)

<sup>25</sup>see Ross (1995), pp. 163

We generally argue statisticians believe that each model  $\mathcal{M} \in \{1, \dots, K\}$  (they construct) could be the "true" model for  $Y = (Y_1, \dots, Y_T)'$ .<sup>26</sup> Of course, this is an extremely hard assumption that might only be appropriate to natural but not to social science. However, a violation of this assumption does not destroy the basic ideas we want to present.

## 5 Frequentist Statistics

In this dissertation we term all statistical inference theory that tries to find *point estimates*  $\hat{\theta}$  for the "true" parameters  $\theta^*$  of the DGP  $L_y(\theta^*) = f_Y(y|\theta^*)$  as *frequentist* (or *classical*) *statistics*. Probability formulation is reserved for uncertain variables  $Y$ . Data generating parameters  $\theta^*$  are assumed to be deterministic and consequently fix but unfortunately unknown. Probability is a rather restricted, nevertheless complex, concept which is loosely speaking the relative frequency of realized  $Y_t = y_t$  (for discrete  $Y$ ) when  $Y_t$  could be repeatedly observed  $n \rightarrow \infty$  times.

### 5.1 Parameter Estimation

One of the earliest and still most popular techniques for the parameter estimation in frequentist statistics is the *least square* (LS) *estimation* which was proposed by Carl Friedrich Gauß in the early 19th century.<sup>27</sup> The basic concept is simple. Assume we believe the observations  $y = (y_1, \dots, y_T)'$  can be approximated by a deterministic function  $y_t \approx \hat{y}_t = g_t(y_{-t}, x, \theta^*)$  which might depend on other observations  $y_{-t}$  (except  $y_t$ ) of the same variable  $Y$ , on observed realization of other variables  $X$ , and on an unknown parameter vector  $\theta^*$ . Then the least square estimates are the parameters that minimize the quadratic error between our model estimates and the observations

$$\theta^{LS} = \arg \min_{\theta} \sum_{t=1}^T [y_t - g_t(y_{-t}, x, \theta)]^2.$$

In the case of  $\theta$  linear in  $g_t(y_{-t}, x, \theta)$ , there exists an analytical solution for the least square estimators. However, when  $g_t(y_{-t}, x, \theta)$  is non-linear, numerical techniques are needed, like a Taylor approximation of  $g_t(y_{-t}, x, \theta)$  in  $\theta$  to linearize  $g_t$ , the steepest-descent (or gradient) method, or the maximum neighborhood method (= interpolation between Taylor and Gradient method, see Marquardt 1963).

---

<sup>26</sup>A prominent counter-example is the quasi-maximum likelihood (QML) estimation, where estimation is deliberately based on a "wrong" model. See page 17 for more details.

<sup>27</sup>See Mood et al. (1974), pp. 498, Hartung (1999), pp. 574, Bamberg & Baur (2001), pp. 151, and especially for-time series Schlittgen & Streitberg (1989), pp. 193, for more details.

Another extremely popular and dominant method is the *maximum likelihood* (ML) estimation introduced by Ronald Aylmer Fisher in the early 20th century. In contrast to the least square estimation the ML estimation needs an assumption on the distribution (in our terms a *model*) of  $Y$  as it maximizes the likelihood function of model  $\mathcal{M}$  in respect of the parameter vector  $\theta$

$$\theta^{ML} = \arg \max_{\theta} L_y(\theta).$$

Unfortunately, the likelihood function relatively fast goes to zero or to extremely large values. Consequently, ML estimation is applied by maximizing the log-likelihood function  $l_y(\theta) = \ln L_y(\theta)$  which is just a monotone transformation.

For basic likelihoods there are analytical solutions to the ML estimators. In the special case of  $y = (y_1, \dots, y_T)'$  stemming from a Gaussian variable  $Y$ , the ML estimators coincide with the least square estimators. For more complex models, numerical or simulation techniques are needed.

An alternative to ML estimation is the *quasi-maximum likelihood* (QML) estimation which is rather a more honest understanding of the world than a technique. The QML estimation admits that the observations  $y = (y_1, \dots, y_n)'$  stem from an *unknown* (and maybe highly complex) DGP  $f_Y^*(y)$  which cannot be maximized by ML.<sup>28</sup> Consequently, the modeler should look for a "simple" but reasonable approximating quasi-likelihood  $L_y^{QML}(\theta) = g_Y(y|\theta)$  given a parameter vector  $\theta$ . The parameters are calibrated by minimization of the *Kullback-Leibler Information Criterion* (KLIC)<sup>29</sup>

$$KLIC(f_Y, g_Y|\theta) = \int \ln \frac{f_Y^*(y)}{L_y^{QML}(\theta)} f_Y^*(y) dy,$$

which measures the misspecification and equals zero for  $f_Y^*(y) = L_y^{QML}(\theta)$ . As the quasi-likelihood function only depends on the parameter vector, the KLIC can be rearranged and the maximization of

$$\int \ln L_y^{QML}(\theta) f_Y^*(y) dy = E_{f_Y^*}(\ln L_y^{QML}(\theta)) = \frac{1}{T} E_{f_Y^*} \left( \sum_{t=1}^T \ln g_Y(Y_t | Y_{<t}, \theta) \right)$$

is equivalent. In empirical estimation, the expectation operator is ignored as the "true" DGP  $f_Y^*(y)$  is unknown and the maximization solely bases on the observations  $y = (y_1, \dots, y_T)'$ . Hence, the QML estimators are  $\theta^{QML} = \arg \max_{\theta} \frac{1}{T} \sum_{t=1}^T \ln g_Y(y_t | y_{<t}, \theta)$ .

<sup>28</sup>Our explanation mainly bases on Kuan (2007), chapter 9.

<sup>29</sup>The idea of the KLIC is to measure the information gain or loss when our probability assessment on the realization of an event  $A$  is updated from  $p^{old}$  to  $p^{new}$  (Kuan 2007, p. 234). An increase (decrease) of the realization probability,  $p^{new} (> (<)) p^{old}$ , increases (reduces) our information level of  $A$ . The information content changes disproportionately for shifting probability assessments. In the case of  $n$  mutually exclusive events  $A_1, \dots, A_n$  the expected shift in the information level is  $\sum_i p_i^{new} \ln(p_i^{new}/p_i^{old})$ . The KLIC is the continuous version.

Sometimes, the QML estimation is understood more trivial as a ML parameter estimation based on a *deliberately* misspecified likelihood function, e.g. assuming Gaussian errors when they are not. Again, the reduction in complexity of the estimation via QML is not up to a different technique but up to the estimation of a *simpler* approximating likelihood.<sup>30</sup>

In contrast to the QML estimation there also exist "true" estimation techniques different from ML, like *generalized method of moments* (GMM, see Hansen 1982), *simulated method of moments* (SMM), or *efficient method of moments* (EMM, see Andersen et al. 1999) estimation.

The GMM works the following way. Assume you have a model with  $S = 2$  conditions  $E(Y_t) = h_{1t}(\theta^*)$  and  $Var(Y_t) = h_{2t}(\theta^*)$ , for  $t = 1, \dots, T$ , which can be rearranged to  $g_{1t}(\theta) = E(Y_t) - h_{1t}(\theta)$  and  $g_{2t}(\theta) = Var(Y_t) - h_{2t}(\theta)$ . Given the data generating process with parameter vector  $\theta^*$ , it follows  $\bar{g}_k(\theta^*) = T^{-1} \sum_{t=1}^T g_{st}(\theta^*) \rightarrow 0$ ,  $s = 1, 2$ . The GMM estimator is  $\theta^{GMM} = \arg \min_{\theta} (\bar{g}(\theta)' \Psi \bar{g}(\theta))$ , where  $\bar{g}(\theta) = (\bar{g}_1(\theta), \bar{g}_2(\theta))'$  and  $\Psi$  is an empirical positive definite weighting matrix for the conditions. A requirement is that the number of conditions  $S$  should at least equal to the number of parameters  $J$ . For  $\theta$  non-linear in  $g_{st}(\theta)$  the GMM estimator can only be obtained numerically.

The simulated method of moments (SMM), proposed by McFadden (1989) and extended to time-series by Lee & Ingram (1991), is an extension of the GMM approach when moment condition functions  $g_{st}(\theta)$  have no analytical representation. Then SMM allows for a simulation of  $g_{st}(\theta)$  via Monte Carlo techniques.

One of the most important problems in ML estimation is the parameter estimation in models where observations  $y$  critically depend on *latent*, i.e. *unobservable variables*  $X$ . Assume a structural model  $Y_t = X_t \varepsilon_t$ , where  $\varepsilon_t \sim N(0, 1)$  and  $X_t$  is a latent process driven by the parameter vector  $\theta$ . Unfortunately, the vector  $y = (y_1, \dots, y_T)'$  is the only observation we have access to. Consequently, ML parameter estimation needs to maximize the *unconditional* likelihood

$$L_Y(\theta) = f_Y(y|\theta) = \int f_{YX}(y, x|\theta) dx = \int f_Y(y|x, \theta) f_X(x|\theta) dx,$$

which requires to solve the  $T$ -dimensional integral that is generally intractable. In this context  $f_Y(y|x, \theta)$  is called *auxiliary density* where the latent variable  $X$  is treated as observation.

The EMM approach is a technique that allows to maximize such complex likelihoods. In a first step a quasi-score function is calculated from the auxiliary density by  $s(y_t, x_t) = \frac{\partial}{\partial x} \ln f_{Y_t}(y_t|x, \theta) = \frac{\partial}{\partial x_t} \phi(y_t|0, x_t)$  which ignores the parameter vector.<sup>31</sup> Via QML we find

<sup>30</sup>In a log-variance stochastic volatility model, Harvey et al. (1994) assume Gaussian errors for the logarithm of the squared log-returns although they are log- $\chi^2$  distributed. Because of the high frequency of financial data, they found no huge misspecification.

<sup>31</sup> $\phi(\cdot|m, s)$  is the pdf of a Gaussian distribution with mean  $m$  and standard deviation  $s$ .

latent realizations  $\hat{x} = (\hat{x}_1, \dots, \hat{x}_T)'$  maximizing  $T^{-1} \sum_{t=1}^T s(y_t, x_t)$ . Under some regularities  $\hat{x} \rightarrow x_0$  with  $x_0$  called as *quasi-true values* of  $x$  (see Andersen et al. 1999, p. 66). The model implies that  $s(y_t, x_t)$  is induced by the probability measure  $f_{Y_t}(y_t|\theta)$  and features a vector of conditional means  $m(x_0, \theta)$ . These are approximated by a SMM step via

$$\hat{m}(\hat{x}, \theta) = (\hat{m}_1(\hat{x}_1, \theta), \dots, \hat{m}_T(\hat{x}_T, \theta))',$$

where  $\hat{m}_t(\hat{x}_t, \theta) = W^{-1} \sum_{w=1}^W s(\hat{y}_t^{(w)}(\theta), \hat{x}_t)$  and  $\hat{y}^{(w)}(\theta) = (\hat{y}_1^{(w)}(\theta), \dots, \hat{y}_T^{(w)}(\theta))'$ , for  $w = 1, \dots, W$ , is a series simulated from the structural model given a parameter vector  $\theta$  and the QML estimates  $\hat{x}$ . The EMM estimators are calculated by minimizing

$$\theta^{EMM} = \arg \min_{\theta} \left( \hat{m}(\hat{x}, \theta)' \hat{\Sigma}^{-1} \hat{m}(\hat{x}, \theta) \right),$$

where  $\hat{\Sigma}$  is the respective covariance matrix which can be calculated in the QML or SMM step.

## 5.2 Quality of Estimation Results

In frequentist statistics, the "true" parameter  $\theta^*$  is a deterministic, fixed, non-random constant<sup>32</sup> while the parameter estimator  $\hat{\Theta}$  is a random variable with different realizations  $\hat{\theta}$  for several observations  $y$  originating from the same population. Based on these assumptions statistical inference is mostly based on *hypothesis testing* and *confidence intervals*.

In (*Neyman-Pearson*) *hypothesis testing* a null hypothesis,  $H_0 : \theta \in \theta_{(0)}$ , and an alternative hypothesis,  $H_1 : \theta \in \theta_{(1)}$ , are stated, where  $\theta_{(0)} \cap \theta_{(1)} = \emptyset$  and the parameter space is often  $\theta_{(0)} \cup \theta_{(1)}$ . It is analyzed whether the observed estimator value  $\hat{\theta}$  is compatible with  $y$  driven by the parameter vector  $\theta \in \theta_{(0)}$ .

Generally, a test is conducted via a test statistics,  $t = \tau(y)$ ,  $t \in \mathbb{R}$ ,<sup>33</sup> which rejects the null hypothesis when  $y \in y_{(1)}$ . The range of  $y_{(1)}$  depends on the desired *significance level*  $\alpha$ , normally  $\alpha \in \{0.1, 0.05, 0.01\}$ , where

$$P(Y \in y_{(1)} | \theta \in \theta_{(0)}) \leq \alpha.$$

Empirical studies often list another statistics called *p-value* originating from *Fisher hypothesis testing theory*, which only states a null hypothesis. The *p-value*

$$p = \begin{cases} P\left(\tau(Y) \begin{matrix} \geq \\ \leq \end{matrix} \tau(y) \mid \theta \in \theta_{(0)}\right), & \text{for one-sided hypothesis} \\ P(|\tau(Y)| \geq |\tau(y)| \mid \theta \in \theta_{(0)}), & \text{for two-sided hypothesis} \end{cases},$$

<sup>32</sup>For a more intuitive understanding in this section we assume that the parameter vector  $\theta$  consists of one parameter, only.

<sup>33</sup>A test statistic for the mean parameter  $\mu$  is the arithmetic mean  $\tau_{\mu}(y) = \sum_t y_t / T$  while the variance  $\sigma^2$  is generally estimated by the test statistic  $\tau_{\mu}(y) = \sum_t [y_t - \tau_{\mu}(y)]^2 / [T - c]$ , where  $c = 0, 1$ .

is the "objective" probability (relative frequency in the limit) of obtaining a  $\tau(Y)$  that is at least as extreme as the observed  $\tau(y)$  if  $\theta \in \theta_{(0)}$ . In statistical software packages like EViews,  $p$ -values are often implemented for the null hypothesis  $H_0 : \theta = 0$  to test whether the estimated parameter is significant or not. As long as the null hypothesis cannot be rejected there is evidence that the parameter can be ignored. Otherwise, the estimator  $\hat{\theta}$  is accepted as "true" parameter value  $\theta^*$ .

Another common statistic is the *confidence interval*  $CI_{y|\beta} = [CI_{y|\beta}^{low}, CI_{y|\beta}^{up}]$  reflecting reliability of the parameter estimator  $\hat{\theta}$ . For multiple samples from the same population the confidence interval contains the *fixed* "true" parameter value  $\theta^*$  with a desired "objective" probability which is called *confidence level*  $\beta$ , where  $P(\theta^* \in CI_{y|\beta}) = \beta$ . Confidence intervals are typically stated for confidence levels 90%, 95%, and 99%. Generally, this means the higher the confidence level the larger the confidence interval. The abstract interpretation of the confidence interval is due to the fact that the confidence interval itself is a random variable in contrast to the "true" but unknown parameter  $\theta^*$ .

## 6 Basic Idea of Bayesian Statistics

In the following we understand Bayesian statistics as the quantification of uncertainty on a parameter vector  $\Theta = (\Theta_1, \dots, \Theta_J)'$  resulting from a model  $\mathcal{M}$  when prior *subjective* uncertainty on  $\Theta$  is rationally updated by observations  $y = (y_1, \dots, y_T)'$  originating from the data generating process  $f_Y(y|\theta^*)$  where  $\theta^*$  is the "true" parameter vector.<sup>34</sup> In contrast to frequentist statistics where a point estimate  $\hat{\theta}$  is calculated which is hoped to be a good proxy for  $\theta^*$  the output of Bayesian statistics is a *stochastic variable*  $\Theta$  describing the updated uncertainty.<sup>35</sup>

### 6.1 Bayesian Parameter Estimation

Assume the modeler has set up a model  $\mathcal{M}$  for the stochastic variables  $Y = (Y_1, \dots, Y_T)'$ . The likelihood (joint density function)  $L_y(\theta) = f_Y(y|\theta)$  is hoped to explain the observations  $y = (y_1, \dots, y_T)'$ . Assume further that the model  $\mathcal{M}$  is correct however the "true" data generating process (DGP) is the special case of  $\mathcal{M}$  with  $\theta = \theta^*$ . Unfortunately,  $\theta^*$  is unknown.

---

<sup>34</sup>Of course, we could also assume that the "true" joint density  $f_Y^*(y)$  is unknown, and we just want to calibrate model  $\mathcal{M}$  with density  $f_Y(y|\theta)$  to the observations as good as possible. However, this is a bit more abstract and could distract from the basic idea of Bayesian statistics.

<sup>35</sup>In the following,  $\Theta$  refers to a stochastic parameter vector whereas  $\theta$  is its realizations. However, for practicable reasons, all other stochastic parameters, written in Greek letters, are not typed in capital letters. Hence, there is no optical difference between random variable  $\mu$  and its realization  $\mu$ .



The Bayesian parameter estimation of  $\theta^*$  is conducted via several steps:

*prior distributions:* In a first step we need to state, purely subjectively, so called prior distributions for each of the  $J$  model parameters  $\Theta = (\Theta_1, \dots, \Theta_J)'$  in form of a joint prior density  $\pi(\theta)$ . As the definition of a joint prior can be very complicated and is often not crucial, prior independence,  $\pi(\theta) = \prod_{j=1}^J \pi(\theta_j)$ , is mostly assumed.

The marginal prior  $\pi(\theta_j)$  conveys two basic information: First, we express our purely subjective beliefs on the domain of  $\Theta_j$  by choosing suitable standard distribution, e.g. Gaussian distribution for  $\Theta_j \in \mathbb{R}$  (mean value), inverse gamma distribution for  $\Theta_j \in \mathbb{R}_+$  (variance), or uniform distribution for  $\Theta_j \in [-1, 1]$  (correlation coefficient). A remarkable advantage of Bayesian statistics is the straightforwardness of incorporating restrictions on parameters. Undesired realizations during the estimation process are simply ruled out. Second, we can adjust the respective standard distribution (to some extent) to express our prior beliefs.

*posterior distribution:* The result of Bayesian statistics is the posterior parameter distribution of  $\Theta|y$  which is the prior distribution of  $\Theta$  updated by the observations  $y = (y_1, \dots, y_T)'$ . The joint posterior density is formulized by the *Bayes rule* (or *theorem*)

$$\pi(\theta|y) = \frac{f_Y(y|\theta) \pi(\theta)}{f(y)} = \frac{L_y(\theta) \pi(\theta)}{\int L_y(\theta) \pi(\theta) d\theta}$$

Unfortunately, the  $J$ -dimensional integral in the denominator is mostly intractable. Fortunately, the form of the posterior is not effected by  $f(y)$ . It is just a constant scaling factor. Hence, the posterior is mostly presented as

$$\pi(\theta|y) \propto L_Y(\theta) \pi(\theta),$$

where  $\propto$  marks proportionality.

There are two important facts on the posterior, which are not necessarily apparent: (1) The Bayes rule shows that the form of the posterior  $\pi(\theta|y)$  is defined by three factors: the prior  $\pi(\theta)$ , the likelihood  $L_y(\theta)$  (which is determined by the distributional assumption on  $Y$ ), and the actual observations  $y$ . The updating from the prior to the posterior is a purely normative process, based on the Kolmogorov axioms. It describes how people should update and not how they actually do! Indeed, when we mention subjective uncertainty in context with the posterior we mean *normatively updated subjective uncertainty*. (2) Theoretically, the posterior  $\pi(\theta|y)$  is the final Bayesian parameter estimation result. Practically, statistics like  $E(\Theta|y)$  or  $Var(\Theta|y)$  are calculated from the posterior to offer a better understanding. Nevertheless, these statistics just describe the *Bayesian parameter estimator*,  $\pi(\theta|y)$ , that is a distribution and normally not a vector of point estimates.

An advantage of Bayesian statistics is the relative simplicity of the incorporation of latent (state) variables  $X$  which cannot be observed. In frequentist statistics we would try to find point estimates  $\hat{\theta}$  that maximize the unconditional likelihood of  $y$  which is the solution of a  $T$ -dimensional integral,  $L_y(\theta) = \int f_Y(y|x, \theta) f_X(x|\theta) dx$ .

Instead, Bayesian statistics "simply" augments the parameter vector  $\tilde{\Theta} = (\Theta', X')'$ , where  $\Theta = (\Theta_1, \dots, \Theta_J)'$  and  $X = (X_1, \dots, X_J)'$ . The resulting posterior is

$$\pi(\tilde{\theta}|y) \propto L_y(\tilde{\theta}) \pi(\tilde{\theta}) = f_Y(y|x, \theta) f_X(x|\theta) \pi(\theta),$$

where  $f_X(x|\theta)$  is the conditional joint pdf of the latent variables  $X$ . We avoid to call it "prior".

## 6.2 Uninformative Priors

We distinguish three different groups of statisticians dealing with Bayesian statistics: (1) frequentist statisticians applying Bayesian techniques because of their efficiency,<sup>36</sup> (2) objective Bayesians, and (3) (subjective) Bayesians.

Like (subjective) Bayesians, objective Bayesian statisticians favor a stochastic interpretation of parameters  $\Theta$  over that of point estimates  $\hat{\theta}$ . However, they oppose the formulation of informative priors  $\pi(\theta)$ . When there is a low number of observations, the prior specification can have a substantial influence on the posterior. This is undesirable for non-intuitive parameters in complex models. Unfortunately, complete ignorance cannot be expressed on all kinds of parameters. A prime example is the volatility parameter  $\sigma$ . Expressing ignorance via an uniform distribution on the interval  $(0, \sigma_{up}]$ , where  $\sigma_{up} < \infty$ , would result in an informative prior on the variance  $\sigma^2$ , which is just a simple transformation of the volatility. Nevertheless, the definition of *improper priors*, i.e. non-finite  $\int \pi(\theta) d\theta$ , sometimes results in proper posteriors, i.e.  $\int \pi(\theta|y) d\theta < \infty$ , which is the crucial point. Popular improper priors are *Jeffrey's priors*  $\pi(\theta) = |I(\theta)|^{1/2}$  where

$$I(\theta) = -E \left( \frac{\partial^2 \ln f_y(y|\theta)}{\partial \theta \partial \theta'} \right).$$

is the Fisher's information matrix for  $\theta$  (Rachev et al. 2008, p. 26). In a model with Gaussian errors, the respective prior for the location parameter  $\mu$  is  $\pi(\mu) \propto 1$ , which expresses a complete ignorance, while the prior  $\pi(\sigma) \propto \sigma^{-1}$  on the scale parameter  $\sigma$  is rather approximatively uninformative.

---

<sup>36</sup>Andersen et al. (1999) find that the Bayesian MCMC sampler (see section 7) results in a lower mean squared error for stochastic volatility models than efficient method of moments (EMM), generalized methods of moments (GMM), quasi maximum likelihood (QML), and simulated method of moments (SMM).

However, keep in mind, that uninformative priors should only be applied if the resulting posterior is proper. E.g., Bauwens & Lubrano (1998), pp. 25, present a proof showing that a prior on  $(2, \infty)$  for the degree of freedom in a GARCH model with Student's t distributed errors results in an improper posterior.

In this context Ronald Aylmer Fisher proposed the *fiducial inference* in 1930 (see Pedersen 1978 and Aldrich 2000). Similar to Bayesian statistics and in opposition to frequentist statistics it regards parameters  $\Theta$  as stochastic variables. However, in contrast to Bayesian statistics, the fiducial posteriors are solely derived from the data  $y$ . No prior distribution is needed for the parameters. Consequently the *fiducial posterior* (often cited by the obsolete term *inverse probability*) needs another formulation than the Bayes rule. Fisher only gave a representation  $f(\theta|y) \propto -\partial F_Y(y|\theta)/\partial\theta$  for a model defined by a single parameter  $\theta$  and with a single observation  $y$  where the likelihood (often cited by the obsolete term *direct probability*) is defined by the derivation of a cdf which is a negative function in  $\theta$  (see Hannig 2006, p. 2).

A very simple example, we take from Hannig (2006), pp. 6, may help to understand the basic idea of Fisher's fiducial argument. Assume, we believe the observation  $y$  stems from the model  $Y = \mu + 1 \cdot Z$ , where  $\mu$  is the uncertain mean parameter and  $Z \sim N(0, 1)$ . This model can be rearranged to  $\mu = Y - Z$ . As we have observed  $Y = y$  we can profit from the fact that the conditional distribution (= inverse probability or posterior) is Gaussian with  $\mu|y \sim N(y, 1)$ . Now it is simple to construct fiducial intervals with  $P(\mu \in [a, b]) = \Phi(b - y) - \Phi(a - y)$ . Such intervals have a natural interpretation like Bayesian credible intervals (see section 6.3).

In the following we ignore fiducial inference as (1) there is no general formulation for a fiducial inverse probability, (2) it has been marginalized,<sup>37</sup> and (3) often obtains similar results compared to Bayesian inference with Jeffrey's priors.

### 6.3 Bayesian Posterior Inference

As already stated the Bayesian parameter estimator is the posterior distribution  $\pi(\theta|y)$ , which requires different analysis techniques compared to frequentist point estimates  $\hat{\theta}$ . Beside the calculation of statistics like  $E(\Theta|y)$  or  $Var(\Theta|y)$ , *credible intervals* and *Bayesian hypotheses comparison* are widespread approaches.<sup>38</sup>

The concept of confidence intervals does not apply since parameters  $\Theta = (\Theta_1, \dots, \Theta_J)'$  are stochastic variables. Instead,  $J$  different credible intervals  $\mathcal{CR}_j^{(\beta)} = (\theta_j^{low}, \theta_j^{up})$  can be con-

<sup>37</sup>See Pedersen (1978). However, Hannig (2006) sees some rediscovering for problems where frequentist statistics fails.

<sup>38</sup>We base our descriptions on Rachev et al. (2008), pp. 32.

structured that guarantee

$$P\left(\Theta_j \in (\theta_j^{low}, \theta_j^{up}) \mid \theta_{-j}, y\right) = \int_{\theta_j^{low}}^{\theta_j^{up}} \pi(\theta_j \mid \theta_{-j}, y) d\theta_j \stackrel{!}{=} \beta,$$

for  $j = 1, \dots, J$  and  $\theta_{-j}$  is the parameter vector without parameter  $\theta_j$ . Credible intervals express the posterior level of uncertainty - measured by credible level  $\beta$  - on the unknown "true" parameter  $\theta_j^*$ . This is a far more intuitive interpretation than that of confidence intervals.

A helping concept for data-based scenario weighting is Bayesian hypotheses comparison (Rachev et al. 2008, pp. 33). In contrast to frequentist hypotheses testing it does not reject hypotheses but compares their posterior probabilities. Assume that we compare the null hypothesis  $H_0 : \Theta \epsilon \theta^{(0)}$  with the alternative hypotheses  $H_l : \Theta \epsilon \theta^{(l)}$ , where  $\theta^{(0)} \cap \theta^{(1)} \cap \dots \cap \theta^{(L)} = \emptyset$  and  $l = 1, \dots, L$ . Instead of (not) rejecting hypotheses, their absolute and relative performance can directly be measured, i.e.  $P\left(\Theta \epsilon \theta^{(l)} \mid y\right)$  and  $P\left(\Theta \epsilon \theta^{(l)} \mid y\right) / P\left(\Theta \epsilon \theta^{(k)} \mid y\right)$ , for  $l \neq k$ .

## 6.4 Bayesian Predictive Inference

The different kind of the Bayesian parameter estimator  $\pi(\theta \mid y)$  influences the way of risk assessment. In frequentist statistics the predictive distribution of  $Y_{T+1}$  is described by  $f_{Y_{T+1}}(y_{T+1} \mid y, \hat{\theta})$ , where  $\hat{\theta}$  is the vector of non-rejected point estimates. In contrast the (*unconditional*) *Bayesian predictive density* is

$$\pi_{Y_{T+1}}(y_{T+1} \mid y) = \int f_{Y_{T+1}}(y_{T+1} \mid y, \theta) \pi(\theta \mid y) d\theta.$$

The formulation above shows that Bayesian statistics distinguishes two different levels of risk assessments on  $Y$ :

- *data generating process (DGP) of  $Y$* : The model  $\mathcal{M}$  presumes that the DGP of  $Y$  is  $L_y(\theta^*) = f_Y(y \mid \theta^*)$  with unknown  $\theta^*$ . The frequentist risk assessment of  $Y_{T+1}$ ,  $f_{Y_{T+1}}(y_{T+1} \mid y, \hat{\theta})$ , is an estimation of the DGP given that the parameter estimates  $\hat{\theta}$  equal the "true" parameters  $\theta^*$ . The Bayesian pendant is the conditional Bayesian predictive density  $f_{Y_{T+1}}(y_{T+1} \mid y, \theta)$ . Its prediction relevance is weighted by  $\pi(\theta \mid y)$ .
- (*subjective*) *uncertainty on  $Y$* : As the Bayesian parameter estimator is the posterior distribution  $\pi(\theta \mid y)$ , the Bayesian risk assessment of  $Y$  is a description of the updated uncertainty on  $Y$  when there is uncertainty on  $\theta^*$ . Generally,  $\pi_{Y_{T+1}}(y_{T+1} \mid y)$  should not be understood as a direct estimation of the "true" DGP,  $f_Y(y \mid \theta^*)$ , but a rational updated uncertainty on the "true" DGP.

It is important to keep in mind that frequentist statistics only accounts for *model induced risk* whereas Bayesian statistics additionally incorporates *parameter risk* (or *estimation risk*, Johannes & Polson 2003, p. 6). The model induced risk describes the randomness of the model given an arbitrary parameter vector  $\theta$  while parameter risk describes the added uncertainty if the "true" parameter vector  $\theta^*$  is not known with certainty.

As the Bayesian risk assessment of  $Y$  reflects model induced as well as parameter risk we have to distinguish two levels of dependence structures: (1) The dependence structure of the DGP, reflected in the likelihood  $L_y(\theta^*) = f_Y(y|\theta^*)$ . (2) The dependency of the Bayesian predictive density on the observations  $y$ .

Assume we know the DGP generates independent samples  $Y = y$ . Then the conditional Bayesian predictive densities reduce to  $f_{Y_{T+1}}(y_{T+1}|\theta)$ . Nevertheless, the Bayesian risk assessment on  $Y_{T+1}$

$$\pi_{Y_{T+1}}(y_{T+1}|y) = \int f_{Y_{T+1}}(y_{T+1}|\theta) \pi(\theta|y) d\theta.$$

depends on past observations  $y = (y_1, \dots, y_T)'$  as the posterior  $\pi(\theta|y)$  is the updated prior.

## 7 Markov Chain Monte Carlo Simulation

As already pointed out the Bayesian estimator is the posterior  $\pi(\theta|y) \propto L_y(\theta) \pi(\theta)$ , which is a distribution and not a point estimate. Theoretically, we need no estimation procedure as  $\pi(\theta|y)$  is just proportional to the factorization of the likelihood and the priors.

Unfortunately, parameter vectors are often of higher dimensions and therefore the posterior  $\pi(\theta|y)$  is an  $J$ -dimensional distribution which can only be plotted for  $J \leq 2$ . In the case of an augmented parameter space the dimension of the posterior  $\pi(\theta, x|y) \propto f_Y(y|\theta, x) f_X(x|\theta) \pi(\theta)$  is even larger. Generally, risk managers and statisticians cannot gain much information from the raw formula proportional to  $\pi(\theta, x|y)$ .

Fortunately, there exists the *Clifford-Hammersley theorem*<sup>39</sup> which states that any continuous (augmented) joint posterior distribution  $\pi(\theta, x|y)$  is completely characterized by the conditional posteriors

$$\begin{aligned} \pi(\theta|x, y) &\propto f_Y(y|\theta, x) \cdot \pi(\theta|x) \quad \& \\ \pi(x|\theta, y) &\propto f_Y(y|\theta, x) \cdot f_X(x|\theta). \end{aligned}$$

---

<sup>39</sup>We could not get a copy of the unpublished paper John M. Hammersley and Peter Clifford (1971): Markov fields on finite graphs and lattices. However, you can find a proof of Clifford-Hammersley theorem and premises in Besag (1974).

This theorem can be applied again to find the *full* or *complete conditional posteriors*<sup>40</sup>

$$\begin{aligned}\pi(\theta_j | \theta_{-j}, x, y) &\propto f_Y(y | \theta, x) \cdot \pi(\theta_j | \theta_{-j}, x), \text{ for } j = 1, \dots, J, \\ \pi(x_i | \theta, x_{-i}, y) &\propto f_Y(y | \theta, x) \cdot f_X(x_i | \theta, x_{-i}), \text{ for } i = 1, \dots, \dim(x).\end{aligned}$$

The formulation  $\theta_{-j}$  and  $x_{-j}$  describe the vectors  $\theta$  and  $x$  without  $\theta_j$  and  $x_j$ .

Thanks to the Clifford-Hammersley theorem, Bayesian parameter estimation of the joint posterior  $\pi(\theta, x | y)$  is simplified to the estimation of the full conditional posteriors  $\pi(\theta_j | \theta_{-j}, x, y)$ ,  $j = 1, \dots, J$ , and  $\pi(x_i | \theta, x_{-i}, y)$ ,  $i = 1, \dots, \dim(x)$ .

The full conditional posteriors are "simple" univariate distributions we could easily plot. This is however not a sensible strategy, as we would need to decide on which values  $\Theta_{-j} = \theta_{-j}$  and  $X = x$  we should condition. Fortunately, there is the concept of *Markov Chain Monte Carlo* (MCMC) simulation which offers an ingenious solution to this problem. It works the following way:

1. Set start values for all parameters  $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_J^{(0)})'$  and latent variables  $x^{(0)} = (x_1^{(0)}, \dots, x_{\dim(x)}^{(0)})'$ . Set  $w = 1$ .
2. Sample all parameters  $\theta_j^{(w)}$ , where  $j = 1, \dots, J$ , one by one from the respective full conditional posterior  $\pi(\theta_j^{(w)} | \theta_{<j}^{(w)}, \theta_{>j}^{(w-1)}, x^{(w-1)}, y)$ .
3. Sample all latent variables  $x_i^{(w)}$ , where  $i = 1, \dots, \dim(x)$ , one by one from the respective full conditional posterior  $\pi(x_i^{(w)} | \theta_i^{(w)}, x_{<i}^{(w)}, x_{>i}^{(w-1)}, y)$ .<sup>41</sup>
4. Set  $w = w + 1$ . If  $w \leq W$  go back to step 2.
5. Dump all sampled values  $(\theta^{(w)}, x^{(w)})'$ ,  $w \leq B < W$ , for which the MCMC series has not yet converged. This we term *burn-in phase*. For inference, we use  $W - B$  repetitions, only.

After convergence, the output of the MCMC sampler  $\theta_j = (\theta_j^{(B+1)}, \dots, \theta_j^{(W)})'$  and  $x_i = (x_i^{(B+1)}, \dots, x_i^{(W)})'$ , for  $j = 1, \dots, J$  and  $i = 1, \dots, \dim(x)$ , approximates the (augmented) posterior  $\pi(\theta, x | y)$  (see section 7.4). Dependence structures between the different parameters  $\Theta = (\Theta_1, \dots, \Theta_J)'$  can be analyzed via covariance matrices  $Cov(\Theta) \approx \widehat{Cov}(\theta)$  or bivariate plots via scatter plots.

<sup>40</sup>In the case of independent priors it follows  $\pi(\theta_j | x) = \pi(\theta_j | \theta_{-j}, x)$ .

<sup>41</sup>The Kalman filter is an alternative which allows to sample latent variables simultaneously. See Murphy 1998/2004 for more details.

Moreover, by means of histograms or Kernel estimators, the parameter path  $\theta_j^{(B+1)}, \dots, \theta_j^{(W)}$  can approximate the *marginal (parameter) posterior*<sup>42</sup>

$$\pi(\theta_j|y) = \int \int \pi(\theta_j, \theta_{-j}, x|y) d\theta_{-j} dx.$$

Usually, statisticians do not merely present the approximated distribution. They calculate a bulk of statistics like

$$\begin{aligned} E(\Theta_j|y) &= \int \theta_j \pi(\theta_j|y) d\theta_j \approx \hat{E}(\Theta_j|y) = \frac{1}{W-B} \sum_{w=B+1}^W \theta_j^{(w)}, \\ \text{Var}(\Theta_j|y) &= E([\Theta_j - E(\Theta_j|y)]^2|y) \\ &\approx \widehat{\text{Var}}(\Theta_j|y) = \frac{1}{W-B-1} \sum_{w=B+1}^W [\theta_j^{(w)} - \hat{E}(\Theta_j|y)]^2. \end{aligned}$$

If  $\hat{E}(\theta_j|\hat{\theta}_{-j}^{(w)}, x, y)$  can be calculated analytically, a *Rao-Blackwellization*

$$E(\Theta_j|y) \approx \hat{E}(\Theta_j|y) = \frac{1}{W-B} \sum_{w=B+1}^W \hat{E}(\theta_j|\theta_{-j}^{(w)}, x, y),$$

can reduce the variance of the estimator  $\hat{E}(\Theta_j|y)$  (see Gelfand & Smith 1990, p. 440, or Rachev et al. 2008, p. 30).

In the following we describe the Gibbs sampler, Griddy Gibbs sampler, and Metropolis-Hastings sampler as dominant MCMC samplers. In the appendix A.3 you can find a short description of the *sampling-importance-resampling* (SIR) which is the most important alternative to the MCMC samplers.

## 7.1 Gibbs Sampler

The *Gibbs sampler* is the special case of the MCMC sampler where all full conditional posteriors  $\pi(\theta_j|\theta_{-j}, x, y)$  and  $\pi(x_i|\theta, x_{-i}, y)$ , for  $j = 1, \dots, J$  and  $i = 1, \dots, \dim(x)$ , are standard distributions which are mostly implemented in standard statistical software.<sup>43</sup> Consequently, all parameters  $\Theta_j|\theta_j, x, y$  and latent variables  $X_i|\theta, x_{-i}, y$  can *directly* be sampled. This has

<sup>42</sup>In the special case of a Gaussian model with uncertain mean and variance where we assume a Gaussian and an inverse gamma prior, the marginal posterior for the mean (variance is integrated out) is a Student's t distribution. See Rachev et al. (2008), pp. 48, for more details.

<sup>43</sup>When the joint posterior  $\pi(\theta|x, y)$  is a standard multivariate distribution, the simultaneous sampling of the  $J$  parameters is also a Gibbs sampler. Indeed, such a multivariate Gibbs sampling is often preferable as speed of convergence is improved.

often two important advantages: (1) the computational implementation is relatively easy and (2) processing speed is fast.

The implementation of the Gibbs sampler is mostly linked to the curiosity of *conjugate priors* (see Fink 1997 for more details). There are some rare but extreme important cases where the prior and posterior follow the same distribution, although the distributional parameters differ. Two points are crucial for conjugate priors: First, the correct choice of the likelihood  $L_y(\theta)$  (distributional assumption on  $y$ ).<sup>44</sup> Second, the respective parameters need special priors given the likelihood. Some important conjugate priors are summarized in table 1.

$y$	$\Theta$	$\Theta y$	posterior parameters
$\sim N(\mu, \sigma)$	$N(m_0, s_0)$	$\mu \sim N(m_*, s_*)$	$m_* = s_*^2 \left[ m_0/s_0^2 + \sigma^{-2} \sum_t^T y_t \right]$ $s_*^2 = [s_0^{-2} + T\sigma^{-2}]^{-1}$
$\sim N(\mu, \sigma)$	$IG(a_0, b_0)$	$\sigma^2 \sim IG(a_*, b_*)$	$a_* = a_0 + T/2$ $b_* = b_0 + 0.5 \sum_t^T [y_t - \mu]^2$
$\sim Exp(\lambda)$	$G(a_0, b_0)$	$\lambda \sim G(a_*, b_*)$	$a_* = a_0 + T$ $b_* = b_0 + \sum_t^T y_t$
$\sim Bin(T, \kappa)$	$Beta(a_0, b_0)$	$\kappa \sim Beta(a_*, b_*)$	$a_* = a_0 + T$ $b_* = b_0 + T - \sum_t^T 1_{[y_t=1]}$

Table 1: conjugate priors

## 7.2 Griddy Gibbs Sampler

The *Griddy Gibbs sampler* approximates a non-standard full conditional posterior on a restricted number of points. It works the following way

1. Set  $S$  fixed points  $\tilde{\theta}_j^{(s)}$ , for  $s = 1, \dots, S$ , for parameter  $\Theta_j$ , where  $\tilde{\theta}_j^{(s-1)} < \tilde{\theta}_j^{(s)} < \tilde{\theta}_j^{(s+1)}$ .
2. Compute the respective posteriors  $w_s = \pi\left(\tilde{\theta}_j^{(s)} \mid \theta_{-j}, x, y\right)$ .
3. Normalize  $w_s$  to unity by  $\sum_{s=1}^S w_s \cdot \Delta_s$ , where  $\Delta_s$  is the distance between the fixed points.
4. Approximate the posterior cdf by  $\tilde{F}_{\Theta_j|y}\left(\tilde{\theta}_j^{(l)}\right) = \sum_{s=1}^l w_s$ , for  $l = 1, \dots, S$ .
5. Generate  $u$  from  $U \sim Uni(0, 1)$ . Find that  $\tilde{F}_{\Theta_j|y}\left(\tilde{\theta}_j^{(l)}\right)$  closest to  $u$  without exceeding it and set  $\theta_j^{(w)} = \tilde{\theta}_j^{(l)}$ .

<sup>44</sup>For the following discussion we ignore latent variables  $X$ .



The Griddy Gibbs sampler is rather fast as in every sample step the proposed parameter is accepted. However, it is only a rough approximation of the "true" posterior. Rachev et al. (2008), pp. 209/ pp. 226, apply the Griddy Gibbs sampler for the degree of freedom  $\nu$  of Student's t distributed errors. In such a case the restriction to integers  $\nu \in \{2, 3, \dots, 50\}$  seems not problematic. However, in the case of the mean value or the volatility of log-returns this might not be the case.

### 7.3 Metropolis-Hastings Sampler

The advantage of Gibbs samplers is that each sampled parameter  $\theta_j^{(w)}$  is accepted as a realization of the parameter  $\Theta_j$ . This is why Gibbs samplers sample from the "true" posterior directly.<sup>45</sup> When Gibbs samplers fail, the dominant MCMC sampler is the *Metropolis-Hastings* (M-H) *sampler*. For reducing the notational demands we introduce two additional terms:

- *target distribution*: The target distribution is the distribution we want to sample from. In the MCMC case this is the respective full conditional posterior  $\tau(\theta_j) = \pi(\theta_j | \theta_{-j}, y)$ . We write the target distribution solely as a function of the uncertain parameter  $\Theta_j$  as all other parameters and observations are fixed for the pure Metropolis-Hastings sampler.
- *proposal distribution*: A proposal, or candidate-generating, distribution  $\rho(\tilde{\theta}_j | \theta_j^{(w-1)})$  is generally a standard distribution we sample from as the target distribution is too complex. In the general form the distribution of the proposal parameter  $\tilde{\Theta}_j$  depends on the current value  $\theta_j^{(w-1)}$ .

In contrast to the Gibbs sampler the Metropolis-Hastings sampler generates proposals  $\tilde{\theta}_j^{(w)}$ ,  $w = 1, \dots, W$ , which are not automatically assumed to stem from the target (posterior) distribution. In exchange the requirements for the Metropolis-Hastings sampler are very low. The target distribution can feature any complex non-standard form as long as the ratio  $\tau(a)/\tau(b)$  can be evaluated, where  $a$  and  $b$  are defined on the domain of  $\Theta_j$ .

The exact Metropolis-Hastings algorithm is the following

1. Sample a proposal value  $\tilde{\theta}_j^{(w)}$  from the (standard) proposal distribution  $\rho(\tilde{\theta}_j | \theta_j^{(w-1)})$ .
2. Set MCMC sample  $\theta_j^{(w)} = \tilde{\theta}_j^{(w)}$  with *acceptance probability*

$$\alpha\left(\theta_j^{(w-1)}, \tilde{\theta}_j^{(w)}\right) = \min\left(\frac{w\left(\tilde{\theta}_j^{(w)}, \theta_j^{(w-1)}\right)}{w\left(\theta_j^{(w-1)}, \tilde{\theta}_j^{(w)}\right)}, 1\right),$$

---

<sup>45</sup>Of course, the Griddy Gibbs sampler generates from the discretized posterior.

where the weight  $w(a, b) = \tau(a) / \rho(a|b)$  is the ratio of target and proposal distribution. Otherwise keep the old MCMC sample, i.e.  $\theta_j^{(w)} = \theta_j^{(w-1)}$ .

Thanks to the acceptance probability function  $\alpha(\cdot, \cdot)$  the MCMC sample path  $\theta_j^{(w)}$ ,  $w = 1, \dots, W$ , converges to the target distribution  $\tau(\theta_j)$ . Interestingly, the Gibbs sampler is a special case of the Metropolis-Hastings algorithm where the proposal equals the target distribution. This results in a constant acceptance probability of  $\alpha(\cdot, \cdot) = 1$  as  $w(\cdot, \cdot) = 1$ . Often, there are models where some parameters can be generated by a Gibbs sampler while others need a Metropolis-Hastings sampler. Such algorithms are termed *hybrid MCMC samplers*.

The M-H algorithm is abstract as long as the proposal distribution is not specified. In the following we concentrate on two dominant and extremely flexible concepts, independence Metropolis-Hastings and random walk Metropolis-Hastings sampler.

### 7.3.1 Independence Metropolis-Hastings Sampler

The *independence Metropolis-Hastings* (iM-H) *sampler* bases on the concept of *rejection sampling* (see appendix A.2). The target density  $\tau(\theta_j)$  is simulated by an approximating standard proposal density  $\rho(\theta_j)$  which does not depend on the last value  $\theta_j^{(w-1)}$  of the MCMC sampler. The acceptance probability reduces to

$$\alpha\left(\theta_j^{(w-1)}, \tilde{\theta}_j^{(w)}\right) = \min\left(w\left(\tilde{\theta}_j^{(w)}\right) / w\left(\theta_j^{(w-1)}\right), 1\right),$$

where  $w(\theta_j) = \tau(\theta_j) / \rho(\theta_j)$ . In contrast to rejection sampling the iM-H sampler does not require the proposal density to be an envelope of the target density, meaning  $\tau(\theta_j) \leq c\rho(\theta_j)$ , for  $c > 1$ . Nevertheless, some basic conditions can fundamentally improve the efficiency of the iM-H sampler (see figure 5): The proposal distribution should be centered at the mode of the target density and the shapes should be matched. Often, the proposal density needs to dominate the target density in the tails ( $\tau(\theta_j) \leq \rho(\theta_j)$ ) to guarantee convergence of the iM-H sampler in finite time (see also 7.4). Consequently, a Student's t proposal density may often outperform a Gaussian proposal density.

Convergence of the iM-H sampler can be slow when several parameters feature strong dependencies. This problem can be reduced by applying multivariate target densities (posteriors)  $\tau(\theta)$ , where  $\theta = (\theta_1, \dots, \theta_J)'$ . A standard choice for the proposal density  $\rho(\theta)$  is the  $J$ -dimensional multivariate Gaussian distribution  $N_J(\theta^{ML}, \Sigma)$  where  $\theta^{ML}$  are the maximum likelihood estimators and  $\Sigma = -\left[\partial^2 \ln \tau(\theta) / \partial \theta^2\right]^{-1} \Big|_{\theta^{\text{mod}}}$  is the covariance matrix evaluated at the mode  $\theta^{\text{mod}}$  of  $\tau(\cdot)$ . The acceptance probability can be improved by inflating the covariance,  $c \cdot \Sigma$ , for  $c > 1$ .<sup>46</sup>

<sup>46</sup>Geweke (1995) sets  $c = 1.2^2$  in the case of a Student's t distributed proposal density.

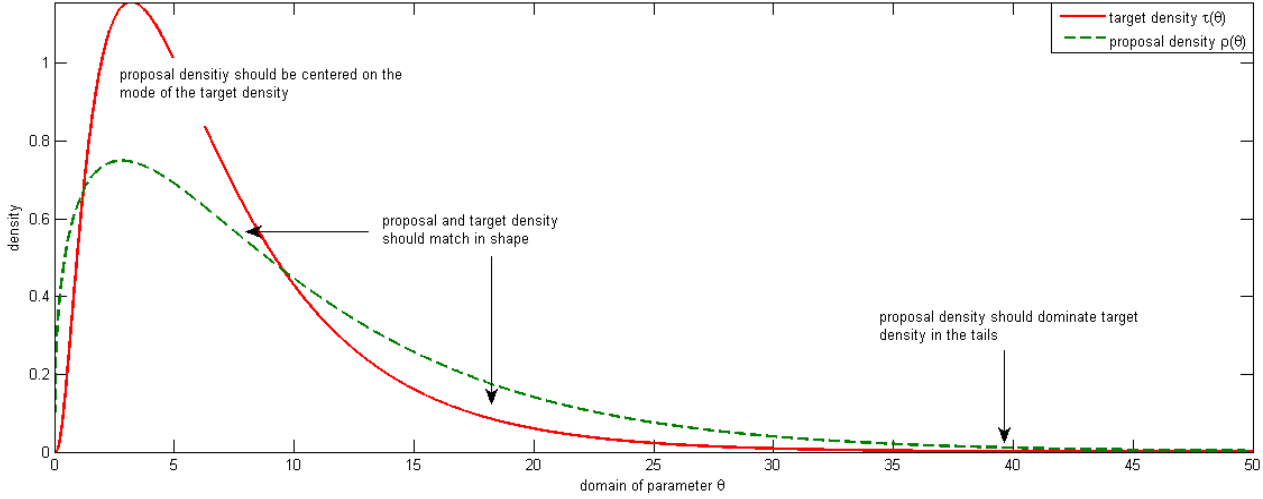


Figure 5: independence Metropolis-Hastings sampler

To produce fat tails a non-central  $J$ -dimensional multivariate Student's  $t$  distribution<sup>47</sup>

$$t_v(\theta^{ML}, \Sigma \cdot [v - 2] / v)$$

could be used instead. Again,  $\Sigma$  is the covariance matrix of the proposal parameters  $\tilde{\theta}$  and  $v$  is the degree of freedom, e.g.  $v = 5$ .

### 7.3.2 Random Walk Metropolis Sampler

The original Metropolis(-Hastings) sampler is the *random walk Metropolis (RWM) sampler* (see Metropolis et al. 1953). Its basic idea is that the proposed parameter value  $\tilde{\theta}_j^{(w)}$  is just the last MCMC sample  $\theta_j^{(w-1)}$  plus a random term  $\varepsilon^{(w)}$  with zero mean

$$\tilde{\theta}_j^{(w)} = \theta_j^{(w-1)} + \varepsilon^{(w)}.$$

This results in a symmetrical proposal density  $\rho(\tilde{\theta}_j^{(w)} | \theta_j^{(w-1)}) = \rho(\theta_j^{(w-1)} | \tilde{\theta}_j^{(w)})$  and, consequently, in the simplified acceptance probability

$$\alpha(\theta_j^{(w-1)}, \tilde{\theta}_j^{(w)}) = \min\left(\tau(\tilde{\theta}_j^{(w)}) / \tau(\theta_j^{(w-1)}), 1\right).$$

<sup>47</sup>The non-central multivariate Student's  $t$  distribution is often not implemented in statistical software. It can simply be simulated by generating samples  $z$  from a central multivariate Student's  $t$  distribution,  $Z \sim t_v(\mathbf{0}, I)$ , where  $\mathbf{0}$  is a  $J$ -dimensional vector of zeros and  $I$  is a  $J \times J$  identity matrix. Then  $\tilde{\theta} = \theta^{ML} + z \cdot \Sigma \cdot [v - 2] / v$  stems from the desired distribution.

As a rule of thumb, a suitable proposal density results in an acceptance frequency of 20% up to 50% (see Johannes & Polson 2003, p. 18/ 28, or Rachev et al. 2008, p. 70). A lower acceptance frequency indicates an extreme dominance of the proposal density in the tails of the target density, which results in a long computing time. In contrast, acceptance frequencies far above 50% could mean that the proposal density does not sample on the whole domain of the target density.

The random term is commonly modelled Gaussian, i.e.  $\varepsilon^{(w)} \sim N(0, s)$ . Generally, a higher  $s$  reduces the acceptance rate (when RWM sampler has converged). Fat tails can be produced by applying a central Student's t distribution for  $\varepsilon^{(w)}$ .

In the case of dependent parameters a multivariate RWM sampler  $\tilde{\theta}^{(w)} = \theta^{(w-1)} + \varepsilon^{(w)}$  seems preferable where  $\tilde{\theta}$  and  $\theta$  are  $J$ -dimensional vectors and  $\varepsilon^{(w)} \sim N_J(\mathbf{0}, \Sigma)$ . Similar to the multivariate iM-H sampler, the covariance matrix  $\Sigma$  can be calculated as the negative inverse of the Hessian evaluated at the mode of the target density. Again, to improve convergence  $\Sigma$  can be scaled by a factor  $c > 1$  or  $\varepsilon^{(w)}$  could be sampled from a central multivariate Student's t distribution. Roberts et al. (1997) show that under some general conditions the asymptotically "optimal" acceptance rate for multivariate RWM samplers is 23.4% (see also Gelman et al. 1996).

## 7.4 Convergence of MCMC Sampler

We have simply assumed the MCMC samplers converge to the posterior. Roughly speaking, this means the MCMC samples  $\theta_j^{(w)}$ ,  $j = 1, \dots, J$ , (and  $x_i^{(w)}$ ,  $i = 1, \dots, \dim(x)$ ) describe the (augmented) posterior  $\pi(\theta, x|y)$ . The mathematics proofing this statement is highly complicated. Hence, we only present the basic facts<sup>48</sup> and, for the sake of simplicity and without loss of generality, we ignore the latent variables  $X$ .

As the name of the MCMC sampler suggests, the generated parameter path  $\theta_j^{(w)}$ ,  $w = 1, \dots, W$ , is a Markov chain of order  $m$ . This means the MCMC distribution of  $\Theta_j^{(w)}$  does not only depend on the observations  $y$  and the other parameters  $\theta_{<j}^{(w)}$  and  $\theta_{>j}^{(w-1)}$  but also on the preceding  $m$  realizations  $\theta^{(w-1)}$  up to  $\theta^{(w-m)}$ . Hence, there is a danger that the start value  $\theta_j^{(0)}$  of the MCMC chain affects the conditional distribution  $\Upsilon_w \left( \Theta_j^{(w)} = \theta_j \mid \theta_j^{(0)} \right)$ , which we term the  $w$ -step transition probability.

Fortunately, under some mild conditions, the MCMC chain is *ergodic* as the transition distribution converges to an unique stationary distribution

$$\lim_{w \rightarrow \infty} \Upsilon_w \left( \Theta_j^{(w)} = \theta_j \mid \theta_j^{(0)} \right) = \hat{\pi}(\theta_j \mid \theta_{-j}, y).$$

---

<sup>48</sup>see Johannes & Polson (2003), pp. 19, or Winkler 2004, pp. 75, for more details

This equilibrium distribution is (1) independent of the initial parameter value  $\theta_j^{(0)}$  and (2) equals the "true" posterior  $\hat{\pi}(\theta_j | \theta_{-j}, y) = \pi(\theta_j | \theta_{-j}, y)$ .<sup>49</sup> This implies the posterior is the *invariant* distribution of the converged MCMC chain.<sup>50</sup>

Beside convergence of the whole distribution, there are two forms of convergence for real transformations  $g(\theta_j^{(w)})$ , where  $g$  is a real-valued function and  $\int g(\theta_j) \pi(\theta_j | \theta_{-j}, y) d\theta_j < \infty$ .<sup>51</sup> The *law of large numbers (ergodic averaging)* guarantees

$$\frac{1}{W} \sum_{w=1}^W g(\theta_j^{(w)}) \rightarrow E(g(\theta_j))$$

almost surely, whereas the *central limit theorem* states

$$\sqrt{W} \left[ \frac{1}{W} \sum_{w=1}^W g(\theta_j^{(w)}) - E(g(\theta_j)) \right] \rightarrow N(0, s)$$

in distribution where  $s \in \mathbb{R}_+$  and  $E(g(\theta_j)) = \int_{\theta_j} g(\theta_j) \pi(\theta_j | \theta_{-j}, y) d\theta_j$  (see Johannes & Polson 2003, pp. 23, Jones 2004, and Winkler 2004, pp. 88).

Although convergence is mostly guaranteed, the speed of convergence is of relevance. It can be measured via the inequality<sup>52</sup>

$$\left\| \Upsilon_w \left( \Theta^{(w)} = \theta | \theta^{(0)} \right) - \pi(\theta | y) \right\| \leq h(w),$$

where  $\|\cdot\|$  could be any number of norm. The dominant convergence measure is *geometric convergence*,  $h_{geom}(w) = K \cdot \lambda^w$ , where  $K < \infty$  and  $\lambda > 1$ . The MCMC samplers often feature geometric convergence.<sup>53</sup> However, geometric convergence does not guarantee convergence in finite time. A stronger concept is *polynomial convergence*,<sup>54</sup> which poses more restrictions on the MCMC sampler.<sup>55</sup>

<sup>49</sup>See Johannes & Polson (2003), pp. 21, for a basic but intuitive proof. For more details see Winkler (2004), pp. 75, Roberts & Tweedie (1996), or Mengersen & Tweedie (1996).

<sup>50</sup>Invariance means  $\pi(\Theta_j^{(w)} = b | \cdot, y) = \int_a \Upsilon_1(b|a) \pi(\Theta_j^{(w-1)} = a | \cdot, y) da$  where  $\Upsilon_1(b|a)$  is the one-step transition probability.

<sup>51</sup>e.g.,  $g_1(\theta_j) = \theta_j$  or  $g_2(\theta_j) = [\theta_j - g_1(\theta_j)]^2$

<sup>52</sup>see Johannes & Polson (2003), pp. 23, or Roberts & Tweedie (1996), p. 99, and Mengersen & Tweedie, pp. 103/ p. 105, for a similar formulation

<sup>53</sup>E.g., Mengersen & Tweedie (1996) show that geometric convergence can be guaranteed (1) for a random-walk Metropolis-Hastings sampler when "true" posterior has geometric tails or (2) for a independence Metropolis-Hastings sampler when the proposal distribution dominates the posterior in the tails.

<sup>54</sup>See Jarner & Roberts (2002) and Moulines et al. (2001), p. 13, who also discuss sub-exponential convergence.

<sup>55</sup>E.g., MCMC samplers with log-concave proposal distributions feature a polynomial convergence (Frieze et al. 1994).

There are several concepts to analyze the convergence of a sampled parameter path  $\theta_j^{(w)}$ ,  $w = 1, \dots, W$ . A basic but widespread approach is a simple *visual inspection* of the path by  $J$  trace plots on  $\left\{w, \theta_j^{(w)}\right\}_{w=1, \dots, W}$ . When the statistician believes that the parameter path becomes stationary (no fundamental trend in the trace plot), he or she discards the  $B$  samples  $\theta_j^{(w)}$ ,  $w \leq B$ , that do not seem to feature stationarity.

The optical scan can be improved (1) by running  $L$  parameter paths starting from different starting points  $\theta_{jl}^{(0)}$ ,  $l = 1, \dots, L$ , and (2) by plotting the  $L$  parameter paths in one trace plot. When, after a certain sample step ( $w = B + 1$ ), there is no structural difference between the different parameter paths, all samples  $\theta_{jl}^{(w)}$ , for  $w > B$  and  $l = 1, \dots, L$ , can be used for inference.

Gelman & Rubin (1992) propose a concept which reduce the information of  $L$  chains, with an after-burn-in length of  $W - B$ , to the *potential scale reduction factor*  $\mathcal{F}_j^{(GR)}$ . It requires to calculate the within-chain-variance

$$\hat{\sigma}_W^2 = \frac{1}{L \cdot [W - B]} \sum_{l=1}^L \sum_{w=B+1}^W \left[ \theta_{jl}^{(w)} - \bar{\theta}_{jl} \right]^2$$

where  $\bar{\theta}_{jl} = [W - B]^{-1} \sum_{w=B+1}^W \theta_{jl}^{(w)}$ , and the between-chain-variance

$$\hat{\sigma}_B^2 = \frac{W - B}{L - 1} \sum_{l=1}^L \left[ \bar{\theta}_{jl} - \bar{\theta}_j \right]^2,$$

for  $\bar{\theta}_j = L^{-1} \sum_{l=1}^L \bar{\theta}_{jl}$ . The posterior variance of  $\Theta_j$  is the weighted average

$$\widehat{Var}(\Theta_j) = \left[ 1 - \frac{1}{W - B} \right] \hat{\sigma}_W^2 + \frac{1}{W - B} \hat{\sigma}_B^2,$$

which is an unbiased estimator. The idea is that when the  $L$  chains are started from far-apart initial values the between-chain-variance dominates the within-chain-variance. After convergence the between-chain-variance should be negligible and consequently the potential scale reduction factor  $\mathcal{F}_j^{(GR)} \approx \sqrt{\widehat{Var}(\Theta_j) / \hat{\sigma}_W^2} \approx 1$ . Commonly, the MCMC sampler is assumed to be converged when  $\mathcal{F}_j^{(GR)} < \alpha$ , where  $\alpha$  is 1.2 or 1.1. In the case of  $J$  parameters the MCMC sampler has reached convergence when  $\max_{j=1}^J \mathcal{F}_j^{(GR)} < \alpha$ .

Another simple graphical tool is the *cumsum convergence criterion* (see Bauwens & Lubrano 1998, pp. 31)

$$CS_j(\varpi) = \left[ \frac{1}{\varpi} \sum_{w=1}^{\varpi} \theta_j^{(w)} - \hat{\mu}(\theta_j) \right] / \hat{\sigma}(\theta_j), \quad \varpi = 50, \dots, W,$$

where  $\hat{\mu}(\theta_j)$  and  $\hat{\sigma}^2(\theta_j)$  are the empirical mean and variance of the complete parameter path  $\theta_j^{(w)}$ ,  $w = 1, \dots, W$ . The interpretation of  $CS_j(\varpi) = \alpha$  is that the difference between the mean of the full sample path and the sample path up to  $\varpi$  is  $\alpha \cdot 100\%$  of the standard deviation of the full path. Bauwens & Lubrano (1998) assume convergence for the path when  $CS_j(\varpi)$  converges smoothly to zero and  $|CS_j(\varpi)|$  is below 0.05 or 0.1.

An important test for convergence of the parameter path  $\theta_j^{(w)}$ ,  $w = 1, \dots, W$ , is the *Geweke Z-scores test* (Geweke 1992). The Geweke Z-score  $\mathcal{Z}_j$  is the standardized difference of the sub-sample means  $\bar{\theta}_j^{(1)} = w_{(1)}^{-1} \sum_{w=1}^{w_{(1)}} \theta_j^{(w)}$  and  $\bar{\theta}_j^{(2)} = w_{(2)}^{-1} \sum_{w=W-w_{(2)}+1}^W \theta_j^{(w)}$ , where  $w_{(1)} < W - w_{(2)} + 1 < W$ . When the MCMC chain reached convergence (the null hypothesis) then  $\mathcal{Z}_j \sim N(0, 1)$ . As MCMC paths generally feature auto-correlation, standardization needs a correction. To eliminate auto-correlation Bonamente et al. (2004) propose to apply Geweke's Z-score on the *average* parameter samples  $\tilde{\theta}_j^{(m)}$  for  $m = 1, \dots, [W - B] / \varsigma$ , where  $\varsigma$  is the step width.

The *Heidelberger-Welch diagnostic* is a multiple-step test (see Heidelberger & Welch 1983). First, a Cramer-von Mises test is run for the full MCMC parameter chain  $\theta_j^{(w)}$ ,  $w = 1, \dots, W$ . The null hypothesis is whether all samples come from the same distribution. When this is rejected the first 10% of all samples are dumped and the Cramer-von Mises test is run again. This is repeated until the test could not be rejected or 50% of the initial samples has been dumped. In the next step, the samples which have passed the test are used to calculate their mean  $\bar{\theta}_j$  and a credible interval  $CR_j^{(1-\alpha)}$  for  $\bar{\theta}_j$ . The MCMC path is assumed to have converged when the ratio of half the width of  $CR_j^{(1-\alpha)}$  and  $\bar{\theta}_j$  is lower than a pre-defined  $\varepsilon$ , e.g.  $\varepsilon = 0.1$ .

It is important to keep in mind that convergence and auto-correlation of the MCMC path  $\theta_j^{(w)}$ ,  $w = 1, \dots, W$ , does not mean the same. Often, the path can be assumed to be converged however auto-correlations  $AC_j(l) = AC(\theta_j^{(w)}, \theta_j^{(w-l)})$  for  $l = 1, 2, \dots$ , often fall slowly. This can be visualized by *auto-correlation plots*. As a rule of thumb: the higher the auto-correlation the less information is captured in the chain. A converged chain with  $W - B = 1.000$  samples and nearly no auto-correlation could better approximate the true posterior than a chain with  $W - B = 10.000$  samples but large auto-correlation. Consequently, an efficient MCMC algorithm is not only defined by the computing time for a fixed number of repetitions but also by other criteria like auto-correlation. However, the concept of auto-correlation should always be taken with a pinch of salt. It is not correct to conclude that a MCMC path with nearly no auto-correlation must have converged.<sup>56</sup>

---

<sup>56</sup>See the discussion on the witch's hat distribution, broad flat brim and a centered sharp peak. The MCMC sampler is geometric convergent,  $K \cdot \lambda^w$ , but  $\lambda$  is close to one. There is practically no chance for the sampler to move from the brim to the center in finite computing time even with low auto-correlation (Johannes & Polson 2003, p. 24, and Geyer 1992).

## 8 Model Weighting

In contrast to frequentist statistics, Bayesian statistics concentrates on model weighting rather than model rejecting and accepting. First, we present different approaches to calculate commonly used Bayes factors and posterior model probabilities. Then we describe how information criteria can be used to weight models.

### 8.1 Posterior Model Probability/ Bayes Factor

The MCMC approach offers an efficient parameter estimation algorithm. Besides this Bayesian statistics allows for a direct weighting of models  $\mathcal{M} = 1, \dots, K$ , via the *posterior model probability (posterior odds ratio)*

$$\begin{aligned} \pi(\mathcal{M}_k|y) &= \frac{f_Y(y|\mathcal{M}_k)\pi(\mathcal{M}_k)}{\sum_{l=1}^K f_Y(y|\mathcal{M}_l)\pi(\mathcal{M}_l)} \\ &= \frac{\int f_Y(y|\theta, \mathcal{M}_k)\pi(\theta|\mathcal{M}_k)d\theta \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K \int f_Y(y|\theta, \mathcal{M}_l)\pi(\theta|\mathcal{M}_l)d\theta \cdot \pi(\mathcal{M}_l)}, \end{aligned} \quad (1)$$

where  $\mathcal{M}_k$  is short for  $\mathcal{M} = k$ , and

- $\pi(\mathcal{M}_k)$  ... prior probability of model  $k$
- $\pi(\theta|\mathcal{M}_k)$  ... prior of the parameter vector  $\theta = (\theta_1, \dots, \theta_{\dim(\theta)})'$  given model  $k$
- $f_Y(y|\theta, \mathcal{M}_k)$  ... conditional joint density given model  $k$
- $f_Y(y|\mathcal{M}_k)$  ... unconditional joint density given model  $k$ .

Alternatively, and more common,  $K$  models can be compared via  $K[K-1]/2$  *Bayes factors*

$$BF_{kl} = \frac{f_Y(y|\mathcal{M}_k)}{f_Y(y|\mathcal{M}_l)} = \frac{\int f_Y(y|\theta, \mathcal{M}_k)\pi(\theta|\mathcal{M}_k)d\theta}{\int f_Y(y|\theta, \mathcal{M}_l)\pi(\theta|\mathcal{M}_l)d\theta}. \quad (2)$$

In the case of two competing models the relationship between posterior model probability and Bayes factor is simply (Kass & Raftery 1995, p. 776)

$$\text{posterior model probability} = \text{Bayes factor} \times \text{prior odds}.$$

Raftery (1996), p. 252/ table 1, suggests an interpretation scale for the Bayes factors (see table 2).<sup>57</sup>

---

<sup>57</sup>The expression  $2 \cdot \ln BF_{kl}$  converts the Bayes factor to the same scale as common deviance and likelihood ratio tests (Spiegelhalter et al. 2002).



$BF_{kl}$	$2 \cdot \ln BF_{kl}$	evidence for model $M_k$
$< 1$	$< 0$	negative
$1 - 3$	$0 - 2.2$	not worth more than a mention
$3 - 20$	$2.2 - 6$	positive
$20 - 150$	$6 - 10$	strong
$> 150$	$> 10$	very strong

Table 2: Bayes factors

Unfortunately, the Bayesian advantage of model comparison or weighting via posterior model probabilities or Bayes factors is computational extremely demanding. The integrals in equations 1 and 2 are only analytically tractable in a restricted number of exceptions (Dellaportas et al. 2002, p. 27).

A basic methodology to calculate the model probability and the Bayes factors is the *Laplace method* (see Tierney & Kadane 1986 or Kass & Raftery 1995, pp. 777). It approximates  $I(\theta) = f_Y(y|\theta, \mathcal{M}_k) \pi(\theta|\mathcal{M}_k)$  by a multivariate Gaussian distribution  $\phi_{\dim(\theta)}(y|\mu_\theta, [-H]^{-1})$  with  $\mu_\theta = \arg \max_\theta I(\theta)$  as mean vector and  $H$  as the respective Hessian matrix. Integrating the Gaussian approximation leads to  $[2\pi]^{\dim(\theta)/2} |-H|^{-1/2} I(\mu_\theta)$  which is hoped to be a good approximation to  $\int I(\theta) d\theta$ . The Laplace method is only sensible for  $I(\theta)$  highly peaking about their maximum and for low-dimensional problems (Kass & Raftery 1995, pp. 777).

### 8.1.1 Sampling from the Joint Posterior

In Bayesian statistics the model  $\mathcal{M} \in \{1, \dots, K\}$  can be understood as uncertain variable where the posterior probability of model  $k$

$$\pi(\mathcal{M}_k|y) \propto f_Y(y|\mathcal{M}_k) \pi(\mathcal{M}_k), \text{ for } \sum_{k=1}^K \pi(\mathcal{M}_k|y) = 1,$$

is described by the relative frequency

$$\pi(\mathcal{M}_k|y) \approx \frac{1}{W} \sum_{w=1}^W 1_{[\mathcal{M}^{(w)}=k]}.$$

the variable  $\mathcal{M}$  realizes as  $k$  in a MCMC sampler.

In the Bayesian way of thinking we need to account for parameter uncertainty. Assuming that model  $k$  is only determined by the parameter vector  $\theta_k$ , for  $\theta = (\theta_1, \dots, \theta_{k-1}, \theta_k, \theta_{k+1}, \dots, \theta_K)'$ , the posterior model probability is the solution of a complicated integral over the joint pos-

teriors<sup>58</sup>

$$\pi(\mathcal{M}_k | y) = \int \pi(\theta_k, \mathcal{M}_k | y) d\theta_k. \quad (3)$$

Research has concentrated on MCMC based simulations that approximate

$$\pi(\theta_k, \mathcal{M}_k | y) \propto f_Y(y | \theta_k, \mathcal{M}_k) \cdot \pi(\theta_k | \mathcal{M}_k) \cdot \pi(\mathcal{M}_k).$$

The joint posterior  $\pi(\theta_k, \mathcal{M}_k | y)$  is mostly a complex multivariate density with a dimension depending on  $K$ , the number of models, and  $D_k = \dim(\theta_k)$ , the dimension of the parameter vector  $\theta_k$ . Hence, there are several approaches to approximate  $\pi(\theta_k, \mathcal{M}_k | y)$  like (see Dellaportas et al. 2002, pp. 27, Robert & Marin, pp. 1, or Chen et al. 2008):

- *independence sampler*: The density  $\pi(\theta_k, \mathcal{M}_k | y)$  is sampled by the independence sampler (Tierney 1994). It uses a multivariate proposal density  $\rho(\theta_k, \mathcal{M}_k)$  which is independent from the last sampled values  $\mathcal{M}_k^{(w-1)}$  and  $\theta_k^{(w-1)}$  with  $w$  as the current step of the sampler. The independence sampler is efficient as long as the proposal density is a reasonable approximation to  $\pi(\theta_k, \mathcal{M}_k | y)$  which is only feasible for low dimensional problems (Dellaportas et al. 2002, p. 28).
- *Metropolis-Hastings sampler*: Because of the complexity of  $\pi(\theta_k, \mathcal{M}_k | y)$  a proposal density,

$$\rho(\mathcal{M}_k, \theta_k | \mathcal{M}_k^{(w-1)}, \theta_k^{(w-1)}),$$

depending on the current values  $\mathcal{M}_k^{(w-1)}$  and  $\theta_k^{(w-1)}$ , is more favorable. However, the Metropolis-Hastings acceptance probability cannot be calculated for models with different parameter dimensions (Dellaportas et al. 2002, p. 28).

- *reversible jump*: A modification the samplers above is the reversible jump technique of Green (1995) which first samples the model  $\mathcal{M}_k$  and then applies a proposal density with the dimension of the respective parameter vector  $\theta_k$ , only.<sup>59</sup>

---

<sup>58</sup>Indeed, the integral

$$\begin{aligned} \pi(\mathcal{M}_k | y) &\propto \int f_Y(y | \theta, \mathcal{M}_k) \pi(\theta | \mathcal{M}_k) d\theta \cdot \pi(\mathcal{M}_k) \\ &= \int f_Y(y | \theta_k, \mathcal{M}_k) \pi(\theta_k | \mathcal{M}_k) \cdot \pi(\mathcal{M}_k) d\theta_k \cdot \int \prod_{l \neq k} \varphi_k(\theta_l) d\theta_{l \neq k} \\ &\propto \int \pi(\theta_k, \mathcal{M}_k | y) d\theta_k \cdot 1 \end{aligned}$$

accounts for linking densities  $\varphi_k(\theta_l)$ . See section 8.1.2 for more details.

<sup>59</sup>see section 8.1.2 for a better understanding of the basic concept

### 8.1.2 Linking Density Approach

As aforementioned, the posterior model probability  $\pi(\mathcal{M}_k|y) = \int \pi(\theta, \mathcal{M}_k|y) d\theta$  is often the solution of an integral on the joint posterior  $\pi(\theta, \mathcal{M}_k|y)$  which is too complex to sample from directly (Gibbs sampler) or indirectly (Metropolis-Hastings sampler). Fortunately, the Clifford-Hammersley theorem guarantees that  $\pi(\theta, \mathcal{M}_k|y)$  can be described by sampling in succession from  $\pi(\mathcal{M}_k|\theta, y)$  and  $\pi(\theta|y, \mathcal{M}_k)$ . Consequently, you are allowed to sample  $\mathcal{M}|y$ , with  $\mathcal{M} \in \{1, \dots, K\}$ . However, this approach requires to fully describe  $\pi(\theta, \mathcal{M}_k|y)$  instead of  $\pi(\theta_k, \mathcal{M}_k|y)$ , only.

Assume you have generated  $\mathcal{M}_k$ , meaning  $\mathcal{M} = k$ , from  $\pi(\mathcal{M}_k|\theta, y)$ . In the next step you need to sample from  $\theta|y, \mathcal{M}_k$ . Again applying Clifford-Hammersley, this can be described by sampling from  $\pi(\theta_k|\theta_{-k}, y, \mathcal{M}_k)$  and  $\pi(\theta_{-k}|\theta_k, y, \mathcal{M}_k)$ , where  $\theta_{-k}$  is  $\theta$  without  $\theta_k$ . Of course,  $\pi(\theta_k|\theta_{-k}, y, \mathcal{M}_k)$  is equivalent to  $\pi(\theta_k|y, \mathcal{M}_k)$ . Nevertheless, a formulation for  $\pi(\theta_{-k}|\theta_k, y, \mathcal{M}_k)$  is needed which only requires that it is a *proper* density.

The posterior model probabilities can be approximated by Monte Carlo techniques

$$\pi(\mathcal{M}_k|y) = \int \pi(\mathcal{M}_k|\theta_k, y) \pi(\theta|y) d\theta \approx \frac{1}{W} \sum_{w=1}^W \pi(\mathcal{M}_k|\theta^{(w)}, y)$$

The equation above is the same as in equation 3 if the likelihood function under model  $k$  only depends on the parameters  $\theta_k$ . However the simulation approach bears a difficulty. The full parameter vector  $\theta^{(w)}$  needs to be sampled from the joint posterior  $\pi(\theta|y)$  which features dependency between  $\theta_k$  and  $\theta_{-k}$  (Robert & Marin 2008). The joint posterior can be formulized by

$$\begin{aligned} \pi(\theta|y) &= \sum_{k=1}^K \pi(\theta, \mathcal{M}_k|y) \\ &\propto \sum_{k=1}^K f_Y(y|\theta_k, \mathcal{M}_k) \cdot \prod_{l=1}^K \pi(\theta_l|\mathcal{M}_k) \cdot \pi(\mathcal{M}_k) \\ &= \sum_{k=1}^K f_Y(y|\theta_k, \mathcal{M}_k) \pi_k(\theta_k|\mathcal{M}_k) \cdot \pi(\mathcal{M}_k) \cdot \prod_{l \neq k} \varphi_k(\theta_l), \end{aligned}$$

where  $\varphi_k(\theta_l)$  are so called proper *linking densities* (*pseudo-priors*) for parameters outside model  $k$  (Carlin & Chip 1995, pp. 475) and  $\int \prod_{l \neq k} \varphi_k(\theta_l) d\theta_{l \neq k} = 1$ .<sup>60</sup> As the joint posterior is too complex Clifford-Hammersley can be applied and the multi-model MCMC algorithm of Carlin & Chip (1995), pp. 475, can be used:

---

<sup>60</sup>Linking densities are not an intuitive concept but rather a technical necessity. Nevertheless, it can be understood as harmonization of model uncertainties. Assume you have two models  $\mathcal{M}_1$  and  $\mathcal{M}_2$  with  $J_1 = 1$  and  $J_2 = 2$  parameters. As long as we separately estimate both models we only need to account for the

1. Sample all model specific parameter vectors  $\theta_k$ , for  $k = 1, \dots, K$ , from

$$\pi(\theta_k | \theta_{-k}, y, \mathcal{M}) \propto \begin{cases} f_Y(y | \theta_k, \mathcal{M}_k) \cdot \pi(\theta_k | \mathcal{M}_k), & \mathcal{M} = k \\ \varphi_{\mathcal{M} \neq k}(\theta_k), & \mathcal{M} \neq k \end{cases}.$$

2. Sample a model  $\mathcal{M} \in \{1, \dots, K\}$  from

$$\pi(\mathcal{M}_k | \theta, y) = \frac{\vartheta_k(\theta_k | y) \cdot \prod_{l \neq k} \varphi_k(\theta_l) \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K \vartheta_l(\theta_l | y) \cdot \prod_{l \neq k} \varphi_l(\theta_s) \cdot \pi(\mathcal{M}_l)},$$

with the kernel  $\vartheta_k(\theta_k | y) = f_Y(y | \theta_k, \mathcal{M}_k) \cdot \pi(\theta_k | \mathcal{M}_k)$ . The posterior model probability is simply the relative frequency of  $\mathcal{M} = k$  that has been sampled.

The crucial point in this algorithm is the choice of the linking densities. Although the shape of the linking densities  $\varphi_k(\theta_l)$  is not restricted as long as they are still densities, an efficient implementation of this method requires linking densities to be close to the conditional posteriors  $\pi(\theta_l | \theta_{-l}, y, \mathcal{M}_l)$  (Dellaportas et al. 2002, p. 29). Nevertheless, the simultaneous MCMC estimation of  $K$  competing models leads to a very complex algorithm. It would be desirable (1) to separately run MCMC estimation on each of the  $K$  models and then (2) to combine the results.

Scott (2002) and Congdon (2006) propose methodologies where the conditional posterior of model  $k$  solely depends on the MCMC samples of  $\theta_k$ , which is computationally appealing. This implicitly assumes independence between the model parameter vectors  $\theta_k$ ,  $k = 1, \dots, K$ . Such an approximation

$$\pi(\theta | y) = \sum_{k=1}^K \pi(\theta, \mathcal{M}_k | y) \stackrel{?}{\approx} \prod_{k=1}^K \pi(\theta_k | y, \mathcal{M}_k)$$

may not achieve good results.

Based on some empirical analyses, Robert & Marin (2008) found that the proposals of Scott (2002) and Congdon (2006) feature some bias although the bias of Congdon (2006) is far less and the approximation is often very close or at least in the same magnitude as the "true" model posterior. Hence we ignore the approach of Scott (2002) and introduce that of

---

parameter uncertainty of one, respectively two, parameters. However, when we compare both models we need to include the uncertainty on the two parameters of model  $\mathcal{M}_2$  in the evaluation of model  $\mathcal{M}_1$  (and vice versa). This is why we are not allowed to ignore uncertainty. In Bayesian model weighting, we should not discriminate model  $\mathcal{M}_2$  because it acknowledges uncertainty on two parameters in contrast to one parameter in model  $\mathcal{M}_1$ . In fact, under model  $\mathcal{M}_1$  we are not sure about the two parameters of model  $\mathcal{M}_2$ . We just ignore them!

Congdon (2006). Finally, we present the improved algorithm of Congdon (2007) which still assumes independence but defines proper instead of improper linking densities.

### Marginal approaches with improper linking densities (Congdon06)

The idea of the Congdon (2006) model (*Congdon06*) is to run MCMC estimations separately on all  $K$  competing models and then to calculate the model posterior by

$$\pi(\mathcal{M}_k|y) = \int \pi(\mathcal{M}_k|\theta, y) \pi(\theta|y) d\theta \approx \frac{1}{W} \sum_{w=1}^W \pi(\mathcal{M}_k|\theta_k^{(w)}, y),$$

assuming improper linking densities  $\varphi_k(\theta_l) \propto 1$ , for  $l \neq k$ , and

$$\pi(\mathcal{M}_k|\theta_k^{(w)}, y) = \frac{\vartheta_k(\theta_k^{(w)}|y) \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K \vartheta_l(\theta_l^{(w)}|y) \cdot \pi(\mathcal{M}_l)}.$$

Again,  $\vartheta_k(\theta_k^{(w)}|y) = f_Y(y|\theta_k^{(w)}, \mathcal{M}_k) \cdot \pi(\theta_k^{(w)}|\mathcal{M}_k)$  stands for the posterior kernel of model  $k$  in MCMC step  $w \in \{1, \dots, W\}$ .

So far, the approach of Congdon (2006) has already been applied by Chen et al. 2008. However, it should be kept in mind that it features a bias which is however mostly modest (Robert & Marin 2008). Nevertheless, Congdon06 is a trade off between simplicity and the danger of a misspecification.

### Marginal approach with diffuse but proper linking densities (Congdon07)

Congdon (2007) presents an extension (*Congdon07*) of *Congdon06* which avoids diffuse linking densities. The posterior model probability is approximated by

$$\pi(\mathcal{M}_k|y) = \int \pi(\mathcal{M}_k|\theta, y) \pi(\theta|y) d\theta \approx \frac{1}{W} \sum_{w=1}^W \pi(\mathcal{M}_k|\theta^{(w)}, y),$$

for

$$\pi(\mathcal{M}_k|\theta^{(w)}, y) = \frac{\vartheta_k(\theta_k^{(w)}|y) \cdot \prod_{l \neq k} \varphi_k(\theta_l^{(w)}) \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K \vartheta_l(\theta_l^{(w)}|y) \cdot \prod_{s \neq l} \varphi_l(\theta_s^{(w)}) \cdot \pi(\mathcal{M}_l)},$$

and  $\vartheta_k(\theta_k^{(w)}|y) = f_Y(y|\theta_k^{(w)}, \mathcal{M}_k) \cdot \pi(\theta_k^{(w)}|\mathcal{M}_k)$ .

As we can arbitrarily define the linking densities as long as they are proper, Congdon (2007), p. 146, assumes *diffuse* homogeneous linking densities for each parameter

$$\tilde{\varphi}_{kd} = \varphi_1(\theta_{kd}) = \dots = \varphi_{k-1}(\theta_{kd}) = \varphi_{k+1}(\theta_{kd}) = \dots = \varphi_K(\theta_{kd}),$$

for  $k = 1, \dots, K$ ,  $d = 1, \dots, D_k$ , and  $D_k = \dim(\theta_k)$ .

Multiplying  $\pi(\mathcal{M}_k | \theta^{(w)}, y)$  by  $\tilde{\varphi}/\tilde{\varphi}$ , for  $\tilde{\varphi} = \prod_{i=1}^K \tilde{\varphi}_i = \prod_{i=1}^K \prod_{d=1}^{D_i} \tilde{\varphi}_{id}$ , results in<sup>61</sup>

$$\pi(\mathcal{M}_k | \theta^{(w)}, y) = \frac{H_k^{(w)} \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K H_l^{(w)} \cdot \pi(\mathcal{M}_l)},$$

where  $H_k^{(w)} = \vartheta_k(\theta_k^{(w)} | y) / \prod_{d=1}^{D_k} \tilde{\varphi}_{kd}^{(w)}$ ,  $\tilde{\varphi}_{kd}^{(w)} = \tilde{\varphi}_{kd}(\theta_{kd}^{(w)})$ , and  $\theta_{kd}^{(w)}$  is sampled from the distribution described by the pdf  $\tilde{\varphi}_{kd}$ .

For an optimal convergence of the multi-model MCMC algorithm of Carlin & Chib (1995), pp. 475, the linking densities need to be close to the conditional posteriors of the respective parameters. This we try to approximate by calibrating (truncated) Gaussian distributions to the parameter paths generated by our MCMC samplers. To reduce the variance of  $H_k$  and  $\pi(\mathcal{M}_k | \theta, y)$ , the tails of the linking densities should be heavy tailed relative to the respective parameter posteriors.<sup>62</sup> Therefore, we scale-up the standard deviations of the MCMC parameter paths by a factor  $c > 1$ .

In our separate MCMC implementation (1) we store the respective  $W$  kernel values resulting from the sampled parameter vectors  $\theta_k^{(w)}$  for each model  $k = 1, \dots, K$  and (2) we generate  $W$  samples from the respective linking densities  $\tilde{\varphi}_{ld}$  for the  $\sum_{l=1}^K D_l$  linking parameters  $\theta_{ld}^{(w)}$ . For computational reasons we store log-values and transform them later

$$\pi(\mathcal{M}_k | \theta^{(w)}, y) = \frac{\exp(c_k)}{\sum_{l=1}^K \exp(c_l)} = \frac{1}{1 + \sum_{l \neq k} \exp(c_l - c_k)},$$

where  $c_k = \ln H_k^{(w)} + \ln \pi(\mathcal{M}_k)$ .

### Marginal approach with diffuse but proper priors for latent variables

In the following sections we calibrate different stochastic processes. Some of these processes feature latent (unobservable) variables  $X = (X_1, \dots, X_T)'$  which need to be integrated out

---

<sup>61</sup>The formulation of Scott (2002) is

$$\pi(\mathcal{M}_k | y) = \frac{\frac{1}{W} \sum_{w=1}^W f_Y(y | \theta^{(w)}, \mathcal{M}_k) \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K \frac{1}{W} \sum_{\tilde{w}=1}^{\tilde{W}} f_Y(y | \theta^{(\tilde{w})}, \mathcal{M}_l) \cdot \pi(\mathcal{M}_l)},$$

which is the special case of Congdon07 where the linking densities equal the respective priors. This approach has already been applied from other authors, e.g. Bunnin et al. (2002) and Pollard (2007).

<sup>62</sup>Heavy tailed Student's t linking densities ( $\nu \approx 2$ ) do not always minimize the variation (Congdon 2007, p. 152).

beside the parameter vector  $\theta$ . The posterior model probability complicates to

$$\begin{aligned}\pi(\mathcal{M}_k|y) &= \int \pi(\mathcal{M}_k|\theta, x, y) \pi(\theta, x|y) d(\theta, x)' \\ &\approx \frac{1}{W} \sum_{w=1}^W \pi(\mathcal{M}_k|\theta^{(w)}, x^{(w)}, y),\end{aligned}\tag{4}$$

where  $\theta$  and  $x$  are sampled from the joint posterior  $\pi(\theta, x|y)$ . Consequently, the conditional posterior model probability is

$$\pi(\mathcal{M}_k|\theta^{(w)}, x^{(w)}, y) = \frac{H_k^{(w)} \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K H_l^{(w)} \cdot \pi(\mathcal{M}_l)},$$

where  $H_k^{(w)} = \vartheta_k(\theta^{(w)}|x^{(w)}, y) / \tilde{\varphi}_k$  with  $\tilde{\varphi}_k = \prod_{d=1}^{D_k} \tilde{\varphi}_{kd}$ . The kernel is defined by

$$\begin{aligned}\vartheta_k(\theta^{(w)}|x^{(w)}, y) &= f_{XY}(x_{(k)}^{(w)}, y|\theta_k^{(w)}, \mathcal{M}_k) \cdot \pi(\theta_k^{(w)}|\mathcal{M}_k) \\ &= f_Y(y|\theta_k^{(w)}, x_{(k)}^{(w)}, \mathcal{M}_k) \cdot f_X(x_{(k)}^{(w)}|\theta_k^{(w)}, \mathcal{M}_k) \cdot \pi(\theta_k^{(w)}|\mathcal{M}_k) \\ &= f_Y(y|\theta_k^{(w)}, x_{(k)}^{(w)}, \mathcal{M}_k) \cdot \pi(\theta_k^{(w)}|\mathcal{M}_k)\end{aligned}$$

as we assume - in contrast to the parameters - that we know the latent variables  $x = x^{(w)}$  with certainty.<sup>63</sup> The estimation error for the latent variables  $X$  is reflected by the samples  $x^{(w)}$ ,  $w = 1, \dots, W$ , from the MCMC sampler (see equation 4).

## 8.2 AIC & BIC Model Weights

The posterior model probabilities and the Bayes factors are nice features of the Bayesian statistics as they allow a direct statement how better or worse a model  $k$  is compared to a model  $l$ . However this concept does not account for our desire for parsimonious models with a low number of parameters.<sup>64</sup> For instance, assume that we want to compare the Geometric Brownian Motion with Jumps (GBMJ) and its nested model, the standard Geometric Brownian motion (GBM), for a series of log-returns. The GBMJ has more flexibility to adjust to the data but entails some parameters more to estimate.<sup>65</sup>

<sup>63</sup>Here, we argue that the latent variables  $X$  are no uncertain parameters. Our MCMC algorithm simulates  $x^{(w)}$ ,  $w = 1, \dots, W$ , and we accept them as estimates for the unobservable  $X$ . Indeed, we state the probability for the model given the sampled  $x^{(w)}$  as we assume they are the correct values. Consequently, we ignore the likelihood under the model that  $x^{(w)}$  realizes.

<sup>64</sup>Of course, this opposes the intuitive idea of linking densities. However, the concepts presented in this section reward parsimonious models according to clear and well-thought-out rules. This we think is reasonable.

<sup>65</sup>Here, we ignore the larger uncertainty under the GBMJ because it is defined by three more parameters.

To meet these concerns we apply the *Akaike's information criterion* (AIC) and the *Bayesian information criterion* (BIC, Congdon 2007, p. 149)

$$\begin{aligned} AIC_k^{(w)} &= \ln f_Y \left( y | \theta_k^{(w)}, \mathcal{M}_k \right) - D_k \\ BIC_k^{(w)} &= \ln f_Y \left( y | \theta_k^{(w)}, \mathcal{M}_k \right) - 0.5 \cdot D_k \cdot \ln T, \end{aligned}$$

for  $k = 1, \dots, K$ ,  $D_k = \dim(\theta_k)$  as the number of parameters in model  $k$ , and  $T$  as the number of observations. We present the (arithmetic) mean Akaike as well as Bayes information criteria  $\overline{AIC}_k$  and  $\overline{BIC}_k$  for a pure ordinal model comparison. Additionally, we calculate the common *mean deviance*  $MD_k = -2 \sum_{w=1}^W \ln f_Y \left( y | \theta^{(w)}, \mathcal{M}_k \right) / W$  where lower deviance indicates a better fit (Pollard Aug. 2007).

Since we are interested in cardinal performance measures we transform the ordinal AIC and BIC to model probabilities (Congdon 2007, p. 149)

$$\begin{aligned} MP_{AIC}^{(w)}(k) &= \frac{1}{1 + \sum_{l \neq k}^K \exp \left( AIC_l^{(w)} - AIC_k^{(w)} \right)} \\ MP_{BIC}^{(w)}(k) &= \frac{1}{1 + \sum_{l \neq k}^K \exp \left( BIC_l^{(w)} - BIC_k^{(w)} \right)} \end{aligned}$$

with  $\overline{MP}_{AIC}(k)$  and  $\overline{MP}_{BIC}(k)$  as their empirical means.

## 9 EU ETS

### 9.1 Overview

To fight Climate Change the Kyoto Protocol was signed in 1997 by 175 countries.<sup>66</sup> Industrial (annex I) countries agreed to reduce their green house gas (GHG) emissions by 5.2% by 2012 compared to their levels in 1990. Moreover the European Union (EU) pledged to lower its overall emissions by at least 20% until 2020.<sup>67</sup> The protocol does not commit developing countries to emission reductions but they are allowed to host emission reduction projects of industrial countries.

The Kyoto Protocol offers three main mechanisms: Clean Development Mechanism (CDM), Joint Implementation (JI), and Emission Trading. CDM and JI allow annex I countries to offset their emissions by emission reduction projects in non-annex I countries and in

<sup>66</sup>The following introduction of the EU Emissions Trading Scheme (EU ETS) relies on EU (2009).

<sup>67</sup>If other developed countries plan comparable reductions, the EU is willing to reduce emissions by 30%.



other industrial countries. The Emission Trading mechanism provides the opportunity that countries can trade emission credits.

Following the idea of emission trading, the EU set up the EU Emissions Trading Scheme (EU ETS) which is an international, mandatory, company-level cap-and-trade system (Directive 2003/87/EC) for emission allowance units (European Union Allowance, EUA) certifying the right to emit one ton of CO<sub>2</sub> or CO<sub>2</sub> equivalent GHGs.<sup>68</sup> At the moment, the EU ETS covers about 11,000 heavy energy-consuming installations<sup>69</sup> in power/ heat generation and manufacturing in the 27 EU member states, Iceland, Liechtenstein, and Norway.

The idea is to foster a cost efficient emission reduction by directing the reduction into sectors with the lowest marginal reduction costs. The EU hopes to meet its reduction target under the Kyoto Protocol at a cost of below 0.1% of GDP.

Every installation covered by the EU ETS requires a permit from the competent authority for their CO<sub>2</sub> (equivalent) emissions. A permit is granted when the installation is able to annually monitor and report its emissions. Every April installations have to surrender allowances corresponding to their verified emissions. Redundant allowances can be sold while missing allowances can be bought from other companies or at a stock exchange.<sup>70</sup> If an installation fails to submit the adequate number of EUAs it is obliged to hand them in one year later. Additionally, it is punished by a fine of 100€/ ton (40€ before 2008).

The EU ETS started with a pilot phase (2005-2007) to verify actual emissions and to allow market participants to understand the mechanism. As the EU ETS has to bear most of the burdens of EU emission reduction goals, the EU has cut the numbers of EUAs by 6.5% below the 2005 level for the the second trading period (2008-2012).<sup>71</sup> The long-lasting third trading period (2013-2020) will offer more predictability for long-term investments. Additionally, there will be some fundamental changes in the system: National caps defined by the respective national allocation plans (NAPs) will be replaced by a single European cap following a linear annual 1.74% reduction until 2020 — a 21% reduction below the 2005 level by 2020. On average, companies will have to buy at least 50% (100% in 2027) of their allowances at auctions. The auction share of companies apart from power generators will

---

<sup>68</sup>The EU ETS is closely linked to the other Kyoto mechanisms as companies can use *certified emission reductions* (CER) and *emission reduction units* (ERU) from CDM and JI projects for compliance where 1EUA= 1CER= 1ERU.

<sup>69</sup>They account for around 40% (50%) of the EU's total GHG (CO<sub>2</sub>) emissions. More sectors (e.g. aviation) will be covered in 2012 increasing the share to 43%.

<sup>70</sup>Prices at exchanges (e.g. European Climate Exchange (ECX), BlueNext, Powernext, Nordpool, European Energy Exchange (EEX), Energy Exchange Austria (EXAA), Climex) are determined by supply and demand and trading is open to all people and institutions that want to invest. The trading volume has constantly risen from 1 billion EUAs in 2006 to 3.1 billion in 2008.

<sup>71</sup>The second trading period corresponds to the first commitment period of the Kyoto Protocol.

gradually rise from 20% in 2013 to 70% in 2020.<sup>72</sup> At the moment, at least 90% (95% before 2008) of the allowances are allocated for free.

## 9.2 Data

In finance, stochastic processes are calibrated to spot price data,  $s = (s_1, \dots, s_T)'$ . The dominant and most liquid spot exchange for European Union Allowances (EUAs) is BlueNext (www.bluenext.eu). However, spot trading on phase II EUAs started on February 26, 2008 which is rather short for parameter estimation. In contrast, the European Climate Exchange (ECX, www.ecx.eu) - the dominant and most liquid future exchange - has been trading on phase II futures (maturity December 2008/ 2009) since April 22, 2005. We have access to the ECX data from April, 22 2005 up to May 07, 2009 with 1,034 settlement prices.

Uhrig-Homburg & Wagner (2007) find support for the basic *cost-of-carry relationship*<sup>73</sup>  $S_t^{(I)} = F_{tT}^{(I)} \cdot \exp(-r \cdot [T - t])$  for *intra*-phase (Dec06) future contracts  $F_{tT}^{(I)}$  and phase I spot prices  $S_t^{(I)}$ . The constant  $r$  is the risk-free interest rate and  $T$  is the date of maturity of the future. Uhrig-Homburg & Wagner (2007), p. 11, assume a cost-of-carry relationship without convenience yields due to economic reasoning as the storage of EUAs does nearly cost nothing and because of the prohibition of inter-phase banking 2007/8. Consequently, the stockpiling of EUAs would have been irrational for compliance in the second trading period.<sup>74</sup>

Unfortunately, Uhrig-Homburg & Wagner (2007) conclude that the basic cost-of-carry relationship is extremely violated for *inter*-phase future contracts (spot prices before 2008 vs. future price for December 2008 delivery (Dec08)). Daskalakis et al. (2009) come to the same result. They run an extremely sophisticated parameter estimation where they conciliate phase II futures and phase I spot prices by augmenting their pricing models by a stochastic mean reverting convenience yield. We fully agree with their findings that the spot prices in phase I are not linked via a basic cost-of-carry formula to phase II future prices (e.g. Dec08 and Dec09). Phase II EUAs are no substitutes for phase I EUAs as they cannot be used for compliance in phase I. Hence, the approach of Daskalakis et al. (2009) seems not logical and complicates the parameter estimation substantially.

We follow the idea of Wagner (2007), pp. 97, and propose the generation of *synthetic spot*

<sup>72</sup>Exceptions are allowed for energy intensive industries exposed to fierce international competition.

<sup>73</sup>It is justified by the no-arbitrage argument (see Hull 2006, pp. 99).

<sup>74</sup>In the case of physical storage costs, the relationship between spot and future prices is assumed to be  $S_t = F_{tT} \cdot \exp(-[r - c] \cdot [T - t])$  where  $c$  is the *convenience yield*. The argumentation is, that the future buyer avoids the storage which cannot be ignored in a no-arbitrage market. Additionally, if the market fears a liquidity squeeze in  $T$ , traders prefer to lock in the market before. This induces an upwards pressure on spot prices. This is equivalent to  $c > 0$ .

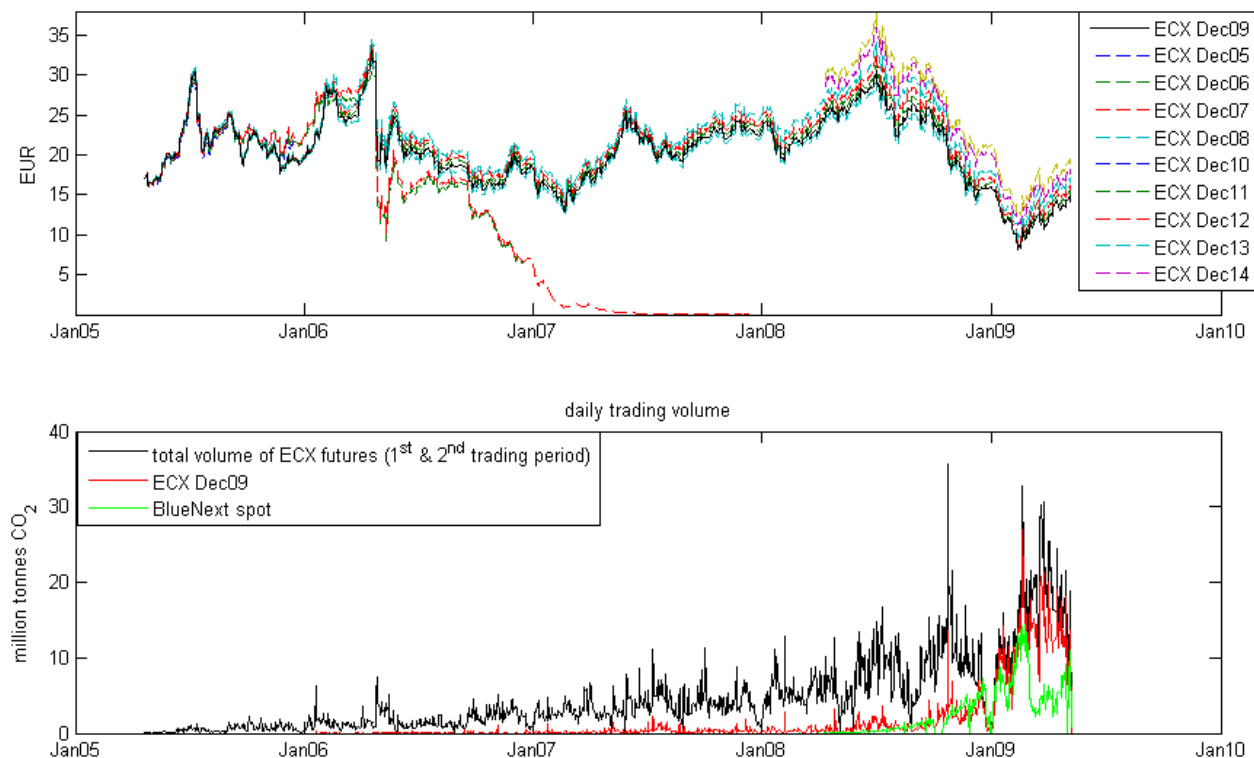


Figure 6: EU ETS market - future prices and market volume

prices from ECX phase II futures via the basic cost-of-carry relationship.<sup>75</sup> Our argument is that future markets are generally far more liquid than spot markets. This is also true for the ETS (lower subplot in figure 6). Moreover, the phase II future prices look highly related (upper subplot in figure 6). All futures seem to be determined by the total traded volume on all futures. Consequently, the future markets seem to rule the spot prices. In the following, we assume that (1) price finding only takes place in the future markets and (2) there are no physical storage costs and no fear of a market squeeze in phase II. Then, we do not require the existence of a spot market as we can generate synthetical spot rates  $\tilde{S}_t = F_{tT} \cdot \exp(-r \cdot [T - t])$ .

For logical reasons, the cost-of-carry approach should be valid for phase II spot and future prices. Therefore, we compare BlueNext phase II spot prices  $s_t^{Blue}$  and *synthetic Dec09 spot prices*  $s_t^{ECX} = f_{tT}^{ECX} \cdot \exp(-r \cdot [T - t])$  generated from ECX Dec09 futures.<sup>76</sup> The time-to-maturity  $T - t$  is measured in years. As a proxy for the annual risk-free interest

<sup>75</sup>In fact, Wagner (2007), pp. 97 - analyzing the EUA prices for the period April 23, 2005 up to December 15, 2006 - calculates synthetic spot prices because of the better liquidity of future markets.

<sup>76</sup>We have decided to generate synthetical spot prices from ECX Dec09 futures as they have been traded

rate  $r^{p.a.} = \ln(1 + \bar{r}_{EURIBOR})$ , we calculate an unweighted arithmetic average of EURIBOR rates with maturities from one week up to twelve month.<sup>77</sup> As there are approximately 250 trading days each year we convert the risk-free rate by  $r = r^{p.a.}/360.25 \cdot 250$ . We find a good match for future and spot prices. The mean absolute difference between the log-returns  $y_t = \ln(s_{t+1}/s_t)$  is  $T^{-1} \sum_{t=1}^T |y_t^{Blue} - y_t^{ECX}| = 0.0058$ .

In the upper subplot of figure 7 you can find the observed ECX Dec09 future prices (black line) and our synthetic Dec09 spot prices (red line). Because of the outliers in April/ May 2006 (see middle subplot of figure 7) we also analyze a corrected time-series without prices from April 24 up to May 25, 2006 (dashed blue lines). In table 3, we have summarized some basic statistics of the log-returns (sample mean  $\hat{\mu}$ , volatility  $\hat{\sigma}$ , skewness  $\hat{\gamma}_1$ , and excess kurtosis  $\hat{\gamma}_2$ ).

	$y_{Dec09}^{Fut}$	$y_{Dec09}$	$y_{Dec09}^{corr}$
$T$	1033	1033	1009
$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$	$-1.3 \cdot e^{-4}$	$-7.3 \cdot e^{-5}$	$1.7 \cdot e^{-4}$
$\hat{\mu}^{p.a.} = \hat{\mu} \cdot 250$	-0.033	-0.018	0.044
$\hat{\sigma} = m_2$	0.030	0.030	0.027
$\hat{\sigma}^{p.a.} = \hat{\sigma} \cdot \sqrt{250}$	0.468	0.478	0.430
$\hat{\gamma}_1 = \frac{m_3}{m_2^{3/2}}$	-0.920	-0.867	-0.376
$\hat{\gamma}_2 = \frac{m_4}{m_2^2} - 3$	11.464	10.738	3.025
$m_k = \frac{1}{T} \sum_{t=1}^T [y_t - \hat{\mu}]^k$			

Table 3: basic statistics of the log-returns

The sample volatility  $\hat{\sigma}$  in the table 3 is an unconditional estimator. Indeed, it is easy to see in the lower subplot of figure 7 that the log-returns of the synthetic spot prices are heteroskedastic.<sup>78</sup>

Similar to many other financial time-series the sample skewness  $\hat{\gamma}_1$  is negative for all log-returns (= more large negative outliers). The positive excess kurtosis  $\hat{\gamma}_2$  indicates that

---

since 2005.

<sup>77</sup>Of course, our proxy for the risk-free interest rate is criticizable. However, there is no consensus on an optimal approximation (see Hull 2006, business snapshot 4.1). On the one hand, treasury bonds of best rated countries like the U.S. and Germany could be below the risk-free interest rate since financial investors are partly forced by regulation to invest in such bonds. On the other hand, EURIBOR or LIBOR rates correspond to best rated companies still bearing some default risk.

<sup>78</sup>We generated the estimators of the conditional variance by *variance smoothing*  $\hat{\sigma}_t^2 = \lambda \cdot y_t^2 + [1 - \lambda] \cdot y_{t-1}^2$ , we use  $\lambda = 0.25$ , and *moving average*  $\hat{\sigma}_t^2 = K^{-1} \sum_{k=0}^{K-1} [y_{t-k} - \bar{y}_{tK}]^2$ , where  $\bar{y}_{tK} = K^{-1} \sum_{k=0}^{K-1} y_{t-k}$  and  $t \geq K$ . We choose  $K = 5$  to calculate the variance of the last trading week.

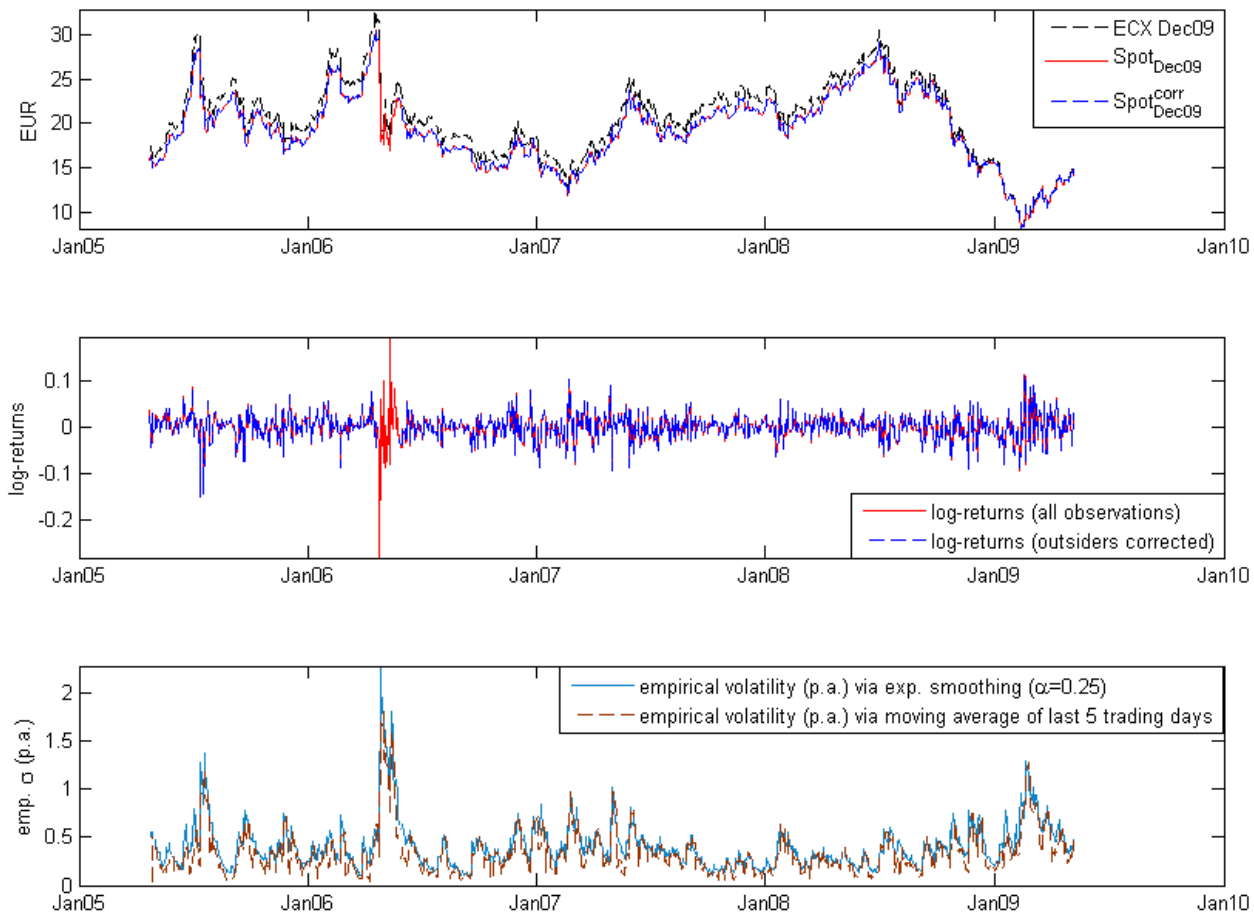


Figure 7: CO2 price data

log-returns feature leptokurtosis meaning an extremely peaked center and fat tails compared to a Gaussian distribution (see figure 8).

In the coming sections we calibrate popular stochastic processes to the Dec09 spot prices respectively to their log-returns. Our preliminary analysis of the data seems to favor models with fat tails and variable variance. As models with variable variance can generate leptokurtosis and negative skewness, stochastic processes seem most promising. Nevertheless, we also estimate simple processes like Geometric Brownian motion. We believe the suitability for daily use of a model is an important criteria beside "correctness" in model selection.

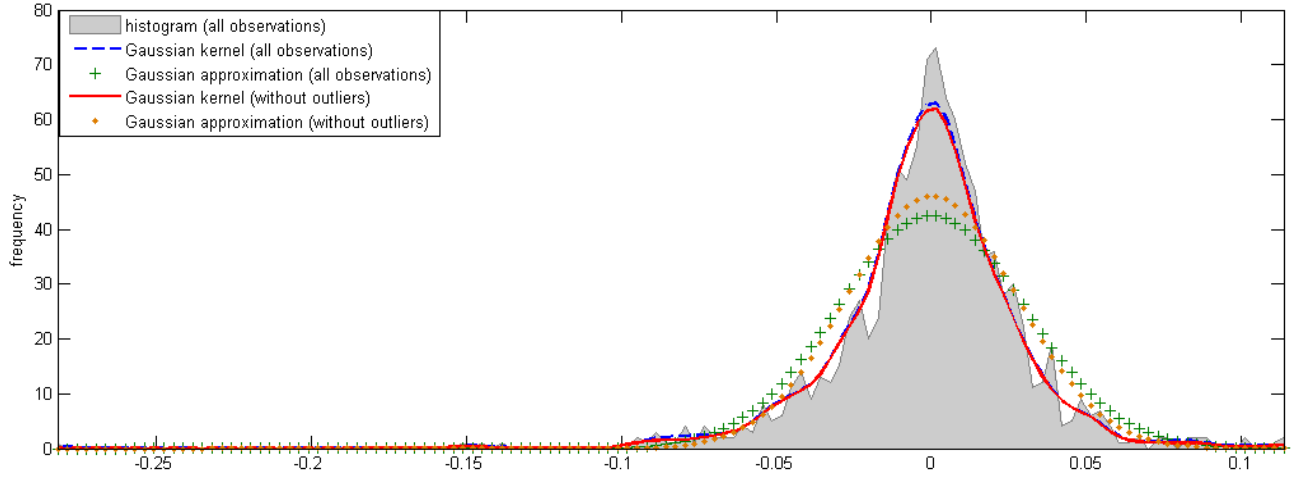


Figure 8: log-returns of synthetic spot prices

## 10 CO2 Price Risk Assessment

The "...knowledge of the statistical distribution of emission trading allowances, and its forecastability, becomes crucial in constructing optimal hedging and purchasing strategies in the carbon market" (Paolella & Taschini 2008, abstract). Therefore, we try to quantify the risk of the EUA price  $S_t$  by means of *stochastic differential processes (with jumps)* (SDP(J)).<sup>79</sup> Most of these processes are nested in the basic formulation<sup>80</sup>

$$dS_t = S_{t+dt} - s_t = \underbrace{\mu(t, s_t, x_t, \theta)}_{\text{drift}} \cdot dt + \underbrace{\sigma(t, s_t, x_t, \theta)}_{\text{volatility}} \cdot dW_t + d \left( \underbrace{\sum_{i=1}^{N_t} Z_i}_{\text{jump}} \right).$$

The variables  $\mu(t, s_t, x_t, \theta)$  and  $\sigma(t, s_t, x_t, \theta)$  can be functions of time  $t$ , the current (known) EUA price  $s_t$ , the realization of an unobservable vector of *latent variables*  $X_t = x_t$ , and a vector of unknown parameters  $\theta = (\theta_1, \dots, \theta_J)'$ . Continuous risk enters the process by the differential of the continuous *Wiener process (standard Brownian motion)*  $dW_t \sim N(0, \sqrt{dt})$ ,

<sup>79</sup>There have been only a few similar attempts (see Borak et al. 2006, Benz & Trück 2009, Paolella & Taschini 2008, or Daskalakis et al. 2009) that all rely on frequentist statistics. Furthermore, our analysis is different from previous research. Either other studies estimate on relatively short times series that are dramatically affected by the price jumps in April/ May 2006 or there are curious attempts to link future markets of the 2nd trading period to the spot markets of the 1st one (see Daskalakis et al. 2009). In contrast, we base our estimations on four years of observations. Consequently, we abstain from a comparison of the estimation results.

<sup>80</sup>In the rest of the paper we will assume that all regularities are met for well-defined solutions. See Eraker (2001), p. 178, for more details.

where  $W_0 = 0$ . Although the SDPJ requires  $S_t$  to be defined in continuous time,  $S_t$  itself does not need to be continuous if a jump term is assumed. Then the *number of jumps* (or *jump times*) from  $t$  to  $t + dt$  is modelled by the counting process  $N_t$ , normally a Poisson process with intensity  $\lambda$ .<sup>81</sup> The *jump width* (or *amplitude*)  $Z_i$  can be a constant  $Z_i = z_i$  or can follow a distribution  $Dist_i(\cdot)$  that depends on some of the unknown parameters and the EUA price. Both number of jumps and jump times are unobservable. Consequently,  $N_t$  and  $Z_i$  are latent variables. The parameters  $\theta$  and the variables  $W_t$ ,  $N_t$ , and  $Z_i$  are all defined under the *objective measure*  $\mathbb{P}$  driving the EUA prices in the "real world".

Although SDP(J)'s are defined on continuous time, EUA prices  $S = s = (s_1, \dots, s_T)'$  are only calculated and published in discrete time. The basic formulation of the SDP(J) changes to an integral equation (see Johannes & Polson 2003, p. 34)

$$S_{t+1} - s_t = \int_t^{t+1} \mu(\tau, s_\tau, x_\tau, \theta) d\tau + \int_t^{t+1} \sigma(\tau, s_\tau, x_\tau, \theta) dW_\tau + \sum_{i=N_t+1}^{N_{t+1}} Z_i.$$

Beside special cases (Geometric Brownian motion, Ornstein-Uhlenbeck, or Cox-Ingersoll-Ross process) there is no analytical solution for the integrals above. Therefore, we apply the *Euler discretization*

$$S_{t+1} - s_t \approx \mu(t, s_t, x_t, \theta) \cdot \Delta + \sigma(t, s_t, x_t, \theta) \cdot \sqrt{\Delta} \cdot \varepsilon_t + \sum_{i=1}^{\xi_t^{Pois}} Z_i,$$

where  $\varepsilon_t \sim N(0, 1)$ ,  $\xi_t^{Pois} \sim Poi(\lambda \cdot \Delta)$ , and  $Z_i$  follows an adequate distribution.

The quality of the approximation crucially depends on the time step  $\Delta$ . The shorter, the better! In this paper all variables and parameters are standardized to an annual basis whereas we have observations on each trading day. Therefore, the time span  $t$  up to  $t + 1$  equals roughly  $\Delta = 1/250 = 0.004$  years.

We further simplify the model by assuming the number of jumps per trading day cannot exceed one. This results in

$$S_{t+1} - s_t \approx \mu(t, s_t, x_t, \theta) \cdot \Delta + \sigma(t, s_t, x_t, \theta) \cdot \sqrt{\Delta} \cdot \varepsilon_t + Z \cdot \xi_t, \quad (5)$$

where  $\xi_t \sim Ber(\lambda \cdot \Delta)$ .<sup>82</sup>

---

<sup>81</sup>The intensity can also be a function  $\lambda(t, s_t, x_t, \theta)$ .

<sup>82</sup>The binomial distribution  $Bin(n, p)$  converges to the Poisson distribution,  $\lim_{n \rightarrow \infty} Bin(n, p) = Pois(\lambda)$ , for a constant  $\lambda = np$ . The Bernoulli distribution is the special case of the binomial distribution for  $n = 1$ ,  $Ber(p) = Bin(1, p)$ . Although Bernoulli and Poisson distributed variables are defined on different domains,  $\xi^{Ber} \in \{0, 1\}$  and  $\xi^{Pois} \in \mathbb{Z}_+$  this might be ignored for a low daily jump intensity ( $\approx$  daily jump probability)  $\tilde{\lambda} = \lambda\Delta = p\Delta$

We estimate models nested in the SDPJ formulation above.<sup>83</sup> The Bayesian MCMC sampler is an excellent algorithm to run the estimation efficiently. Moreover it can separate model induced risk (= Brownian motion), model risk, and parameter risk. This seems important as the risk management is prone to point estimation errors. E.g. Merton (1980), p. 355, believes the development of models accounting for variance estimation errors is the most important direction of research.

Since we estimate twelve competing models, we need a model selection procedure. Normally, frequentist statistics cannot reject all but one model. Although they could be graded by information criteria (e.g. AIC and BIC) or by the p-value calculated from the probability integral transform (Rosenblatt 1952), a model weighing is generally not possible. In contrast, Bayesian statistics offers posterior model probabilities which can be used for generating hybrid-models. This is far more than a basic ranking.

In the next subsections, we first present the nested models we want to estimate. Then we present the estimation results. Subsequently, we calculate model probabilities.

## 10.1 Models

### 10.1.1 Geometric Brownian Motion (GBM)

The *Geometric Brownian motion* (GBM) is defined by

$$dS_t = \nu s_t dt + \sigma s_t dW_t.$$

By *Ito's lemma* this simplifies to (see appendix A.4)

$$d \ln S_t = \mu dt + \sigma dW_t,$$

	$\tilde{\lambda} = 0.1\%$	$\tilde{\lambda} = 1.0\%$	$\tilde{\lambda} = 5.0\%$	$\tilde{\lambda} = 10.0\%$	$\tilde{\lambda} = 15.0\%$	$\tilde{\lambda} = 30.0\%$
$P(\xi^{Ber} = 0) =$	99.9%	99.0%	95.0%	90.0%	85.0%	70.0%
$P(\xi^{Pois} = 0) =$	99.9%	99.0%	95.1%	90.5%	86.1%	74.1%
$P(\xi^{Ber} = 1) =$	0.1%	1.0%	5.0%	10.0%	15.0%	30.0%
$P(\xi^{Pois} = 1) =$	0.1%	1.0%	4.8%	9.0%	12.9%	22.2%
$P(\xi^{Pois} > 1) =$	$5.0 \cdot 10^{-5}\%$	$5.0 \cdot 10^{-3}\%$	0.1%	0.5%	1.0%	3.3%

In the table above, it can be seen that for low jump intensities,  $\tilde{\lambda} \leq 15\%$ , the Bernoulli distribution is an acceptable approximation as the probability for more than one jump per trading day is low. In the case of  $\tilde{\lambda} \approx 30.0\%$  we need carefully interpret estimation results.

<sup>83</sup>There are two exceptions: the Student's t error GARCH process and the Student's t error log-variance process (see sections 10.1.6 and 10.1.9). For both we assume Student's t distributed errors. However, we can transform them to a Gaussian model with a gamma distributed latent precision.



where  $\mu = \nu - \sigma^2/2$ . For the parameter estimation we apply the Euler discretization  $Y_t = \ln S_{t+1} - \ln s_t = \tilde{\mu} + \tilde{\sigma} \cdot \varepsilon_t \sim N(\tilde{\mu}, \tilde{\sigma})$ , where  $\tilde{\mu} = \mu\Delta$  and  $\tilde{\sigma} = \sigma\sqrt{\Delta}$ .

The Bayesian parameter estimation aims to calculate the posterior distribution  $\pi(\theta|y) \propto L_y(\theta) \cdot \pi(\theta)$ , for the parameter vector  $\theta = (\tilde{\mu}, \tilde{\sigma}^2)'$  given the observations  $y = (y_1, \dots, y_T)'$ . The likelihood of the GBM is simply  $L_y(\theta) = \prod_{t=1}^T f_{Y_t}(y_t|\theta) = \prod_{t=1}^T \phi(y_t|\tilde{\mu}, \tilde{\sigma})$ , where  $\phi(y_t|\tilde{\mu}, \tilde{\sigma})$  is a Gaussian pdf with mean  $\tilde{\mu}$  and standard deviation  $\tilde{\sigma}$ . For simplicity we assume independent priors, i.e.  $\pi(\theta) = \pi(\tilde{\mu}) \cdot \pi(\tilde{\sigma}^2)$ .

The posterior  $\pi(\theta|y)$  is a bivariate distribution. We reduce complexity by applying the Clifford-Hammersley theorem:  $\pi(\theta|y)$  is completely characterized by the conditional posteriors

$$\pi(\tilde{\mu}|\tilde{\sigma}^2, y) \propto L_y(\tilde{\mu}) \cdot \pi(\tilde{\mu}) \quad \& \quad \pi(\tilde{\sigma}^2|\tilde{\mu}, y) \propto L_y(\tilde{\sigma}^2) \cdot \pi(\tilde{\sigma}^2).$$

As we use a Gaussian prior  $\pi(\tilde{\mu}) = \phi(\tilde{\mu}|m_0, s_0)$  and an inverse-gamma prior  $\pi(\tilde{\sigma}^2) = IG(\tilde{\sigma}^2|a_0, b_0)$ , the posteriors are conjugate priors which we know exactly. Hence, we can implement the Gibbs sampler (see appendix A.7.1).

### 10.1.2 Geometric Brownian Motion with Jumps (GBMJ)

Beside the standard Geometric Brownian motion we estimate the *jump-diffusion model* (GBMJ) of Merton (1976)

$$dS_t = \left[ \mu + \frac{1}{2}\sigma^2 \right] s_t \cdot dt + \sigma s_t \cdot dW_t + [e^{J_t} - 1] s_t dN_t,$$

where  $N_t$  is an independent Poisson process. Moreover,  $dN_t$  and  $dW_t$  are assumed to be independent. We apply Ito's lemma on  $\ln S_t$  and the resulting process is still a jump-diffusion (see appendix A.4)

$$d \ln S_t = \mu dt + \sigma dW_t + J_t \cdot dN_t$$

We use the Euler discretization and approximate  $dN_t \sim Pois(\lambda_J \cdot \Delta)$  by a Bernoulli variable  $\xi_t \sim Ber(\lambda_J \cdot \Delta)$ . Then the log-returns calculate by

$$Y_t = \ln S_{t+1} - \ln s_t = \tilde{\mu} + \tilde{\sigma} \cdot \varepsilon_t + J_t \cdot \xi_t,$$

where  $\tilde{\mu} = \mu \cdot \Delta$ ,  $\tilde{\sigma} = \sigma \cdot \sqrt{\Delta}$ ,  $\varepsilon_t \sim N(0, 1)$ , and  $J \sim N(\mu_J, \sigma_J)$ . The conditional distribution of  $Y_t$  is

$$Y_t | \xi_t \sim \begin{cases} \tilde{\mu} + \tilde{\sigma} \cdot \varepsilon_t & \sim N(\tilde{\mu}, \tilde{\sigma}), & \text{for } \xi_t = 0 \\ \tilde{\mu} + \mu_J + \tilde{\sigma} \cdot \varepsilon_t + \sigma_J \cdot \eta & \sim N\left(\tilde{\mu} + \mu_J, \sqrt{\tilde{\sigma}^2 + \sigma_J^2}\right), & \text{for } \xi_t = 1 \end{cases},$$

where  $\eta \sim N(0, 1)$ .

The parametrization in GBMJ model is  $\theta = (\tilde{\mu}, \tilde{\sigma}^2, \mu_J, \sigma_J^2, p_J)'$  with  $p_J = \lambda_J \cdot \Delta$  as the daily jump probability. The resulting likelihood function is  $L_y(\theta) = \prod_{t=1}^T f_{Y_t}(y_t|\theta)$  with marginal distributions

$$\begin{aligned} f_Y(y_t|\theta) &= \sum_{k=0}^{\infty} f_{Y_t}(y_t|\theta, N_t = k) \cdot P(N_t = k|\theta) \\ &\approx f_{Y_t}(y_t|\theta, N_t = 0) \cdot P(N_t = 0|\theta) + f_{Y_t}(y_t|\theta, N_t = 1) \cdot P(\Delta N_t = 1|\theta) \\ &= \phi(y_t|\tilde{\mu}, \tilde{\sigma}) \cdot [1 - p_J] + \phi\left(y_t|\tilde{\mu} + \mu_J, \sqrt{\tilde{\sigma}^2 + \sigma_J^2}\right) \cdot p_J, \end{aligned} \quad (6)$$

where  $N_t$  is the number of jumps on trading day  $t$ . Given independent priors for  $\theta = (\tilde{\mu}, \tilde{\sigma}^2, \mu_J, \sigma_J^2, \tilde{p}_J)'$ , the posterior results as

$$\pi(\theta|y) \propto L_y(\theta) \cdot \pi(\tilde{\mu}) \cdot \pi(\tilde{\sigma}^2) \cdot \pi(\mu_J) \cdot \pi(\sigma_J^2) \cdot \pi(p_J),$$

where  $\pi(\tilde{\mu}) = \phi(\tilde{\mu}|m_0, s_0)$ ,  $\pi(\tilde{\sigma}^2) = IG(\tilde{\sigma}^2|a_0, b_0)$ ,  $\pi(\mu_J) = \phi(\mu_J|m_0^{(J)}, s_0^{(J)})$ ,  $\sigma_J^2 = IG(\sigma_J^2|a_0^{(J)}, b_0^{(J)})$ , and  $\pi(p_J) = Beta(p_J|\alpha_0, \beta_0)$ .

For a more efficient estimation via the Gibbs sampler we augment the parameter vector  $\theta$  by the *latent variables* jump width  $J_t$  and number of jumps  $N_t \in \{0, 1\}$  for  $t = 1, \dots, T$ .<sup>84</sup> Therefore, the augmented posterior is

$$\begin{aligned} \pi(\theta, x|y) &\propto L_y(\theta, x) \cdot f_X(x|\theta) \cdot \pi(\theta) \\ &= L_y(\theta, x) \cdot f_{J_t}(j_t|\theta) \cdot f_{N_t}(n_t|\theta) \cdot \pi(\theta) \\ &= L_y(\theta, x) \cdot \phi(j_t|\mu_J, \sigma_J) \cdot Ber(n_t|p_J) \cdot \pi(\theta), \end{aligned}$$

for  $x = (x_1, \dots, x_T)'$  and  $x_t = (j_t, n_t)'$ . The conditional likelihood

$$L_y(\theta, x) = \prod_{t=1}^T \phi(y_t|\tilde{\mu} + j_t \cdot n_t, \tilde{\sigma})$$

assumes both number of jumps and jump width as observations. Based on the augmented posterior we can program a Gibbs sampler (see appendix A.7.2).

### 10.1.3 Ornstein-Uhlenbeck Process for Log-Prices (OU)

A simple mean-reverting process is the *Ornstein-Uhlenbeck process* (OU) which was applied by the *Vasicek model* (see Vasicek 1977) to describe interest rate movements. Although this process is one of the most prevalent stochastic processes to model commodity prices, its

<sup>84</sup>Beside some slight modifications we follow Chan & Wong (2006), pp. 172.

popularity is due to its analytical solution rather than its theoretical implications. The main disadvantage of the Ornstein-Uhlenbeck process is that it allows for negative (commodity) prices, interest rates, etc. Because of this flaw we calibrate the Ornstein-Uhlenbeck process to the log-CO2-prices instead to the basic CO2 prices (see Hull 2006, pp. 719, & Daskalakis et al. 2009)<sup>85</sup>

$$d \ln S_t = \kappa [\tilde{\mu} - \ln s_t] dt + \sigma dW_t,$$

where  $d \ln S_t = \ln S_{t+dt} - \ln s_t$ . The volatility  $\sigma$  is known from the GBM above. The parameter  $\tilde{\mu}$  is the long-term equilibrium of  $\ln s_t$ . The mean reversion rate  $\kappa$  describes the speed of mean reversion. The larger  $\kappa$  the more  $\ln s_t$  is fixed to  $\tilde{\mu}$ . A more intuitive understanding offers the *half-life time*  $\Delta_{1/2} = \ln 2 / \kappa$  which states the time needed to pass half-way from  $\ln s_t$  to the long-term mean  $\tilde{\mu}$ .<sup>86</sup>

The OU process features an analytical solution  $\ln S_{t+1} \sim N(m(t), \varsigma)$  for<sup>87</sup>

$$\begin{aligned} m(t) &= \tilde{\mu} + [\ln s_t - \tilde{\mu}] \exp(-\kappa \Delta), \text{ and} \\ \varsigma &= \sigma \sqrt{[1 - \exp(-2\kappa \Delta)] / 2\kappa}. \end{aligned}$$

Based on this, the posterior is defined by  $\pi(\theta | s) \propto L_s(\theta) \pi(\theta)$ ,<sup>88</sup> where  $\theta = (\kappa, \tilde{\mu}, \sigma)'$

<sup>85</sup>By means of Ito's lemma (see appendix A.4), we can calculate the price process

$$dS_t = \left[ \kappa [\tilde{\mu} - \ln s_t] + \frac{\sigma^2}{2} \right] s_t dt + \sigma s_t dW_t,$$

which results in an always positive CO2 price.

<sup>86</sup>Assume a discretized mean-reverting process  $\Delta \ln S_t = \kappa [\tilde{\mu} - \ln s_t] \Delta + \sigma dW_t$ , where  $\Delta \ln S_t = \ln S_{t+1} - \ln s_t$ . Further assume  $y_{1/2}$  is the half-way from  $\ln s_t$  to the long-term mean  $\tilde{\mu}$ . Then the time  $\Delta_{1/2}$  needed to reach  $y_{1/2}$  is called *half-life time*. After applying the expectation operator it follows (see [www.puc-rio.br/marco.ind/half-life.html](http://www.puc-rio.br/marco.ind/half-life.html))

$$\kappa \cdot \Delta_{1/2} = \frac{E(\Delta \ln S_t)}{[\tilde{\mu} - \ln s_t]} = \int_{\ln s_t}^{y_{1/2}} \frac{dx}{[\tilde{\mu} - x]} = -\ln(\tilde{\mu} - x) \Big|_{\ln s_t}^{y_{1/2}} = -\frac{\ln(\tilde{\mu} - y_{1/2})}{\ln(\tilde{\mu} - \ln s_t)} = -\frac{\ln([\tilde{\mu} - \ln s_t] / 2)}{\ln(\tilde{\mu} - \ln s_t)}.$$

Hence, it holds  $\Delta_{1/2} = \ln 2 / \kappa$ . The respective half-life times are

	$\kappa = 0$	$\kappa = 0.1$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 5$	$\kappa = 10$	$\kappa = 100$	$\kappa = 1000$
$\Delta_{1/2}$ for $\Delta = 1/250$	never	6y 11m	1y 5m	8m 7d	1m 14d	17d	1d 7h	1h 44min
$\Delta_{1/2}$ for $\Delta = 1/365.5$	never	6y 11m	1y 5m	8m 10d	1m 20d	25d	2d 12h	6h

For  $\Delta = 1/250$  we use an average month (m) with 21 trading days (d) and 10 opening hours (h) a day (ECX opens from 7:00 to 17:00 UK local time). The tags "y" and "min" mean years and minutes.

<sup>87</sup>Nevertheless, we weight the OU process according to  $d \ln S_t = \kappa [\tilde{\mu} - \ln s_t] dt + \sigma dW_t$ .

<sup>88</sup>Of course, we could also run a MCMC estimation based on the Euler discretized process where  $\Delta \ln S_t \sim N(\kappa [\tilde{\mu} - \ln s_t] \Delta, \sigma \sqrt{\Delta})$ . The differences in estimation results are minor (Liu 2006). However, the MCMC algorithm is even fast for the "exact" approach. There is no need for the approximation.

and  $L_s(\theta) = \prod_{t=1}^{T-1} \phi(\ln s_{t+1} | m(t), \varsigma)$ .<sup>89</sup> We assume the following independent priors:  $\pi(\kappa) = \phi(\kappa | m_\kappa, \varsigma_\kappa)|_{\kappa \geq 0}$ ,  $\pi(\tilde{\mu}) = \phi(\tilde{\mu} | m_{\tilde{\mu}}, \varsigma_{\tilde{\mu}})|_{\mu \in [\tilde{\mu}_{\min}, \tilde{\mu}_{\max}]}$ , and  $\pi(\sigma) = \phi(\sigma | m_\sigma, \varsigma_\sigma)|_{\sigma > 0}$ .<sup>90</sup> Unfortunately, the posterior above possesses no standard marginal distributions.<sup>91</sup> Hence, we need to construct a hybrid Metropolis-Hastings sampler (see appendix A.7.3).

#### 10.1.4 Cox-Ingersoll-Ross (CIR) Process

The *Cox-Ingersoll-Ross* (CIR) process (see Cox et al. 1985) is a mean-reverting extension of the Ornstein-Uhlenbeck process with variable volatility term

$$dS_t = \kappa [\mu - s_t] dt + \sigma \sqrt{s_t} dW_t, \quad (7)$$

where  $\mu$  is the long term mean level of the CO2 price,  $\kappa$  is the speed of reversion to  $\mu$ , and  $\sigma$  is the instantaneous volatility. After Euler discretization this can be rearranged to

$$S_{t+1} = \alpha + \beta s_t + \sigma \sqrt{s_t} \Delta \epsilon_t, \quad \epsilon_t \sim N(0, 1),$$

where  $\alpha = \kappa \mu \Delta$  and  $\beta = 1 - \kappa \Delta$ . For Bayesian estimation with a restricted number of observations it is critical whether we model priors on  $(\kappa, \mu, \sigma)'$  or on  $(\alpha, \beta, \sigma)'$ . We decided to define priors for  $(\kappa, \mu, \sigma)'$ . For a better model comparison we transform the CIR process into the log-price notion by Ito's lemma (see appendix A.4)

$$d \ln S_t = \frac{1}{s_t} \left[ \kappa [\mu - s_t] - \frac{1}{2} \sigma^2 \right] dt + \frac{\sigma}{\sqrt{s_t}} dW_t,$$

<sup>89</sup>In the case of the short time-series, without the outliers in April/ May 2006, we use the likelihood

$$L_y(\theta) = \prod_{t=1}^{T^{(1)}-1} \phi(\ln s_{t+1} | m(t), \varsigma) \cdot \prod_{\tau=T^{(2)}}^{T-1} \phi(\ln s_{\tau+1} | m(\tau), \varsigma),$$

where  $T^{(1)}$  is the last trading day before and  $T^{(2)}$  the first trading day after the break.

<sup>90</sup>The function  $\phi(x | \cdot, \cdot)|_{x \in [x_{\min}, x_{\max}]}$  is the pdf of a truncated Gaussian distribution. See appendix A.6 for generating samples from a truncated distribution.

<sup>91</sup>Of course, we could rearrange the OU process to

$$\ln S_{t+1} = \alpha + \beta \ln s_t + \tilde{\sigma} \varepsilon_t \sim N(\alpha + \beta \ln s_t, \tilde{\sigma}),$$

where  $\alpha = \kappa \Delta \mu$ ,  $\beta = 1 - \kappa \Delta$ , and  $\tilde{\sigma} = \sigma \sqrt{\Delta}$ . Consequently, a Gibbs sampler could be applied for this simple regression (see Rachev et al. 2008, chapter "Bayesian Linear Regression Model"). We avoid this approach as it demands to specify conjugate priors to  $\alpha$ ,  $\beta$ , and  $\tilde{\sigma}$ . Unfortunately, it needs some effort to guarantee non-negative  $\kappa$  and  $\mu$ .

which we use for MCMC estimation. The joint posterior is  $\pi(\theta|s) \propto L_s(\theta) \cdot \pi(\theta)$  given the CO2 prices  $s = (s_1, \dots, s_T)'$ . The likelihood function is defined by  $L_s(\theta) = \prod_{t=1}^T \phi(y_t|m(t), \varsigma_t)$ ,<sup>92</sup> where  $y_t = \ln s_{t+1} - \ln s_t$ ,  $m(t) = [\kappa[\mu - s_t] - \frac{1}{2}\sigma^2] \Delta/s_t$ , and  $\varsigma_t = \sigma\sqrt{\Delta/s_t}$ . We assume the same independent prior distributions like in the OU model:  $\pi(\kappa) = \phi(\kappa|m_\kappa, \varsigma_\kappa)|_{\kappa \geq 0}$ ,  $\pi(\mu) = \phi(\mu|m_\mu, \varsigma_\mu)|_{\mu \in [0.01, \mu_{\max}]}$ , and  $\pi(\sigma) = \phi(\sigma|m_\sigma, \varsigma_\sigma)|_{\sigma > 0}$ . As the hybrid Metropolis-Hastings sampler has the same structure like that of the OU model we refer to our remarks there.

### 10.1.5 Constant Elasticity of Variance (CEV) Process

The *constant elasticity of variance* (CEV) process with mean reversion<sup>93</sup>

$$dS_t = \kappa[\mu - s_t]dt + \sigma s_t^\gamma dW_t, \quad (8)$$

nests the basic OU process for  $\gamma = 0$  and the CIR process for  $\gamma = 1/2$ . A  $\gamma < 0$  reflects a negative dependency between the variance of  $S_t$  and  $S_t$ , while  $\gamma > 0$  describes a positive dependency. By Ito's lemma the log-CO2-price process is

$$d \ln S_t = \left[ \kappa[\mu - s_t] - \frac{1}{2}\sigma^2 s_t^{2\gamma-1} \right] \frac{1}{s_t} dt + \sigma s_t^{\gamma-1} dW_t,$$

which we apply for the model selection.

<sup>92</sup>The CIR differential equation possesses the non-central  $\chi^2$ -distribution as closed-form solution. Its pdf is

$$f_{S_t|s_{t-\Delta}}(s_t) = c \cdot \exp(-u - v) \left[ \frac{v}{u} \right]^{q/2} Bess_q(2\sqrt{uv}),$$

where

$$c = \frac{2\kappa}{\sigma^2 [1 - \exp(-\kappa\Delta)]}, \quad u = c \cdot s_t \exp(-\kappa\Delta), \quad v = \frac{2\kappa\mu}{\sigma^2} - 1.$$

The modified Bessel function of the first kind of order  $q$  is defined by

$$Bess_q(x) = \sum_{k=0}^{\infty} \frac{[x/2]^{2k+q}}{k! \Gamma(q+k+1)},$$

where  $\Gamma(\cdot)$  is the gamma function (see Cox et al. 1985, pp. 391). The mean and variance are

$$\begin{aligned} E(S_t|s_{t-\Delta}) &= \mu [1 - \exp(-\kappa\Delta)] + \exp(-\kappa\Delta) s_{t-\Delta} \\ Var(S_t|s_{t-\Delta}) &= \frac{\sigma^2}{\kappa} \left[ s_{t-\Delta} [\exp(-\kappa\Delta) - \exp(-2\kappa\Delta)] + \frac{\mu}{2} [1 - \exp(-\kappa\Delta)]^2 \right]. \end{aligned}$$

For computational reasons, this paper applies a Gaussian approximation. See Lee (2006) for a comparison of exact and approximated estimation.

<sup>93</sup>An alternative would be to define the basic CEV process  $dS_t = \mu dt + \sigma s_t^\gamma dW_t$  with constant mean which nests the GBM.

The joint posterior for  $\theta = (\kappa, \mu, \sigma, \gamma)'$  is  $\pi(\theta|s) \propto L_s(\theta) \cdot \pi(\theta)$  where the likelihood  $L_s(\theta) = \prod_{t=1}^{T-1} \phi(s_{t+1}|m(t), \varsigma_t)$  is defined by  $m(t) = \kappa[\mu - s_t]\Delta + s_t$  and  $\varsigma_t = \sigma s_t^\gamma \sqrt{\Delta}$ .<sup>94</sup> We assume the same independent priors like for the OU and the CIR process:  $\pi(\kappa) = \phi(\kappa|m_\kappa, \varsigma_\kappa)|_{\kappa \geq 0}$ ,  $\pi(\mu) = \phi(\mu|m_\mu, \varsigma_\mu)|_{\mu \in [0.01, \mu_{\max}]}$ , and  $\pi(\sigma) = \phi(\sigma|m_\sigma, \varsigma_\sigma)|_{\sigma > 0}$ . Furthermore, we define a Gaussian prior  $\pi(\gamma) = \phi(\gamma|m_\gamma, \varsigma_\gamma)$  for  $\gamma$ .

In the main, the CEV hybrid Metropolis-Hastings sampler has the same structure like that of the OU process. Hence, we refer to our remarks above. However, there is one difference. Generally, the parameters  $\sigma$  and  $\gamma$  are highly correlated. This results in auto-correlations of their MCMC sample paths close to one, which means a bad mixing. Even a thinning factor of 100 could not improve mixing performance in simulation studies. Therefore we run a *bivariate* random walk Metropolis sampler which generates  $(\sigma, \gamma)'$  from a bivariate Gaussian distribution with a covariance matrix retrieved from the ML estimation.<sup>95</sup> As the dependency is actually non-linear we still have to thin the samples.

---

<sup>94</sup>Our hybrid Metropolis-Hastings sampler bases on the CEV process in equation 8 as it is more efficient according to its computing time.

<sup>95</sup>Please keep in mind, that the posteriors of  $\sigma$  and  $\gamma$  are dependent. There is no theoretical problem if we assume independent priors. This just reflects that we believe, prior to our estimation, that the parameters  $\sigma$  and  $\gamma$  are independent.

### 10.1.6 Student's t GARCH (1, 1)

The inspection of the EUA log-returns  $y_t = \Delta \ln s_t = \ln s_{t+1} - \ln s_t$  suggests that the assumption of constant volatility is not suitable.<sup>96</sup> Therefore we estimate different models with variable variance. We start with a stochastic differential process on the log-returns where the latent variance process is described by a GARCH (1, 1) model.<sup>97</sup> We base our MCMC sampler on the algorithm of Rachev et al. (2008), pp. 203, which also includes a regression term of external factors. Although we are curious about the influence of oil and coal prices or marginal greenhouse gas abatement costs we could not implement them in our analysis. There is partly no homogeneous good and therefore no general market price and partly there is no daily data. Instead, we implement a regression term analyzing the influence of the mean EURIBOR interest rate  $r = (r_1, \dots, r_T)'$  on the log-returns. This seems interesting as we use  $r$  to calculate the synthetic spot rates. Consequently, our GARCH (1, 1) model can be formulized in discrete time by

$$\begin{aligned} Y_t &= \gamma_0 + \gamma_1 \cdot r_t + \tilde{\sigma}_t \cdot \varepsilon_t \quad (\text{observation process}) \\ \tilde{\sigma}_t^2 &= \omega + \alpha \cdot u_{t-1}^2 + \beta \cdot \tilde{\sigma}_{t-1}^2 \quad (\text{latent process}) \end{aligned}$$

for  $u_t = y_t - \gamma_0 + \gamma_1 \cdot r_t$ . To account for fat tails we assume  $\varepsilon_t \sim t_v(0, 1)$ , where  $v$  is the degree of freedom.

The implementation of a GARCH process requires to check some properties (see Rachev 2008, pp. 190): There is no guarantee that the process at least features weak stationarity, i.e. the existence of finite means, variances, and covariances that do not change with time.<sup>98</sup> In our Student's t GARCH (1, 1) model the log-returns  $Y_t$  are stationary<sup>99</sup> if the *process persistence parameter*

$$\alpha \frac{v}{v-2} + \beta < 1.$$

The larger it becomes the more slowly volatility shocks  $u_t^2$  die out. Only if this condition holds there exists a finite, long-term (unconditional) log-return variance  $\tilde{\sigma}_{u.c.}^2 \cdot v / [v - 2]$ , where

$$\tilde{\sigma}_{u.c.}^2 = \frac{\omega}{1 - \alpha - \beta}.$$

<sup>96</sup>The GBMJ and the CIR/CEV processes feature a heteroskedastic variance, too. However, there has not been an explicit modelling of the volatility term.

<sup>97</sup>Bollerslev (1986) has proposed the GARCH model as an extension of the ARCH model of Engle (1982) which reduces number of lags fundamentally.

<sup>98</sup>In this context, covariance stationarity means that the auto-covariance can be a function of the time difference  $\Delta$  but not of the time  $t$  itself,  $Cov(Z_t, Z_{t+\Delta}) = g(\Delta)$ .

<sup>99</sup>We ignore problems arising from the EURIBOR interest rates.

In basic GARCH models there is no auto-correlation between the log-returns. This is in coincidence with the Efficient Market Hypothesis (EMH, Fama 1970). GARCH models can however describe volatility clustering as the volatility shocks are positive correlated by

$$\text{Corr}(U_t^2, U_{t+\Delta}^2) = [\alpha + \beta]^\Delta \frac{\alpha [1 - \alpha\beta - \beta^2]}{[\alpha + \beta] [1 - 2\alpha\beta - \beta^2]},$$

which declines for increasing  $\Delta$ .

In our specification the Student's t *GARCH* (1,1) model, the joint posterior  $\pi(\theta|y) \propto L_y(\theta) \cdot \pi(\theta)$  is defined by the parameter vector  $\theta = (\gamma_0, \gamma_1, \omega, \alpha, \beta, v)'$ . The likelihood function

$$L_y(\theta) = \prod_{t=1}^T t_v(y_t | \gamma_0 + \gamma_1 \cdot r_t, \tilde{\sigma}_t)$$

is an independent multivariate Student's t distribution which is conditioned on the latent conditional variances  $\tilde{\sigma}^2 = (\tilde{\sigma}_1^2, \dots, \tilde{\sigma}_T^2)'$ .

To improve efficiency of our MCMC algorithm we follow Rachev et al. (2008), pp. 206, and introduce *mixing variables*  $\eta_t | v \sim \text{Gam}(v/2, 2/v)$  for the precision of the log-returns to transform the Student's t likelihood to a Gaussian likelihood.<sup>100</sup> The likelihood function simplifies to<sup>101</sup>

$$\tilde{L}_y(\theta) = \prod_{t=1}^T \phi(y_t | \gamma_0 + \gamma_1 \cdot r_t, \tilde{\sigma}_t / \sqrt{\eta_t})$$

but the model is augmented by the latent precision vector  $\eta = (\eta_1, \dots, \eta_T)'$ . Consequently, the posterior is

$$\begin{aligned} \pi(\theta, \eta | y) &\propto \tilde{L}_Y(\theta) \cdot \prod_{t=1}^T \text{Gam}(\eta_t | v/2, 2/v) \\ &\quad \cdot \pi(v) \cdot \pi(\gamma_0, \gamma_1) \cdot \pi(\omega) \cdot \pi(\alpha) \cdot \pi(\beta), \end{aligned}$$

where  $\text{Gam}(\eta_t | a, b)$  is the pdf of a gamma distribution. We assume the following priors:  $\pi(v) = \text{Exp}(v | \lambda_0)$  and  $\pi(\omega) = \pi(\alpha) = \pi(\beta) = \text{Uni}(\cdot | \vartheta_{\min}, \vartheta_{\max})$ . The joint prior for the regression parameters is bivariate Gaussian  $\pi(\gamma_0, \gamma_1) = \phi_2(\gamma_0, \gamma_1 | m_0, \Sigma_0)$  where  $m_0$  reflects our prior assumption on the mean of both regression parameters while the covariance matrix  $\Sigma_0 = [z'z]^{-1} c$ , with  $c$  as a scaling factor, tries to capture dependency structure between the regression parameters.<sup>102</sup>

<sup>100</sup> Assume  $Y \sim t_v(m, s)$ . Then the Student's t pdf can be approximated by  $t_v(y | m, s) = \int \phi(x | m, s/\sqrt{\eta}) \cdot \text{Gam}(\eta | v/2, 2/v) d\eta \approx \frac{1}{W} \sum_{w=1}^W \phi(x | m, s/\sqrt{\eta^{(w)}})$ .

<sup>101</sup> For model weighting we will use the original likelihood function  $L_y(\theta)$ .

<sup>102</sup> In appendix A.7.4 you can find a short description of the hybrid Metropolis-Hastings algorithm we programmed for this model.



### 10.1.7 Markov Switching (MS) GARCH (1, 1)

Additionally, we apply a *Markov Switching model* (MS, see Hamilton 1989) by introducing two regimes for the GARCH process. Time periods with different volatility levels can be described by different GARCH processes which can reduce instationarity of the GARCH process fundamentally. As standard frequentist (ML) instruments computationally fail to estimate MS models with regime-dependent parameters within the GARCH process (the number of variance paths explodes exponentially with the number of observations) less flexible ARCH-type MS structures have been implemented (see Rachev et al. 2008, p. 214). Hamilton & Susmel (1994), p. 317, multiply the time-dependent GARCH variance  $\tilde{\sigma}_t^2$  by a state-dependent factor,  $Y_t = \sqrt{\lambda_{x_t}} \tilde{\sigma}_t \varepsilon_t$ , for  $\lambda_{x_t} > 0$  and  $x_t = 1, 2, \dots$  Cai (1994), p. 310, only models an ARCH process for the variance, e.g.

$$\begin{aligned} Y_t &= \gamma_0 + \gamma_1 \cdot x_t + \tilde{\sigma}_t \varepsilon_t, \\ \tilde{\sigma}_t^2 &= \eta_0 + \eta_1 \cdot x_t + \sum_{j=1}^J \alpha_j \cdot u_{t-j}^2, \end{aligned}$$

where  $Y_t = \ln S_{t+1} - \ln s_t$  and  $x_t \in \{0, 1\}$ .

In contrast, MCMC allows for an efficient estimation as state-dependent variance paths are simply sampled given all other parameters. The observation space is augmented by the regime path  $x = (x_1, \dots, x_T)'$ . Our estimation approach bases mainly on Haas et al. (2004), Henneke et al. (2006), Bauwens & Rombouts (2007), and Rachev et al. (2008), pp. 214. Because of our restricted number of observations we implement some slight modifications. We drop the Student's t distribution assumption and any external risk factors. Our model can be formalized by

$$\begin{aligned} Y_t | \theta &= \mu_{X_t} + \tilde{\sigma}_t \varepsilon_t, \varepsilon_t \sim N(0, 1) \\ \tilde{\sigma}_t^2 &= \omega_{X_t} + \alpha_{X_t} \cdot u_{t-1}^2 + \beta_{X_t} \cdot \tilde{\sigma}_{t-1}^2, \end{aligned}$$

where  $u_t = y_t - \mu_{X_t}$  and  $X_t \in \{0, 1\}$  is the latent regime variable.

The joint posterior in the MS GARCH model

$$\begin{aligned} \pi(\theta | y) &= \int \pi(\theta, x | y) dx = \int \pi(\theta | x, y) \cdot f_X(x | \theta, y) dx \\ &\approx \sum_{w=1}^W \pi(\theta | x^{(w)}, y) \end{aligned}$$

is augmented by the latent regime states  $X = (X_1, \dots, X_T)'$  which are sampled from the observation based density  $f_X(x | \theta, y)$ . The transition of the regime variable can be described

by the *transition probability matrix*

$$\Pi = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

where  $p_{ij}$  is the transition probability from state  $i$  to state  $j$  and  $p_{i2} = 1 - p_{i1}$ , for  $i = 1, 2$ . For simplicity we assume the *Markov property* for the transition probabilities, i.e.  $p_{ij} = P(X_t = j | x_{t-1} = i)$ .

A crucial step in this model is to sample the regime path  $x = (x_1, \dots, x_T)'$  out of  $K^T = 2^T$  potential paths. Because of the complexity we separately sample each state of the regime

$$\begin{aligned} P(X_t = k | \theta, x_{-t}, y) &= \frac{f_Y(y | \theta, X_t = k, x_{-t}) P(X_t = k, x_{-t} | \theta)}{P(x_{-t}, y | \theta)} \\ &= \frac{f_Y(y | \theta, X_t = k, x_{-t}) P(X_t = k, x_{-t} | \theta)}{\sum_{l=1}^2 f_Y(y | \theta, X_t = l, x_{-t}) P(X_t = l, x_{-t} | \theta)}, \end{aligned}$$

where  $k = 1, 2$  and  $x_{-t}$  is the vector  $x$  without  $x_t$ . As the number of regimes for  $X_t$  is restricted to  $K = 2$  it is more efficient to sample from the posterior instead of the kernel although the posterior follows no standard distribution. Thanks to the Markov property it is

$$P(X_t = k, x_{-t} | \theta) = P(X_{t-1} = l, X_t = k, X_{t+1} = s | \theta) = p_{lk} p_{ks}$$

$k, l, s \in \{0, 1\}$ . Consequently, we get

$$P(X_t = k | \theta, x_{-t}, y) = \frac{\prod_{i=1}^T \phi(y_i | \mu_{x_{ik}}, \tilde{\sigma}_{ik}) p_{x_{t-1}, k} p_{k, x_{t+1}}}{\sum_{l=1}^2 \prod_{i=1}^T \phi(y_i | \mu_{x_{il}}, \tilde{\sigma}_{il}) p_{x_{t-1}, l} p_{l, x_{t+1}}},$$

where  $\mu_{x_{ik}}$  and  $\tilde{\sigma}_{ik}$  are the mean and volatility paths with  $X_t = k$ . For reducing the computational burden we shorten the likelihoods by

$$\begin{aligned} P(X_t = k | \theta, x_{-t}, y) &= \frac{\prod_{i=1}^{t-1} \phi(y_i | \mu_{x_i}, \tilde{\sigma}_i)}{\prod_{i=1}^{t-1} \phi(y_i | \mu_{x_i}, \tilde{\sigma}_i)} \cdot \frac{\prod_{i=t}^T \phi(y_i | \mu_{x_{ik}}, \tilde{\sigma}_{ik}) p_{x_{t-1}, k} p_{k, x_{t+1}}}{\sum_{l=1}^2 \prod_{i=t}^T \phi(y_i | \mu_{x_{il}}, \tilde{\sigma}_{il}) p_{x_{t-1}, l} p_{l, x_{t+1}}} \\ &= \frac{\prod_{i=t}^T \phi(y_i | \mu_{x_{ik}}, \tilde{\sigma}_{ik}) p_{x_{t-1}, k} p_{k, x_{t+1}}}{\sum_{l=1}^2 \prod_{i=t}^T \phi(y_i | \mu_{x_{il}}, \tilde{\sigma}_{il}) p_{x_{t-1}, l} p_{l, x_{t+1}}}, \end{aligned} \quad (9)$$

as the first  $t - 1$  observations are not influenced by the different regimes. Unfortunately, the density values of the observations  $y_\tau$ ,  $\tau > t$ , depend on  $x_t$  as the volatility path is a function of  $x_t$ .

Given the sampled regime path  $x = (x_1, \dots, x_T)'$  and the augmented parameter vector  $\theta = (\{\mu_k, \omega_k, \alpha_k, \beta_k, p_{kk}\}_{k=1,2})'$  we can set up the augmented joint posterior

$$\begin{aligned} \pi(\theta | x, y) &= L_{xy}(\theta) \cdot \pi(\theta) \\ &= L_{xy}(\theta) \cdot \prod_{k=1}^2 \pi(\mu_k) \pi(\omega_k) \pi(\alpha_k) \pi(\beta_k) \pi(p_{kk}). \end{aligned}$$

The augmented likelihood is  $L_{xy}(\theta) = \prod_{t=1}^T \phi(y_t | \mu_{x_t}, \tilde{\sigma}_t)$ . We assume the following priors:  $\pi(\mu_k) = \phi(\mu_k | m_0, s_0)$ ,  $\pi(\omega_k) = \pi(\alpha_k) = \pi(\beta_k) = Uni(\vartheta_{\min}, \vartheta_{\max})$ , and  $\pi(p_{kk}) = Beta(p_{kk} | a_{0k}, b_{0k})$ .<sup>103</sup>

### 10.1.8 Basic Stochastic Log-Variance (BLV)

Beside the class of GARCH processes, variable variances can be modelled more flexible by *stochastic variance processes*. First, we start with the basic *stochastic log-variance* (BLV) process<sup>104</sup>

$$\begin{aligned} d \ln S_t &= \mu dt + \exp(h_t/2) dW_t^S \quad (\text{observation process}) \\ dH_t &= \kappa[\lambda - h_t] dt + \varpi dW_t^H, \quad (\text{latent process}) \end{aligned}$$

which assumes *independent* Wiener processes  $dW_t^S$  and  $dW_t^H$ . After Euler discretization we get

$$\begin{aligned} Y_t &= \tilde{\mu} + \exp(h_t/2) \epsilon_t \\ H_{t+1} &= h_t + \tilde{\kappa}[\lambda - h_t] + \tilde{\tau} \eta_{t+1}, \end{aligned} \quad (10)$$

where  $Y_t = \ln S_{t+1} - \ln S_t$  and  $\epsilon_t, \eta_t \stackrel{i.i.d.}{\sim} N(0, 1)$ . The formulation above is defined on daily basis, i.e.  $\Delta = 1$ . For comparison with the other models, we state the annual counterparts  $\mu = \tilde{\mu}/250$ ,  $\kappa = \tilde{\kappa}/250$ , and  $\tau = \tilde{\tau}/\sqrt{250}$ . Our MCMC estimation follows the efficient BUGS algorithm of Meyer & Yu (2000) which is defined on the re-parameterized processes

$$\begin{aligned} Y_t &= \tilde{\mu} + \exp(h_t/2) \epsilon_t \\ H_{t+1} &= \lambda + \psi[h_t - \lambda] + \tilde{\tau} \eta_{t+1}, \end{aligned} \quad (11)$$

where  $\psi = 1 - \tilde{\kappa}\Delta = 1 - \tilde{\kappa}$  measures the persistence in the volatility which we restrict to  $\psi \in (-1, 1)$ , i.e.  $\tilde{\kappa} \in [0, 2]$  ( $\kappa \in [0, 500]$ ). Consequently, the joint posterior is  $\pi(\theta, h | y) \propto L_y(\theta) \cdot f_H(h | \theta) \cdot \pi(\theta)$ , where  $\theta = (\tilde{\mu}, \lambda, \psi, \tilde{\tau}^2)'$  and<sup>105</sup>

$$\begin{aligned} L_y(\theta) &= \prod_{t=1}^T \phi(y_t | \tilde{\mu}, \exp(h_t/2)), \\ f_H(h | \theta) &= \phi(h_0 | \lambda, \tilde{\tau}) \cdot \prod_{t=1}^T \phi(h_t | \lambda + \psi[h_{t-1} - \lambda], \tilde{\tau}). \end{aligned}$$

We assume the following independent priors:  $\pi(\tilde{\mu}) = \phi(\tilde{\mu} | m_{\tilde{\mu}}, s_{\tilde{\mu}})$ ,  $\pi(\lambda) = \phi(\lambda | m_{\lambda}, s_{\lambda})$ ,  $\pi(\tilde{\tau}^2) = IG(\tilde{\tau}^2 | a_{\tilde{\tau}}, b_{\tilde{\tau}})$ , and  $\pi(\psi^*) = Beta(\psi^* | a_{\psi^*}, b_{\psi^*})$ , for  $\psi^* = [1 + \psi]/2$ .

<sup>103</sup> A description of our MCMC sampler can be found in the appendix A.7.5.

<sup>104</sup> By applying Ito's lemma this is equivalent to  $dS_t/S_t = [\mu + \frac{1}{2} \exp(h_t)] dt + \exp(h_t/2) dW_t^S$ .

<sup>105</sup> For computational reasons we truncate the log-variance process. As we do this far in the tails we will ignore this in the further analysis.

### 10.1.9 Stochastic Log-Variance with t-Errors (tLV)

To account for fatter tails we modify our basic log-variance model (BLV) in 11 to a *Student's t log-variance* (tLV) model by assuming t-distributed errors (see Meyer & Yu 2000)

$$\begin{aligned} Y_t &= \tilde{\mu} + \exp(h_t/2) \epsilon_t, \epsilon_t \sim t_v(0, 1) \\ H_{t+1} &= \lambda + \psi[h_t - \lambda] + \tilde{\tau} \eta_{t+1}, \eta_{t+1} \sim N(0, 1), \end{aligned}$$

where  $\tilde{\mu}$ ,  $\tilde{\tau}$  as well as  $\tilde{\kappa} = 1 - \psi$  are defined on a daily basis and  $v$  is the degree of freedom of a Student's t distribution.

Similar to the BLV model, we need the augmented joint posterior  $\pi(\theta, h|y) \propto L_y(\theta) \cdot f_H(h|\theta) \cdot \pi(\theta)$ , where  $\theta = (\tilde{\mu}, v, \lambda, \psi, \tilde{\tau}^2)'$ ,  $L_y(\theta) = \prod_{t=1}^T t_v(y_t | \tilde{\mu}, \exp(h_t/2))$ ,<sup>106</sup> and<sup>107</sup>

$$f_H(h|\theta) = \phi(h_0 | \lambda, \tau) \cdot \prod_{t=1}^T \phi(h_t | \lambda + \psi[h_{t-1} - \lambda], \tau).$$

We assume the same independent priors like in the BLV model. Additionally, we define  $\pi(v) = \text{Exp}(v | a_v) |_{v \in (2, 50)}$  which is the pdf of a truncated exponential distribution with a mean value of  $a_v$ .<sup>108</sup>

### 10.1.10 Correlated Log-Variance (CLV)

So far, we have presented the basic log-variance (BLV) model which allows to describe excess kurtosis (i.e. heavy tails). Additionally, we introduced an extension where the errors of the observation process are Student's t distributed. Even fatter tails can be modelled. However, both approaches can explain symmetrical log-returns, only. In the following, we describe two log-variance processes which both can model skewness — a common feature of log-returns. The advantage of these models is that they achieve this improvement by an intuitive modification which replicates the *leverage effect* reported in empirical research (see Engle & Ng 1993). The leverage effect is the typical negative relationship between prices/returns and volatility in financial data.

Although there is no discussion on the existence of a leverage effect, the correct specification is not without controversy. The mainstream favors the formulation of Harvey & Shephard (1996) who assume an *inter-temporal dependency* (*asymmetric stochastic variance*, ASV1)

<sup>106</sup>The function  $t_v(x|m, s)$  is a non-central Student's t distribution with mean  $m$  and variance  $s^2v/[v-2]$ .

<sup>107</sup>For computational reasons we truncate the log-variance process. As we do this far in the tails we will ignore this in the further analysis.

<sup>108</sup>Here, we differ from Meyer & Yu (2000) who apply a  $\chi^2$ -prior. Unfortunately, our posterior for  $v$  is highly sensitive in respect of the prior. Alternatively, we use an exponential prior which we make rather non-informative by setting  $a_v = 1000$ .

between the log-return  $Y_t = \ln S_{t+1} - \ln s_t$  of the current period  $t$  and the log-variance  $H_{t+1}$  of the next period  $t + 1$ . In contrast, there is the proposal of Jacquier et al. (2004) who model an *intra*-temporal relationship (ASV2) of the current log-return  $Y_t$  and log-variance  $H_t$  in  $t$ . The differences seem subtle but they have huge theoretical and empirical implications (see Yu 2005, p. 166): A negative correlation parameter  $\rho$  in the ASV1 model corresponds to the leverage effect while there is no clear relationship in the ASV2 model. Furthermore, (Bayesian) empirical research on S&P500 data favors ASV1 (Yu 2005, p. 166).

In the following, we present the ASV1 as well as the ASV2 model, according to the specification of Yu (2005), p. 166 and Yu (2002), which we implemented in WinBugs. Both models base on the observation and latent log-variance processes

$$d \ln S_t = \exp(h_t/2) dW_t^S \quad (12)$$

$$dH_t = \kappa[\lambda - h_t] dt + \varpi dW_t^H. \quad (13)$$

### *Inter-temporal Correlation (ASV1)*

The ASV1 model results from a simple Euler discretization of equation 12 and 13

$$\begin{aligned} Y_t &= \exp(h_t/2) \varepsilon_t \\ H_{t+1} &= \alpha + \psi h_t + \tilde{\tau} \eta_{t+1}, \end{aligned} \quad (14)$$

where  $Y_t = \ln S_{t+1} - \ln s_t$ ,  $\psi = 1 - \tilde{\kappa} \Delta = 1 - \tilde{\kappa}$ ,  $\alpha = \lambda \tilde{\kappa} \Delta = \lambda \tilde{\kappa} = \lambda [1 - \psi]$ . We formulize the model on a daily basis, i.e.  $\Delta = 1$ . The special characteristic of the ASV1 model is its *inter*-temporal correlation

$$\text{Corr}(\varepsilon_t, \eta_{t+1}) = \rho.$$

The ASV1 model can be rearranged to (see Yu 2005, p. 170)

$$\begin{aligned} H_{t+1} | h_t, \alpha, \psi, \tilde{\tau} &\sim N(\alpha + \psi h_t, \tilde{\tau}) \ \& \\ Y_t | h_{t+1}, h_t, \alpha, \psi, \tilde{\tau}, \rho &\sim N(m_{y_t}, s_{y_t}), \end{aligned}$$

where  $m_{y_t} = \rho \exp(h_t/2) [h_{t+1} - \alpha - \psi h_t] / \tilde{\tau}$  and  $s_{y_t} = \exp(h_t/2) \sqrt{1 - \rho^2}$ .

For computational reasons priors are defined for  $\psi^* = [\psi + 1] / 2$  and  $\lambda = \alpha / [1 - \psi]$  instead of  $\alpha$  and  $\psi$  (see Kim et al. 1998). The augmented joint posterior is  $\pi(\theta, h | y) \propto L_y(\theta) \cdot f_H(h | \theta) \cdot \pi(\theta)$ , where  $\theta = (\lambda, \psi, \tilde{\tau}, \rho)'$ ,  $L_y(\theta) = \prod_{t=1}^T \phi(y_t | m_{y_t}, s_{y_t})$ , and<sup>109</sup>

$$f_H(h | \theta) = \phi\left(h_0 \mid \frac{\alpha}{1 - \psi}, \frac{\tilde{\tau}}{1 - \psi^2}\right) \cdot \prod_{t=1}^{T+1} \phi(h_t \mid \alpha + \psi h_{t-1}, \tilde{\tau}).$$

<sup>109</sup>For computational reasons we truncate the log-variance process. As we do this far in the tails we will ignore this in the further analysis.

We assume the same independent priors like in the BLV model. Additionally, we define an uniform prior  $\pi(\rho) = Uni(\rho - 1, 1)$  for the correlation parameter.

### *Intra-temporal Correlation (ASV2)*

Jacquier et al. (2004) propose a slightly different asymmetric stochastic log-variance model (ASV2) which results from the Euler discretization of  $d \ln S_t$  and  $dH_{t-1}$  in equations 12 and 13

$$\begin{aligned} Y_t &= \exp(h_t/2) \varepsilon_t \\ H_t &= \alpha + \psi h_{t-1} + \tilde{\tau} \eta_t. \end{aligned}$$

Again, we formulize the model on a daily basis, i.e.  $\Delta = 1$ . In contrast to the specification above, the ASV2 model assumes an *intra-temporal* correlation

$$Corr(\varepsilon_t, \eta_t) = \rho.$$

The ASV2 model can be rearranged to (see Yu 2005, p. 170)

$$\begin{aligned} H_{t+1} | h_t, \alpha, \psi, \tilde{\tau} &\sim N(\alpha + \psi h_{t-1}, \tilde{\tau}) \ \& \\ Y_t | h_{t+1}, h_t, \alpha, \psi, \tilde{\tau}, \rho &\sim N(m_{y_t}, s_{y_t}), \end{aligned}$$

where  $m_{y_t} = \rho \exp(h_t/2) [h_t - \alpha - \psi h_{t-1}] / \tilde{\tau}$  and  $s_{y_t} = \exp(h_t/2) \sqrt{1 - \rho^2}$ . This this corresponds to  $L_y(\theta) = \prod_{t=1}^T \phi(y_t | m_{y_t}, s_{y_t})$  and<sup>110</sup>

$$f_H(h | \theta) = \phi\left(h_0 | \frac{\alpha}{1 - \psi}, \frac{\tilde{\tau}}{1 - \psi^2}\right) \cdot \prod_{t=1}^T \phi(h_t | \alpha + \psi h_{t-1}, \tilde{\tau}),$$

where  $\theta = (\lambda, \psi, \tilde{\tau}, \rho)'$ . Additionally, we assume the same priors like in the ASV1 model.

#### 10.1.11 Correlated Heston Model

Although log-variance processes are extremely flexible they are no affine models. This has two main disadvantages (see Johannes & Polson 2003, p. 58): (1) The calculation of option prices is numerically costly and (2) the volatility of the log-variance is constant. An alternative is the *Heston model*

$$\begin{aligned} d \ln S_t &= \mu dt + \sqrt{h_t} dW_t^S \\ dH_t &= \kappa [\lambda - h_t] dt + \tau \sqrt{h_t} dW_t^H, \end{aligned}$$

<sup>110</sup>For computational reasons we truncate the log-variance process. As we do this far in the tails we will ignore this in the further analysis.

where  $dW_t^S$  and  $dW_t^H$  are *correlated* Wiener processes with constant correlation coefficient  $\rho$ .<sup>111</sup> Euler discretization leads to the *inter*-temporal correlated model

$$\begin{aligned} Y_t &= \mu\Delta + \sqrt{h_t\Delta}\varepsilon_t \\ H_{t+1} &= h_t + \kappa[\lambda - h_t]\Delta + \tau\sqrt{h_t\Delta}\eta_{t+1}, \end{aligned}$$

where  $Y_t = \ln S_{t+1} - \ln s_t$ ,  $\varepsilon_t, \eta_t \sim N(0, 1)$ , and *inter*-temporal correlation  $\text{Corr}(\varepsilon_t, \eta_{t+1}) = \rho$ .

For the MCMC implementation we follow Li et al. (2008) but do not incorporate jumps because of the low number of EUA price observations. The joint posterior is  $\pi(\theta, h|y) \propto L_y(\theta) \cdot f_H(h|\theta) \cdot \pi(\theta)$ , where  $\theta = (\mu, \kappa, \lambda, \omega, \varkappa)'$ . Following Jacquier et al. (1994) we do not directly construct priors for  $\tau$  and  $\rho$  but for  $\omega = \tau^2[1 - \rho^2]$  and  $\varkappa = \rho\tau$ .<sup>112</sup> The likelihood function is  $L_y(\theta) = \prod_{t=1}^T \phi(y_t | \mu\Delta, \sqrt{h_t\Delta})$ , while the joint density for the latent log-variances is<sup>113</sup>

$$f_H(h|\theta) = \prod_{t=0}^T \phi\left(h_{t+1} | h_t + \kappa[\lambda - h_t]\Delta, \tau\sqrt{h_t\Delta}\right).$$

We assume the the following priors:  $\pi(\mu) = \phi(\mu | m_\mu, s_\mu)$ ,  $\pi(\kappa) = \phi(\kappa | m_\kappa, s_\kappa)|_{\kappa>0}$ ,  $\pi(\lambda) = \phi(\lambda | m_\lambda, s_\lambda)|_{\lambda>0}$ ,  $\pi(\omega) = IG(\omega | a_\omega, b_\omega)$ , and  $\pi(\varkappa | \omega) = \phi(\varkappa | 0, \sqrt{\omega/2})$ .

## 10.2 Parameter Estimation Results

In this subsection we present the MCMC estimation results of all models presented above for the full time-series. We analyze the convergence and interpret the simulated posteriors. Finally, we compare the stability of the results to the outcome of our estimations for the short time-series without the outliers in April/ May 2006.

### 10.2.1 Geometric Brownian Motion (GBM)

We start parameter estimation with the Geometric Brownian motion (GBM) on the full time-series (April 22, 2005 up to May 7, 2009). We sample  $W = 51,000$  times the parameter

<sup>111</sup>This can be transformed to

$$\begin{pmatrix} d \ln S_t \\ dH_t \end{pmatrix} = \begin{pmatrix} \mu \\ \kappa[\lambda - h_t] \end{pmatrix} dt + \sqrt{h_t} \begin{pmatrix} 1 & 0 \\ \rho\tau & \sqrt{1 - \rho^2}\tau \end{pmatrix} \begin{pmatrix} dW_t^S \\ dW_t^* \end{pmatrix},$$

where  $dW_t^S$  and  $dW_t^*$  are *independent* Wiener processes.

<sup>112</sup>For model weighting we calibrate a truncated Gaussian distribution to  $\tau$  and  $\rho$ .

<sup>113</sup>The variance  $h_0$  is unknown. Hence, we cannot calculate  $\phi(h_1 | h_0 + \kappa[\lambda - h_0]\Delta, \tau\sqrt{h_0\Delta})$ . Fortunately, we can sample  $h_1$  from its posterior  $\pi(h_1 | h_2, h_0, \theta, y) \propto \pi(h_1 | h_0, \theta) \cdot \pi(h_2 | h_1, \theta) \cdot f_{Y_1}(y_1 | h_1, \theta)$  which depends on  $h_2$ . We ignore the dependence on  $h_0$ .

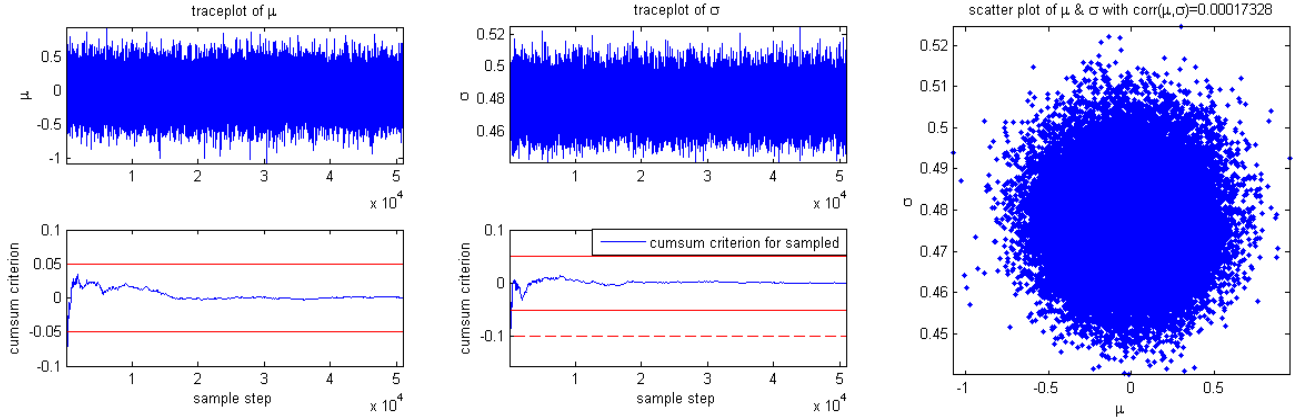


Figure 9: GBM - trace, cumsum, &amp; scatter plots

vector  $(\tilde{\mu}, \tilde{\sigma}^2)'$  and dump the first 1,000 as burn-in.<sup>114</sup> We assume the following priors:  $\tilde{\mu} \sim N(\bar{y} = -7.275 \cdot 10^{-5}, 1000 \cdot \Delta)$  and  $\tilde{\sigma}^2 \sim IG(2.0001, 9.1241 \cdot 10^{-4})$ . This latter specification results in a prior mean of  $E(\tilde{\sigma}^2) = \widehat{Var}(y) = 9.1231 \cdot 10^{-4}$  and a prior variance of  $Var(\tilde{\sigma}^2) = 9 \cdot \widehat{Var}(y)$ .

A first visual inspection of the trace plots and cumsum plots for  $\mu = \tilde{\mu}/\Delta$  and  $\sigma = \tilde{\sigma}/\sqrt{\Delta}$  (see figure 9) indicates convergence for both parameters which seem uncorrelated ( $\widehat{Corr}(\mu, \sigma) = 0$ ) and feature no mentionable auto-correlation.

The marginal posteriors of  $\mu$  and  $\sigma$  have fundamentally down-scaled and shifted from their priors (see upper subplots in figure 10). For a better understanding of the parameter uncertainty we add quantile plots (credible level  $\beta = 1\%, 2\%, \dots, 99\%$ ) showing the lower and upper credible bounds for both posteriors (see lower subplots in figure 10). Although the mean posterior of  $\mu$  ( $-0.019$ ) is centered around zero,  $\mu$  features a huge standard deviation (0.235) resulting in annual drift rates even likely to be around  $-0.5$  or  $0.5$ . To a less extent, this statement is also true for the posterior of  $\sigma$  featuring a highly informative (concentrated) posteriors - a general finding for scale parameters.

We apply the same Gibbs sampler to the short time-series where we dump the observations from the turbulent period April 24 to May 25, 2006.<sup>115</sup> The posterior of  $\mu$  has slightly shifted to the right and features a lower standard deviation while the posterior of  $\sigma$  is more

<sup>114</sup>Indeed, we thin the sampled data by a factor three, meaning that we store every third generated parameter vector, only. Consequently, we sample a total of  $W \cdot 3 = 153,000$  parameter vectors.

<sup>115</sup>We assume the following priors:  $\tilde{\mu} \sim N(\bar{y} = -1.742 \cdot 10^{-4}, 1000 \cdot \Delta)$  and  $\tilde{\sigma}^2 \sim IG(2, 0001, 7.4113 \cdot 10^{-4})$ . This specification results in a prior mean of  $E(\tilde{\sigma}^2) = \widehat{Var}(y) = 7.4107 \cdot 10^{-4}$  and a prior variance of  $Var(\tilde{\sigma}^2) = 9 \cdot \widehat{Var}(y)$ . With a thinning factor of three we store 50,000 samples after a burn-in phase of 1,000 generations.



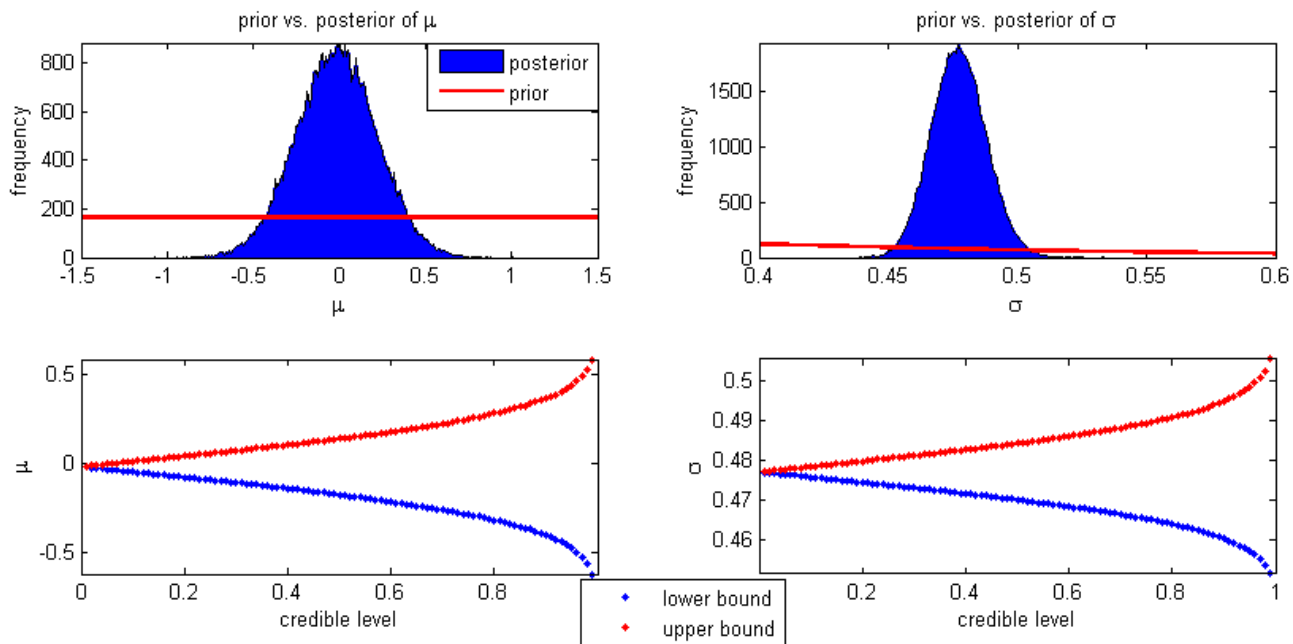


Figure 10: GBM - priors/ posteriors &amp; credible intervals

to the left with nearly no change in its standard deviation (see figure 11). In table 4 we have summarized some basic statistics on the parameters for the full and the short time-series.

	mean	std. deviation	skewness	excess kurtosis
$\mu$	-0.019	0.235	-0.010	0.017
$\sigma$	0.477	0.010	0.119	0.019
$\mu_{short}$	0.041	0.215	0.004	-0.017
$\sigma_{short}$	0.430	0.010	0.105	0.064

Table 4: results for GBM model

Additionally, we analyze the lower and upper credible bounds (0.5% and 99.5%) of the log-returns given the mean posteriors  $\hat{E}(\mu|y)$  and  $\hat{E}(\sigma|y)$  (see figure 12). The lower and upper bound for the full time-series (dashed red line) is far below and above the outliers which are extremely unlikely under the GBM. All in all 2.61% of all observations are outside the 99% credible interval. Even for the short time-series the GBM seems not be a suitable model (green line). There are still 2.58% of the observations outside the credible interval. This key figure is still that high as  $\hat{E}(\sigma|y)$  has shrunk from 0.477 to 0.430 which has narrowed the interval.

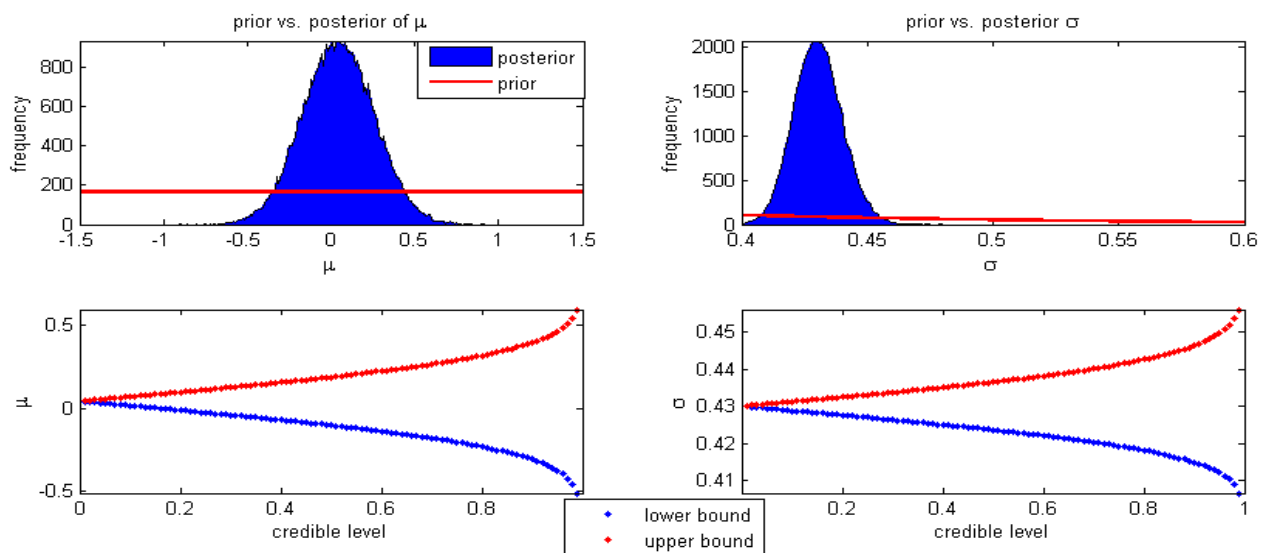


Figure 11: GBM (short) - priors/ posteriors &amp; credible intervals

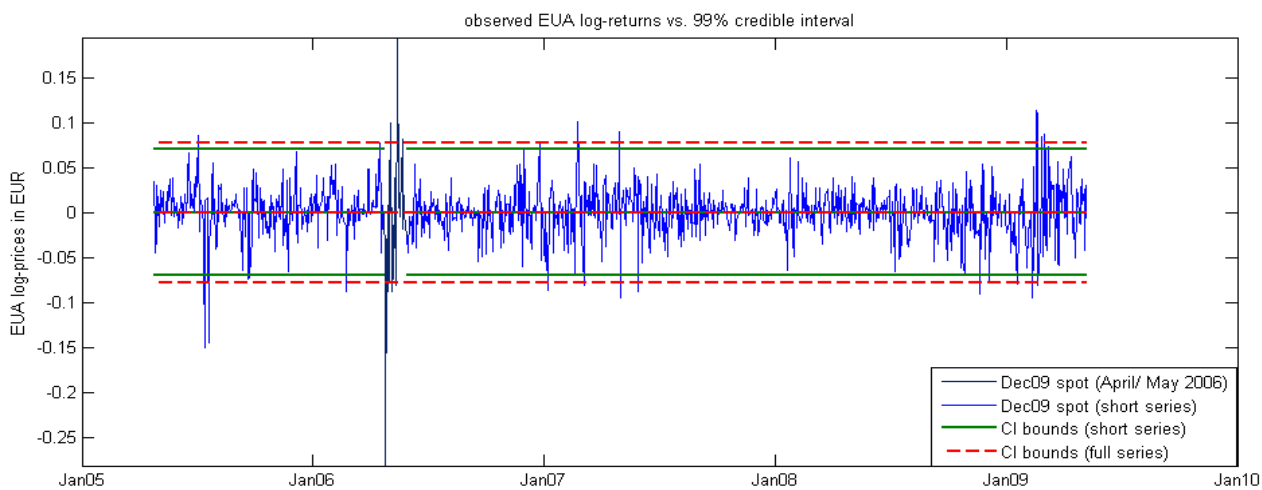


Figure 12: GBM - credible bounds for log-returns

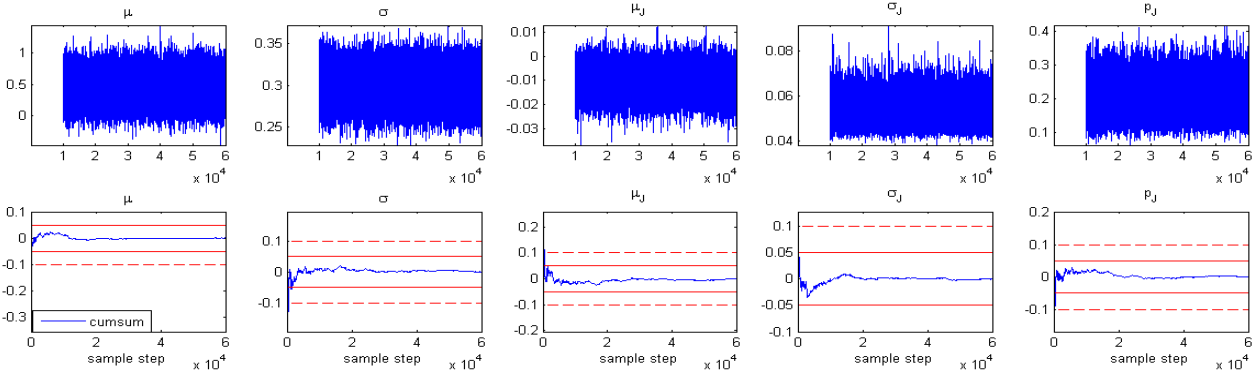


Figure 13: GBMJ - trace &amp; cumsum plots

### 10.2.2 Geometric Brownian Motion with Jumps (GBMJ)

In the next step we run a parameter estimation for the Geometric Brownian motion with jumps (GBMJ). We sample  $W = 60,000$  (burn-in  $B = 10,000$ ) times the parameter vector  $\theta = (\tilde{\mu}, \tilde{\sigma}^2, \mu_J, \sigma_J^2, \tilde{p}_J)'$  with a thinning factor of ten. We assume the following priors:  $\tilde{\mu} \sim N(0, 1000 \cdot \Delta)$ ,  $\tilde{\sigma}^2 \sim IG(a, b)$ , for  $a = 2.0001$  and  $b = 4.8024 \cdot 10^{-4}$ ,<sup>116</sup>  $\mu_J \sim N(0, 0.2)$ ,  $\sigma_J^2 \sim IG(a_J, b_J)$ , for  $a_J = 2.0009$  and  $b_J = 8.5173 \cdot 10^{-3}$ .<sup>117</sup> The jump probability  $p_J \sim Beta(1, 1)$  is modelled by a flat prior on the unity interval.

All MCMC parameter paths have converged according to the trace and cumsum plots (see figure 13). Even by a thinning factor of ten, the parameters  $\tilde{\sigma}$ ,  $\sigma_J^2$ , and  $\tilde{p}_J$  feature some auto-correlation up to a lag of four or five. Moreover, we find some substantial positive and negative correlation for the parameters  $\tilde{\sigma}$ ,  $\tilde{\sigma}_J$ , and  $p_J$  (see table 5).

	$\tilde{\mu}$	$\tilde{\sigma}$	$\mu_J$	$\sigma_J$
$\tilde{\sigma}$	-0.227			
$\mu_J$	-0.323	-0.175		
$\sigma_J$	-0.219	0.612	-0.155	
$p_J$	0.230	-0.821	0.221	-0.748

Table 5: parameter correlations of GBMJ model

<sup>116</sup>This specification results in a prior mean of  $4.8022 \cdot 10^{-4}$  which is the empirical variance of the observed log-returns without the 15% most extreme ones. The prior variance is nine times the prior mean of  $\tilde{\sigma}^2$ .

<sup>117</sup>This specification results in a prior mean of  $8.5092 \cdot 10^{-3}$  which is the empirical variance of the 15% most extreme log-returns minus the mean prior variance of  $\tilde{\sigma}^2$ . The prior variance is nine times the prior mean of  $\sigma_J^2$ .

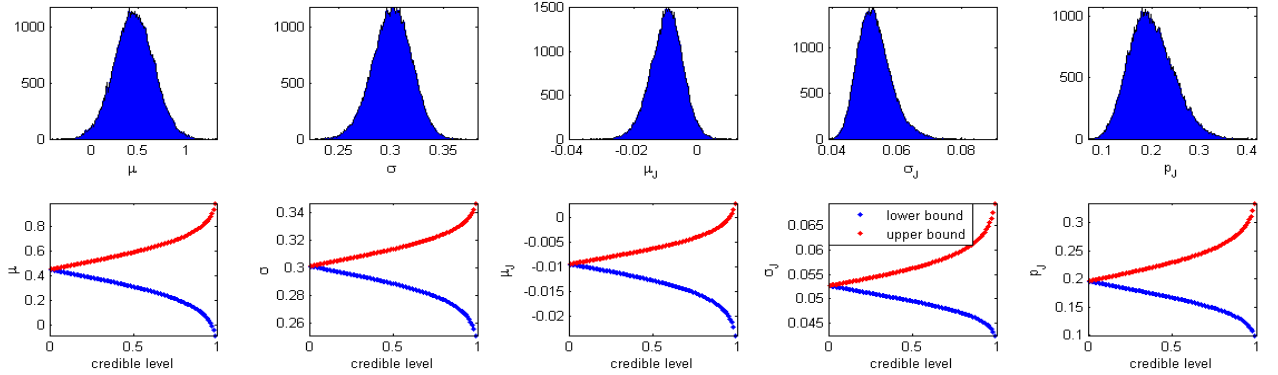


Figure 14: GBMJ - posteriors &amp; credible intervals

Both kinds of correlations reduce information content in the MCMC samples. We could curb auto-correlation by applying a larger thinning factor. However this would increase computational burdens. Alternatively, the correlation problem could be tackled by the construction of a MCMC sampler on the joint posterior  $\pi(\tilde{\sigma}, \sigma_J^2, \tilde{p}_J | \tilde{\mu}, \mu_J, y)$ . Unfortunately, this is rather complex and could not be managed by a Gibbs sampler. The computation time would rise sharply. Therefore, we have decided to accept correlation but we run the Gibbs sampler for a larger number of repetitions.

We apply the same MCMC analysis on the short time-series.<sup>118</sup> The cumsum criterion gives some evidence to be more cautious with convergence. Therefore, we sampled  $W = 100,000$  (burn-in  $B = 50,000$ ) times the parameter values  $\theta = (\tilde{\mu}, \tilde{\sigma}^2, \mu_J, \sigma_J^2, \tilde{p}_J)'$  with a thinning factor of ten (see figure 14 and 15 for the posteriors and quantile plots). In the tables 6 and 7 we have summarized some basic statistics on the parameters for the full and the short time-series.

Similar to the GBM, the drift rate  $\mu$  rises and the volatility  $\sigma$  falls for the short time-series as negative skewed outliers have been removed. A less apparent effect is the increase in the jump probability  $p_J$ . The GBMJ tries to find jumps. As there are less extreme outliers it identifies more frequent medium size outliers as jumps. Consequently, the mean jump width  $\mu_J$  and jump volatility  $\sigma_J$  fall.

Generally, the GBMJ describes the fat-tailed log-returns more closely than the GBM (see

<sup>118</sup>We assume the following priors:  $\tilde{\mu} \sim N(0, 1000 \cdot \Delta)$  and  $\tilde{\sigma}^2 \sim IG(2.0001, 4.5277 \cdot 10^{-4})$ . This results in a prior mean of  $4.5275 \cdot 10^{-4}$  which is the empirical variance of the observed log-returns without the 15% most extreme ones. The prior variance is nine times the prior mean of  $\tilde{\sigma}^2$ . Other priors are  $\mu_J \sim N(0, 0.2)$ ,  $\sigma_J^2 \sim IG(2.0009, 8.5173 \cdot 10^{-3})$ . The prior mean of  $\sigma_J^2$  is  $5.6683 \cdot 10^{-3}$  which is the empirical variance of the 15% most extreme outliers minus the mean prior variance of  $\tilde{\sigma}^2$ . The prior variance is nine times the prior mean of  $\sigma_J^2$ . Again, we defined an uninformative prior  $Beta \sim IG(1, 1)$ .

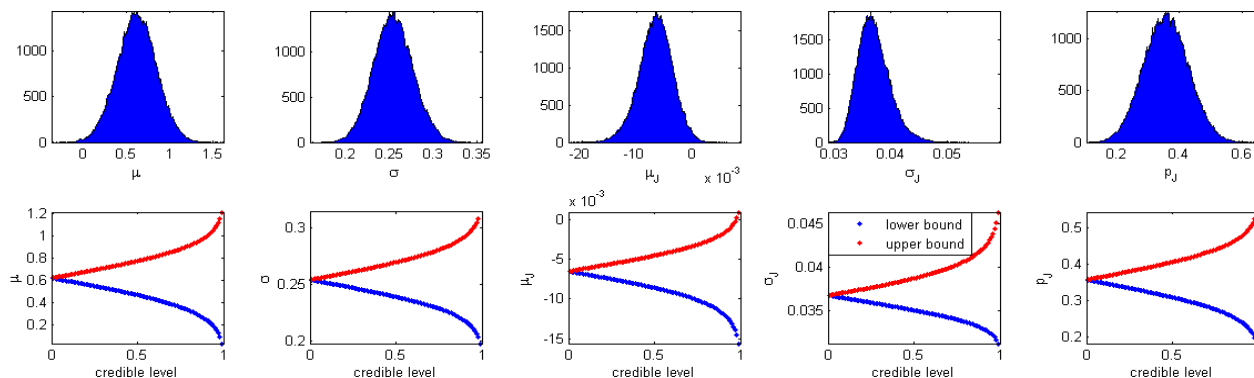


Figure 15: GBMJ (short) - posteriors & credible intervals

full series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.451	0.210	-0.005	0.022
$\tilde{\mu}$	$1.802 \cdot 10^{-3}$	$8.390 \cdot 10^{-4}$	-0.005	0.022
$\sigma$	0.301	0.019	-0.129	0.025
$\tilde{\sigma}$	0.019	$1.179 \cdot 10^{-3}$	-0.129	0.025
$\mu_J$	$-9.638 \cdot 10^{-3}$	$4.967 \cdot 10^{-3}$	-0.219	0.412
$\sigma_J$	0.053	$5.185 \cdot 10^{-3}$	0.631	0.742
$\rho_J$	0.200	0.046	0.407	0.136

Table 6: results of GBMJ model (full series)

upper subplot of figure 16). The better performance of the GBMJ results from a state dependent credible interval. In contrast to the GBM (see figure 12), the credible interval bounds of the GBMJ depend on the fact whether  $y_t$  is in a state with or without jump (see middle subplot in figure 16). All in all, 0.19% (0.30% for the short time-series ) of all observed log-returns fall outside the 99% credible interval. Of course, this is far less than a perfect calibrated frequentist model allows. However, Bayesian statistics reflects parameter uncertainty compensated by a risk premium on the credible interval (wider interval).<sup>119</sup>

The GBMJ seems to be a fundamental improvement compared to the GBM but it is not perfectly suitable for the observed log-returns. It has been developed to describe log-returns following a GBM interrupted by large outliers, once in a while. Our Dec09 log-returns do not follow this pattern. They seem to feature phases with modest and phases with large absolute log-returns. Beside this structural mismatch our Gibbs sampler does not estimate the "true" GBMJ model but an approximation which allows one jump per day, only. The

<sup>119</sup>see section 11 for more on the Bayesian VaR

short series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.620	0.229	$2.526 \cdot 10^{-3}$	0.050
$\tilde{\mu}$	$2.478 \cdot 10^{-3}$	$9.145 \cdot 10^{-4}$	$2.510 \cdot 10^{-3}$	0.050
$\sigma$	0.254	0.022	0.074	-0.029
$\tilde{\sigma}$	0.016	$1.420 \cdot 10^{-3}$	0.074	-0.029
$\mu_J$	$-6.641 \cdot 10^{-3}$	$3.088 \cdot 10^{-3}$	-0.240	0.379
$\sigma_J$	0.037	$2.811 \cdot 10^{-3}$	0.668	0.905
$p_J$	0.357	0.071	0.059	-0.096

Table 7: results of GBMJ model (short series)

approximation is not critical for daily jump probabilities  $\tilde{p}_J$  below 10%. In fact, we estimate mean posterior values of  $p_J$  which are twice up to 3.5 times larger than 10%. Fortunately, in simulation studies, we found no substantial misspecification resulting from our algorithm.

### 10.2.3 Ornstein-Uhlenbeck Process for Log-Prices (OU)

For the Ornstein-Uhlenbeck (OU) process on log-EUA-prices we sample  $W = 60,000$  (burn-in  $B = 10,000$ ) times the parameter values  $\theta = (\kappa, \tilde{\mu}, \sigma)'$  with a thinning factor of ten which reduces auto-correlation to low levels up to lag ten. We assume truncated Gaussian priors for all parameters and set the hyperparameters to the respective ML estimates and modified covariance values.<sup>120</sup> As we believe that the long-term EUA price equilibrium (marginal abatement cost for one metric tonne of CO2 equivalent) will never fall below 0.01€ and will never be above 150€, we truncate the prior for  $\tilde{\mu}$  at  $\tilde{\mu}_{\min} = \ln(0.01)$  and  $\tilde{\mu}_{\max} = \ln(150)$ . The trace and cumsum plots (see figure 17) indicate convergence for all parameter which show low levels of correlation, only.<sup>121</sup> Therefore, we avoid to construct a multivariate random walk Metropolis sampler.

In figure 18 we present the posteriors and their quantile plots. The mean posterior of  $\kappa$  is 1.70. This results in a half-life time of approximately 4 months and 19 trading days. However, as our EUA price series is rather short, there is considerable uncertainty on  $\kappa$  which reaches from  $\kappa = 0$  (no mean reversion at all) up to  $\kappa = 4.5$  (half-life time of approximately 1 month and 18 trading days). Probably, there is much amazement about the curious form of the posterior. This originates from the fact that the prior distribution has not been ruled out by the observations. Even after more than 1000 observations the slight mode of the rather non-informative prior ( $m_\kappa = \kappa^{ML} = 2.181$  and  $\zeta_\kappa^2 = \Sigma_\kappa^{ML} \cdot 1000 = 1016$ ) results in a posterior

<sup>120</sup>We set  $m_\kappa = \kappa^{ML} = 2.181$ ,  $m_{\tilde{\mu}} = \tilde{\mu}^{ML} = 2.9318$ ,  $m_\sigma = \sigma^{ML} = 0.47842$ ,  $\zeta_\kappa^2 = \Sigma_\kappa^{ML} \cdot 1000 = 1016$ ,  $\zeta_{\tilde{\mu}}^2 = \Sigma_{\tilde{\mu}}^{ML} \cdot 1000 = 13.768$ , and  $\zeta_\sigma^2 = \Sigma_\sigma^{ML} \cdot 1000 = 0.11077$ .

<sup>121</sup> $\widehat{Corr}(\tilde{\mu}, \kappa) = 0.169$ ,  $\widehat{Corr}(\tilde{\mu}, \sigma) = 0.004$ , and  $\widehat{Corr}(\kappa, \sigma) = 0.085$

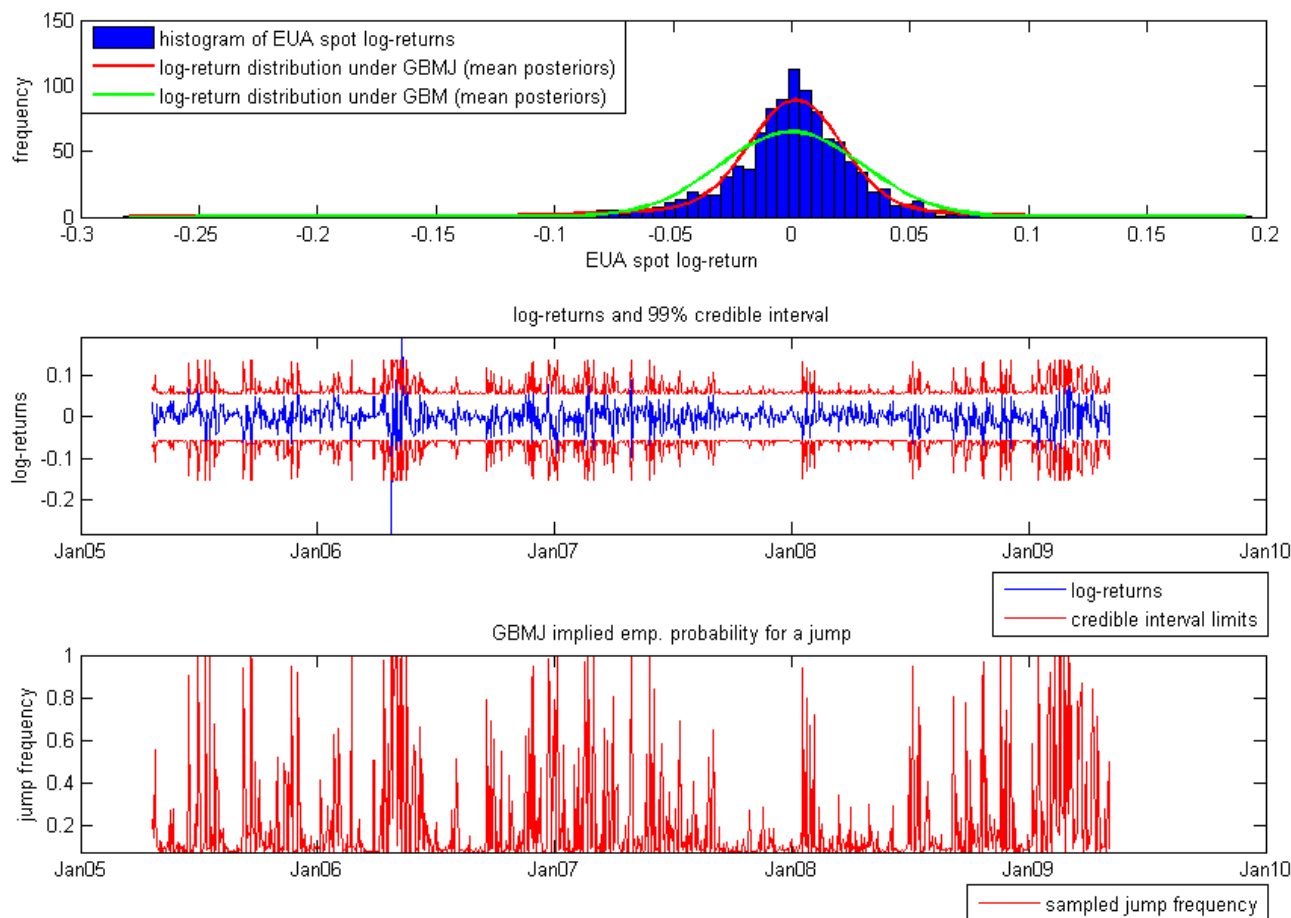


Figure 16: GBMJ - modelling of log-returns

peak. For a far more non-informative prior the posterior of  $\kappa$  would be an unimodal and continuously falling function.

The other extreme is a more informative prior. We run the same model with a truncated prior on  $\kappa$  that matches the ML estimation error ( $m_\kappa = \kappa^{ML} = 2.181$  and  $\varsigma_\kappa^2 = \Sigma_\kappa^{ML} = 1.016$ ).<sup>122</sup> The posterior of  $\kappa$  is by far more concentrated as the prior dominates the posterior (see figure 19). The mean posterior is 1.96. We appreciate that our numerical example above might entail much critics on Bayesian statistics because of its arbitrariness. However, we are convinced that a prudent Bayesian proceeding reveals uncertainty ignored in frequentist statistics.<sup>123</sup> From a Bayesian perspective, the distribution of the ML estimator does not

<sup>122</sup>We also increased the information level on  $\mu$  and  $\sigma$  by  $\varsigma_\mu^2 = \Sigma_\mu^{ML} \cdot 10 = 0.13768$ , and  $\varsigma_\sigma^2 = \Sigma_\sigma^{ML} \cdot 100 = 0.011077$ .

<sup>123</sup>Frequentist statistics acknowledges (ML) estimation error, only. In fact, most users of frequentist sta-

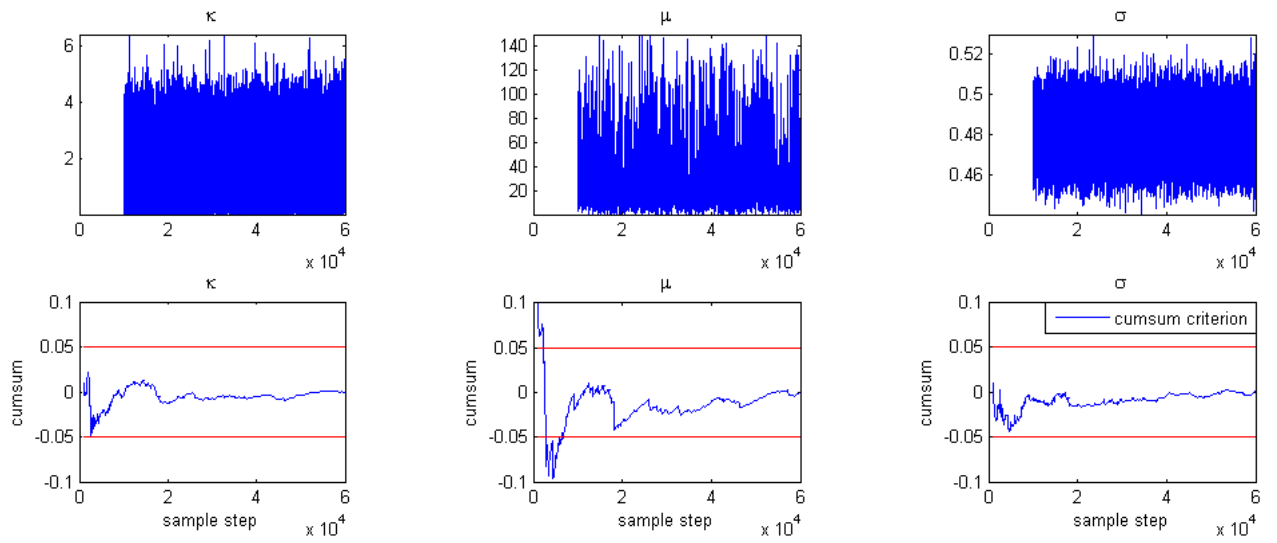


Figure 17: OU - trace &amp; cumsum plots for uninformative priors

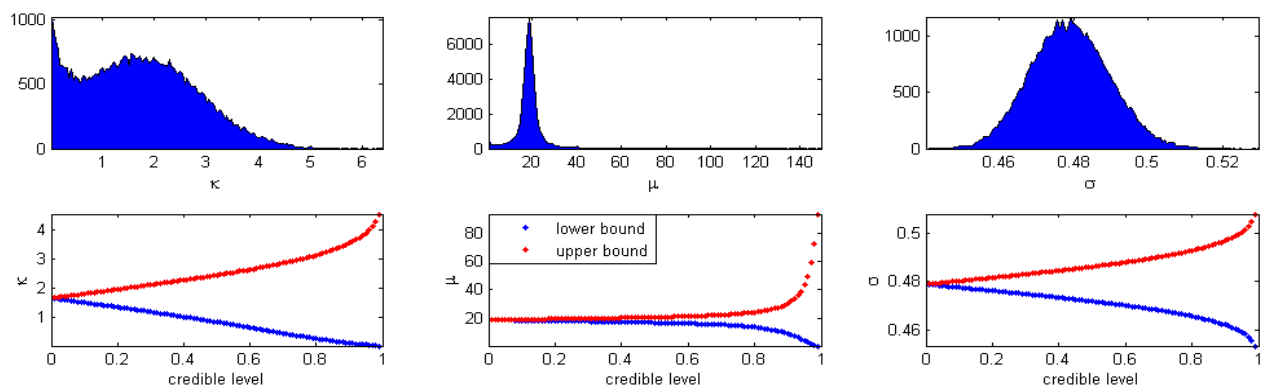


Figure 18: OU - posteriors &amp; credible intervals for uninformative priors



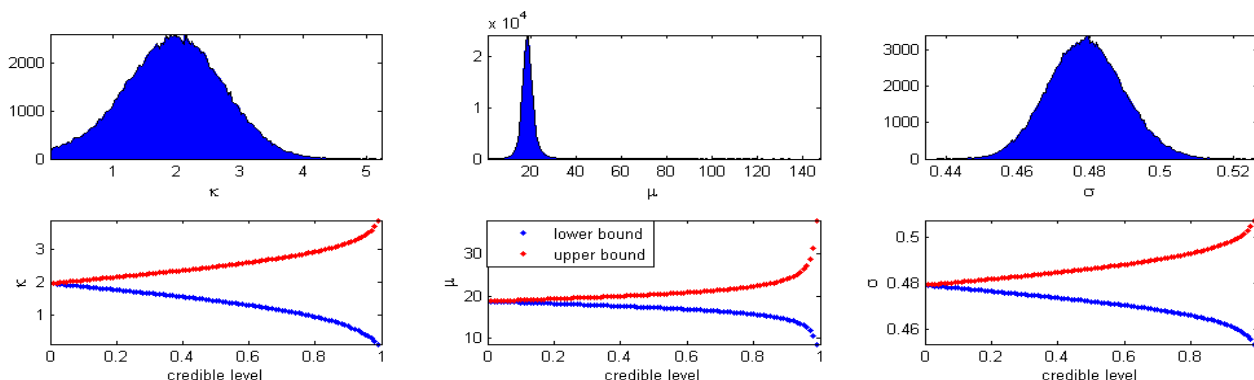


Figure 19: OU - posteriors &amp; credible intervals for informative priors

need to be an adequate prior. First, it does not reflect uncertainty on the parameter  $\kappa$  but on the maximum value of the likelihood function. Second, ML estimation is done under parameter certainty. This is one important reason why the mode for the posterior of  $\kappa$  fundamentally differs from the ML estimate. Consequently, we decide to base our analysis on rather non-informative priors.<sup>124</sup>

Back to the analysis of the estimation results derived from rather non-informative priors. The mean posterior of the long-term CO2 price equilibrium  $\mu = \exp(\tilde{\mu})$  is 19.75€. Parameter values below 10€ and above 30€ are unlikely. As in the previous cases the posterior of the volatility  $\sigma$  features low uncertainty (standard deviation 0.01). With a mean posterior 0.48 the estimated volatility is comparable to the GBM as both models only possess a fixed volatility term to describe variability. Although the GBMJ has the same constant volatility term, its mean posterior is only 0.30 as the jump term explains parts of the observed variability.

We apply the same algorithm to the short time-series.<sup>125</sup> Even after a thinning factor of ten, our MCMC parameter paths have only slowly falling auto-correlations. Fortunately, trace and cumsum criterion show a good mixing and convergence. Again, there is no alarming

---

tistics even ignore this uncertainty beyond parameter testing.

<sup>124</sup>We refuse to apply diffuse (improper) priors for two reasons. First, there can arise problems of improper posteriors that do not integrate to a finite value. Second, our prior is rather non-informative. Even  $\kappa = 1000$  (half-life 1 hour and 44 minutes) features a relatively high likelihood. Nevertheless, posterior values of  $\kappa > 7$  have not been sampled.

<sup>125</sup>We generate  $W = 60,000$  samples with a thinning factor of ten and dump the first 20,000 samples. We use the same priors with hyperparameters resulting from ML estimation:  $m_\kappa = \kappa^{ML} = 1.3448$ ,  $m_{\tilde{\mu}} = \tilde{\mu}^{ML} = 2.9707$ ,  $m_\sigma = \sigma^{ML} = 0.43115$ ,  $\varsigma_\kappa^2 = \Sigma_\kappa^{ML} \cdot 1000 = 921.85$ ,  $\varsigma_{\tilde{\mu}}^2 = \Sigma_{\tilde{\mu}}^{ML} \cdot 1000 = 26.576$ , and  $\varsigma_\sigma^2 = \Sigma_\sigma^{ML} \cdot 1000 = 0.096781$ .

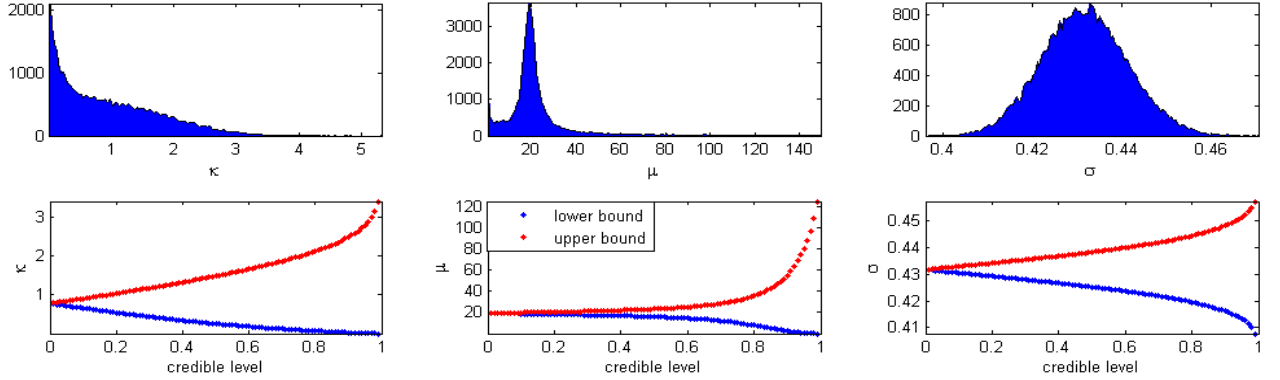


Figure 20: OU (short) - posteriors &amp; credible intervals for uninformative priors

correlation between the parameters.<sup>126</sup> In figure 20 you can find the respective posteriors and in the tables 8 and 9 we have summarized some statistics of the posteriors on the full and the short EUA log-return series.

full series	mean	std. dev.	skewness	excess kurtosis
$\kappa$	1.702	1.071	0.356	-0.466
$\tilde{\mu}$	2.862	0.615	-4.394	33.721
$\mu = \exp(\tilde{\mu})$	19.752	10.686	5.227	40.723
$\sigma$	0.479	0.011	0.135	0.025

Table 8: results of OU model (full series)

short series	mean	std. dev.	skewness	excess kurtosis
$\kappa$	0.959	0.806	0.856	0.147
$\tilde{\mu}$	2.808	1.051	-3.043	13.640
$\mu = \exp(\tilde{\mu})$	22.624	17.900	3.259	14.018
$\sigma$	0.432	$9.606 \cdot 10^{-3}$	0.094	$-6.817 \cdot 10^{-3}$

Table 9: results of OU model (short series)

For the short time-series we estimate a lower level of mean reversion compared to the full time-series. The mean half-life time is 8 months and 14 trading days whereas a half-life time below 2 months and 16 trading days ( $\kappa > 3$ ) is unlikely. This is a result we have not

<sup>126</sup> $\widehat{Corr}(\kappa, \mu) = 0.1801$ ,  $\widehat{Corr}(\kappa, \sigma) = 0.0704$ ,  $\widehat{Corr}(\mu, \sigma) = 0.0169$

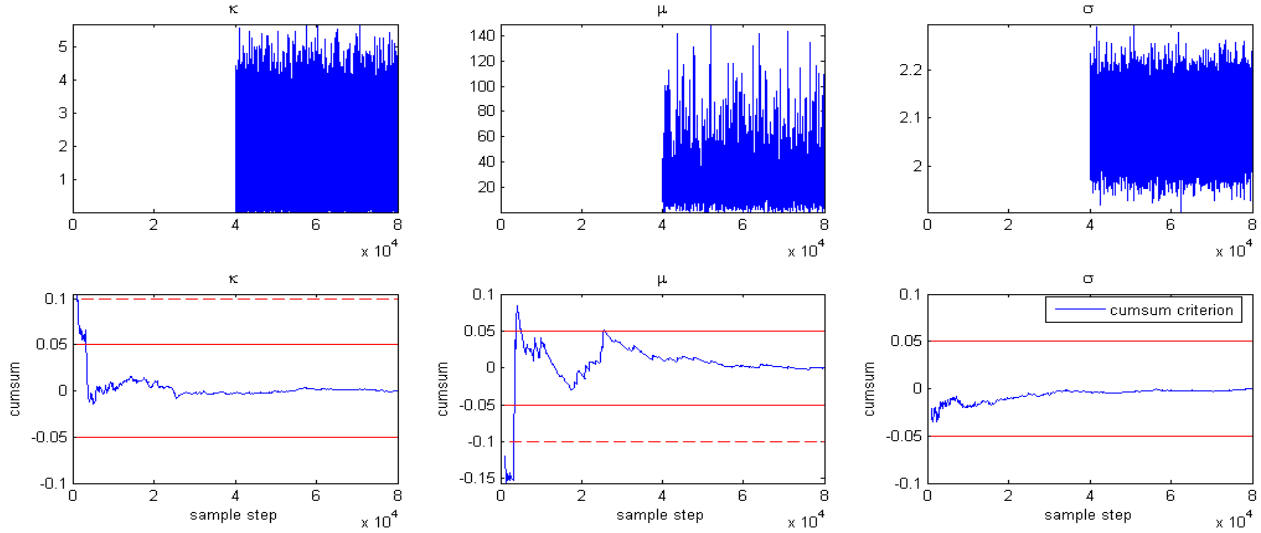


Figure 21: CIR - trace &amp; cumsum plots for uninformative priors

expected but seems reasonable. The full series, inclusive the outliers, needs some level of mean reversion (during our observation time). Otherwise, we would not have identified the outliers as outliers. Another reasonable finding is the higher long-term EUA price equilibrium for the short series because of the negative skewed outliers. As we reduced variability in the short series, the posterior for  $\sigma$  is also shifted towards zeros and close to the posterior of the GBM.

#### 10.2.4 Cox-Ingersoll-Ross (CIR) Process

For the Cox-Ingersoll-Ross (CIR) process on EUA prices we apply a thinning factor of ten to sample  $W = 120,000$  times the parameter vector  $\theta = (\kappa, \mu, \sigma)'$ . We dump the first 40,000 samples and thin the remaining path by a factor of two. Hence, we inference on 40,000 samples. We assume truncated Gaussian priors for all parameters and set the hyperparameters to the respective ML estimates and modified covariance values.<sup>127</sup> Comparable to the OU process we set the limits  $\mu_{\min} = 0.01$  and  $\mu_{\max} = 150$  for the long-term EUA price equilibrium. The trace and the cumsum plots (see figure 21) indicate convergence for all parameters featuring negligible correlations beside  $\widehat{Corr}(\kappa, \mu) = -0.331$ . A simultaneous sampling of  $\kappa$  and  $\mu$  could reduce auto-correlations which are only close to zero for lags larger than five. This could fundamentally improve mixing. However, our MCMC sampler is fast

<sup>127</sup>We set  $m_{\kappa} = \kappa^{ML} = 2.171$ ,  $m_{\mu} = \mu^{ML} = 19.258$ ,  $m_{\sigma} = \sigma^{ML} = 2.079$ ,  $s_{\kappa}^2 = \Sigma_{\kappa}^{ML} \cdot 1000 = 971.14$ ,  $s_{\mu}^2 = \Sigma_{\mu}^{ML} \cdot 1000 = 3854.5$ , and  $s_{\sigma}^2 = \Sigma_{\sigma}^{ML} \cdot 1000 = 1.9427$ .

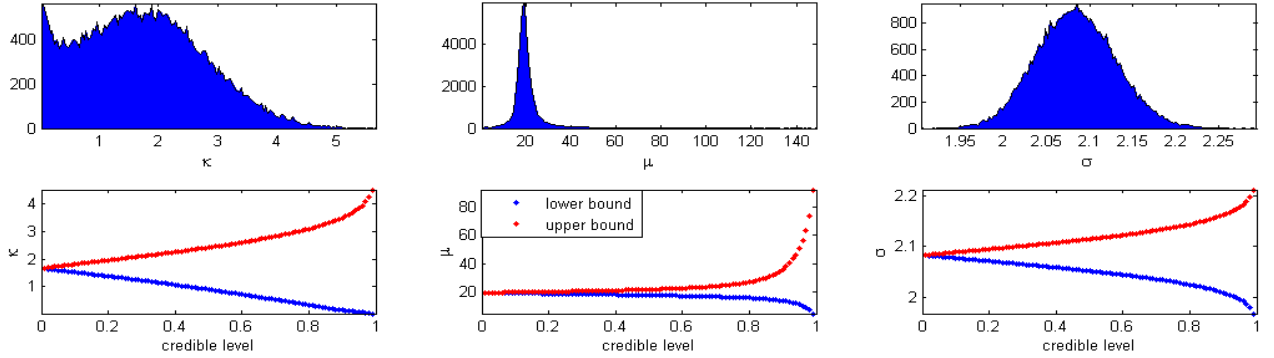


Figure 22: CIR - posteriors & credible intervals for uninformative priors

and we avoid to complicate the algorithm.

The mean posterior for  $\kappa$  is 1.71 corresponding to a half-life time of 4 months and 18 trading days (see figure 22). This is nearly the same mean posterior like in the OU model ( $\kappa = 1.70$ ). Moreover, both posteriors are similarly shaped (standard deviation 1.03 vs. 1.07 in OU). Because of the more pronounced second mode, the CIR posterior of  $\kappa$  seems more dominated by its prior. The mean posterior and the standard deviation of the long-term equilibrium  $\mu$  are 21.67€ and 10.71€. This is slightly higher than in the OU model (19.75€ and 10.69€). According to our posterior, values of  $\mu$  above 70€ (90€) have a probability of 1% (0.5%). The mean posterior of  $\sigma$  is 2.08 which we cannot compare to the GBM or OU process. In the CIR model the variable  $\sigma$  is not a volatility level but a volatility factor. Nevertheless, the posterior of  $\sigma$  is a relatively concentrated distribution.

We applied the same algorithm to the short time-series.<sup>128</sup> In figure 23 you can find the respective posteriors<sup>129</sup> and quantile plots while we have summarized some statistics of the posteriors on the full and the short time-series in the tables 10 and 11.

The CIR estimation results of the short time-series are consistent to those of the OU process. The level of mean reversion  $\kappa$  and the volatility  $\sigma$  is reduced while the long-term EUA price equilibrium  $\mu$  is higher. However, we should keep in mind that our CIR process models mean reversion for the EUA prices in contrast to the OU process where reversion is defined on log-prices. Consequently, the mean priors of  $\hat{\mu}^{CIR}$  and  $\hat{\mu}^{OU}$  are hardly to compare as

<sup>128</sup>We sample the parameters  $W = 120,000$  times with a thinning factor of ten and dump the first 40,000. We thin the remaining samples by a factor of two. This results in a sample of 40,000. We use the same priors with hyperparameters resulting from ML estimation:  $m_\kappa = \kappa^{ML} = 1.1709$ ,  $m_{\tilde{\mu}} = \tilde{\mu}^{ML} = 20.487$ ,  $m_\sigma = \sigma^{ML} = 1.8289$ ,  $\zeta_\kappa^2 = \Sigma_\kappa^{ML} \cdot 1000 = 891.11$ ,  $\zeta_{\tilde{\mu}}^2 = \Sigma_{\tilde{\mu}}^{ML} \cdot 1000 = 11632$ , and  $\zeta_\sigma^2 = \Sigma_\sigma^{ML} \cdot 1000 = 1.6387$ .

<sup>129</sup>Comparable to the OU estimation, auto-correlation of our MCMC parameter paths falls slowly. Fortunately, visual inspection of the trace plots and cumsum criterion support a good mixing and convergence.

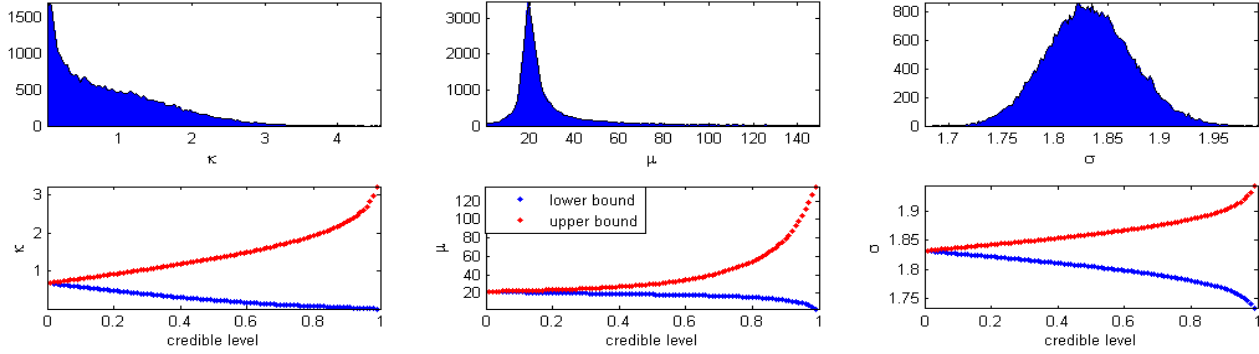


Figure 23: CIR (short) - posteriors &amp; credible intervals for uninformative priors

full series	mean	std. dev.	skewness	excess kurtosis
$\kappa$	1.713	1.034	0.353	-0.410
$\mu$	21.667	10.712	5.014	34.863
$\sigma$	2.084	0.046	0.124	0.046

Table 10: results of CIR model (full series)

generally  $\hat{\mu} \neq \sum_{w=1}^W \exp(\tilde{\mu}_w) / W$ .

### 10.2.5 Constant Elasticity of Variance (CEV) Process

For the constant elasticity of variance (CEV) process we sample  $W = 60,000$  (burn-in  $B = 20,000$ ) times the parameter vector  $\theta = (\kappa, \mu, \sigma, \gamma)'$  with a thinning factor of ten. This could eliminate auto-correlation for lags larger than five. We assume truncated Gaussian priors for all parameters and set the hyperparameters to the respective ML estimates and modified covariance values.<sup>130</sup> Again, we set the limits  $\mu_{\min} = 0.01$  and  $\mu_{\max} = 150$  for the long-term EUA price equilibrium.

A characteristic of the CEV process is the nearly deterministic relationship between  $\sigma$  and  $\gamma$  ( $\widehat{Corr}(\sigma, \gamma) = -0.983$ ). This results in extreme high and long lasting MCMC parameter auto-correlations which prevent convergence of our MCMC algorithm in practicable computing time.<sup>131</sup> Hence, we implement a bivariate random walk Metropolis-Hastings algorithm which generates joint proposals  $(\sigma^*, \gamma^*)'$  from a bivariate Gaussian distribution driven by

<sup>130</sup>We set  $m_\kappa = \kappa^{ML} = 2.0081$ ,  $m_\mu = \mu^{ML} = 19.276$ ,  $m_\sigma = \sigma^{ML} = 1.1109$ ,  $m_\gamma = \gamma^{ML} = 0.70608$ ,  $s_\kappa^2 = \Sigma_\kappa^{ML} \cdot 1000 = 720.64$ ,  $s_\mu^2 = \Sigma_\mu^{ML} \cdot 1000 = 3986.5$ ,  $s_\sigma^2 = \Sigma_\sigma^{ML} \cdot 1000 = 61.352$ , and  $s_\gamma^2 = \Sigma_\gamma^{ML} \cdot 1000 = 5.672$ .

<sup>131</sup>Other parameter correlations are not problematic and reach from  $\widehat{Corr}(\mu, \kappa) = -0.396$  up to  $\widehat{Corr}(\mu, \gamma) = 0.039$ .

short series	mean	std. dev.	skewness	excess kurtosis
$\kappa$	0.863	0.740	0.972	0.515
$\mu$	29.388	22.153	2.715	8.074
$\sigma$	1.833	1.833	0.112	-0.010

Table 11: results of CIR model (short series)

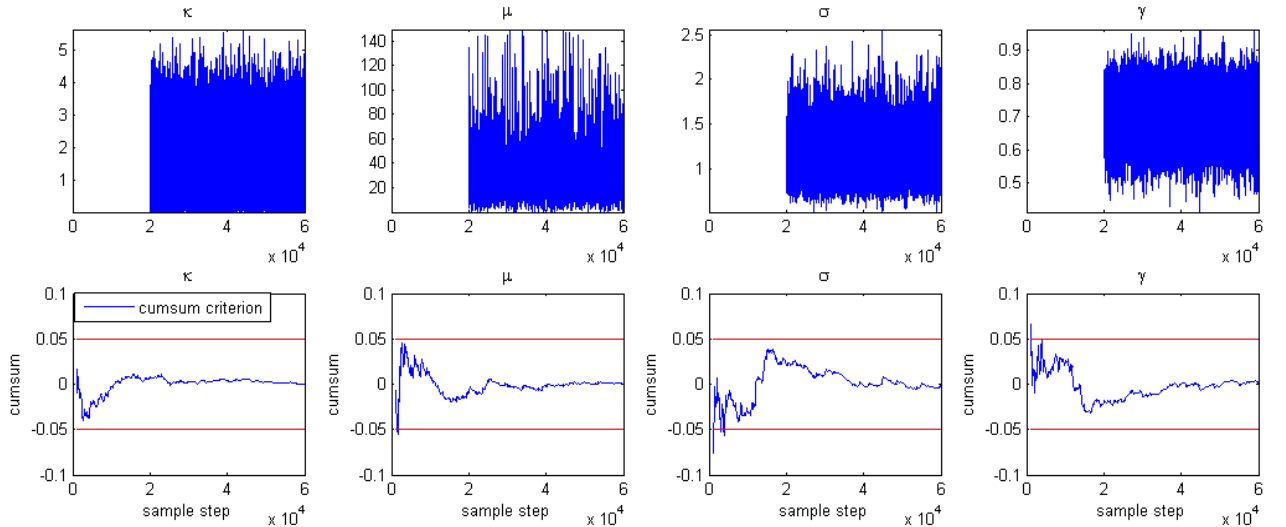


Figure 24: CEV - trace &amp; cumsum plots for uninformative priors

the correlation of the ML estimates ( $\widehat{Corr}_{ML}(\sigma, \gamma) = -0.99535$ ). The trace and cumsum plots indicate that our modification has been successful (see figure 24).

The mean posterior for  $\kappa$  is 1.57 corresponding to a half-life time of 5 months and 6 trading days (see figure 25). This is a lower mean reversion rate than in the OU and CIR model. However, the form of the CEV posterior for  $\kappa$  resembles that of the OU model. The dominance of the prior seems even higher as the two modes are more pronounced. The posterior on the long-term EUA equilibrium  $\mu$  has a mean value of 22.79€ which is substantially above the posterior means of the OU and CIR models (19.75€ and 19.43€). Indeed, the difference between 22.79€ and 19.43€ is larger than the standard deviation of  $\mu$  in the CIR model (2.83€).

The mean posterior of  $\sigma$  is 1.21 which is lower than 2.08 in the CIR model. In contrast, its standard deviation is substantially larger (0.28€ compared to 0.05€ for CIR). This higher level of uncertainty originates from the additional elasticity parameter  $\gamma$ . Its posterior (mean 0.69) is nearly fully above 0.5 corresponding to the CIR model. Consequently, a lower value

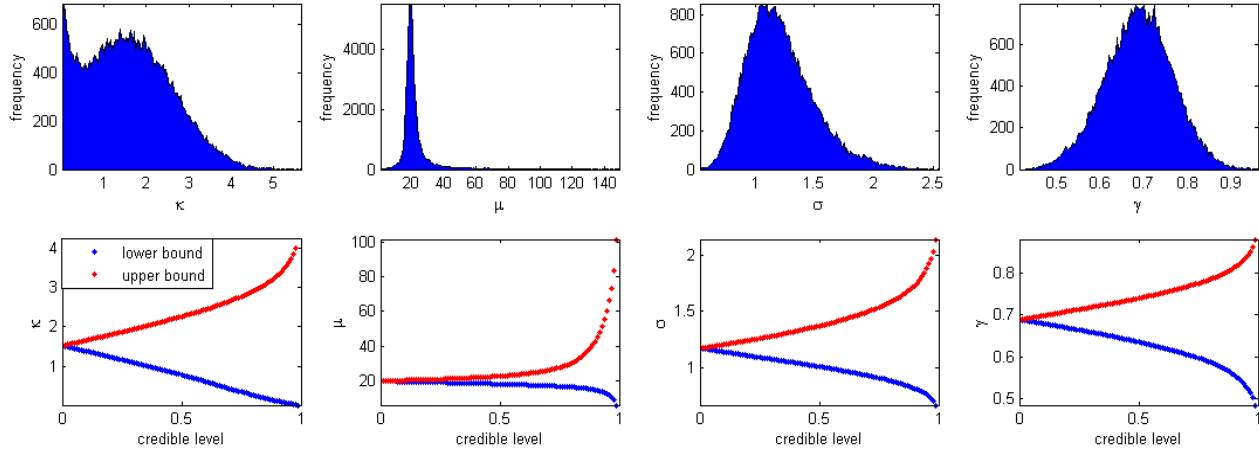


Figure 25: CEV - posteriors & credible intervals for uninformative priors

of  $\sigma$  in the CEV model can express the same volatility in the log-returns. We applied the same algorithm to the short time-series.<sup>132</sup> In figure 26 you can find the respective posteriors and quantile plots while we have summarized some statistics of the posteriors on the full and the short time-series in the tables 12 and 13.

full series	mean	std. dev.	skewness	excess kurtosis
$\kappa$	1.573	0.986	0.389	0.421
$\mu$	22.790	12.449	4.468	26.170
$\sigma$	1.210	0.283	0.759	0.757
$\gamma$	0.687	0.077	-0.074	-0.063

Table 12: results of CEV model (full series)

The statistics for the CEV model are mainly consistent to those of the OU and CIR models. The posterior uncertainty on  $\kappa$  reduces for the short time-series. There is more evidence for a lower mean reversion (half-life time of 9 months and 11 trading days vs. 5 months and 6 trading days). Comparable to the other processes the uncertainty on the long-term EUA price equilibrium  $\mu$  is higher for the short time-series.

In contrast to the other models, the posterior of  $\sigma$  has shifted to a higher level. This could seduce us to conclude that the model finds more volatility for the time-series without

<sup>132</sup>We sample the parameters  $W = 130,000$  (burn-in  $B = 60,000$ ) times with a thinning factor of ten. We use the same priors with hyperparameters resulting from ML estimation:  $m_\kappa = \kappa^{ML} = 1.17149$ ,  $m_{\tilde{\mu}} = \tilde{\mu}^{ML} = 20.486$ ,  $m_\sigma = \sigma^{ML} = 1.8434$ ,  $m_\gamma = \gamma^{ML} = 0.49732$ ,  $\varsigma_\kappa^2 = \Sigma_\kappa^{ML} \cdot 1000 = 739.5$ ,  $\varsigma_{\tilde{\mu}}^2 = \Sigma_{\tilde{\mu}}^{ML} \cdot 1000 = 11482$ ,  $\varsigma_\sigma^2 = \Sigma_\sigma^{ML} \cdot 1000 = 202.49$ , and  $\varsigma_\gamma^2 = \Sigma_\gamma^{ML} \cdot 1000 = 6.8918$ .

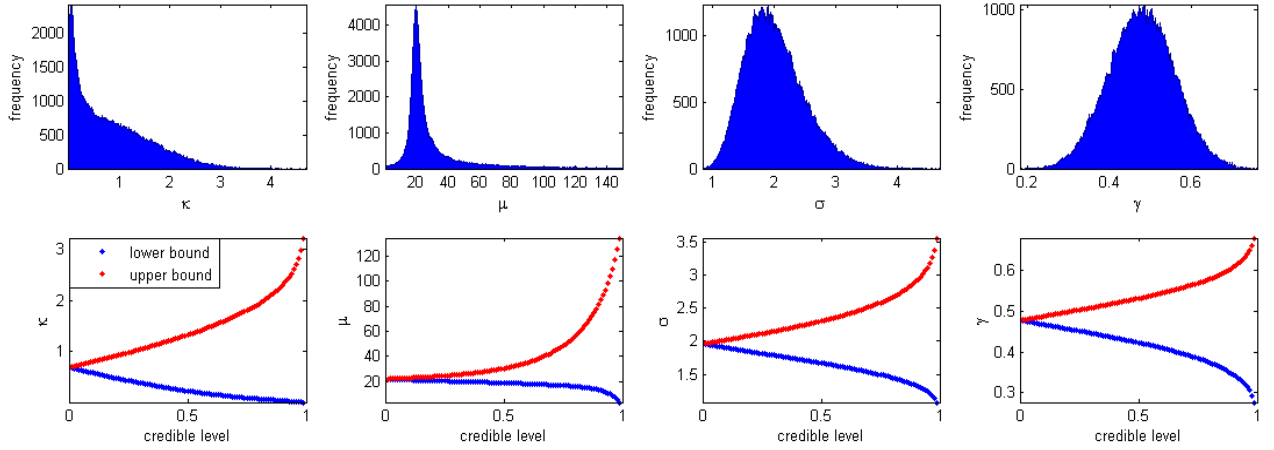


Figure 26: CEV (short) - posteriors &amp; credible intervals for uninformative priors

short series	mean	std. dev.	skewness	excess kurtosis
$\kappa$	0.862	0.741	0.954	0.407
$\mu$	29.827	22.268	2.627	7.446
$\sigma$	2.016	0.483	0.699	0.690
$\gamma$	0.477	0.080	-0.010	-0.116

Table 13: results of CEV model (short series)

outliers. However, this interpretation ignores the fact of a reduced elasticity parameter  $\gamma$  which counteracts this effect. The exact interplay of  $\sigma$  and  $\gamma$  is not intuitive. Hence, we do not try to interpret it.

### 10.2.6 Student's t GARCH

For the Student's t *GARCH* (1, 1) model we sample  $W = 46,000$  (burn-in  $B = 6,000$ ) times the parameter vector  $\theta = (\gamma_0, \gamma_1, \omega, \alpha, \beta, v)'$  without any thinning. We assume uniform priors on  $(0, 2]$  for the GARCH parameters<sup>133</sup> and a prior mean of five, i.e.  $\lambda_0 = 1/5$ , for the exponential prior on the degree of freedom  $v$ .<sup>134</sup> For the joint prior of the regression

<sup>133</sup>The sampled GARCH parameters realized far away from 2 in our simulation studies. Hence, we see no problem for these narrow priors.

<sup>134</sup>For our Griddy Gibbs sampler we create the fixed points 2.1, 2.2, ..., 10.0. In our simulations the sampled  $v$  have been far below 10. Therefore, we see no problem with that low upper limit.



parameters  $\gamma_0$  and  $\gamma_1$ , we set  $m_0 = (0, 0)'$  and<sup>135</sup>

$$\Sigma_0 = \begin{pmatrix} 0.23434 & -6.1184 \\ -6.1184 & 174.14 \end{pmatrix}.$$

Compared to the GBM and GBMJ, our GARCH sample paths are rather short - especially as we do not thin because of the time consuming computing. Unfortunately, our samples show high auto-correlation. Fortunately, the cumsum criterion indicates convergence (see the upper six subplots in figure 27). Moreover, we analyze the behavior of the posterior means

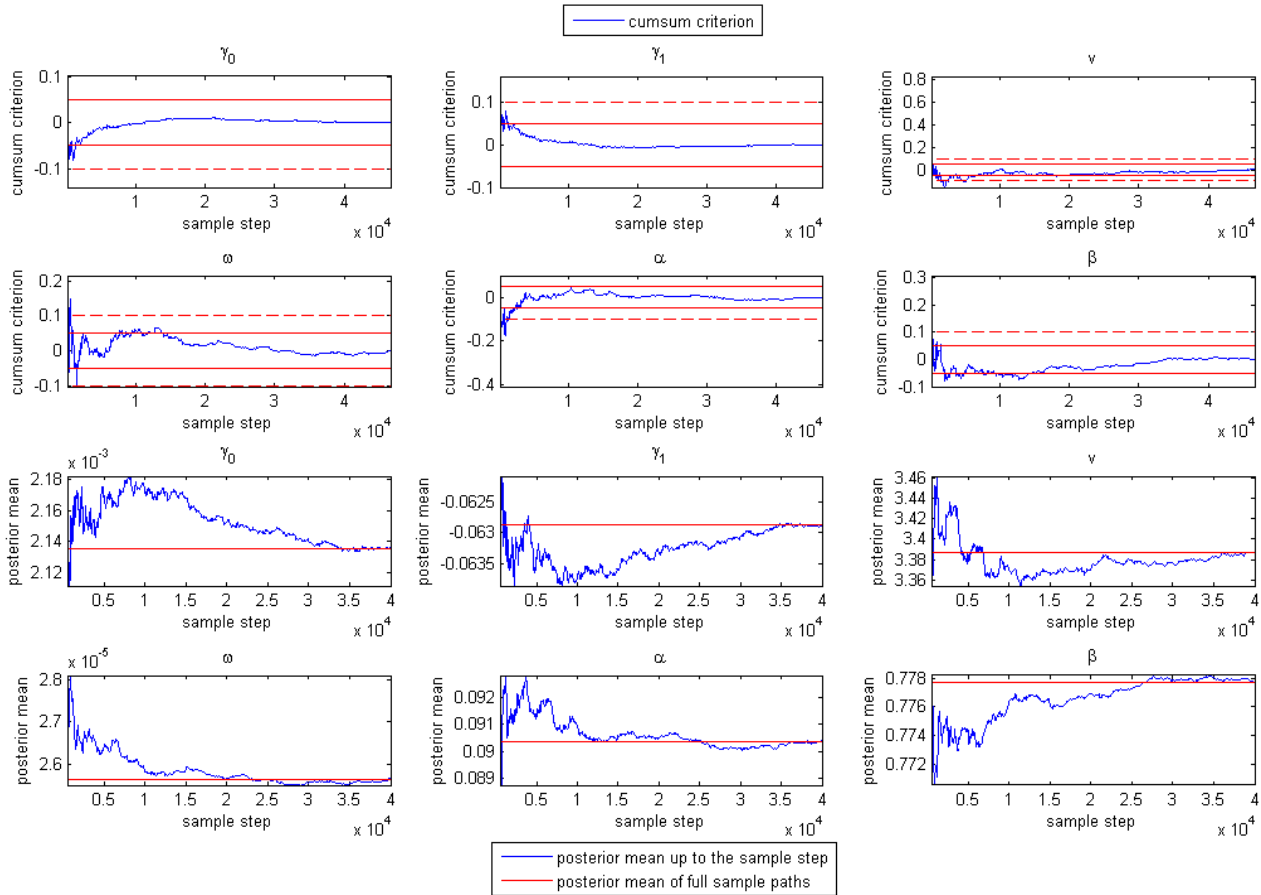


Figure 27: Student's t GARCH - cumsum and posterior mean plots

when we vary the sample size. They appear stable for all parameters with a variation on a

<sup>135</sup> $\Sigma_0$  is based on a conservative (less informative) scaling factor  $c = 1/20$  and corresponds to a parameter correlation of  $-0.96$ .

very low level (see lower six subplots in figure 27). A further sampling seems not necessary. Our MCMC algorithm is rather efficient as it accounts for the substantial correlation between the regression parameters and between the GARCH parameters (see table 14).

	$\gamma_0$	$\gamma_1$	$v$	$\omega$	$\alpha$
$\gamma_1$	-0.965				
$v$	0.004	-0.002			
$\omega$	0.010	-0.012	0.187		
$\alpha$	0.008	-0.006	0.270	0.307	
$\beta$	-0.015	0.013	-0.028	-0.745	-0.748

Table 14: parameter correlation of Student's t GARCH model

An encouraging finding is that the EURIBOR interest rates have a negligible influence on the log-returns (figure 28). This supports our approach with synthetic Dec09 spot rates. The posterior of  $v$  concentrates between 2.5 and 4.5. Hence, the implementation of fat tail errors seems adequate.<sup>136</sup> The current variance is by far more influenced by the last variance than by past shocks in the log-return series ( $\alpha = 0.09$  vs.  $\beta = 0.78$ ). Unfortunately, the log-returns possess no stationarity for 53% of all parameter samples (see upper subplot in figure 29). The long-term annual volatility<sup>137</sup> of the log-returns is below the annual volatility of the GBM but above that of the GBMJ (see lower subplot in figure 29). All things considered, the GARCH process can replicate the variability of the log-returns (see figure 30).

We apply the same algorithm to the short time-series.<sup>138</sup> In figure 31 you can find the respective posteriors<sup>139</sup> and quantile plots while we have summarized some statistics of the posteriors on the full and the short time-series in the tables 15 and 16.

The most remarkable difference between the whole and the short time-series is that the short one has a by far larger annual drift (0.73 vs. 0.53). This result we have expected as the whole series features more left skewness (-0.87 vs. -0.38) than the series without the turbulent April/ May 2006. Consequently, the higher degree of freedom is in the line of

<sup>136</sup>Indeed, the EUA log-returns even more support heavy tails than the MSCI Canadian log-returns (mean posterior 9.2) analyzed by Rachev et al. (2008), pp. 211.

<sup>137</sup>Of course, we can only calculate a (finite) volatility for stationary GARCH processes (persistence measure below one).

<sup>138</sup>We generate  $W = 57,000$  samples without thinning and dump the first 17,000. We use the same priors with the exception that we set

$$\Sigma_0 = \begin{pmatrix} 0.23786 & -6.1823 \\ -6.1823 & 175.29 \end{pmatrix}$$

which corresponds to a correlation of -0.96.

<sup>139</sup>Although there is huge auto-correlation the cumsum criterion and the posterior means favor convergence.

full series	mean	std. dev.	skewness	excess kurtosis
$\gamma_0$	$2.135 \cdot 10^{-3}$	$2.076 \cdot 10^{-3}$	$-8.788 \cdot 10^{-4}$	0.047
$\gamma_0^{p.a.} = \gamma_0/\Delta$	0.534	0.519	$-8.788 \cdot 10^{-4}$	0.047
$\gamma_1$	-0.063	0.054	0.001	0.037
$v$	3.387	0.462	0.499	0.353
$\omega$	$2.565 \cdot 10^{-5}$	$8.530 \cdot 10^{-6}$	0.755	1.035
$\alpha$	0.090	0.021	0.553	0.493
$\beta$	0.778	0.040	-0.536	0.764

Table 15: results of Student's t GARCH model (full series)

short series	mean	std. dev.	skewness	excess kurtosis
$\gamma_0$	$2.926 \cdot 10^{-3}$	$2.082 \cdot 10^{-3}$	$-8.188 \cdot 10^{-4}$	0.034
$\gamma_0^{p.a.} = \gamma_0/\Delta$	0.732	0.521	$-8.188 \cdot 10^{-4}$	0.034
$\gamma_1$	-0.078	0.055	0.015	0.032
$v$	3.655	0.544	0.715	1.048
$\omega$	$2.640 \cdot 10^{-5}$	$9.558 \cdot 10^{-6}$	0.883	1.550
$\alpha$	0.094	0.022	0.453	0.318
$\beta$	0.776	0.044	-0.435	0.459

Table 16: results of Student's t GARCH model (short series)

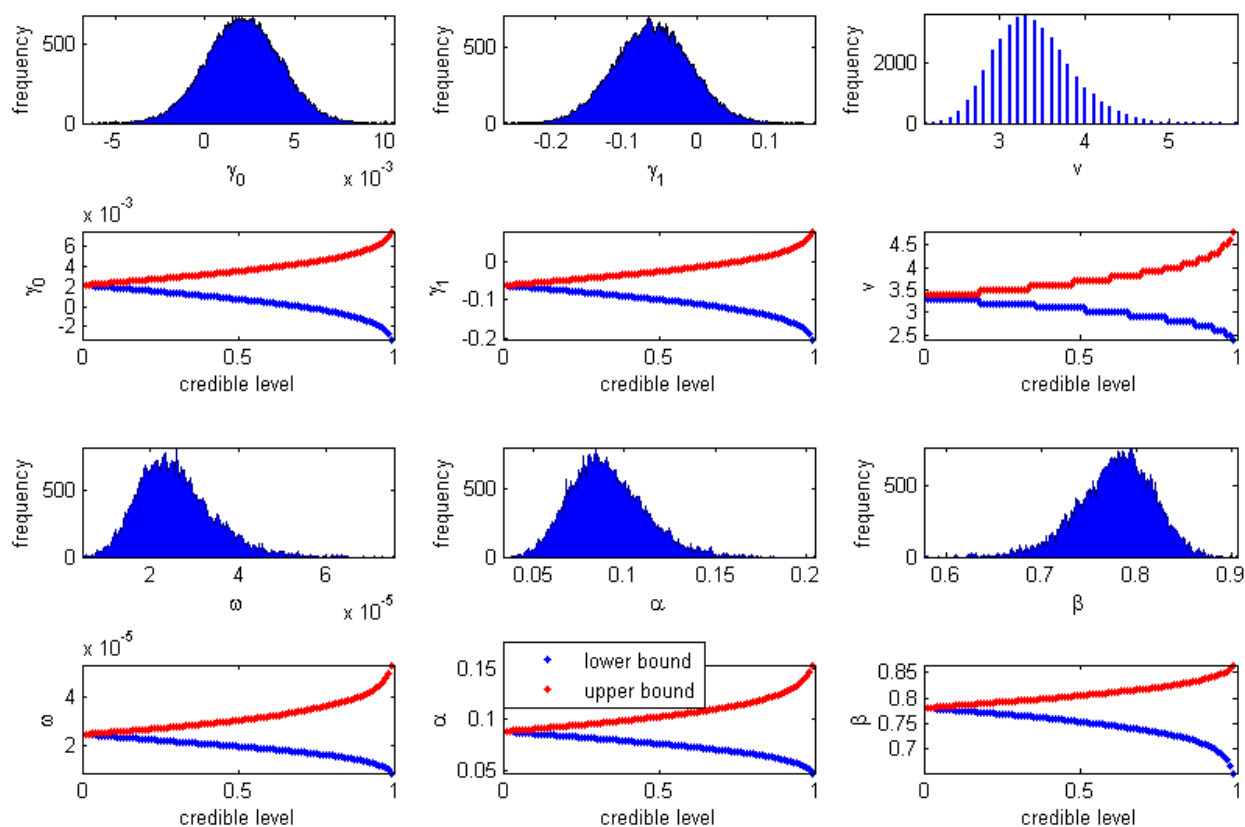


Figure 28: Student's t GARCH - posteriors &amp; credible intervals

this argumentation. Surprisingly, the difference between the posteriors is not large. We are really surprised by the stability of the GARCH estimates which do not change much. The mean long-term annual volatility  $\frac{\omega}{1-\alpha-\beta} \frac{\nu}{\nu-2}$  also changed slightly from 0.350 to 0.339. The non-stationarity of the log-returns reduced from 53% to 41%.

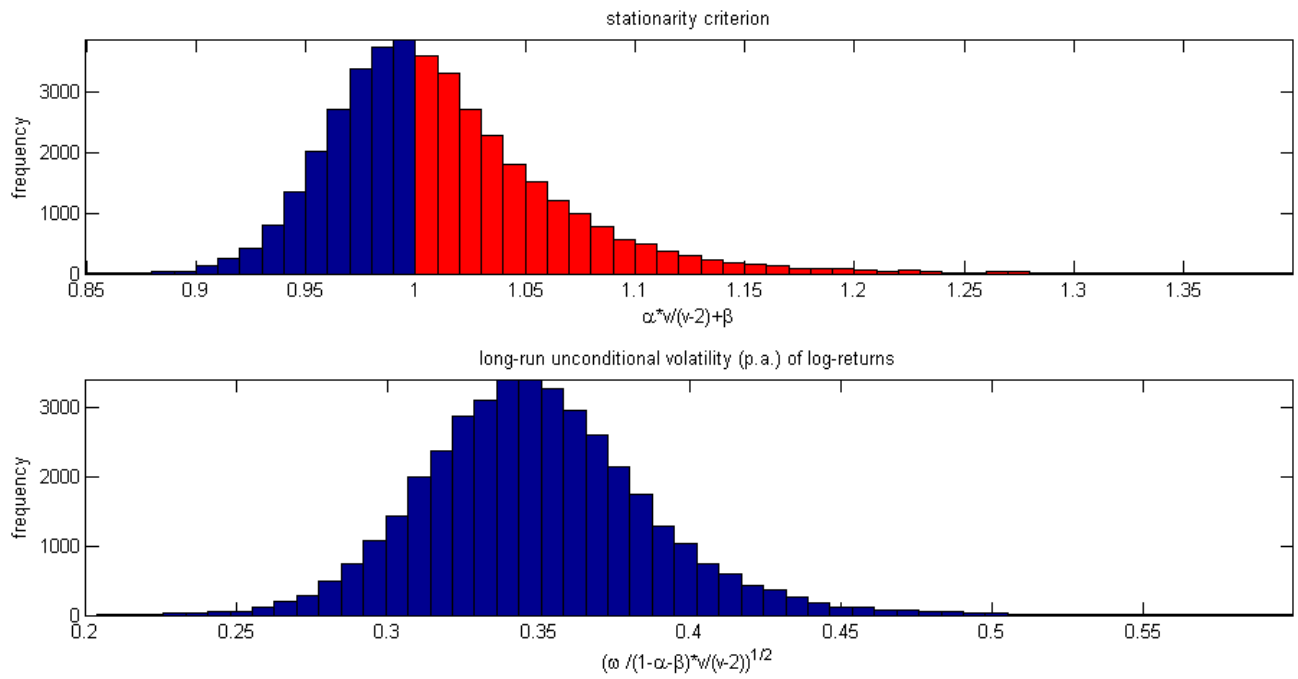


Figure 29: Student's t GARCH - stationarity & long-run unconditional volatility

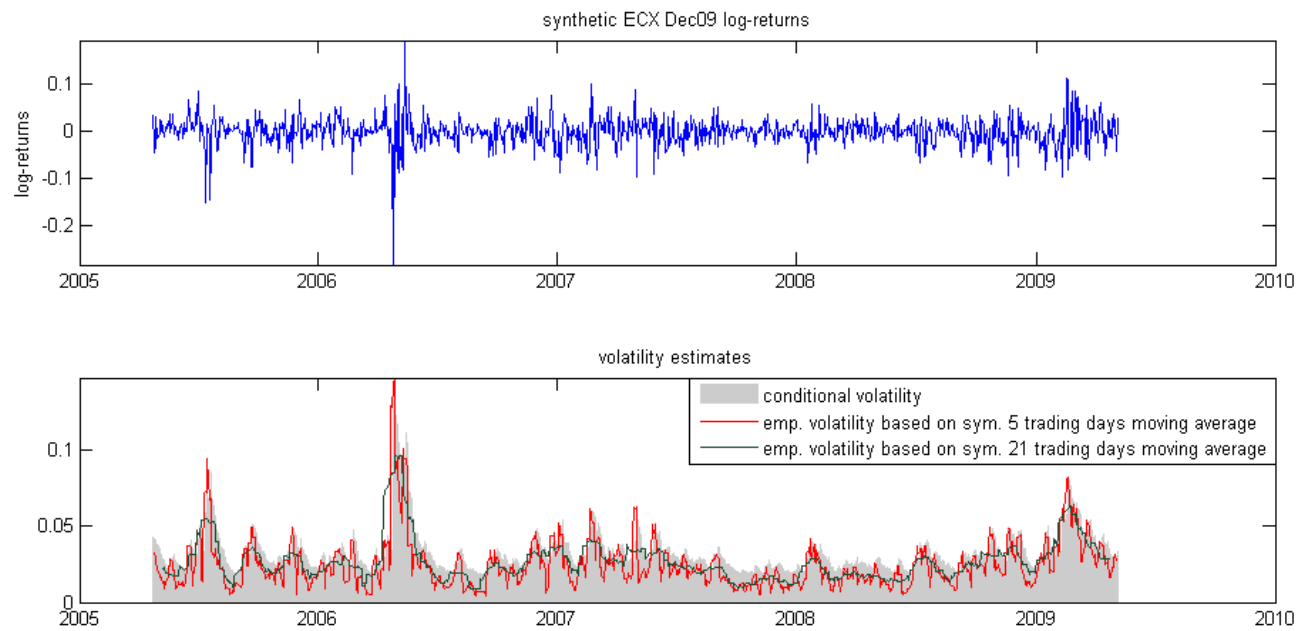


Figure 30: Student's t GARCH - empirical volatility

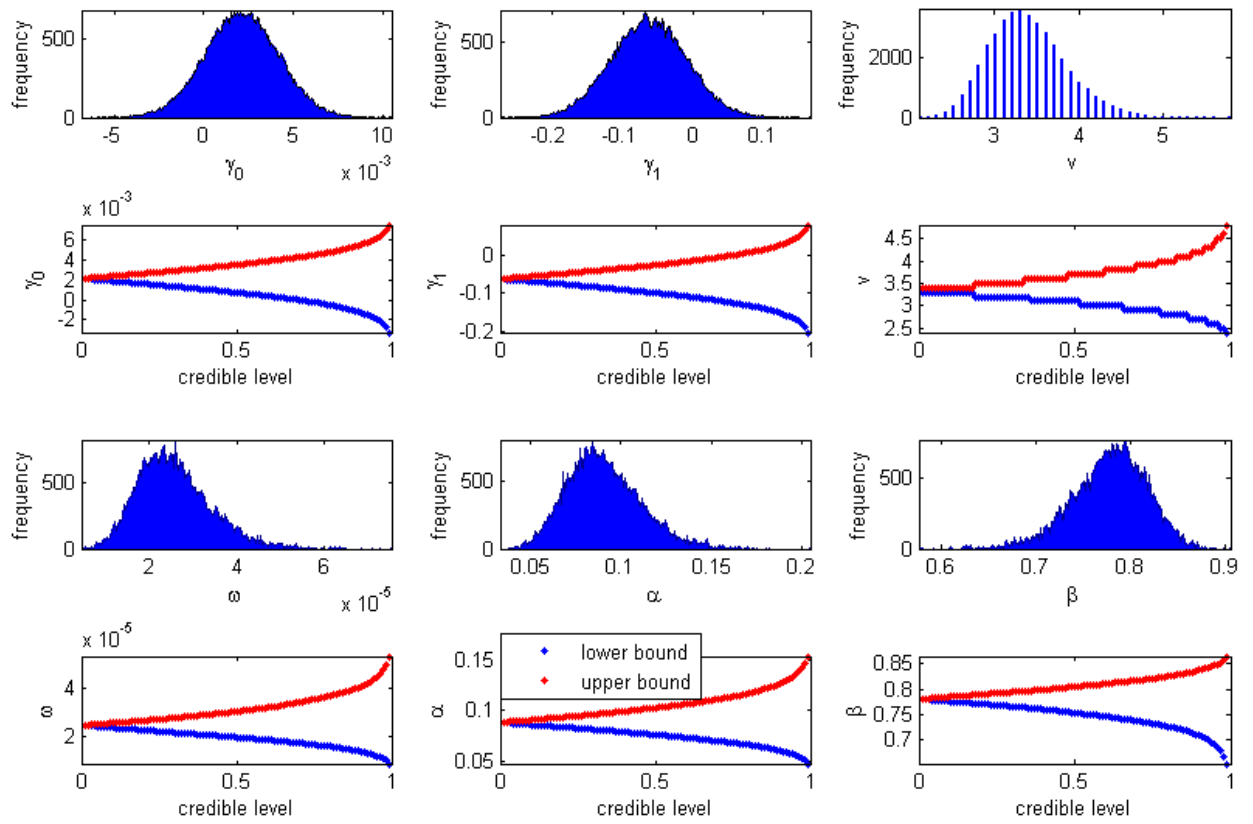


Figure 31: Student's t GARCH (short) - posteriors &amp; credible intervals

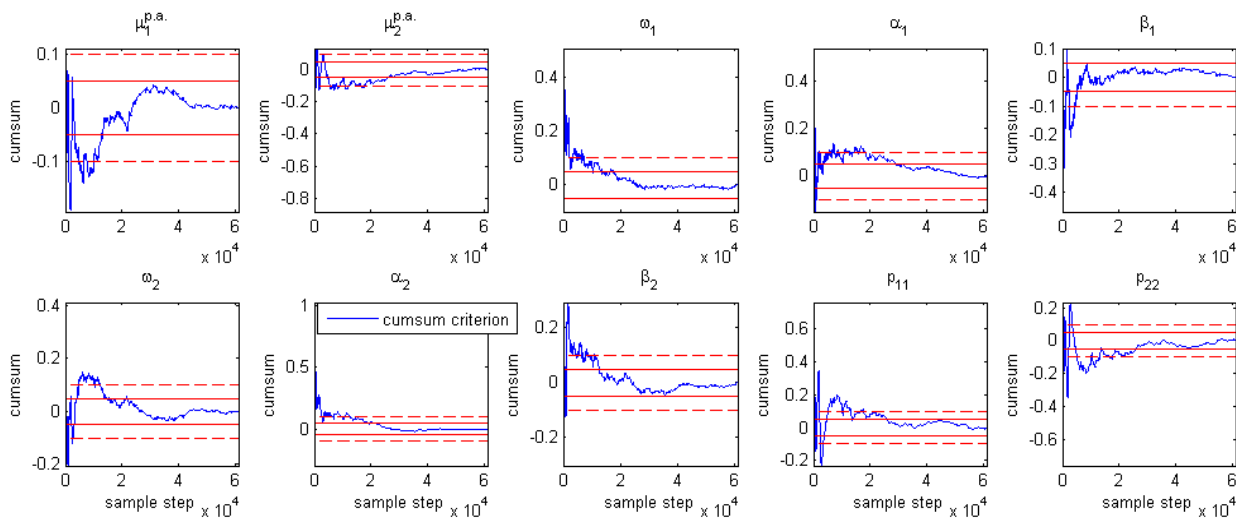


Figure 32: MS GARCH - cumsum plots

### 10.2.7 Markov Switching (MS) GARCH (1, 1)

For the Markov switching GARCH (1,1) model we sample  $W = 60,000$  (burn-in  $B = 20,000$ ) times the parameter values  $\theta = \left( \{\mu_k, \omega_k, \alpha_k, \beta_k, p_{kk}\}_{k=1,2} \right)'$  without any thinning. We assume uniform priors on  $(0, 5]$  for the GARCH parameters<sup>140</sup> and flat priors on  $p_{11}$  and  $p_{22}$  ( $a_{0k} = b_{0k} = 1$ ). For  $\mu_k$ ,  $k = 1, 2$ , we set  $m_0 = 0$  and  $s_0 = 10$ .

Comparable to the Student's t GARCH model we do not thin the MCMC parameter paths for computational reasons. This might be critical as all parameters feature substantial autocorrelation at least up to lag 20. Nevertheless, the cumsum criterion indicates convergence (see figure 32). The plots on the mean posteriors support this finding (see figure 33). For all parameters they appear stable with a variation on a very low level. A further sampling seems not necessary. Of course, we could improve mixing behavior by a simultaneous parameter sampling. We, however, find no critical parameter correlation.<sup>141</sup>

The posteriors of  $\mu_1^{p.a.}$  and  $\mu_2^{p.a.}$  are rather disjunct (see figure 34). The posterior of  $\mu_1^{p.a.}$  (mean 0.50) is nearly fully above zero which means that in "normal" times there is an upwards trend in the prices. However, in "turbulent" times there is a relatively harsh negative trend with  $\mu_2^{p.a.}$  (mean  $-4.56$ ) nearly fully below zero. The posterior of  $p_{11}$  reflects a high level of stability for regime 1 while there is only a 1/4 change ( $p_{22}$ ) to stay in regime 2. As regime

<sup>140</sup>The sampled GARCH parameters realized far away from 5 in our simulations. Hence, we see no problem for these narrow priors.

<sup>141</sup>There is parameter correlation between  $\widehat{Corr}(\omega_1, \beta_1) = -0.667$  and  $\widehat{Corr}(\mu_1, p_{22}) = 0.357$ .

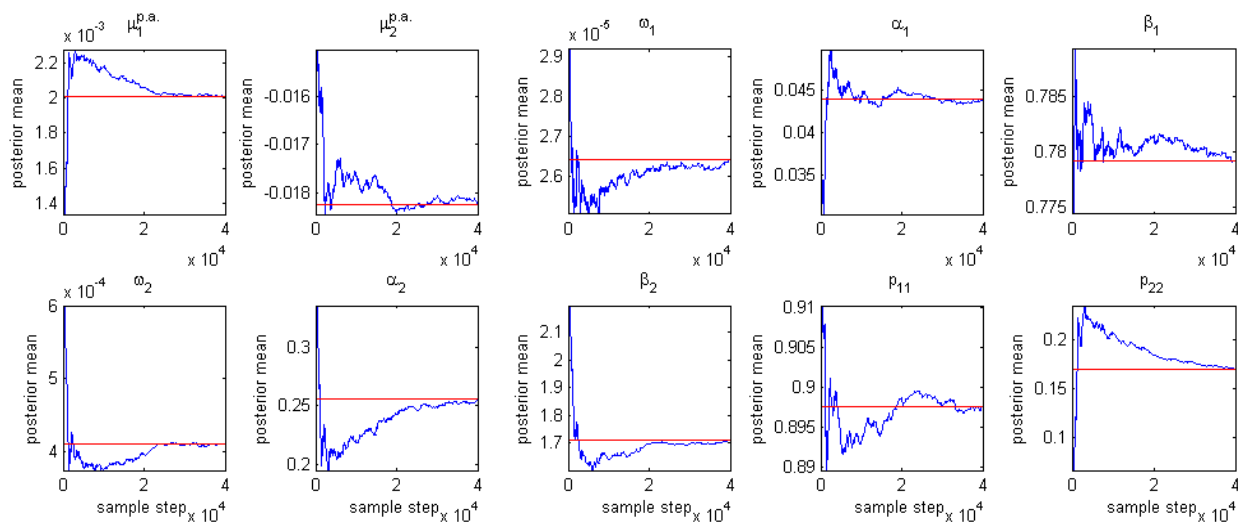


Figure 33: MS GARCH - mean posterior plots

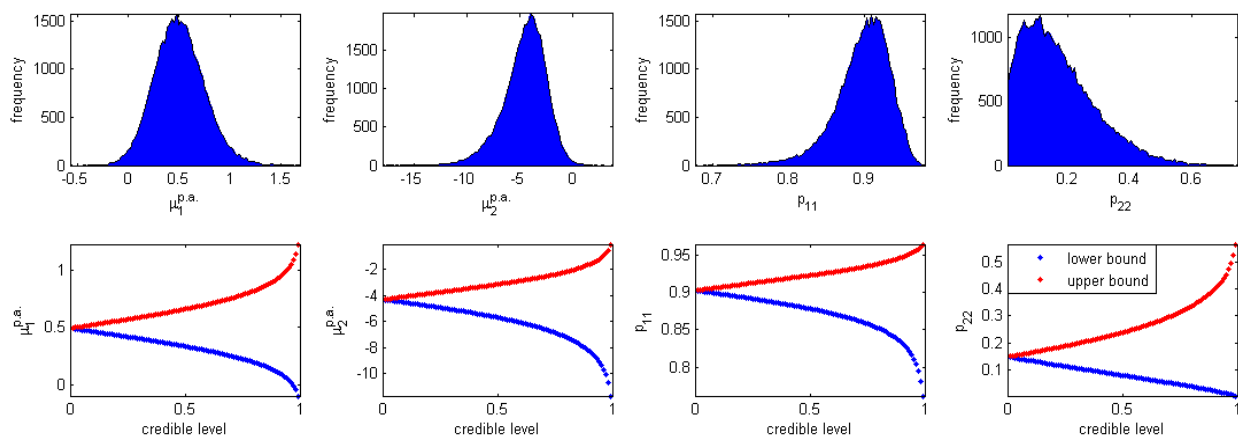


Figure 34: MS GARCH - posteriors & credible intervals for mean log-returns & transition probabilities



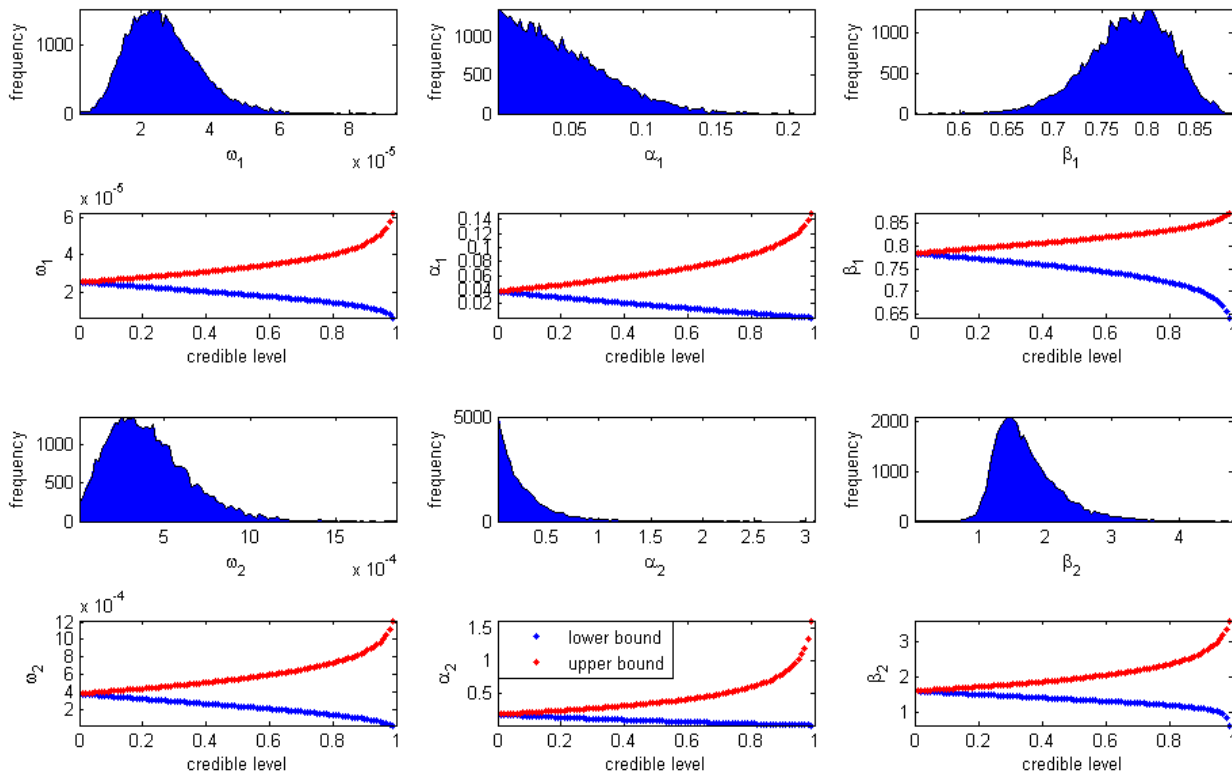


Figure 35: MS GARCH - posteriors & credible intervals for GARCH parameters

2 has been estimated as a rare event, it is a reasonable fact that the uncertainty on  $p_{22}$  is larger than on  $p_{11}$  (std. dev. 0.12 vs. 0.04).

The GARCH posteriors of regime 1 correspond to the GARCH posteriors of the Student's t GARCH model with a slight reduction in the extreme values (see figure 35). In contrast, the posteriors of regime 2 feature extreme non-stationarity and uncertainty (see upper subplot in figure 36). While the persistence measure of regime 1 is nearly fully below one that of regime 2 is nearly fully above. Hence, we can only calculate the long-term variance for regime 1 (see lower subplot in figure 36).

The MS GARCH model is also able to reproduce the variability of the volatility where regime 2 describes the "turbulent" times (see figure 37).

We apply the same algorithm to the short time-series.<sup>142</sup> In the figures 38 and 39 you can find the respective posteriors<sup>143</sup> and quantile plots while we have summarized some statistics

<sup>142</sup>We generate  $W = 90,000$  samples without thinning and dump the first 50,000. We use the same priors as for the full time series.

<sup>143</sup>After roughly 50,000 repetitions the cumsum criterion, the trace and posterior mean plots favor conver-

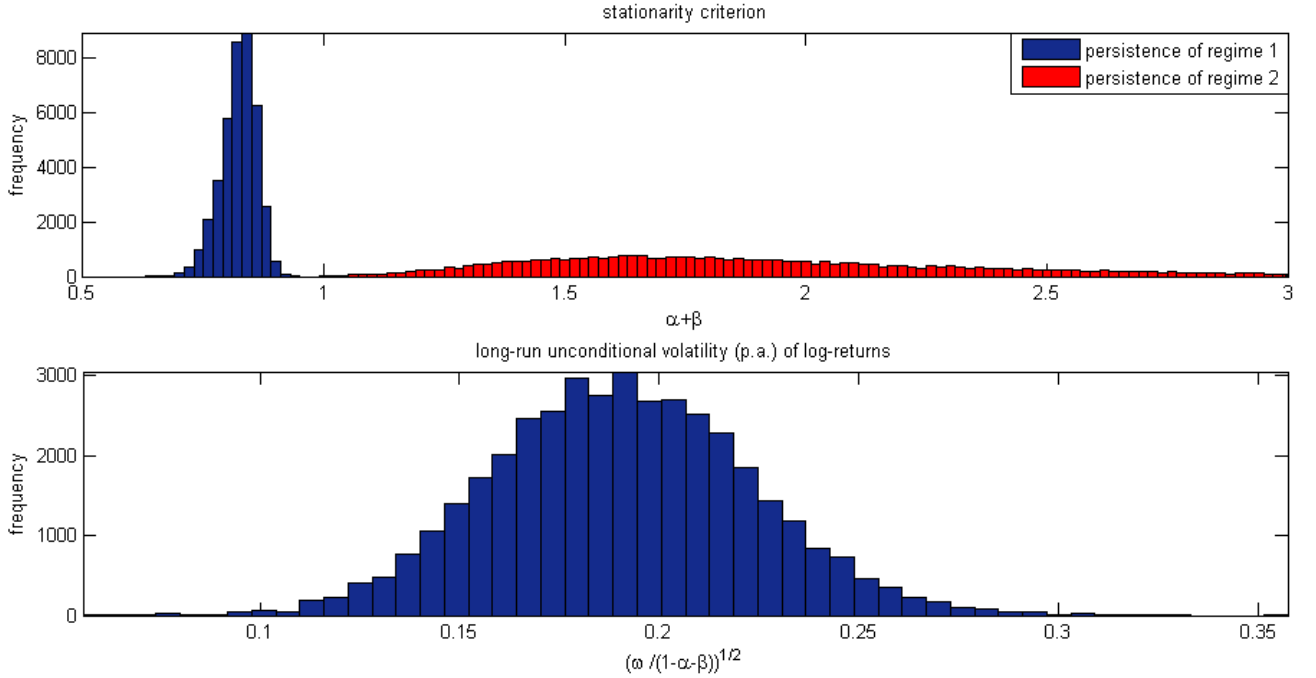


Figure 36: MS GARCH - stationarity & long-run unconditional volatility

of the posteriors on the full and the short time-series in the tables 17 and 18.

A remarkable result is that the annual drift rate  $\mu_1^{p.a.}$  for the short time-series is nearly the same as for the full time-series (0.49 vs. 0.50). This statement also holds for the GARCH parameters  $\omega_1$ ,  $\alpha_1$ ,  $\beta_1$ , and  $p_{11}$ . An interpretation could be that the dumped outliers have been fully allocated to regime 2 in our estimation on the whole time-series. In contrast, the parameters of regime 2 have changed to a larger extent. The annual drift  $\mu_2^{p.a.}$  is substantially less extreme ( $-2.84$  vs.  $-4.55$ ) because negatively skewed outliers have been ignored. The GARCH terms have changed fundamentally. For the whole time-series the variance in  $t$  is dominated by the variance in  $t-1$  while for the short time-series the past estimation error in  $t-1$  is most important. As the level of  $p_{11}$  falls and that of  $p_{22}$  rises regime 2 undertakes the task to describe above average rather than extreme log-returns (see figure 40). Moreover, the persistence measure estimates stationarity for 41% of all sampled parameter tuples (far more than roughly 0% for the full time-series).

---

gence although there are substantial auto-correlations in the parameter paths and parameter correlations in the range from  $\widehat{Corr}(\beta_1, \omega_1) = -0.642$  to  $\widehat{Corr}(p_{11}, \omega_2) = 0.568$ .

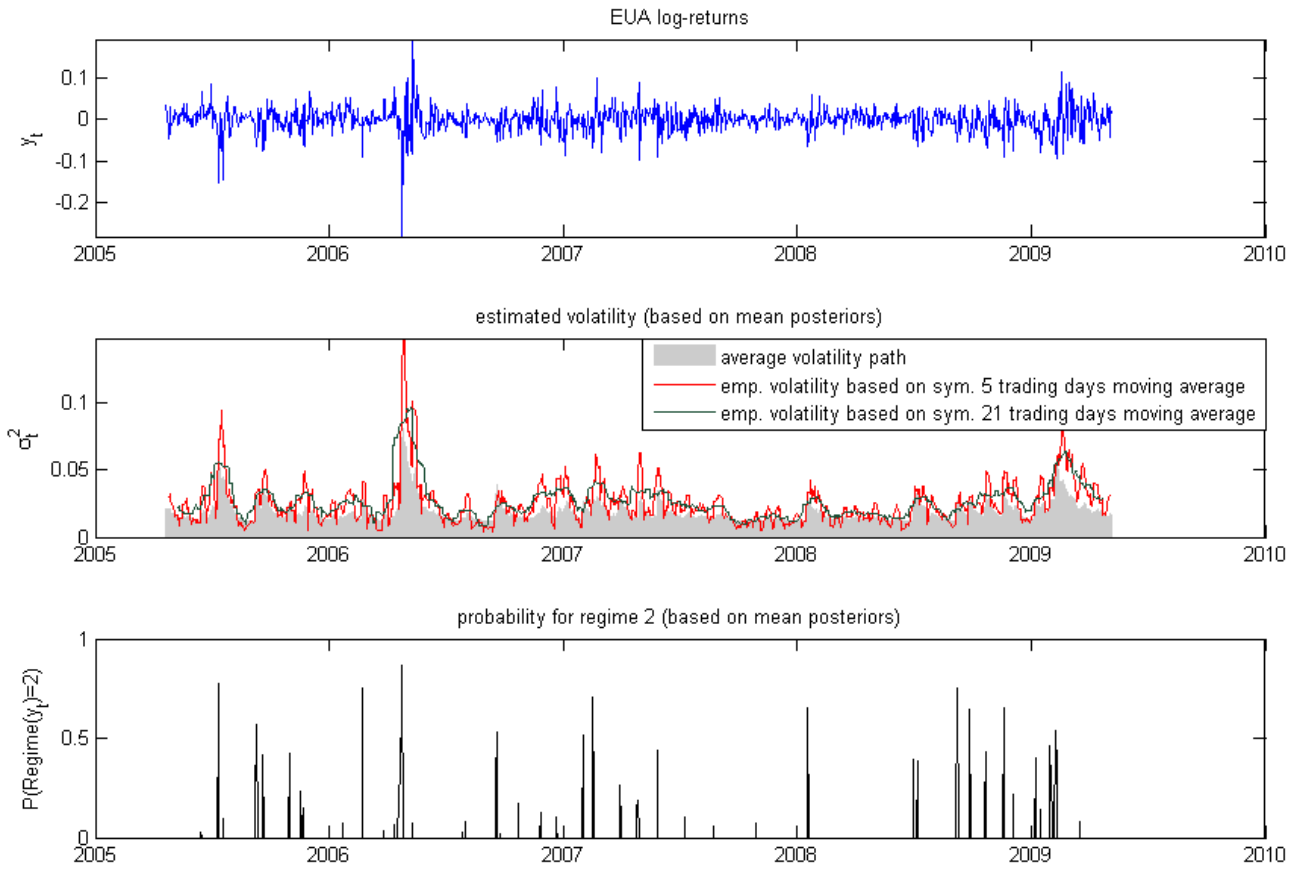


Figure 37: MS GARCH - empirical volatility

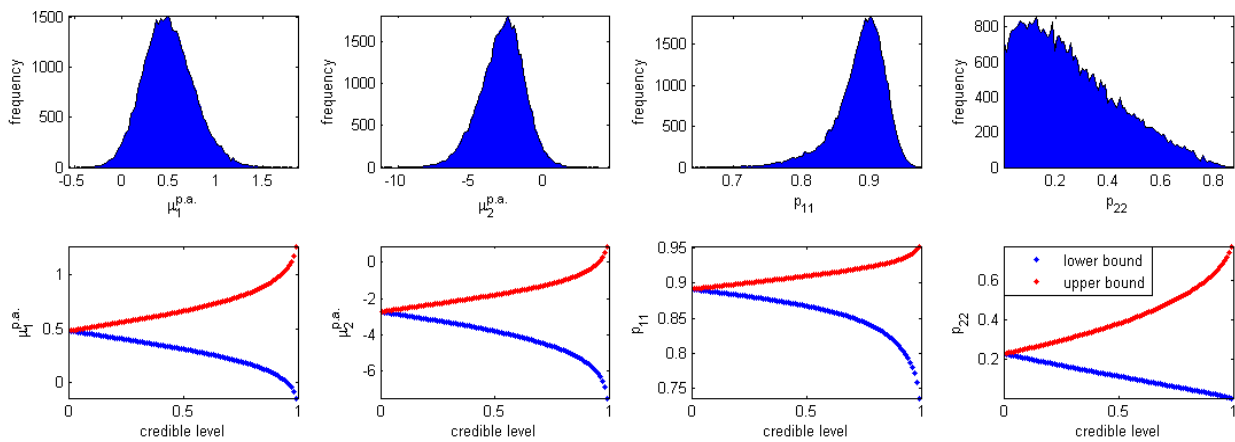


Figure 38: MS GARCH (short) - posteriors & credible intervals for mean log-returns & transition probabilities

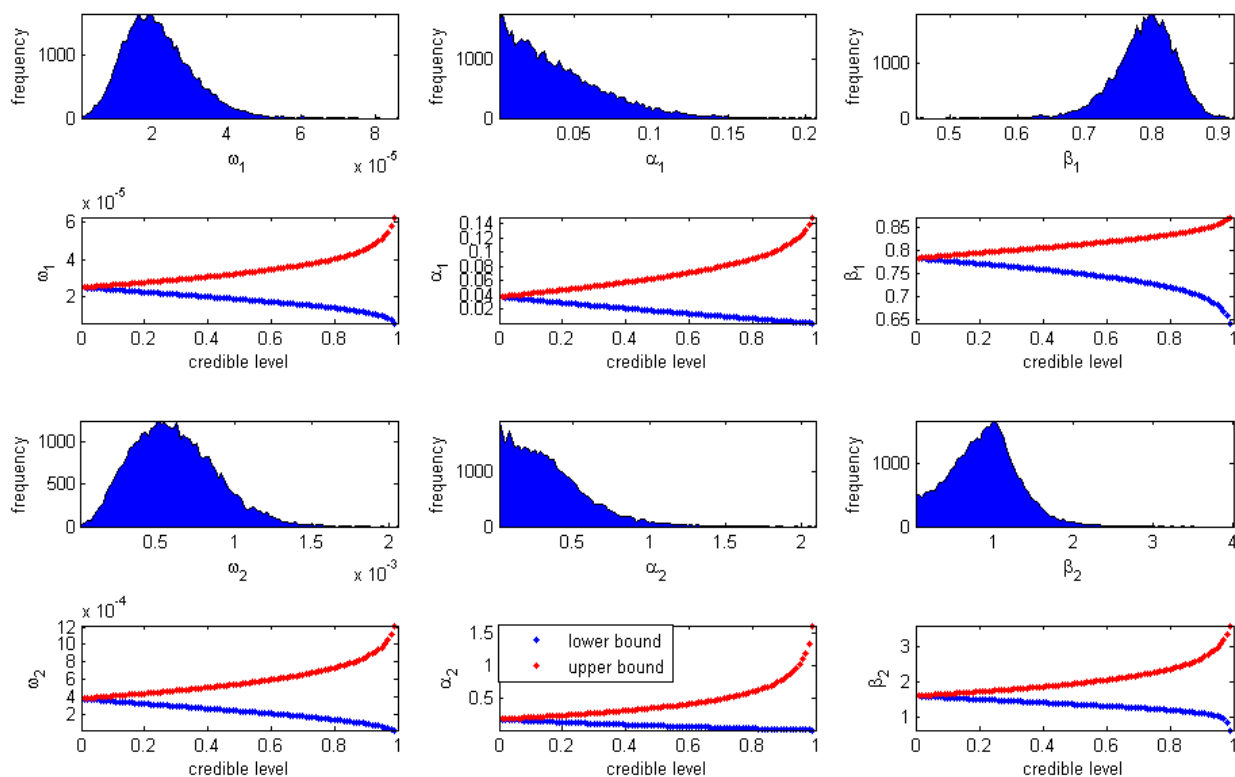


Figure 39: MS GARCH (short) - posteriors &amp; credible intervals for GARCH parameters

full series	mean	std. dev.	skewness	excess kurtosis
$\mu_1^{p.a.}$	0.503	0.247	0.262	0.331
$\mu_2^{p.a.}$	-4.55	2.073	-0.778	1.542
$\omega_1$	$2.641 \cdot 10^{-5}$	$1.051 \cdot 10^{-5}$	0.845	1.402
$\alpha_1$	0.044	0.033	0.900	0.466
$\beta_1$	0.779	0.045	-0.518	0.345
$\omega_2$	$4.104 \cdot 10^{-4}$	$2.386 \cdot 10^{-4}$	0.945	1.366
$\alpha_2$	0.256	0.280	2.585	10.851
$\beta_2$	1.710	0.503	1.176	2.591
$p_{11}$	0.898	0.036	-1.006	1.897
$p_{22}$	0.170	0.119	0.959	0.791

Table 17: results of Markov Switching MS GARCH(1,1) model (full series)

short series	mean	std. dev.	skewness	excess kurtosis
$\mu_1^{p.a.}$	0.493	0.270	0.289	0.256
$\mu_2^{p.a.}$	-2.841	1.543	-0.364	0.519
$\omega_1$	$2.141 \cdot 10^{-5}$	$9.318 \cdot 10^{-6}$	0.853	1.540
$\alpha_1$	0.038	0.031	1.134	1.316
$\beta_1$	0.789	0.045	-0.756	1.733
$\omega_2$	$6.117 \cdot 10^{-4}$	$2.720 \cdot 10^{-4}$	0.544	0.313
$\alpha_2$	0.332	0.263	1.321	2.559
$\beta_2$	0.875	0.433	0.393	0.723
$p_{11}$	0.885	0.038	-1.323	3.060
$p_{22}$	0.261	0.184	0.709	-0.230

Table 18: results of Markov Switching MS GARCH(1,1) model (short series)

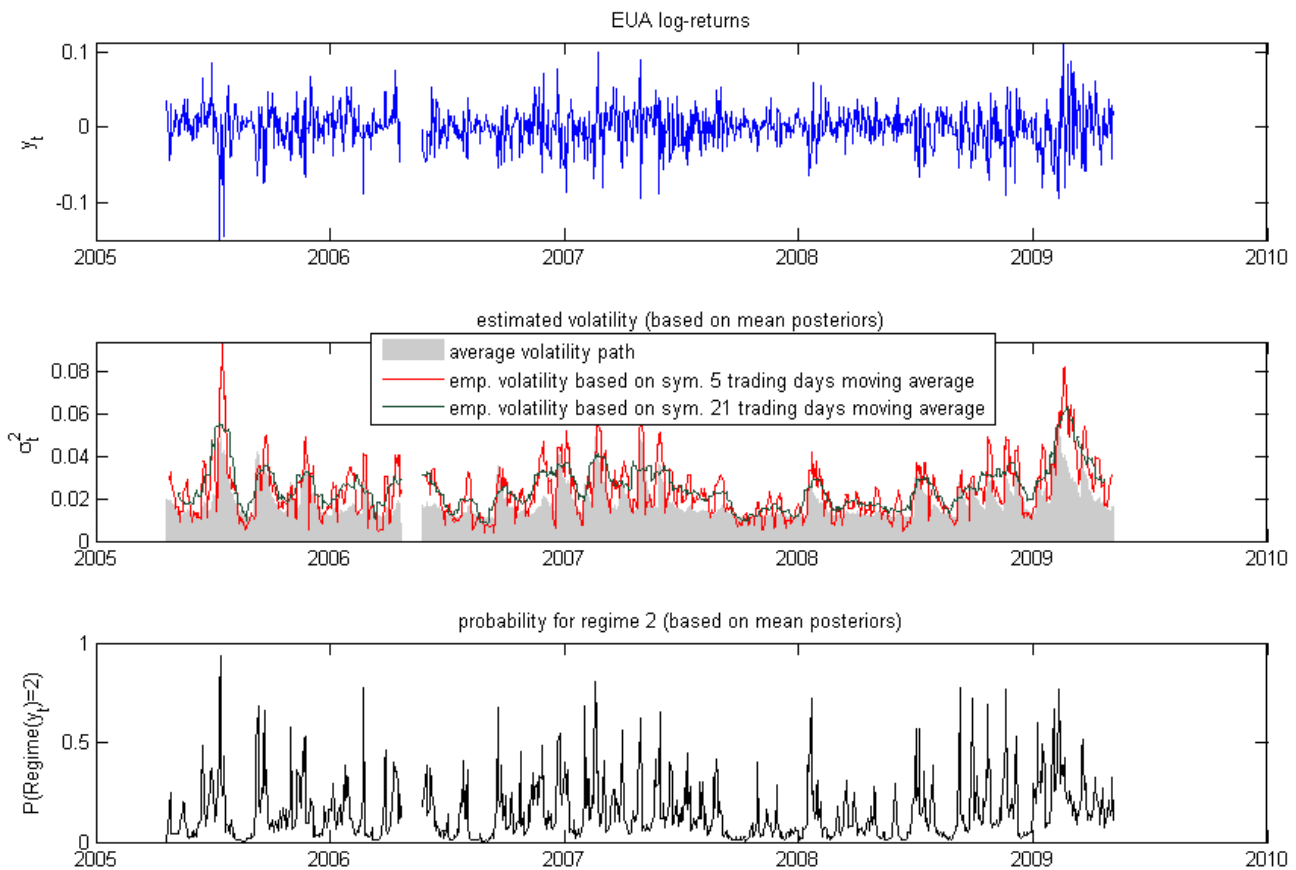


Figure 40: MS GARCH (short) - empirical volatility

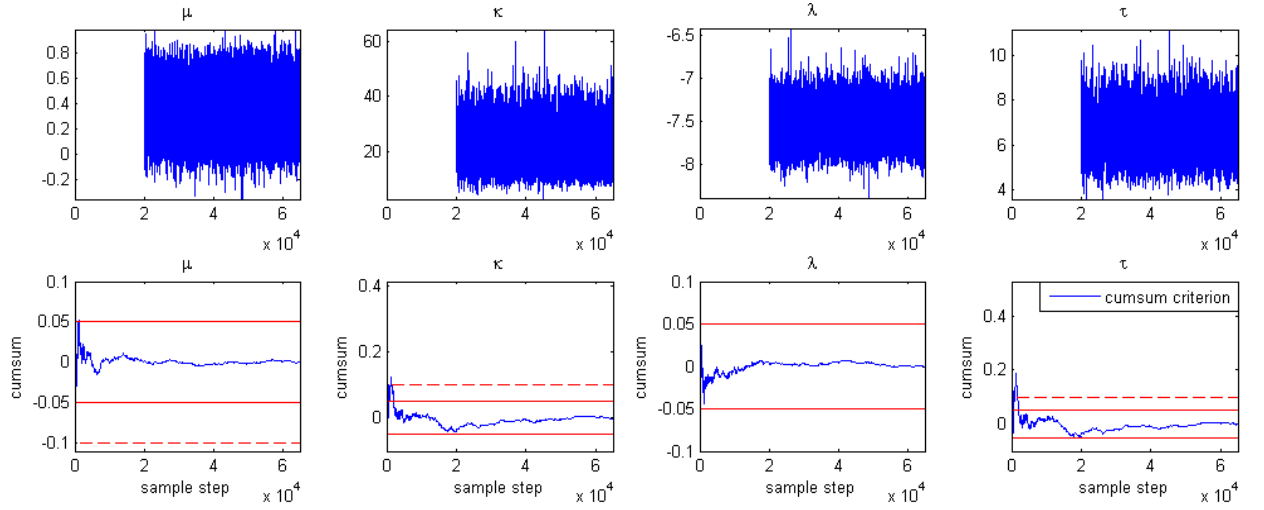


Figure 41: BLV - trace &amp; cumsum plots

### 10.2.8 Basic Stochastic Log-Variance (BLV)

For the basic log-variance (BLV) model we sample  $W = 65,000$  times the parameter vector  $\theta = (\tilde{\mu}, \lambda, \psi, \tilde{\tau}^2)'$  with a thinning factor of ten and dump the first 20,000 samples as burn-in.<sup>144</sup> Because of the thinning we could eliminate all auto-correlation for lags larger than two. Consequently, trace plots and cumsum criterion indicate convergence (see figure 41). We could reduce auto-correlation if we implemented a simultaneous sampling of  $\kappa = \tilde{\kappa}/250$  and  $\tau = \tilde{\tau}/\sqrt{250}$  which feature a high parameter correlation. Other correlations are not problematic ( $|\widehat{Corr}| \leq 0.13$ ).

The posteriors of the parameters are plotted in figure 42. In contrast to the GBM which estimates the posterior of  $\mu$  between  $\pm 0.5$ , the posterior of the BLV model is concentrated between 0 and 0.8. This is close to those of the GARCH model and the GBMJ which describe the huge negative skewed outliers by variability in the variance or by jumps. The mean posterior of  $\kappa$  (21.5) corresponds to a half-life time of around eight trading days. However, we cannot compare this result to the  $\kappa$ 's in the OU, CIR, and CEV model as they are defined on (log-)prices rather than on log-variances. From the mean posterior of  $\lambda$  we can construct the mean long-term equilibrium of the log-returns' volatility,  $\exp(-7.5278/2) = 0.023$  (see figure 43). Similar to  $\kappa$ , we cannot compare the parameter  $\tau$  to  $\sigma$  in the GBM model. While  $\tau$  describes the volatility of the log-variances,  $\sigma$  does the same for the log-returns.

<sup>144</sup>We assume the following hyperparameters:  $m_{\tilde{\mu}} = 0$ ,  $s_{\tilde{\mu}} = \sqrt{1000}$ ,  $m_{\lambda} = 0$ ,  $s_{\lambda} = \sqrt{1000}$ ,  $a_{\psi^*} = b_{\psi^*} = 1$ ,  $a_{\tilde{\tau}} = 0$ , and  $a_{\tilde{\tau}} = \sqrt{1000}$ .

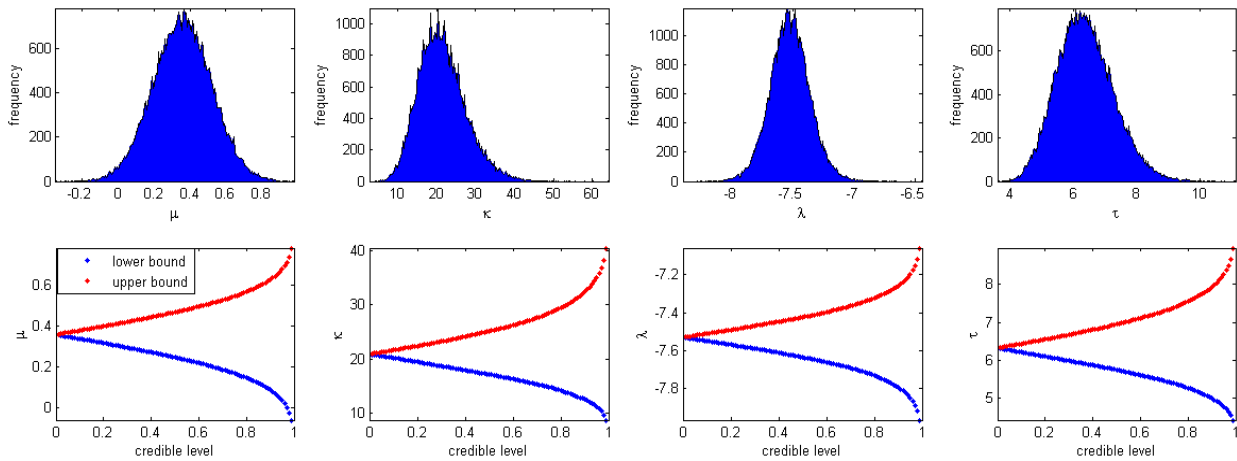


Figure 42: BLV - posteriors & credible intervals

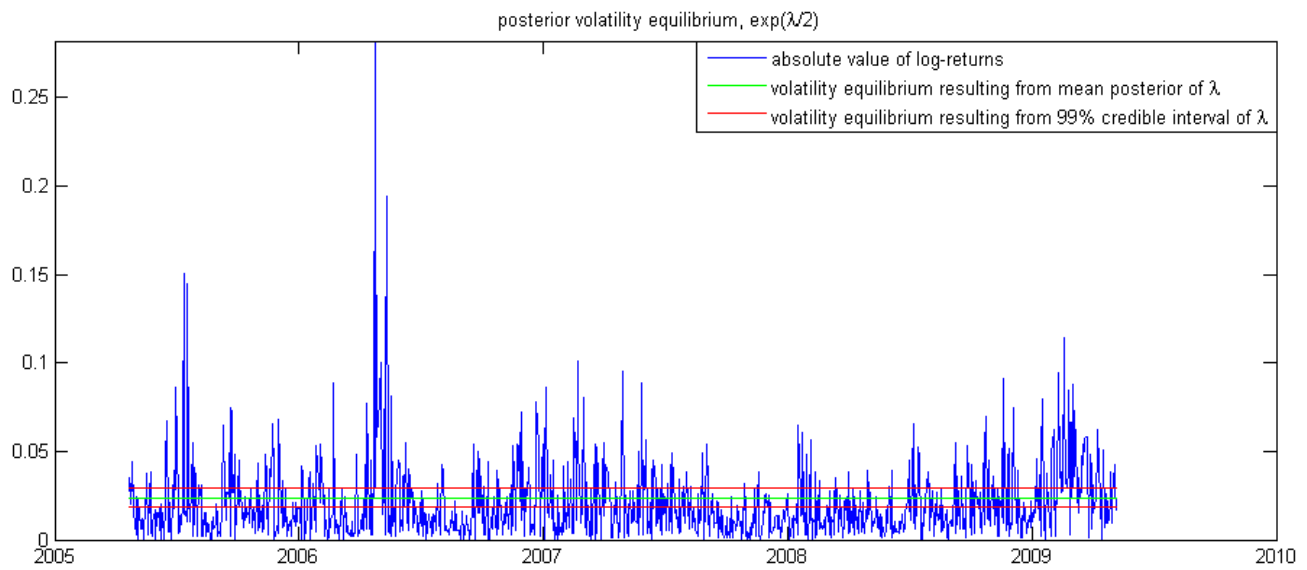


Figure 43: BLV - long-run volatility

We apply exactly the same algorithm to the short time-series. With a thinning factor of ten the auto-correlation is negligible. According to the trace plots and the cumsum criterion we believe the MCMC sampler has converged (see figure 44 for the posterior). We have summarized some basic statistics of the posteriors in the tables 19 and 20.

full series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.358	0.164	-0.012	$-4.435 \cdot 10^{-3}$
$\kappa$	21.452	6.070	0.583	0.650
$\lambda$	-7.528	0.164	0.118	0.646
$\tau$	6.378	0.892	0.368	0.195

Table 19: results of BLV model (full series)

short series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.347	0.166	-0.011	0.023
$\kappa$	23.172	6.826	0.624	0.697
$\lambda$	-7.580	0.146	0.089	0.420
$\tau$	6.013	0.914	0.351	0.134

Table 20: results of BLV model (short series)

Our main finding is that the posteriors are extremely stable with nearly no change of their means. Even the long-term equilibrium for the volatility only slightly changes from 0.0232 to 0.0226.

### 10.2.9 Stochastic Log-Variance with t-Errors (tLV)

For the Student's t log-variance (tLV) model we sample  $W = 65,000$  times the parameter vector  $\theta = (\tilde{\mu}, v, \lambda, \psi, \tau^2)'$  with a thinning factor of ten and dump the first 20,000 samples as burn-in.<sup>145</sup> Because of the thinning we could eliminate all auto-correlation for lags larger than two. Trace plots and cumsum criterion indicate convergence (see figure 45). Similar to the BLV model there is substantial correlation between  $\kappa = \tilde{\kappa}/250$  and  $\tau = \tilde{\tau}/\sqrt{250}$  ( $\widehat{Corr}(\kappa, \tau) = 0.762$ ). Additionally, there is some medium correlation between  $\kappa$  and  $v$  ( $\widehat{Corr}(\tau, v) = 0.408$ ). A simultaneous sampling of all three parameters could improve mixing of parameter paths. However, our MCMC sampler is fast enough for thinning.

Compared to the BLV model the posterior of  $\mu$  (mean posterior 0.353 vs. 0.358) has nearly not changed (see figure 46). Roughly 80.1% of posterior  $v$  is below 30 which means an

<sup>145</sup>We assume the following hyperparameters:  $m_{\tilde{\mu}} = 0$ ,  $s_{\tilde{\mu}} = \sqrt{1000}$ ,  $m_{\lambda} = 0$ ,  $s_{\lambda} = \sqrt{1000}$ ,  $a_{\psi^*} = b_{\psi^*} = 1$ ,  $a_{\tilde{\tau}} = 0$ ,  $a_{\tilde{\tau}} = \sqrt{1000}$ , and  $a_v = 1/1000$ .



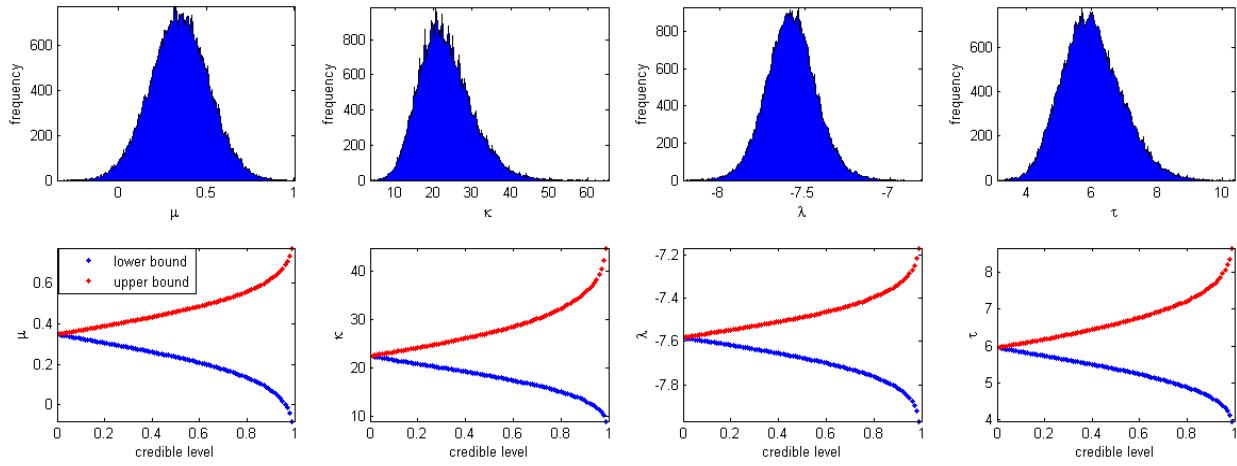


Figure 44: BLV (short) - posteriors & credible intervals

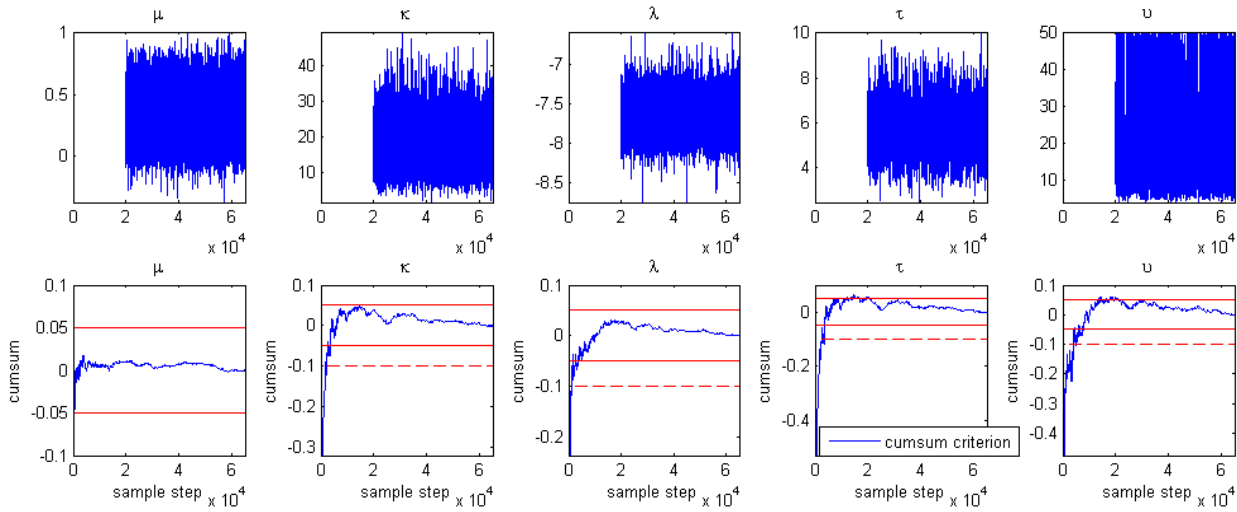


Figure 45: tLV - trace & cumsum plots

approximation by a Gaussian distribution (like in the BLV) model is not suitable. This leads to some remarkable changes in the posteriors of the log-variance parameters. The posterior of the mean reversion parameter  $\kappa$  is slightly more left than in the BLV model. Its mean is 17.83 corresponding to a half-life time of about ten trading days. The mean reversion is lower as outliers are partly explained by the Student's t errors which require short term peaks of the variances to a less extent. This statement is supported by a lower mean posterior of  $\tau$  (5.54 vs. 6.38 in the BLV model). Nevertheless, the mean long-term equilibrium of the log-variances  $\lambda$  ( $-7.63$ ) is rather stable (long-term volatility of log-returns  $\exp(-7.6337/2) = 0.022$  vs.  $0.023$  in the BLV model).

We apply exactly the same algorithm to the short time-series and generate  $W = 65,000$  samples with a thinning factor of ten. Because of a slower convergence rate we dump the first 30,000 samples (see figure 47 for the posteriors). We have summarized some basic statistics of the posteriors for the full and the short series in the tables 21 and 22.

full series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.353	0.162	$-9.399 \cdot 10^{-3}$	0.043
$\kappa$	17.829	5.561	0.590	0.690
$\lambda$	$-7.634$	0.181	0.011	0.552
$\tau$	5.543	0.914	0.269	0.250
$\nu$	19.516	11.624	0.931	$-0.174$

Table 21: results of tLV model (full series)

short series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.344	0.164	$7.708 \cdot 10^{-3}$	$8.787 \cdot 10^{-3}$
$\kappa$	19.42	6.537	0.622	0.784
$\lambda$	$-7.670$	0.165	$6.892 \cdot 10^{-3}$	0.453
$\tau$	5.280	0.982	0.207	0.126
$\nu$	22.782	12.325	0.545	$-0.868$

Table 22: results of tLV model (short series)

There are no surprising changes in the posteriors. Because of the elimination of the outliers a higher mean degree of freedom  $\nu$  estimates the log-returns more Gaussian (71.3% of posterior below 30). Similar to the BLV model, the mean posterior of  $\kappa$  increases, while the means of  $\tau$  and  $\lambda$  decrease.

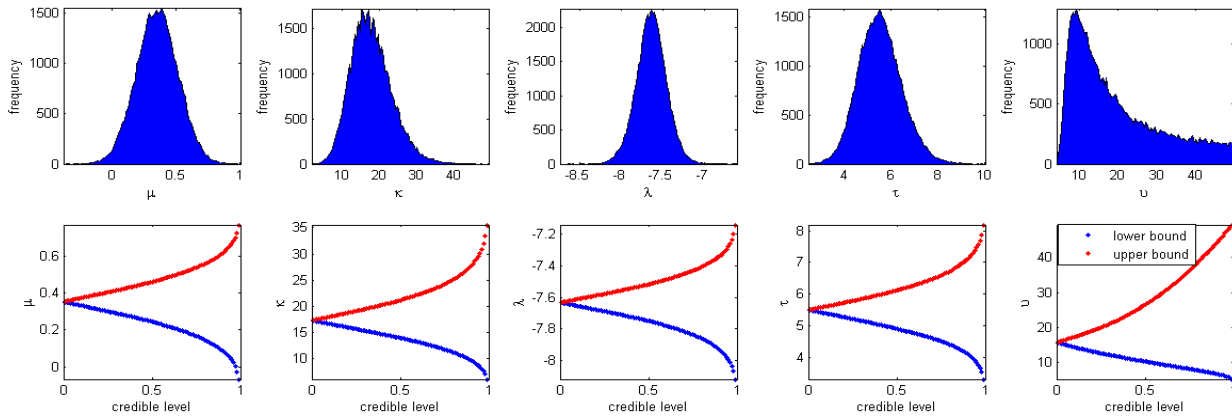


Figure 46: tLV - posteriors & credible intervals

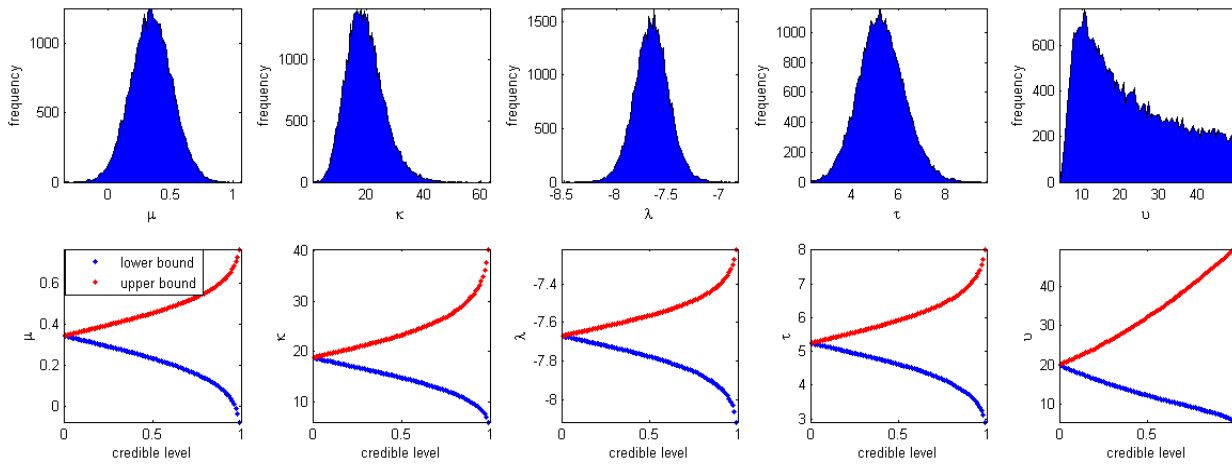


Figure 47: tLV (short) - posteriors & credible intervals

### 10.2.10 Correlated Log-Variance (CLV)

In the following we present our estimation results for the inter-temporal (ASV1) and the intra-temporal (ASV2) correlated log-variance models. In favor of a more straightforward MCMC algorithm we did not implement the drift  $\mu$ . In fact we just correct the log-returns  $y_t$  by their empirical mean  $\bar{y} = \sum_{t=1}^T y_t/T$ . Hence, we apply both CLV models on the observations  $\tilde{y}_t = y_t - \bar{y}$ .

#### ASV1

For the inter-temporal correlated ASV1 model we sample  $W = 57,000$  times the parameter vector  $\theta = (\lambda, \psi, \tilde{\tau}, \rho)'$  with a thinning factor of ten and dump the first 17,000 samples as burn-in.<sup>146</sup> Even with a thinning factor of ten our parameter paths feature substantial auto-correlation up to lag 20. Nevertheless, trace plots and cumsum criterion show convergence (see figure 48). Similar to the BLV and tLV models there is substantial correlation between

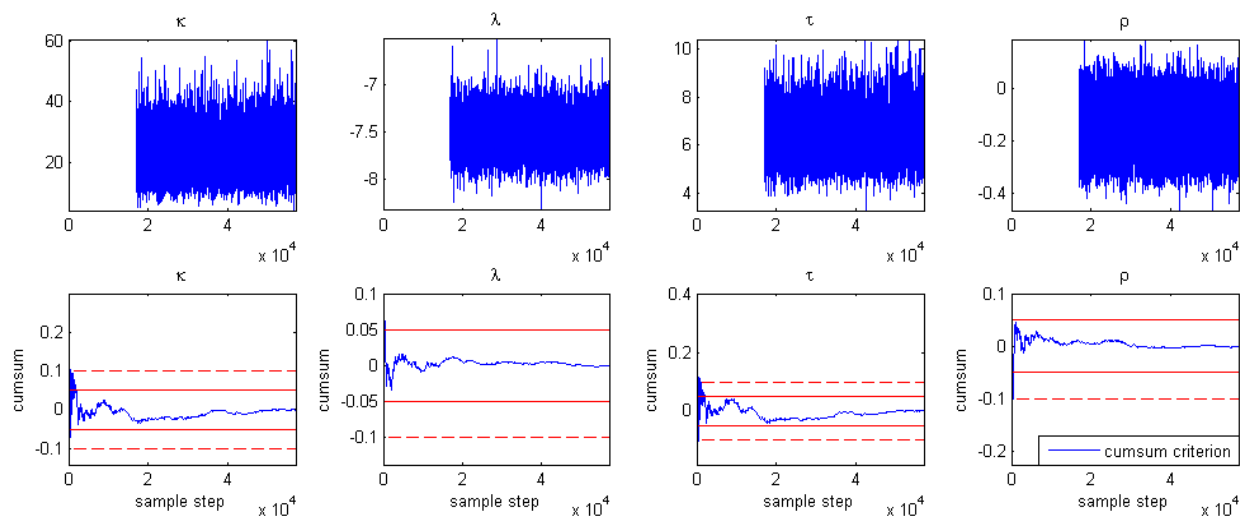


Figure 48: ASV1 - trace & cumsum plots

$\kappa = \tilde{\kappa}/250$  and  $\tau = \tilde{\tau}/\sqrt{250}$  ( $\widehat{Corr}(\kappa, \tau) = 0.764$ ). All other parameter combinations seem not critical with correlations between  $-0.110$  and  $-0.042$ .

The posteriors of  $\kappa$ ,  $\lambda$ , and  $\tau$  are close to those of the BLV model (see figure 49 and the table below). This is reasonable as the BLV model is nested in the ASV1 model (ignoring the drift parameter  $\mu$ ) for  $\rho = 0$ . In fact, we estimate a low leverage effect with a correlation between 0 and 0.3. The mean posterior is  $-0.151$ .

<sup>146</sup>We assume the following hyperparameters:  $m_\lambda = 0$ ,  $s_\lambda = \sqrt{1000}$ ,  $a_{\psi^*} = b_{\psi^*} = 1$ , and  $a_{\tilde{\tau}} = 0$ ,  $a_{\tilde{\tau}} = \sqrt{1000}$ .

We apply exactly the same algorithm to the short time-series and generate  $W = 60,000$  (burn-in  $B = 20,000$ ) samples with a thinning factor of ten (see figure 50 for the posteriors). We have summarized some basic statistics of the posteriors for the full and the short series in the tables 23 and 24.

full series	mean	std. dev.	skewness	excess kurtosis
$\bar{y}$	$-7.275 \cdot 10^{-5}$			
$\kappa$	22.590	6.091	0.603	0.707
$\lambda$	-7.514	0.154	0.090	0.525
$\tau$	6.349	0.874	0.377	0.296
$\rho$	-0.151	0.080	0.083	-0.056

Table 23: results of ASV1 model (full series)

short series	mean	std. dev.	skewness	excess kurtosis
$\bar{y}$	$1.742 \cdot 10^{-4}$			
$\kappa$	23.548	6.735	0.658	0.871
$\lambda$	-7.575	0.141	0.053	0.513
$\tau$	5.921	0.901	0.367	0.283
$\rho$	-0.122	0.082	0.076	-0.041

Table 24: results of ASV1 model (short series)

The changes of the posteriors are similar to those of the BLV and tLV models. The leverage effect is reduced (lower mean posterior of  $\rho$ ).

## ASV2

For the intra-temporal correlated ASV2 model we sample  $W = 57,000$  times the parameter vector  $\theta = (\lambda, \psi, \tilde{\tau}, \rho)'$  with a thinning factor of ten and dump the first 17,000 samples as burn-in.<sup>147</sup> Even with a thinning factor of ten our parameter paths feature substantial auto-correlation at least up to lag 20. Nevertheless, trace plots and cumsum criterion show convergence (see figure 51). Similar to the BLV, tLV, and ASV1 models, the substantial correlation between  $\kappa = \tilde{\kappa}/250$  and  $\tau = \tilde{\tau}/\sqrt{250}$  ( $\widehat{Corr}(\kappa, \tau) = 0.746$ ) deteriorate the mixing quality of the sampler.<sup>148</sup>

The posteriors of the ASV2 model (see figure 52) are rather close to those of the ASV1

<sup>147</sup>We assume the following hyperparameters:  $m_\lambda = 0$ ,  $s_\lambda = \sqrt{1000}$ ,  $a_{\psi^*} = b_{\psi^*} = 1$ , and  $a_\tau = 0$ ,  $a_\tau = \sqrt{1000}$ .

<sup>148</sup>All other correlations are between  $-0.195$  and  $0.204$ .

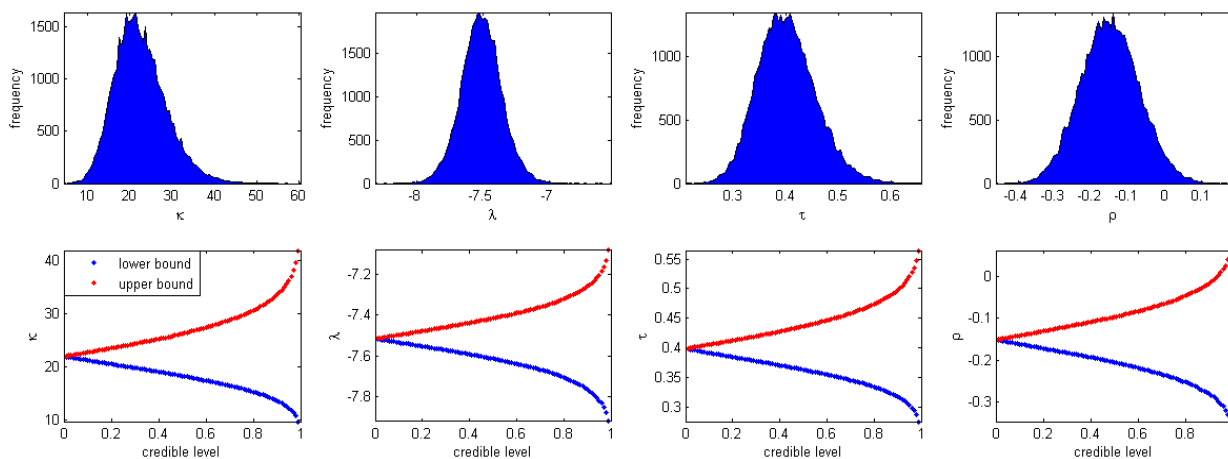


Figure 49: ASV1 - posteriors & credible intervals

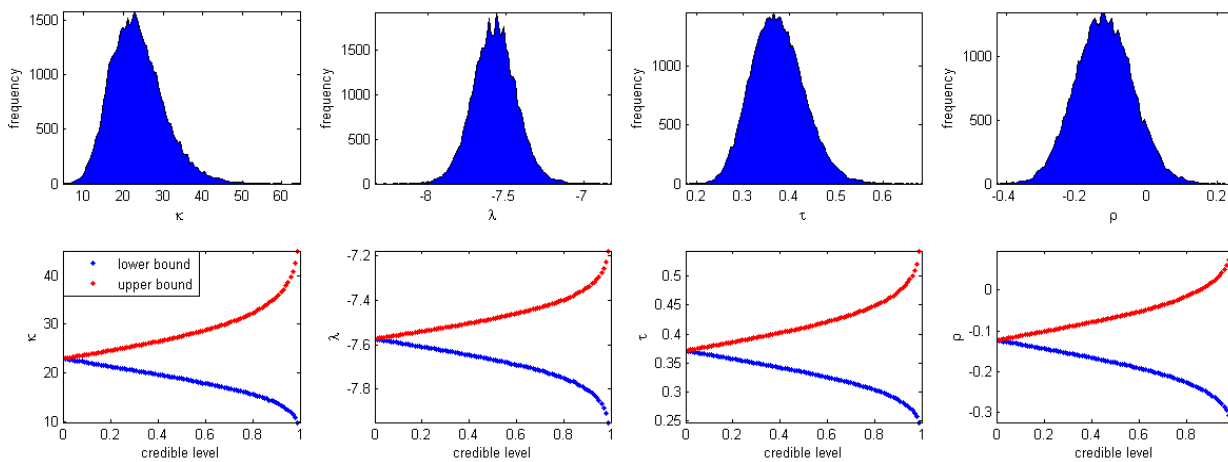


Figure 50: ASV1 (short) - posteriors & credible intervals

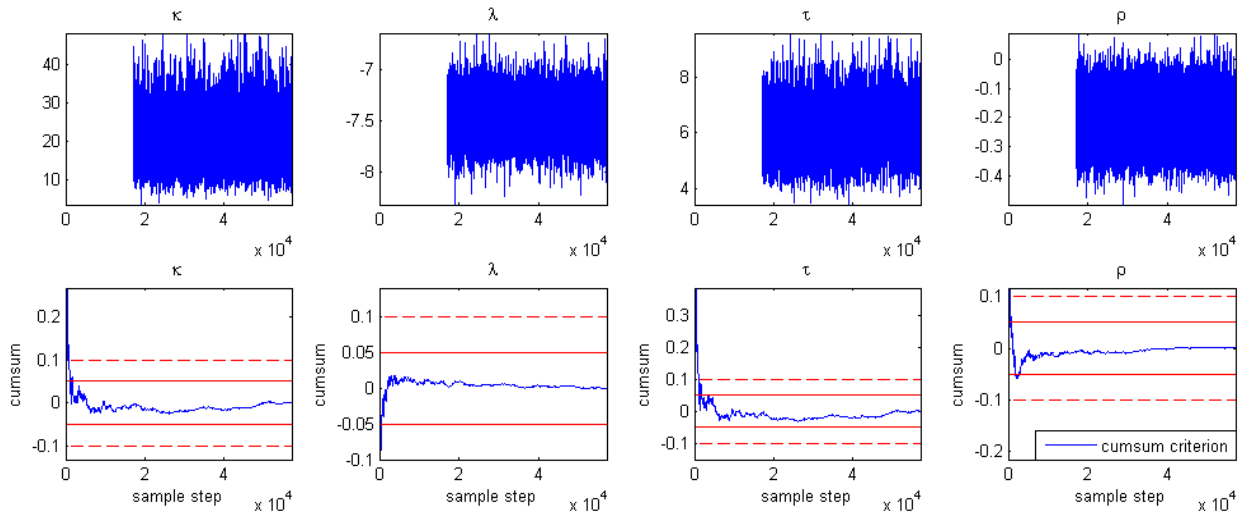


Figure 51: ASV2 - trace & cumsum plots

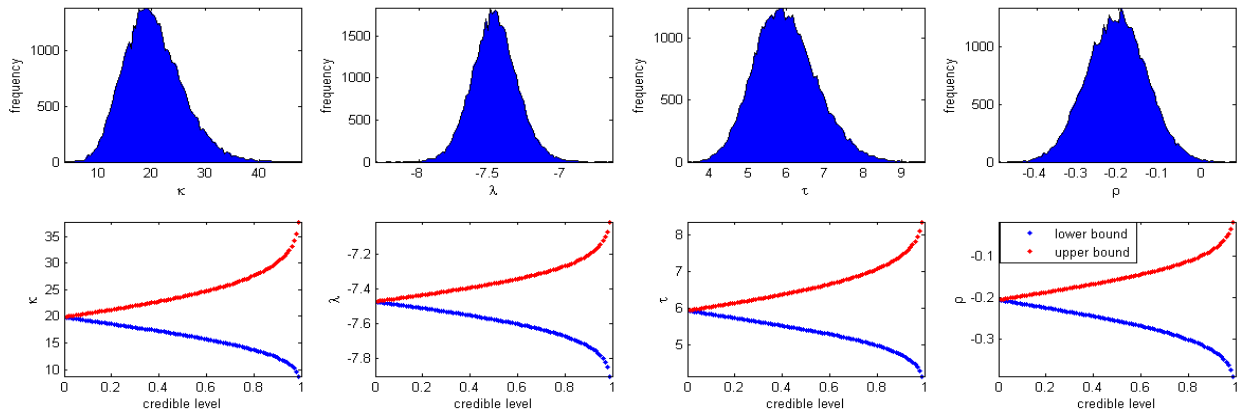


Figure 52: ASV2 - posteriors & credible intervals

model. All posteriors (with exception of  $\lambda$ ) are slightly shifted to the left and feature a lower standard deviation.

We apply exactly the same algorithm to the short time-series and generate  $W = 60,000$  (burn-in  $B = 20,000$ ) samples with a thinning factor of ten (see figure 53 for the posteriors). We have summarized some basic statistics of the posteriors for the full and the short series in the tables 25 and 26.

The changes of the posteriors are similar to those of the BLV, tLV, and ASV1 models.

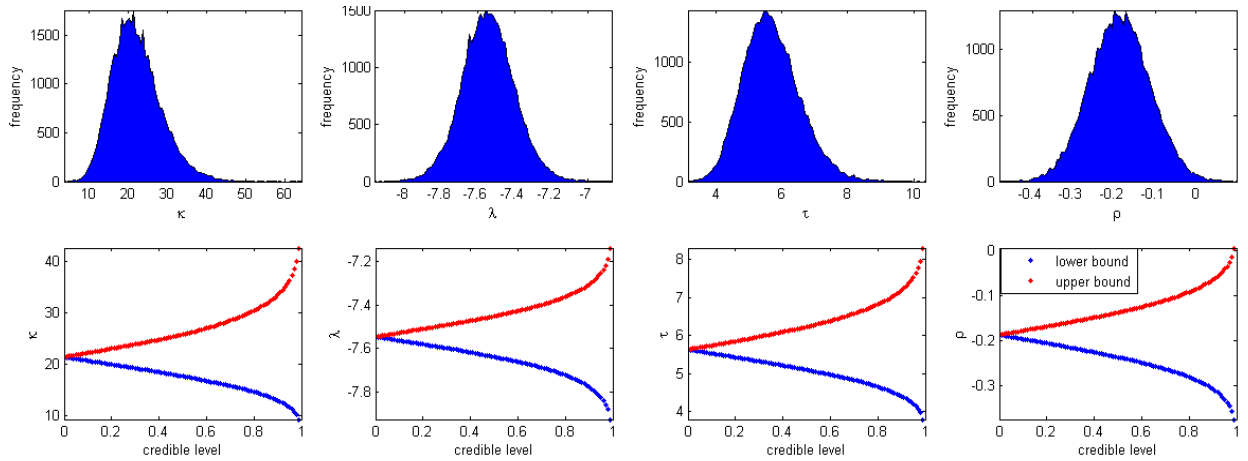


Figure 53: ASV2 (short) - posteriors & credible intervals

full series	mean	std. dev.	skewness	excess kurtosis
$\bar{y}$	$-7.275 \cdot 10^{-5}$			
$\kappa$	20.356	5.505	0.558	0.580
$\lambda$	-7.471	0.161	0.102	0.541
$\tau$	5.989	0.816	0.365	0.153
$\rho$	-0.206	0.073	0.018	-0.075

Table 25: results of ASV2 model (full series)



short series	mean	std. dev.	skewness	excess kurtosis
$\bar{y}$	$1.742 \cdot 10^{-4}$			
$\kappa$	22.124	6.298	0.697	1.044
$\lambda$	-7.545	0.144	0.097	0.443
$\tau$	5.694	0.856	0.462	0.481
$\rho$	-0.186	0.074	0.036	0.004

Table 26: results of ASV2 model (short series)

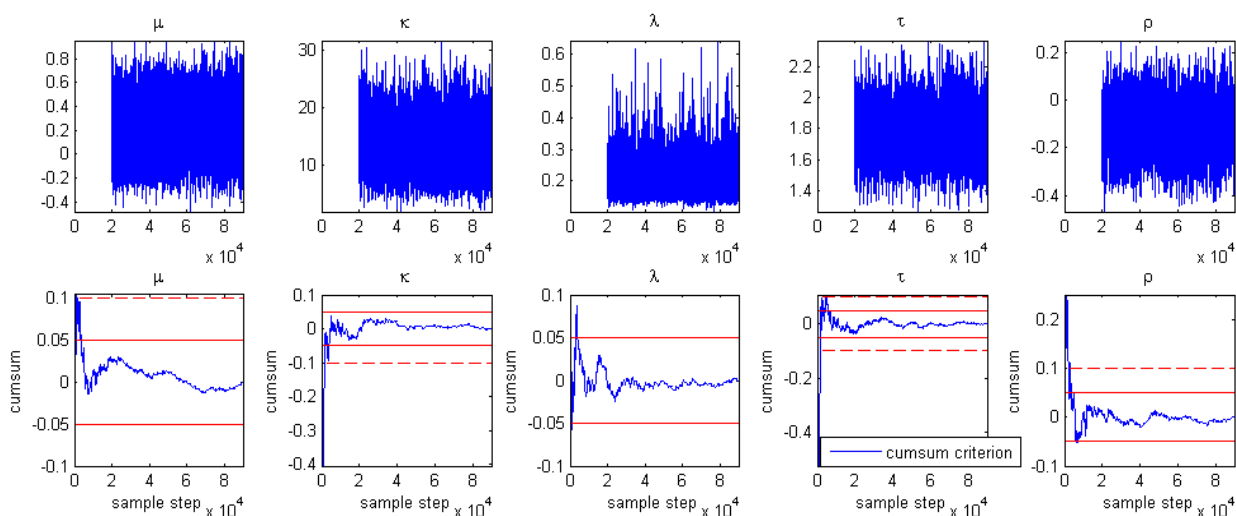


Figure 54: Heston - trace &amp; cumsum plots

### 10.2.11 Correlated Heston Model

For the correlated Heston model we sample  $W = 90,000$  times the parameter vector  $\tilde{\theta} = (\mu, \kappa, \lambda, \omega, \varkappa)'$ , where  $\tau = \sqrt{\varkappa^2 + \omega}$  and  $\rho = \varkappa/\omega$ .<sup>149</sup> Because of the computational burdens we do not thin the sample paths. We just dump the first 20,000 samples as burn-in. Although there is substantial auto-correlation in the parameter paths at least up to lag 20, trace plots and cumsum criterion indicate convergence (see figure 54). Moreover, the mean posteriors seem rather stable, too (see figure 55).

Our Heston Gibbs sampler generates some remarkable parameter correlation between  $\lambda$  and  $\kappa$  ( $\widehat{Corr}(\kappa, \lambda) = -0.594$ ) and between  $\tau$  and  $\kappa$  ( $\widehat{Corr}(\kappa, \tau) = 0.498$ ). We could construct a

<sup>149</sup>We assume the following hyperparameters:  $m_\mu = 0$ ,  $s_\mu = 10$ ,  $m_\kappa = 0$ ,  $s_\kappa = 100$ ,  $m_\lambda = 0$ , and  $s_\lambda = 10$ . The parameters  $a_\omega = 1$ , and  $b_\omega = 1/70$  imply a flat prior for  $\rho$  between  $\pm 0.9$ . Values close to  $\pm 1$  are rather unlikely.

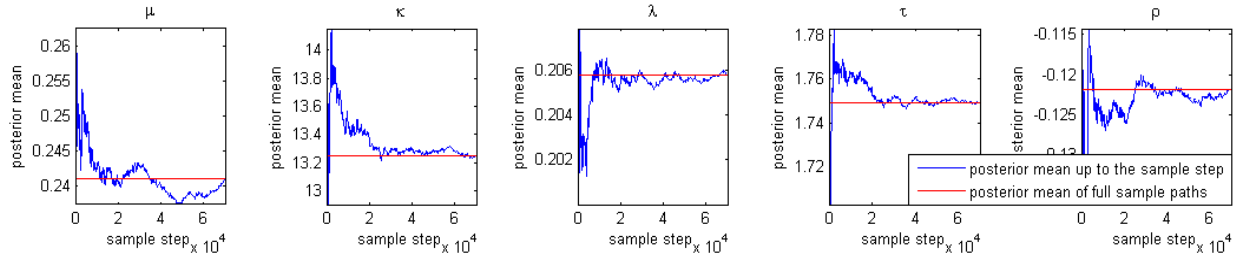


Figure 55: Heston - mean posterior plots

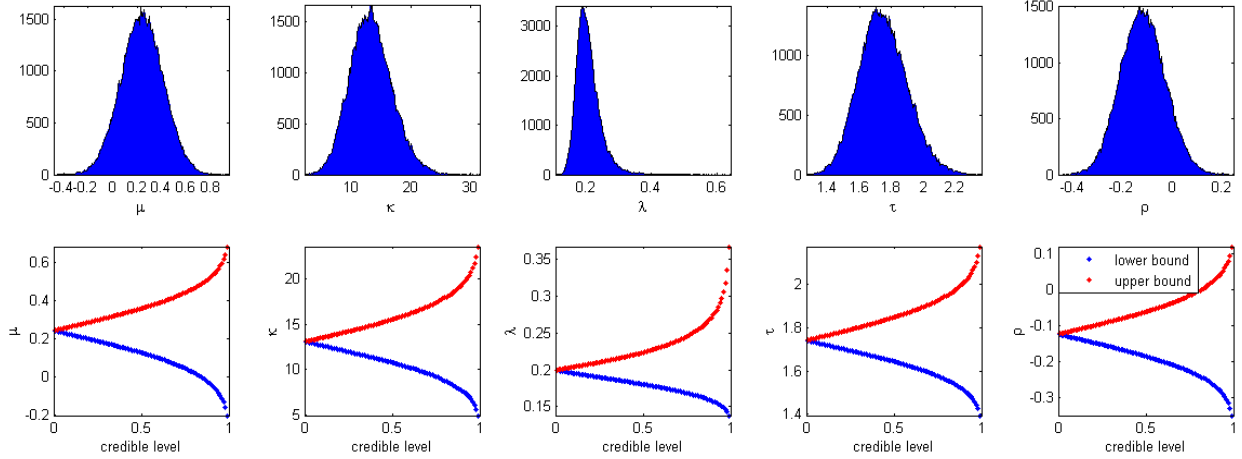


Figure 56: Heston - posteriors &amp; credible intervals

multivariate Metropolis-Hastings sampler to improve mixing behavior. However, this would reduce the acceptance rate of the generated parameters. Simulation studies could analyze this trade-off.

For the interpretation of the posteriors (see figure 56) it is important to keep in mind that the Heston model directly assumes a CIR process for the variance while the log-variance processes assume an OU process for the log-variances. Hence, we should be careful with comparisons of the parameters.

Similar to the GBM and the Student's  $t$  model, the posterior of  $\mu$  (mean 0.241) does not rule out a negative trend. Roughly speaking, the Heston posterior of  $\mu$  is concentrated between the posterior of the GBM and those of the GBMJ, Student's  $t$  GARCH, MS GARCH, BLV, and tLV models. An interpretation is that the Heston model is less able to explain large negative log-returns by its variable volatility term relatively to the other models.<sup>150</sup>

<sup>150</sup>Of course, the GBMJ explains large positive and negative log-returns by jumps.

The level of the posterior of  $\kappa$  (mean 13.25) is lower compared to those of the log-variance models. This can be explained by the assumed CIR process which allows for a varying volatility term accelerating mean reversion. The posterior of  $\lambda$  (mean  $\bar{\lambda} = 0.206$ ) cannot directly be compared to the log-variance processes. Instead, we calculate the mean long-term equilibrium  $\sqrt{\lambda\Delta}$  of the log-returns. It is 0.029 (see figure 57) compared to 0.023 (ASV1) and 0.024 (ASV2).

The posterior of  $\tau$  (mean 1.749) cannot directly be compared to those of the log-variance models, too. As it is a parameter of a CIR process, it is a factor rather than a level like in an OU process. The posterior of  $\rho$  is in the same magnitude as those of the ASV1 and ASV2 models although its mean is  $-0.122$  — below those of ASV1 ( $-0.151$ ) and ASV2 ( $-0.206$ ) model. As expected it is closer to the ASV1 model as both assume an inter-temporal correlation.

We apply exactly the same algorithm to the short time-series and generate 110,000 (burn-in  $B = 20,000$ ) samples without any thinning (see figure 58 for the posteriors). We have summarized some basic statistics of the posteriors for the full and the short series in the table 27 and 28.

full series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.241	0.173	$-0.008$	0.007
$\kappa$	13.246	3.569	0.307	0.219
$\lambda$	0.206	0.039	1.706	6.954
$\tau$	1.749	0.152	0.251	0.030
$\rho$	$-0.122$	0.091	0.088	$-0.028$

Table 27: results of Heston model (full series)

short series	mean	std. dev.	skewness	excess kurtosis
$\mu$	0.249	0.174	$-0.023$	0.011
$\kappa$	13.778	3.855	0.407	0.325
$\lambda$	0.181	0.031	1.334	4.163
$\tau$	1.567	0.157	0.190	$-0.041$
$\rho$	$-0.116$	0.094	0.044	$-0.097$

Table 28: results of Heston model (short series)

Similar to the BLV and tLV models, the posterior of  $\mu$  is extremely stable. All other parameters change in the same manner as in the ASV1 and ASV2 models.

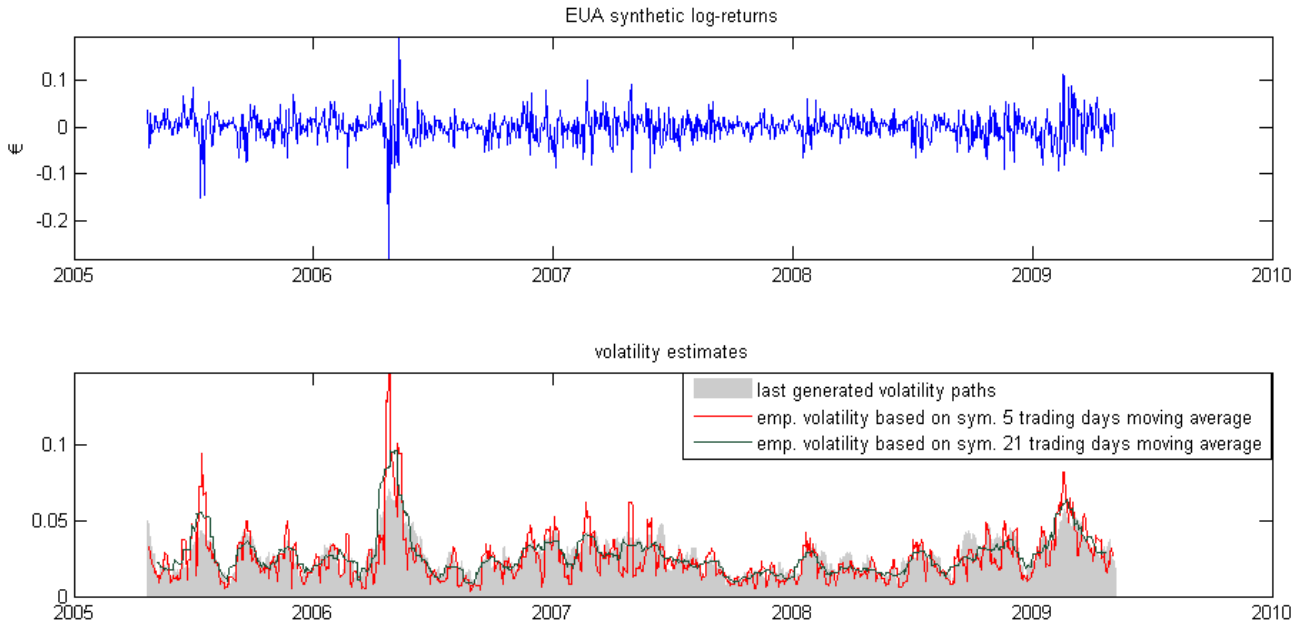


Figure 57: Heston - empirical volatility

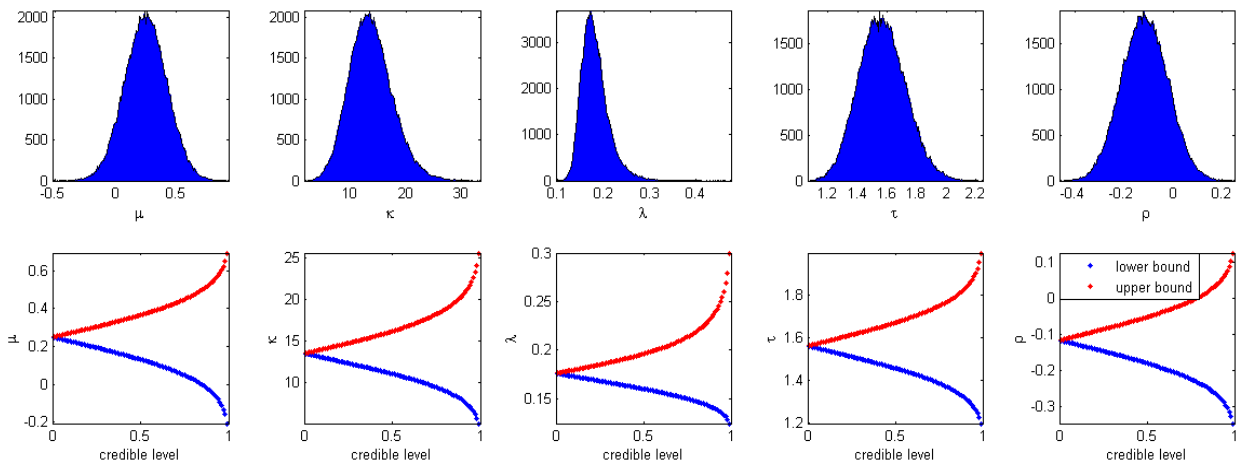


Figure 58: Heston (short) - posteriors & credible intervals

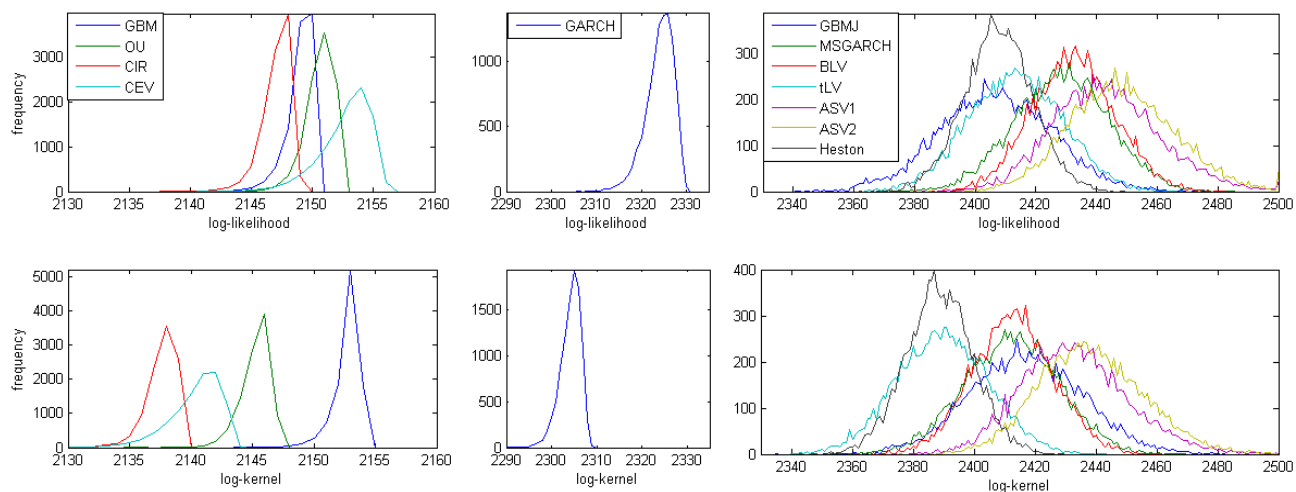


Figure 59: model weighting - log-likelihoods &amp; log-kernels

### 10.3 Model Weighting Results

For model comparison we calculate Akaike's and Bayesian information criterion (AIC and BIC), mean deviance, as well as AIC and BIC based model probabilities (see section 8). Because of the large number of partly rather complex models we avoid the multi-model MCMC algorithm of Carlin & Chib (1995), pp 475. Instead, we apply the approximation of Congdon (2007) allowing for a separate MCMC estimation of different models. Of course, these approaches seem to be rather crude but Robert & Marin (2008) found the results of Congdon (2006) normally in the same magnitude compared to the exact results. As Congdon (2007) is an improvement of the Congdon (2006) algorithm we hope for an even better approximation.

Our Bayesian model comparison for the full time-series is based on  $W = 10,000$  MCMC samples from all models we have estimated. To reach 10,000 samples we thin the MCMC outputs as much as possible and dump redundant samples where necessary. In the figures 59 and 60 we present the plots of the sampled log-likelihoods and log-kernels as well as their mean values and standard deviations. It is easy to see that we can separate three different groups of models: (1) *plain vanilla models* GBM, OU, CIR, and CEV, (2) Student's t GARCH model, and (3) *sophisticated models* GBMJ, MSGARCH, BLV, tLV, ASV1, ASV2, and Heston.

All plain vanilla models feature low log-likelihoods and log-kernels with low standard deviations. In contrast, the sophisticated models have by far larger log-likelihoods and log-kernels accompanied by a substantial larger standard deviation which originates from the simulation

	<i>GBM</i>	<i>GBMJ</i>	<i>OU</i>	<i>CIR</i>	<i>CEV</i>	<i>GARCH</i>	<i>MSGARCH</i>	<i>BLV</i>	<i>tLV</i>	<i>ASV1</i>	<i>ASV2</i>	<i>Heston</i>
<i>dim</i>	2	5	3	3	4	6	10	4	5	4	4	5
<i>mean(logLik)</i>	2149	2406.7	2150.6	2147.1	2152.7	2324	2428.4	2432.2	2413.2	2443.1	2448.5	2407.6
<i>std(logLik)</i>	1.021	18.632	1.168	1.164	2.068	3.106	15.605	13.701	15.477	17.800	17.585	11.290
<i>MD</i>	-4298	-4813	-4301	-4294	-4305	-4648	-4857	-4864	-4826	-4886	-4897	-4815
<i>mean(logKernel)</i>	2152.7	2417	2145.2	2137.6	2140.4	2304.2	2412.3	2412.6	2389.5	2432.9	2438.5	2388.5
<i>std(logKernel)</i>	1.028	18.546	1.188	1.935	2.138	2.268	15.605	13.337	15.018	17.541	17.456	11.097
<i>MP<sub>AIC</sub></i>	0%	1%	0%	0%	0%	0%	4.3%	11.1%	1.2%	33.6%	48.8%	0.1%
<i>MP<sub>AIC</sub></i>	14.5%		24.9%	1%	59.6%							
<i>MP<sub>BIC</sub></i>	0%	0.8%	0%	0%	0%	0%	0.7%	11.8%	1%	35%	50.6%	0%
<i>MP<sub>BIC</sub></i>	73.1%		17.2%	0.6%	9%							
$\hat{\pi}(M_k y)$	0%	0.9%	0%	0%	0%	0%	0.1%	2.7%	0.1%	39.3%	56.7%	0%
$\hat{\pi}(M_k y)$	2.8%		36.8%	3.4%	57%							

Figure 60: model weighting - basic statistics

of the respective latent variables (variance, jumps, jump width). A special group forms the Student's t GARCH model with medium size log-likelihood and log-kernel but low standard deviations.

The mean log-likelihood and the mean deviation express the same statement in two different scales. The log-kernel additionally accounts for our prior uncertainty on the parameters. We have mostly defined relatively non-informative (flat) priors. Hence, the log-kernels are shifted to the left.<sup>151</sup> As a consequence, the GBM - two parameters only - has the "third best" log-likelihood of the plain vanilla models while it has the "best" one of the log-kernels. The mean values of the log-likelihoods and log-kernels only allow for an ordinal comparison of different models. Hence, we calculate model probabilities based on the AIC and BIC allowing to weight the models (see figure 60). The model probabilities clearly favor the asymmetric log-variance models with a slight advantage for the intra-temporal correlated ASV2. Beside these, the basic log-variance model BLV and the MSGARCH model have still remarkable model probabilities. Moreover, we find BIC based probabilities supporting models with less than five parameters. Models with exactly five parameters slightly lose probability compared to the AIC ones whereas the BIC probability is devastating for the MSGARCH model with its ten parameters.

Of course, it seems confusing why the Student's t error log-variance model (tLV) has a by far lower model probability than the BLV model. However, in our specification, the tLV model

<sup>151</sup>A flat prior results in a close to zero density. The resulting log-value is mostly negative.

is not a generalization of the BLV model as we restrict the degree of freedom to  $v \in (0, 50]$ . We are only willing to use a more complex model if it has an advantage compared to the BLV model. Indeed, we have run another MCMC estimation were we have allowed  $v \in (0, 350]$ . In this case, the log-likelihood has been rather close to that of the BLV model.

Although the performance of GBM, OU, CIR, and CEV is far behind all other models they seem interesting because of their simplicity. Therefore, we additionally compare their AIC and BIC based probabilities (see figure 60). While  $MP_{AIC}$  favors the CEV model,  $MP_{BIC}$  prefers the GBM. From a Bayesian perspective, we should treat the  $MP_{BIC}$  results with care. All models feature likelihoods rather close to each other. The GBM could improve its relative performance under  $MP_{BIC}$  as it has a lower number of parameters resulting in a lower degree of uncertainty. Of course, the classical criterion of parameter parsimony is important in Bayesian statistics, too. However, the preference for a model with a low number of parameters should not originate from ignored parameter uncertainty. Intuitively, this means that the CEV model could reach a larger log-likelihood compared to the GBM (2153 vs. 2149) but the uncertainty on two more parameters results in a lower log-kernel (2140 vs. 2153). In the CEV model we acknowledge that we are uncertain about the relationship between the annual and daily volatility while we simply impose  $\sigma_{annual} = \sigma_{daily} \cdot \sqrt{250}$  for the GBM.

We estimate the posterior model probabilities  $\hat{\pi}(\mathcal{M}_k|y)$  by the approximation Congdon (2007) where we assume uniform model priors  $\pi(\mathcal{M}_k) = 1/K$ ,  $k = 1, \dots, K$ . We need to specify linking densities for all parameters of all models. As the approximation is better for linking densities close to the posteriors of the parameters but with fatter tails we calibrate (truncated) Gaussian distributions to the respective posteriors and scale up their standard deviations by a factor of two.<sup>152</sup> The posterior model probabilities are close to the AIC and BIC probabilities (see figure 60). Again, the intra-temporal correlated log-variance process (ASV2) outperforms the inter-correlated log-variance process (ASV1). The GBMJ, MSGARCH, BLV, and tLV models have only some marginal importance. When we compare the model probabilities for the plain vanilla models, the OU (36.8%) and CEV (57%) model dominate. For a more intuitive comparison of the model performances, we present the Bayes factors (according to Congdon 2007) with the wording of Raftery (1996), 252/ table 1 (see figure 61). Additionally, we plot all AIC, BIC, and posterior model probabilities in figure 62.

To analyze the stability of the model performances, we also estimate model probabilities for the short time-series (see figures 63, 64, and 65). Again, we find the identical three groups (1) plain vanilla models, (2) Student's t GARCH model, and (3) sophisticated models. Although the asymmetric log-variance models still have the largest model probabilities, the GBMJ is

---

<sup>152</sup>We have not found substantial shifts in the model probabilities for different factors lower than 10.

<i>BF</i>	<i>GBM</i>	<i>GBMJ</i>	<i>OU</i>	<i>CIR</i>	<i>CEV</i>	<i>GARCH</i>	<i>MSGARCH</i>	<i>BLV</i>	<i>tLV</i>	<i>ASV1</i>	<i>ASV2</i>
<i>GBMJ</i>	very strong										
<i>OU</i>	strong	-									
<i>CIR</i>	~	-	-								
<i>CEV</i>	very strong	-	positive	very strong							
<i>GARCH</i>	very strong	-	very strong	very strong	very strong						
<i>MSGARCH</i>	very strong	-	very strong	very strong	very strong	very strong					
<i>BLV</i>	very strong	~	very strong	very strong	very strong	very strong	strong				
<i>tLV</i>	very strong	positive	very strong	very strong	very strong	very strong	~				
<i>ASV1</i>	very strong	strong	very strong	very strong	very strong	very strong	very strong	positive	very strong		
<i>ASV2</i>	very strong	strong	very strong	very strong	very strong	very strong	very strong	strong	very strong	~	
<i>Heston</i>	very strong	-	very strong	very strong	very strong	very strong	-	-	-	-	-

Figure 61: model weighting - Bayes factors (~ ... similar model performance, - ... model is inferior)

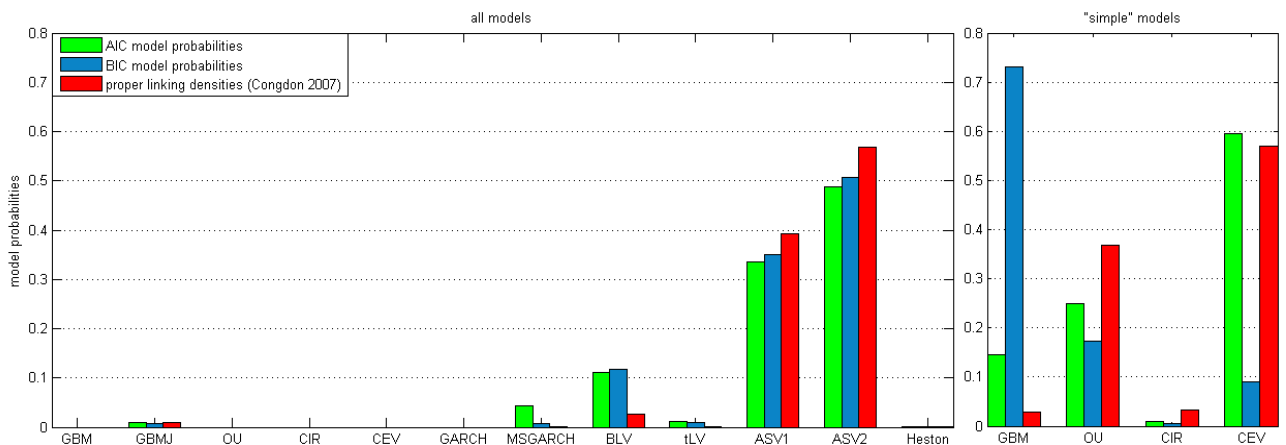


Figure 62: model weighing - AIC, BIC, & posterior model probabilities



	<i>GBM</i>	<i>GBMJ</i>	<i>OU</i>	<i>CIR</i>	<i>CEV</i>	<i>GARCH</i>	<i>MSGARCH</i>	<i>BLV</i>	<i>tLV</i>	<i>ASV1</i>	<i>ASV2</i>	<i>Heston</i>
<i>dim</i>	2	5	3	3	4	6	10	4	5	4	4	5
<i>mean(logLik)</i>	2203.9	2407.6	2201.9	2225.7	2225.2	2302.5	2398.0	2401.3	2385.1	2408.2	2415.4	2383.1
<i>std(logLik)</i>	1.026	19.917	1.055	1.008	1.221	2.844	14.792	13.463	15.414	16.128	16.599	10.982
<i>MD</i>	-4408	-4815	-4404	-4451	-4450	-4605	-4796	-4803	-4770	-4817	-4831	-4766
<i>mean(logKernel)</i>	2207.8	2418.9	2196.5	2216.1	2211.9	2281.3	2381.9	2384.2	2361.6	2398.3	2405.6	2364.5
<i>std(logKernel)</i>	1.029	19.631	1.074	1.020	1.228	2.165	14.792	13.072	14.882	15.831	16.414	10.662
<i>MP<sub>AIC</sub></i>	0%	24.1%	0%	0%	0%	0%	3.2%	9.6%	1.5%	22.2%	39.2%	0.2%
<i>MP<sub>AIC</sub></i>	0%		0%	75%	25%							
<i>MP<sub>BIC</sub></i>	0%	21.4%	0%	0%	0%	0%	0.4%	10.9%	1.3%	23.9%	41.9%	0.2%
<i>MP<sub>BIC</sub></i>	0%		0%	95.3%	4.7							
$\hat{\pi}(M_k y)$	0%	21.9%	0%	0%	0%	0%	0%	3.0%	0.2%	27.4%	47.4%	0%
$\hat{\pi}(M_k y)$	0%		0%	79.3%	20.7%							

Figure 63: model weighting (short) - basic statistics

<i>BF</i>	<i>GBM</i>	<i>GBMJ</i>	<i>OU</i>	<i>CIR</i>	<i>CEV</i>	<i>GARCH</i>	<i>MSGARCH</i>	<i>BLV</i>	<i>tLV</i>	<i>ASV1</i>	<i>ASV2</i>
<i>GBMJ</i>	very strong										
<i>OU</i>	-	-									
<i>CIR</i>	very strong	-	very strong								
<i>CEV</i>	very strong	-	very strong	-							
<i>GARCH</i>	very strong	-	very strong	very strong	very strong						
<i>MSGARCH</i>	very strong	-	very strong	very strong	very strong	very strong					
<i>BLV</i>	very strong	-	very strong	very strong	very strong	very strong	strong				
<i>tLV</i>	very strong	-	very strong	very strong	very strong	very strong	positive	-			
<i>ASV1</i>	very strong	~	very strong	very strong	very strong	very strong	very strong	positive	strong		
<i>ASV2</i>	very strong	~	very strong	very strong	very strong	very strong	very strong	positive	very strong	~	
<i>Heston</i>	very strong	-	very strong	very strong	very strong	very strong	-	-	-	-	-

Figure 64: model weighting (short) - Bayes factors (~... similar model performance, - ... model is inferior)

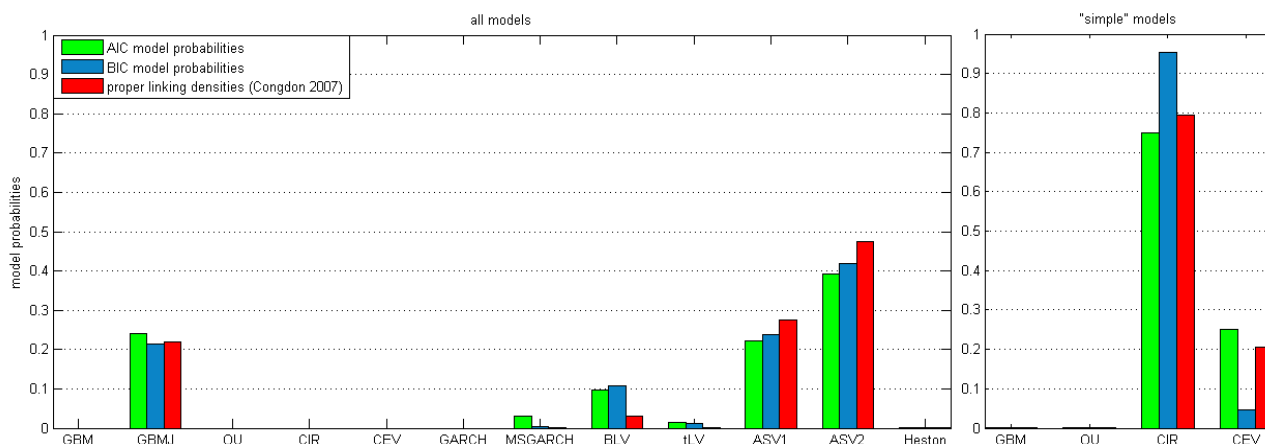


Figure 65: model weighting (short) - AIC, BIC, & posterior model probabilities

now the third dominant model.

All things considered, the asymmetric log-variance models ASV1 and ASV2 seem to be most suitable to describe the synthetic Dec09 spot prices irrespectively whether we analyze the full time-series from 2005 up to 2009 or whether we exclude the turbulent phase in April/May 2006. To some extent the basic log-variance process BLV seems also relevant in a model weighting. At least for the short time-series, we should consider the GBMJ as it features substantial model probability. When we want to restrict to the plain vanilla models there is no consistent result. At least the CEV process features some remarkable model probability for all measures (AIC, BIC, and posterior model probabilities) and data sets (full and restricted time-series). In contrast, for the GBMJ, OU, and CIR processes, we have estimated mixed levels of model probabilities. Hence, we advice not to dump any of these models.

## 11 Bayesian VaR

The *value-at-risk* (VaR) measures the downside risk of an uncertain asset price, here the EUA price  $S_t$ . Roughly speaking, the risk for  $S_t$  to fall below  $VaR(\beta)$  can be quantified by a probability of  $1 - \beta$  for  $\beta \geq 0.5$  and by a probability of  $\beta$  for  $\beta < 0.5$ . The frequentist

version<sup>153</sup>

$$VaR_F(\beta) = VaR_F(\beta|\hat{\theta}) = \begin{cases} \inf \left( s_t | P \left( S_t \leq s_t | \hat{\theta} \right) = 1 - \beta \right), & \text{for } \beta > 0.5 \\ \inf \left( s_t | P \left( S_t \geq s_t | \hat{\theta} \right) = \beta \right), & \text{for } \beta < 0.5 \end{cases}$$

evaluates the asset price distribution  $P \left( S_t \leq s_t | \hat{\theta} \right) = \int_0^{s_t} f_{S_t}(\varsigma | \hat{\theta}) d\varsigma$  given the frequentist point estimate  $\hat{\theta}$ . In contrast, the Bayesian formulation

$$VaR_B(\beta) = \begin{cases} \inf \left( s_t | \int_0^{s_t} \pi(\varsigma | y) d\varsigma = 1 - \beta \right), & \text{for } \beta > 0.5 \\ \inf \left( s_t | \int_{s_t}^{\infty} \pi(\varsigma | y) d\varsigma = \beta \right), & \text{for } \beta < 0.5 \end{cases}$$

evaluates the posterior Bayesian uncertainty on the asset price<sup>154</sup>

$$\pi(s_t | y) = \int f_{S_t}(s_t | \theta) \pi(\theta | y) d\theta$$

which is a posterior weighted average of potential candidates  $f_{S_t}(s_t | \theta)$  of the data generating process  $f_{S_t}(s_t | \theta^*)$ .<sup>155</sup> Again,  $\pi(\theta | y)$  is the posterior of the parameter vector  $\theta$  and  $y$  is a set of observations.

In the following, we first present the ex-post one-day-ahead VaRs for our Dec09 spot prices. For the calculation, we assume that the CO2 price follows the asymmetric log-variance model with intra-temporal correlation (ASV2) - the best performing model according to our Bayesian model weighting (see section 10.3). Subsequently, we estimate ex-ante multi-period-ahead VaRs based on the GBM and OU process.

## 11.1 Bayesian Ex-Post VaR

The concept of ex-post VaRs allows us to quantify the past risk (April 22, 2005 up to May 07, 2009) of the one-day-ahead EUA price  $S_{t+1} | s_t, y$  given  $s_t$  in  $t$  and the full set of observations  $y$  used for parameter estimation.

<sup>153</sup>In contrast to credit risk, we have no loss variable but the CO2 price. Traders long and short in the EU ETS might have different understandings of downside risk. Hence, our definition of the VaR is special.

<sup>154</sup>Beware of a misinterpretation of the Bayesian VaR. It is generally not a posterior weighted average of different conditional VaRs, i.e.  $VaR_B(\beta) \stackrel{\text{generally}}{\neq} \int VaR(\beta | \theta) \pi(\theta | y) d\theta$ .

<sup>155</sup>For a clearer presentation, we ignore model uncertainty. Otherwise the posterior uncertainty would be

$$\pi(s_t | y) = \sum_{k=1}^K \int f_{S_t}(s_t | \theta, \mathcal{M}_k) \pi(\theta | y, \mathcal{M}_k) d\theta \cdot \pi(\mathcal{M}_k | y),$$

where  $f_{S_t}(s_t | \theta, \mathcal{M}_k)$  equals the data generating process when  $\mathcal{M}_k$ ,  $k = 1, \dots, K$ , is the "true" model  $\mathcal{M}_k^*$  and  $\theta$  is the "true" parameter vector  $\theta^*$ .

In the frequentist approach, we generate  $W_1 = 100,000$  samples  $s_{t+1}$ , for  $t = 1, \dots, T - 1$ , from the presumed data generating process  $f_{S_{t+1}}(s_{t+1} | s_t, \hat{\theta})$  where we simply set the point estimates equal to the mean of the posteriors, i.e.  $\hat{\theta} = \sum_{w=1}^W \theta^{(w)} / W$ . We base our VaR estimation on the best performing model - the asymmetric log-variance model with intra-temporal correlation (ASV2, see section 10.1.10). Consequently, we can easily sample according to

$$S_{t+1} = s_t \exp(\bar{y} + \exp(h_t/2) \cdot \varepsilon_t),$$

where  $\bar{y}$  is the empirical mean log-return and  $h_t$  is a realization of the latent log-variance  $H_t = \hat{\alpha} + \hat{\psi} \cdot h_{t-1} + \hat{\tau} \cdot \eta_t$ . We have to account for the correlated errors  $\varepsilon_t$  and  $\eta_t$  following a standard bivariate Gaussian distribution with Pearson's correlation coefficient  $\hat{\rho}$ .

The Bayesian approach is similar but it samples from the (unconditional) Bayesian predictive density (see section 6.4)

$$\pi_{S_{t+1}}(s_{t+1} | s_t, y) = \int f_{S_{t+1}}(s_{t+1} | s_t, \theta) \pi(\theta | y) d\theta.$$

We generate  $W_2 = 100$  samples  $s_{t+1}$  for each of the  $W = 40,000$  parameter vectors  $\theta^{(w)}$ ,  $w = 1, \dots, W$ , simulated by the MCMC technique. This sums up to  $W \cdot W_2 = 4,000,000$  samples  $(s_{t+1}^{(w, w_2)})$  for  $w = 1, \dots, W$  and  $w_2 = 1, \dots, W_2$ .<sup>156</sup>

In the upper subplot of figure 66, you can find our Dec09 spot rates as well as the frequentist and Bayesian VaRs ( $VaR_{(0.99)}$  &  $VaR_{(0.01)}$ ) for the ASV2 model. We additionally present the frequentist VaR for the GBM to roughly quantify the risk ignored by assuming the "simple" Black & Scholes world. You can easily see, that the frequentist GBM VaRs (green lines) underestimate the downwards and upwards risk relatively to the ASV2 model (black & red lines).

The differences between the frequentist and Bayesian ASV2 VaRs are low. We separately depict them in the lower subplot of figure 66. Generally, we find the Bayesian VaR more conservative, i.e.  $VaR_B(0.99) \leq VaR_F(0.99)$  &  $VaR_B(0.01) \geq VaR_F(0.01)$ . This empirical finding has already been proofed theoretically by Pollard (Nov. 2007). In table 29, we have summarized the mean differences between the VaRs (row minus column).

<sup>156</sup>For an efficient sampling of the correlated errors  $\varepsilon_t$  and  $\eta_t$  in Matlab, we favor vector programming over loops. In the Bayesian approach, we stack the MCMC samples  $\theta_{MCMC} = (\theta^{(1)}, \dots, \theta^{(W)})'$   $W_2$  times to  $\tilde{\theta} = (\theta_{MCMC}^{(1)}, \dots, \theta_{MCMC}^{(W_2)})'$ . Then, we generate  $W \cdot W_2$  samples from the standard bivariate Gaussian distribution with  $W$  different correlation coefficients. This we cannot accomplish by the means of the Matlab command *mvnrnd*( $\cdot$ ). Hence, we apply the *Cholesky decomposition*. First, we sample the i.i.d. standard Gaussian variables  $z_{ij}$ , for  $i = 1, \dots, W \cdot W_2$  and  $j = 1, 2$ . Subsequently, we set  $\varepsilon_{ti} = z_{i1}$  and  $\eta_i = \rho \varepsilon_{ti} + \sqrt{1 - \rho^2} z_{i2}$ . This results in an empirical correlation of  $\widehat{corr}(\varepsilon_t, \eta_t) \approx \rho$ .

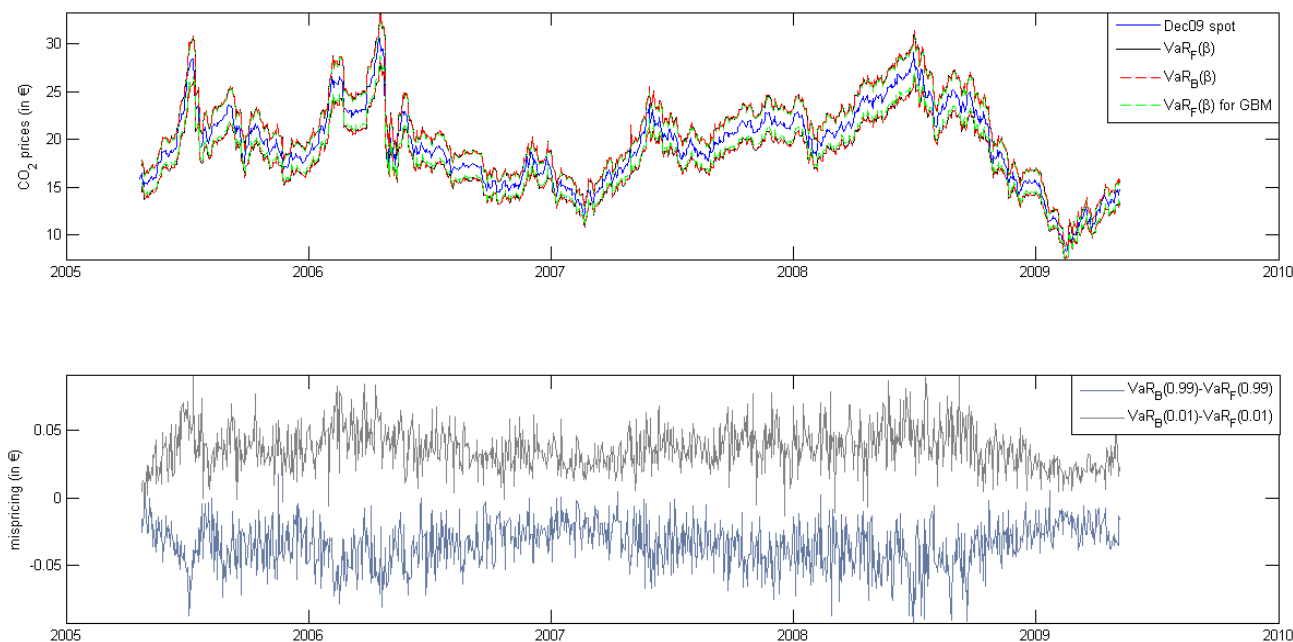


Figure 66: Bayesian ex-post value-at-risk

	$VaR_F^{GBM}(99\%)$	$VaR_F(99\%)$		$VaR_F^{GBM}(1\%)$	$VaR_F(1\%)$
$VaR_F(99\%)$	-0.327€		$VaR_F(1\%)$	0.130€	
$VaR_B(99\%)$	-0.362€	-0.035€	$VaR_B(1\%)$	0.167€	0.037€

Table 29: differences between the VaRs

Indeed, the difference between the frequentist and Bayesian VaR is easy to understand:  $VaR_F$  tries to describe the "true" data generating process  $f_{S_{t+1}}(s_{t+1}|s_t, \theta^*)$  but ignores any parameter uncertainty.<sup>157</sup>  $VaR_B$  instead analyzes the Bayesian predictive density

$$\pi_{S_{t+1}}(s_{t+1}|s_t, y)$$

reflecting the uncertainty on the true parameter vector  $\theta^*$ . Hence, the Bayesian VaR calls for a *risk cushion* or *premium* depending on the parameter uncertainty. Pollard (Nov. 2007) explains the *capital charge puzzle*<sup>158</sup> by a Bayesian (= subjective) risk cushion. By the way, Weitzman (2007) offers the same arguments to solve the *equity premium puzzle*.<sup>159</sup>

<sup>157</sup>Here, we ignore the fact of an uncertain model.

<sup>158</sup>Banks announce too conservative VaRs not compatible with frequentist VaRs.

<sup>159</sup>The observed risk premiums contradict reasonable assumptions on the market's risk aversion as long as parameter uncertainty is ignored.

In table 30, we illustrate the relative number of "hits" (= VaR is crossed by the spot price) for the different VaRs.

	"true" model	$VaR_F^{GBM}(\beta)$	$VaR_F(\beta)$	$VaR_B(\beta)$
$\beta = 99\%$	1%	2.03%	0.87%	0.68%
$\beta = 1\%$	1%	1.45%	1.26%	1.06%

Table 30: relative number of "hits" for different VaRs

The GBM fundamentally underestimates the variability of the EUA price. The respective lower VaR is crossed more than twice compared to the "true" model. The upper VaR is also violated too often. It can be seen that the Bayesian VaR is less often hit relative to the frequentist VaR. In the case of  $VaR_B(99\%)$ , the Bayesian hits are below the "true" 1% - the Bayesian VaR is absolutely conservative. Unfortunately,  $VaR_B(1\%)$  is too low meaning there are too many hits ( $> 1\%$ ). This provides some evidence that even our best performing ASV2 model cannot perfectly mirror the data generating process ("true" model with "true" parameters). Nevertheless, there is still a risk cushion relative to  $VaR_F(1\%)$ .

## 11.2 Bayesian Ex-Ante VaR

In the next step, we estimate some ex-ante multi-periods-ahead (up to one year) VaRs starting from the last Dec09 spot rate (14.77€) on May 7, 2009. To reduce computational burdens, we present VaRs based on the GBM and OU process. We choose these processes because we exactly know their conditional distributions for future EUA prices  $S_{\mathcal{T}}$ ,  $\mathcal{T} > t$ . Both follow a log-Gaussian distribution

$$S_{\mathcal{T}} | s_t, \theta \sim \ln N(m, \varsigma),$$

where  $m_{GBM} = \ln s_t + \mu \cdot \Delta$ ,  $\varsigma_{GBM} = \sigma \cdot \sqrt{\Delta}$ ,  $m_{OU} = \tilde{\mu} + [\ln s_t - \tilde{\mu}] \exp(-\kappa\Delta)$ ,  $\varsigma_{OU} = \sigma \sqrt{[1 - \exp(-2\kappa\Delta)] / 2\kappa}$ , and  $\Delta = \mathcal{T} - t$ .

Thanks to the analytical solution of the GBM and OU process we can directly calculate  $VaR.(99\%)$  for all  $\Delta$  (from five trading days up to one year, see black lines in figure 67).<sup>160</sup> Again, we simply use the mean of the posteriors as point estimates. The Bayesian programming is similar. We simulate  $W_2 = 200$  samples of each conditional CO2 price  $S_{\mathcal{T}} | s_t, \theta^{(w)}$ , for  $w = 1, \dots, W [= 50,000]$ . We approximate the Bayesian VaR with a total of  $W \cdot W_2 = 10,000,000$  samples for each time step  $\mathcal{T} = t + \Delta$  (see red lines in figure 67). In table 31, we have summarized the frequentist misspecification of the CO2 price risk for some key maturities.

<sup>160</sup>Due to a clearer presentation, we solely calculate the downside VaR ( $VaR.(99\%)$ ).

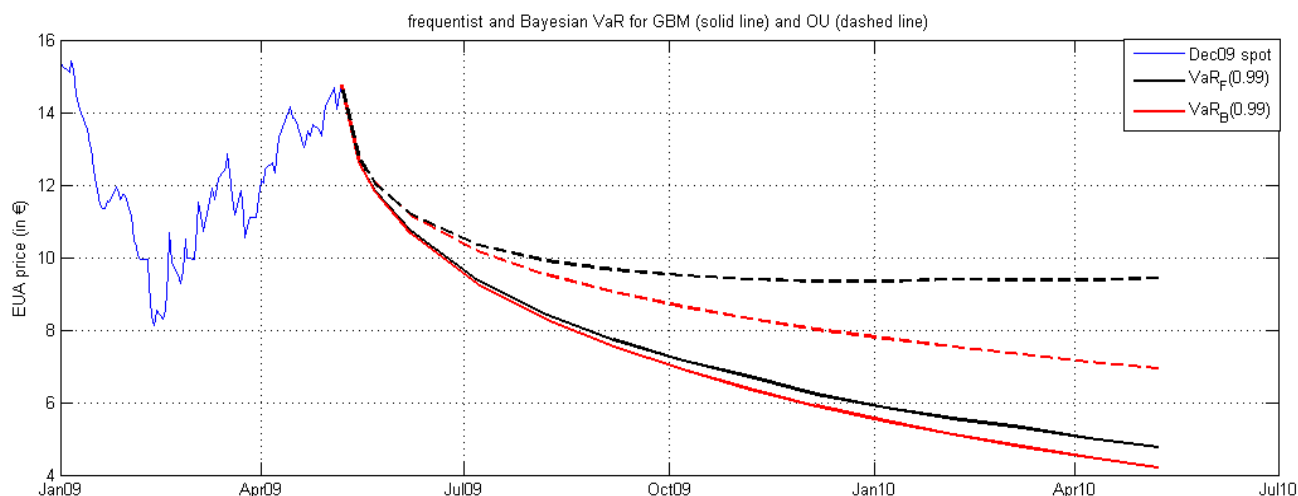


Figure 67: Bayesian ex-ante value-at-risk

$\beta = 99\%$	1 month	3 months	6 months	1 year
$VaR_B^{GBM}(\beta) - VaR_F^{GBM}(\beta)$	-0.037€	-0.142€	-0.306€	-0.575€
$VaR_B^{OU}(\beta) - VaR_F^{OU}(\beta)$	-0.034€	-0.397€	-1.124€	-2.434€

Table 31: frequentist misspecification of the CO2 price risk

Similar to the ex-post VaRs, the ex-ante Bayesian VaRs are more conservative than their frequentist counterparts.

### 11.3 Bayesian Capital Charge

Banks with substantial trading activities are required to insure large losses in their portfolio value by a *capital charge* (Amendment to Capital Accord to Incorporate Market Risks, see BIS 1996). Again, there is a frequentist as well as a Bayesian version of the capital charge. Here, we follow Pollard (Nov. 2007) to compare both.

Although the banks are relatively free in their models to estimate the market risk, the quantification of the capital charge has to base on  $VaR_t(99\%)$  for a  $\Delta = 10$  trading day horizon<sup>161</sup>

$$CC(t) = \max \left( VaR_t(99\%), \frac{Factor(t)}{60} \sum_{\tau=0}^{59} VaR_{t-\tau}(99\%) \right).$$

The capital charge is not only affected by the current VaR but by the average performance

<sup>161</sup>The estimation of the VaR requires at least one year of historical data (see BIS July 2008), pp. 7.

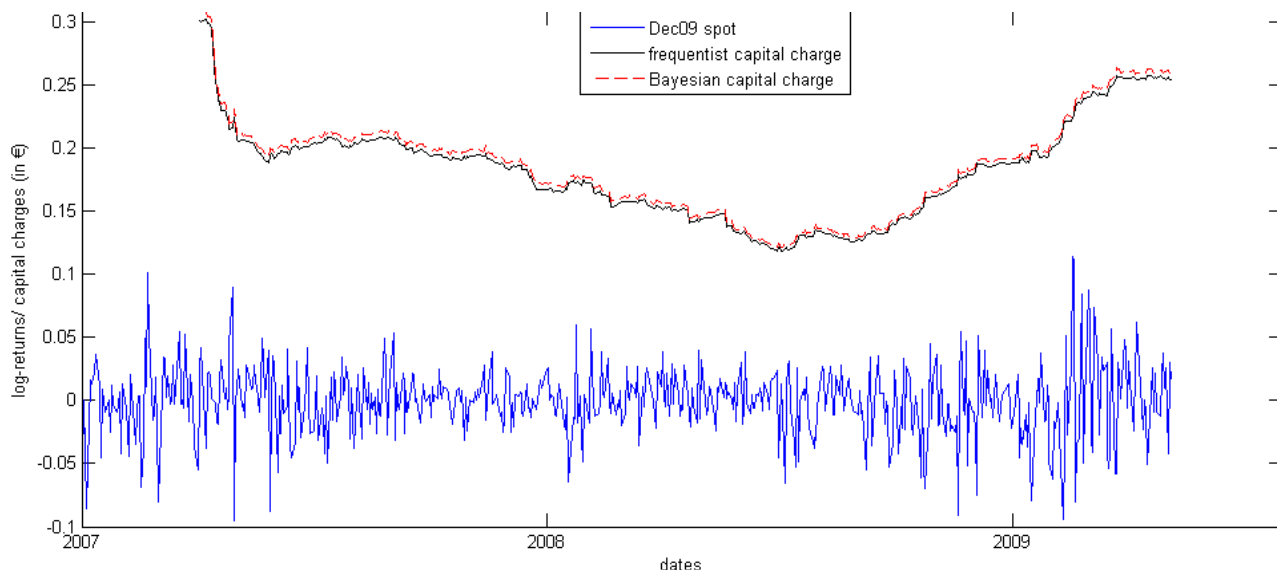


Figure 68: Bayesian capital charge

of the past 60 VaRs. Additionally, there is a correction factor accounting for the number of hits ((log-)return below VaR) during the last trading year

$$Factor(t) = \begin{cases} 3 & \text{for } Hits \leq 4 & \text{(green)} \\ 3 + 0.2 \cdot [N - 4] & \text{for } Hits = 5, 6, 7, 8, 9 & \text{(yellow)} \\ 4 & \text{for } Hits \geq 10 & \text{(red)} \end{cases} ,$$

where a perfectly calibrated model would feature  $Hits = 1$ . If the number of hits is in the red zone, the model is seen as inaccurate and the risk model needs to be improved immediately. We have only calculated frequentist and Bayesian capital charges (per 1€ investment) for the EUA price based on the GBM and the last 250 trading days (see figure 68). As expected, the Bayesian capital charges are higher than the frequentist ones. However, the average risk cushion for the parameter uncertainty is low (0.41€-cents per 1€ investment).

## 12 Derivative Pricing of Some Special Models

In this section, we first present the theoretical concept of Bayesian option pricing and subsequently we exemplarily calculate some Bayesian option prices for plain vanilla European call options on EUA prices.<sup>162</sup> Such option are not purely academic. In a survey (mid 2005)

<sup>162</sup>Wagner (2007) already calculated option prices resulting from a *frequentist* calibration of stochastic processes. He had access to EUA prices from April 22, 2005 up to December 15, 2006.



25 experts<sup>163</sup> believed in a realistic chance for a long-term success of futures/ forwards and - too a lesser extent - of options on the EUA price (Uhrig-Homburg & Wagner 2008). Indeed, the European Climate Exchange (ECX) introduced option instruments in October 2006.

## 12.1 Basic Idea of Option Pricing

In this section we describe Bayesian pricing of plain vanilla call options that guarantee a deterministic payoff  $C(s_{\mathcal{T}}) = C(s_{\mathcal{T}}, \zeta) = \max(s_{\mathcal{T}} - \zeta, 0)$  given the EUA price  $s_{\mathcal{T}}$  in  $\mathcal{T}$  and a strike price  $\zeta$ . As people feature a time preference (people enjoy a payoff now more than a payoff in future) and the future EUA price  $S_{\mathcal{T}}$  is uncertain in  $t < \mathcal{T}$ , the call price  $p_t$  in  $t$  is the discounted value of the payoff  $C(s_{\mathcal{T}})$ . This can be formulized by (see Cochrane 2001)

$$p_t = E_{\mathbb{P}}(D \cdot C(S_{\mathcal{T}})),$$

where  $E_{\mathbb{P}}(\cdot)$  is the expectation in  $t$  under the real world measure  $\mathbb{P}$  and  $D$  is the stochastic discount factor.<sup>164</sup> Without any uncertainty,  $S_{\mathcal{T}} = s_{\mathcal{T}}$ , the time preference can be reflected by discounting with the (continuous) risk free interest rate  $r_t$

$$p_t = E_{\mathbb{P}}(D) \cdot C(s_{\mathcal{T}}) = \exp(-r\Delta) \cdot C(s_{\mathcal{T}}),$$

where  $\exp(-r\Delta) = \exp\left(-\int_t^{\mathcal{T}} r_{\tau} d\tau\right)$  and  $\Delta = \mathcal{T} - t$ .

In the case of an uncertain EUA price in  $\mathcal{T}$ , risk-averse people ask for a risk premium resulting in a lower option price<sup>165</sup>

$$p_t = E_{\mathbb{P}}(D) \cdot E_{\mathbb{P}}(C(S_{\mathcal{T}})) < \exp(-r\Delta) \cdot C(s_{\mathcal{T}}).$$

In such a case we could calibrate a stochastic process and deduce  $E_{\mathbb{P}}(C(S_{\mathcal{T}}))$ . Unfortunately, the risk premium cannot directly be observed as it originates from the utility functions of the market. Generally, a quantification of  $E_{\mathbb{P}}(D)$  is rather arbitrary.

Fortunately, option pricing theory partly allows to avoid the determination of risk premiums. The basic idea is that a risk-neutral and a risk-averse person agree on the same price  $p_t$  if they *disagree* on the expected payoff in  $\mathcal{T}$ , i.e.

$$\exp(-r\Delta) \cdot E_{\mathbb{Q}}(C(S_{\mathcal{T}})) = p_t = E_{\mathbb{P}}(D) \cdot E_{\mathbb{P}}(C(S_{\mathcal{T}})),$$

<sup>163</sup>Experts were representatives of energy providers and a steel producer as well as of banks, energy traders, and exchanges.

<sup>164</sup>The real world measure  $\mathbb{P}$  is the market's perception of the distribution of the data-generating process (ignoring the stochastic discount factor).

<sup>165</sup>The problem becomes even more complicated if the future payoff  $C(S_{\mathcal{T}})$  and the discount factor  $D$  (risk-free interest rate plus risk premium) are stochastically dependent. Then a separate analysis is not possible, i.e.  $E_{\mathbb{P}}(D \cdot C(S_{\mathcal{T}})) \neq E_{\mathbb{P}}(D) \cdot E_{\mathbb{P}}(C(S_{\mathcal{T}}))$ . See Björk (2004), pp. 348, for more details.

where  $E_{\mathbb{Q}}(C(S_{\mathcal{T}})) < E_{\mathbb{P}}(C(S_{\mathcal{T}}))$ . Here,  $\mathbb{Q}$  and  $\mathbb{P}$  are the probability measures reflecting the uncertainty of the risk-neutral and the risk-averse person. To avoid the quantification of the risk adjusted discount factor we need to determine the risk neutral measure  $\mathbb{Q}$ .

This can be done by the calibration of a stochastic process (e.g. GBM, GBMJ, OU, ..., Heston) to the observed CO2 prices  $s_0, \dots, s_t$ . As pricing in the market results from the interaction of risk-averse traders, we can estimate the real world measure  $\mathbb{P}$ , only.<sup>166</sup> In a next step, we need to change the measure from  $\mathbb{P}$  to  $\mathbb{Q}$ . A dominant approach is to find the measure  $\mathbb{Q}$ , that transforms  $\exp(-r\Delta) \cdot S_{\mathcal{T}}$  into a martingale, i.e.  $\exp(-r\Delta) \cdot E_{\mathbb{Q}}(S_{\mathcal{T}}) = s_{\tau}$ .

E.g., the GBM ( $dS_t = \nu s_t dt + \sigma s_t dW_t$ ) has an analytical solution

$$S_{\mathcal{T}} = s_t \exp\left([\nu - \sigma^2/2] \Delta + \sigma [W_{\mathcal{T}} - W_t]\right)$$

with  $E(S_{\mathcal{T}}) = s_t \exp(\nu\Delta)$ . To meet the martingale condition we require  $\nu = r$ . Consequently, we can calculate call prices by

$$p_t = \exp(-r\Delta) \cdot E_{\mathbb{Q}}(C(S_{\mathcal{T}})) \approx \exp(-r\Delta) \frac{1}{W} \sum_{w=1}^W C\left(s_{\mathcal{T}}^{(w)}\right),$$

where we generate  $s_{\mathcal{T}}^{(1)}, \dots, s_{\mathcal{T}}^{(W)}$  from the risk neutral GBM process  $dS_t = r s_t dt + \sigma s_t dW_t$ . Generally, the change of measure relies on the *Radon-Nikodym theorem*<sup>167</sup>

$$E_{\mathbb{Q}}(C(S_{\mathcal{T}}) | \mathcal{F}_t) = \int C(s_{\mathcal{T}}) d\mathbb{Q}(s_{\mathcal{T}}) = \int C(s_{\mathcal{T}}) \frac{d\mathbb{Q}(s_{\mathcal{T}})}{d\mathbb{P}(s_{\mathcal{T}})} d\mathbb{P}(s_{\mathcal{T}}) = \frac{E_{\mathbb{P}}(C(S_{\mathcal{T}}) L_{\mathcal{T}} | \mathcal{F}_t)}{E_{\mathbb{P}}(L_{\mathcal{T}} | \mathcal{F}_t)}, \quad (15)$$

where  $L_{\mathcal{T}} = L_{\mathcal{T}} | \mathcal{F}_{\mathcal{T}} = d\mathbb{Q}(s_{\mathcal{T}}) / d\mathbb{P}(s_{\mathcal{T}})$  is the *Radon-Nikodym derivative* of  $\mathbb{Q}$  with respect to  $\mathbb{P}$  which is  $\mathcal{F}_{\mathcal{T}}$ -measurable but not  $\mathcal{F}_t$ -measurable.<sup>168</sup> As the expectation operator is based on  $\mathcal{F}_t \subseteq \mathcal{F}_{\mathcal{T}}$ , we need to apply Bayes' Theorem after the last equal sign to deal with  $L_{\mathcal{T}} | \mathcal{F}_t$  (see Björk 2004, pp. 440, for a proof).

An important finding is that the Radon-Nikodym derivative is a stochastic process in form of a  $\mathbb{P}$ -martingale  $E_{\mathbb{P}}(L_{\mathcal{T}} | \mathcal{F}_t) = L_t$ . In the case that the EUA prices follow stochastic processes

<sup>166</sup>Of course, the market price of the option results from  $E_{\mathbb{P}}(C(S_{\mathcal{T}}))$ , where the probability measure  $\mathbb{P}$  is based on the aggregated information of the market in  $t$ . In contrast, the measure  $\mathbb{P}$  estimated by the frequentist calibration of stochastic processes relies on the observed asset prices up to  $t$ , only. In Bayesian statistics with informative priors,  $\mathbb{P}$  is additionally influenced by personal information of the statistician or experts who deliver priors.

<sup>167</sup>see Björk (2004), pp. 416, for requirements

<sup>168</sup>The filtration  $\mathcal{F}_t = \mathcal{F}_t^S$  comprises all observed asset prices up to  $t$ . The price  $S_t$  is *adapted* to  $\mathcal{F}_t$  as with the information in  $\mathcal{F}_t$  we know the realization  $s_t$ . Events that can completely be determined by the information in  $\mathcal{F}_t$  are called  $\mathcal{F}_t$ -measurable. E.g., the event  $A = \{S_{\tau} \leq 10, \tau \leq \mathcal{T}\}$  is  $\mathcal{F}_{\mathcal{T}}$ -measurable but not  $\mathcal{F}_t$ -measurable with  $t < \mathcal{T}$ . See Björk (2004) or Oksendal (2007) for more details in measure theory.

that are driven by  $d$ -dimensional Wiener terms, the non-negative  $L_t$  is the solution of the simple GBM  $dL_t = \varphi_t L_t dW_t^{\mathbb{P}}$  where<sup>169</sup>  $L_0 = 1$  and  $\varphi_t$  is a suitable *Girsanov kernel* and we just index  $dW_t^{\mathbb{P}}$  to emphasis the measure  $\mathbb{P}$ . Applying Ito's lemma it follows

$$L_t = \exp \left( \int_0^t \varphi_\tau dW_\tau^{\mathbb{P}} - \int_0^t \varphi_\tau^2 d\tau \right).$$

The *Girsanov theorem* states that a change of measure from  $\mathbb{P}$  to  $\mathbb{Q}$  can be achieved by replacing the Wiener term<sup>170</sup>

$$dW_\tau^{\mathbb{P}} = \varphi_t dt + dW_\tau^{\mathbb{Q}}.$$

Again assume  $dS_t = \nu s_t dt + \sigma s_t dW_\tau^{\mathbb{P}}$ . According to the Girsanov theorem we can set  $dW_\tau^{\mathbb{P}} = \varphi_t dt + dW_\tau^{\mathbb{Q}}$  to determine the  $\mathbb{Q}$ -measure process of  $S_t$ :  $dS_t = [\nu + \sigma \varphi_t] s_t dt + \sigma s_t dW_\tau^{\mathbb{Q}}$ . Finally, we need to determine the suitable Girsanov kernel. It needs to guarantee  $\nu + \sigma \varphi_t = r$ . Consequently, it follows  $\varphi_t = -[\nu - r] / \sigma$  which is the negative of the market price of risk.

## 12.2 Bayesian Option Pricing

The term *Bayesian option pricing* (BOP) is sometimes used for the estimation of an option price  $\hat{p}_t$  based on observed asset prices  $s_{\leq t} = (s_0, \dots, s_t)'$  and option prices  $p_{\leq t} = (p_0, \dots, p_t)'$ . Because of a lack of traded options, the option price  $p_t$  can often be deduced from the observed stock prices  $s_{\leq t} = (s_0, \dots, s_t)'$ , only.

Assume we have access to past EUA prices  $s = (s_1, \dots, s_T)'$  and option prices  $p = (p_1, \dots, p_T)'$  of a simple plain vanilla call with uncertain payoff  $C(S_T) = C(S_T, \zeta) = \max(S_T - \zeta, 0)$  in  $\mathcal{T}$ . Again,  $\zeta$  is the strike price. Given some suitable stochastic models for the asset price  $S_t$  and option price  $P_t$ , we can calculate the posteriors of the parameters  $\pi(\theta | z) \propto L_z(\theta) \pi(\theta)$ , where  $z = (s_{\leq t}, p_{\leq t})'$  and  $\theta$  is the parameter vector of the asset and option price models.<sup>171</sup> The determination of the likelihood function  $L_z(\theta) = \prod_{\tau=0}^t f_{S_\tau, P_\tau}(s_\tau, p_\tau | \theta)$  is the crux of the matter. Generally, the CO2 price is assumed to follow a stochastic process of the form<sup>172</sup>

$$dS_t = \underbrace{\mu(t, s_t, x_t, \theta)}_{drift} \cdot dt + \underbrace{\sigma(t, s_t, x_t, \theta)}_{volatility} \cdot dW_t + \underbrace{d\left(\sum_{i=1}^{N_t} Z_i\right)}_{jump} \quad (16)$$

<sup>169</sup>Again, the martingale property applies, i.e.  $E_{\mathbb{P}}(L_T) = L_0 = 1$ .

<sup>170</sup>See Oksendal (2007), pp. 159, for requirements and details and Björk (2004), likewise pp. 159, for an intuitive derivation.

<sup>171</sup>In the case of competing models  $\mathcal{M}_1, \dots, \mathcal{M}_K$ , the posterior model probabilities are

$$\pi(\mathcal{M}_k | z) = \frac{\int L_z(\theta, \mathcal{M}_k) \pi(\theta | \mathcal{M}_k) d\theta \cdot \pi(\mathcal{M}_k)}{\sum_{l=1}^K \int L_z(\theta, \mathcal{M}_l) \pi(\theta | \mathcal{M}_l) d\theta \cdot \pi(\mathcal{M}_l)}.$$

<sup>172</sup>see section 10 for more details on the process

or some nested model. The likelihood  $L_y(\theta)$  for  $y_t = \ln s_{t+1} - \ln s_t$  can easily be derived from the discretized version of the respective stochastic process. In contrast, the construction of the likelihood for the call option prices is not that straightforward. In fact, the conditional call price  $p_t(\theta)$  is deterministic. Hence, we cannot set up a likelihood function. Moreover, a model would normally be over-determined as mostly  $T > \dim(\theta)$ .

To cope with that problem we can impose a measurement error  $\varepsilon_t \sim N(0, s)$  for the option price<sup>173</sup>

$$\tilde{P}_t = p_t(\theta) + \varepsilon_t.$$

This allows us to set up the likelihood easily if there is an analytical solution for  $p_t(\theta)$ . If the EUA price follows a GBM,  $p_t(\theta)$  is the Black & Scholes formula.<sup>174</sup>

As we have no access to liquid options on EUA contracts, we follow a more simple Bayesian option pricing approach. In return, we have to assume that the market is certain about the "true" model  $\mathcal{M}^*$  and parameters<sup>175</sup>  $\theta^*$  but we - the analysts - are uncertain about the market assessment.<sup>176</sup> In this case we need to weight the conditional option price by our posteriors on the model and the parameters<sup>177</sup> (Bunnin et al. 2002)

$$\begin{aligned} p_t &= \sum_{k=1}^K \int_{\theta_k} p_t(\theta, \zeta, \Delta, s_t, \mathcal{M}_k) \pi(\theta_k | s_{\leq t}, \mathcal{M}_k) d\theta_k \cdot \pi(\mathcal{M}_k | s_{\leq t}) \\ &= \exp(-r\Delta) \cdot \sum_{k=1}^K \int_{s_{\mathcal{T}}} \int_{\theta_k} C(s_{\mathcal{T}}) \Upsilon(s_{\mathcal{T}} | s_t, \theta_k, \mathcal{M}_k) \pi(\theta_k | s_{\leq t}, \mathcal{M}_k) d\theta_k ds_{\mathcal{T}} \cdot \pi(\mathcal{M}_k | s_{\leq t}), \end{aligned}$$

where  $s_{\leq t}$  are the observed EUA prices,  $\Delta = \mathcal{T} - t$ , and  $\Upsilon(s_{\mathcal{T}} | s_t, \theta_k, \mathcal{M}_k)$  is the conditional transition probability in model  $k$  to reach a price  $s_{\mathcal{T}}$  in  $\mathcal{T}$  if it is  $s_t$  in  $t$ .

The advantage of the transition probability representation is the closed form of  $\Upsilon(s_{\mathcal{T}} | \cdot)$  for some simple models  $dS_t = a(s_t, t) dt + b(s_t, t) dW_t$  where the function  $s_{\mathcal{T}} = g(\Delta w_t | s_t, t, \theta_k, \mathcal{M}_k)$  and its inverse  $\Delta w_t = w_{\mathcal{T}} - w_t = g^{-1}(s_{\mathcal{T}} | s_t, t, \theta_k, \mathcal{M}_k)$  exist in analytical form. Then the transition probability can be calculated by a density transformation

$$\Upsilon(s_{\mathcal{T}} | s_t, \theta_k, \mathcal{M}_k) = \left| \frac{\partial g^{-1}(s_{\mathcal{T}} | \cdot)}{\partial s_{\mathcal{T}}} \right| \cdot \phi\left(g^{-1}(s_{\mathcal{T}} | \cdot) \mid 0, \sqrt{\Delta}\right)$$

<sup>173</sup>see Johannes & Polson (2003), pp. 32, Darsinos & Satchell (2001), or Pollard (Aug. 2007) for more details

<sup>174</sup>Alternatively, Malone & Horst (2006) and Gzyl et al. (2006) present a sophisticated but rather academic approach relying on the Radon-Nikodym derivative where the Bayes rule can be used without the assumption of a measurement error.

<sup>175</sup>Anderson et al. (2000) present a rather academic proposal to account for traders' parameter uncertainty.

<sup>176</sup>see Eraker 2004, Eraker et al. 2003, or Bunnin et al. 2002

<sup>177</sup>For a clearer presentation we ignore latent variables  $X$ .

The  $\mathbb{Q}$ -measure transition probability of a GBM is the log-normal pdf

$$\Upsilon(s_{\mathcal{T}}|s_t, \theta_k, \mathcal{M}_k) = \frac{1}{\sigma s_{\mathcal{T}} \sqrt{2\pi\Delta}} \exp\left(-\frac{1}{2\Delta\sigma^2} \left[\ln \frac{s_{\mathcal{T}}}{s_t} + [r - \sigma^2/2] \Delta\right]^2\right), \quad (17)$$

resulting from  $s_{\mathcal{T}} = g(\Delta w_t | \cdot) = s_t \exp([r - \sigma^2/2] \Delta + \sigma \Delta w_t)$  and  $\Delta w_t = g^{-1}(s_{\mathcal{T}} | \cdot) = \left[\ln \frac{s_{\mathcal{T}}}{s_t} - [r - \sigma^2/2] \Delta\right] / \sigma$ . Knowing this transition probability, we can generate  $W$  samples of  $S_{\mathcal{T}}$  and approximate the call option price by  $\exp(-r\Delta) \sum_{w=1}^W C\left(s_{\mathcal{T}}^{(w)}\right) / W$ .

If there is no analytical solution to the transition probability we need to approximate  $\Upsilon(s_{\mathcal{T}}|s_t, \theta_k, \mathcal{M}_k) = \prod_{\tau=t}^{\mathcal{T}-1} \Upsilon(s_{\tau+1}|s_{\tau}, \theta_k, \mathcal{M}_k)$ . This we can do by the following algorithm:<sup>178</sup>

1. Generate  $W_1$  times a price path  $s_{t+1}^{(w_1, w_2)}, \dots, s_{\mathcal{T}}^{(w_1, w_2)}$  from  $\Upsilon\left(s_{l+1}|s_l, \theta^{(w_2)}\right)$  where the parameter vector  $\theta^{(w_2)}$ ,  $w_2 = 1, \dots, W_2$ , is sampled from the posterior  $\pi(\theta|y)$  (see section 10).<sup>179</sup>
2. Approximate the option price by  $\exp(-r\Delta) \sum_{w_1=1}^{W_1} \sum_{w_2=1}^{W_2} C\left(s_{\mathcal{T}}^{(w_1, w_2)}\right) / [W_1 \cdot W_2]$ .

## 12.3 Some Simulation Studies

In the following, we present some simulation studies to analyze the effect on the option prices when parameter uncertainty is incorporated. We compare the mean European call price for EUA contracts resulting from  $W = 10,000$  MCMC samples of the parameter posterior (*Bayesian option pricing*) to the call price originating from the means of the posterior parameters (*frequentist option pricing*)

$$p_t^{Bayes} = \frac{1}{W} \sum_{w=1}^W p_t\left(\theta^{(w)}, r_{\Delta}, \zeta, \Delta, s_t\right) \text{ vs. } p_t^{freq} = p_t\left(\frac{1}{W} \sum_{w=1}^W \theta^{(w)}, r_{\Delta}, \zeta, \Delta, s_t\right),$$

where  $\Delta = \mathcal{T} - t$  is the time to expiry,  $r_{\Delta}$  is the maturity adequate risk-free interest rate (p.a.),  $\zeta$  is the strike price, and  $s_t$  is the current EUA price.

In our simulation studies, we intend to estimate call prices in  $t = \{\text{May 07, 2009}\}$ , the last trading day of our observations. At this day, our synthetical EUA spot price was  $s_t = 14.77\text{€}$ . We analyze different maturities  $\Delta$  and strike prices  $\zeta$  for the options. As a rule of thumb, a data based forecasting should not exceed 20% up to 30% of the period of past observations

<sup>178</sup>For a clearer presentation we ignore the uncertainty on the correct model  $\mathcal{M}_k$ ,  $k = 1, \dots, K$ .

<sup>179</sup>The conditional distribution  $\Upsilon\left(s_{\tau+1}|s_{\tau}, \theta^{(w_2)}\right)$  is mostly a nested model of equation 16 under measure  $\mathbb{Q}$ .

(Vose 2008). With four years of observations we restrict option maturity to a maximum of one year, i.e.  $\Delta \leq 1$ . As a proxy for the risk-free interest rate we use EURIBOR rates offered on May 07, 2009 for maturities between one week and one year. We approximate EURIBOR rates for any  $\Delta \in (0, 1]$  by a polynomial of degree five.<sup>180</sup> We define, strike prices  $\zeta$  between 10€ and 20€ which allows to evaluate *in-the-money* ( $\zeta < s_t$ ), *at-the-money* ( $\zeta = s_t$ ), and *out-of-the-money* ( $\zeta > s_t$ ) call option prices.

### 12.3.1 Geometric Brownian Motion (GBM)

Because of its importance and its simplicity we start with the Geometric Brownian motion (GBM). For European call option pricing, we do not need the analytical solution of the transition probability  $\Upsilon(s_T | s_t, \theta_k, \mathcal{M}_k)$ , we derived in equation 17. In fact, the Black & Scholes formula is the analytical solution of the conditional option price  $p_t(\theta^{(w)}, r_\Delta, \zeta, \Delta, s_t) = BS_t(\sigma^{(w)}, r_\Delta, \zeta, \Delta, s_t)$  where (see Bingham & Kiesel 2004, p. 133)

$$BS_t(\sigma^{(w)}, r, \zeta, \Delta, s_t) = s_t \cdot \Phi(d_1) - \zeta \cdot \exp(-r\Delta) \cdot \Phi(d_2),$$

$$d_1 = [\ln(s_t/\zeta) + [r + \sigma^2/2] \Delta] / \sigma\sqrt{\Delta} \quad \& \quad d_2 = d_1 - \sigma\sqrt{\Delta}.$$

In table 32 we have summarized Bayesian option prices

$$\hat{p}_t^{Bayes} = \frac{1}{W} \sum_{w=1}^W BS_t(\sigma^{(w)}, r_\Delta, \zeta, \Delta, s_t)$$

and their standard deviations<sup>181</sup> (see brackets) for some combinations of maturity and strike price.

call prices	$\Delta = 1/4$	$\Delta = 1/2$	$\Delta = 1$
$\zeta = 14\text{€}$	1.8093€ (0.029056€)	2.389€ (0.041193€)	3.2209€ (0.057482€)
$\zeta = 14.77\text{€}$	1.4245€ (0.030691€)	2.0281€ (0.042994€)	2.8874€ (0.059643€)
$\zeta = 15\text{€}$	1.322€ (0.030892€)	1.9289€ (0.043327€)	2.7935€ (0.060144€)

Table 32: GBM option prices

In the upper plot of figure 69 we visualize the prices from the table above. Additionally, in the lower subplot, we present the mispricing ( $\hat{p}_t^{Bayes} - \hat{p}_t^{freq}$ ) of the frequentist option

<sup>180</sup>We find  $r_\Delta = -0.040141\Delta^5 + 0.084461\Delta^4 - 0.034092\Delta^3 - 0.035873\Delta^2 + 0.03615\Delta + 0.0062165$ .

<sup>181</sup> $\widehat{Std}(p_t(\theta, r_\Delta, \zeta, \Delta, s_t)) = \sum_{w=1}^W [p_t(\theta^{(w)}, r_\Delta, \zeta, \Delta, s_t) - \hat{p}_t] / [W - 1]$

prices  $\hat{p}_t^{freq} = BS_t \left( \frac{1}{W} \sum_{w=1}^W \sigma^{(w)}, r_\Delta, \zeta, \Delta, s_t \right)$ . A nice result under the GBM assumption is the insignificance of parameter uncertainty. We find a mean (absolute) mispricing of 0.01€-cents (0.01€-Cents) with a maximum absolute mispricing of 0.029€-cents. This outcome is not surprising as the single uncertain parameter necessary for option pricing is the volatility  $\sigma$  which has an extremely informative posterior (mean 0.477, std. dev. 0.010).

### 12.3.2 Geometric Brownian Motion with Jumps (GBMJ)

Another important process in finance and in our results for the short time-series is the Geometric Brownian motion with jumps (GBMJ) proposed by Merton (1976)

$$\begin{aligned} dS_t &= \left[ \mu + \frac{1}{2}\sigma^2 \right] s_t dt + \sigma s_t dW_t^{\mathbb{P}} + \left[ e^{J_t^{\mathbb{P}}} - 1 \right] s_t dN_t^{\mathbb{P}}, \text{ or} \\ d \ln S_t &= \mu dt + \sigma dW_t^{\mathbb{P}} + J_t^{\mathbb{P}} dN_t^{\mathbb{P}}, \end{aligned}$$

where  $dW_t^{\mathbb{P}} \sim N(0, \sqrt{dt})$ ,  $J_t^{\mathbb{P}} \sim N(\mu_J, \sigma_J)$ , and  $dN_t^{\mathbb{P}} \sim Pois(\lambda_J \cdot \Delta)$  are independent risk factors (see section 10.1.2).

For option pricing we need to specify the risk-neutral process under measure  $\mathbb{Q}$  (Merton 1976). Such a process has (1) a risk neutral drift  $\varphi$  but the same volatility and jump term<sup>182</sup> and (2) it needs to be a martingale, i.e.  $E_{\mathbb{Q}}(S_{t+\Delta}) = \exp(r\Delta) \cdot s_t$  or

$$E_{\mathbb{Q}}(\Delta S_t) = \varphi s_t \Delta + \sigma s_t E_{\mathbb{Q}}(\Delta W_t^{\mathbb{Q}}) + E_{\mathbb{Q}}\left(e^{J_t^{\mathbb{Q}}} - 1\right) s_t E_{\mathbb{Q}}(\Delta N_t^{\mathbb{Q}}) \stackrel{!}{=} r s_t \Delta,$$

where  $\Delta = \mathcal{T} - t$ ,  $E_{\mathbb{Q}}(\Delta N_t^{\mathbb{Q}}) = \lambda_J \Delta$ , and  $E_{\mathbb{Q}}\left(e^{J_t^{\mathbb{Q}}} - 1\right) = \exp(\mu_J + \sigma_J^2/2) - 1 = \kappa$ . After some rearrangements we get  $\varphi = r - \lambda_J \cdot \kappa$ . Applying Ito's lemma, it follows

$$\Delta \ln S_t = \left[ r - \frac{1}{2}\sigma^2 - \lambda_J \cdot \kappa \right] \Delta + \sigma \Delta W_t^{\mathbb{Q}} + \sum_{\tau=1}^{N_{\mathcal{T}}^{\mathbb{Q}}} J_{\tau}^{\mathbb{Q}},$$

which is equivalent to

$$S_{\mathcal{T}} = s_t \exp \left( \left[ r - \frac{1}{2}\sigma^2 - \lambda_J \cdot \kappa \right] \Delta + \sigma \Delta W_t^{\mathbb{Q}} \right) \prod_{\tau=1}^{N_{\mathcal{T}}^{\mathbb{Q}}} \exp(J_{\tau}^{\mathbb{Q}}), \quad (18)$$

where  $N_{\mathcal{T}}^{\mathbb{Q}}$  is the risk-neutral number of jumps between  $t$  and  $\mathcal{T}$ . Consequently, an European call option could be approximated by  $p_t \approx \exp(-r\Delta) \frac{1}{W} \sum_{w=1}^W C\left(s_{\mathcal{T}}^{(w)}\right)$ .

<sup>182</sup>A risk-neutral person sees the same risk but does it evaluate differently, i.e. no call for a risk premium.

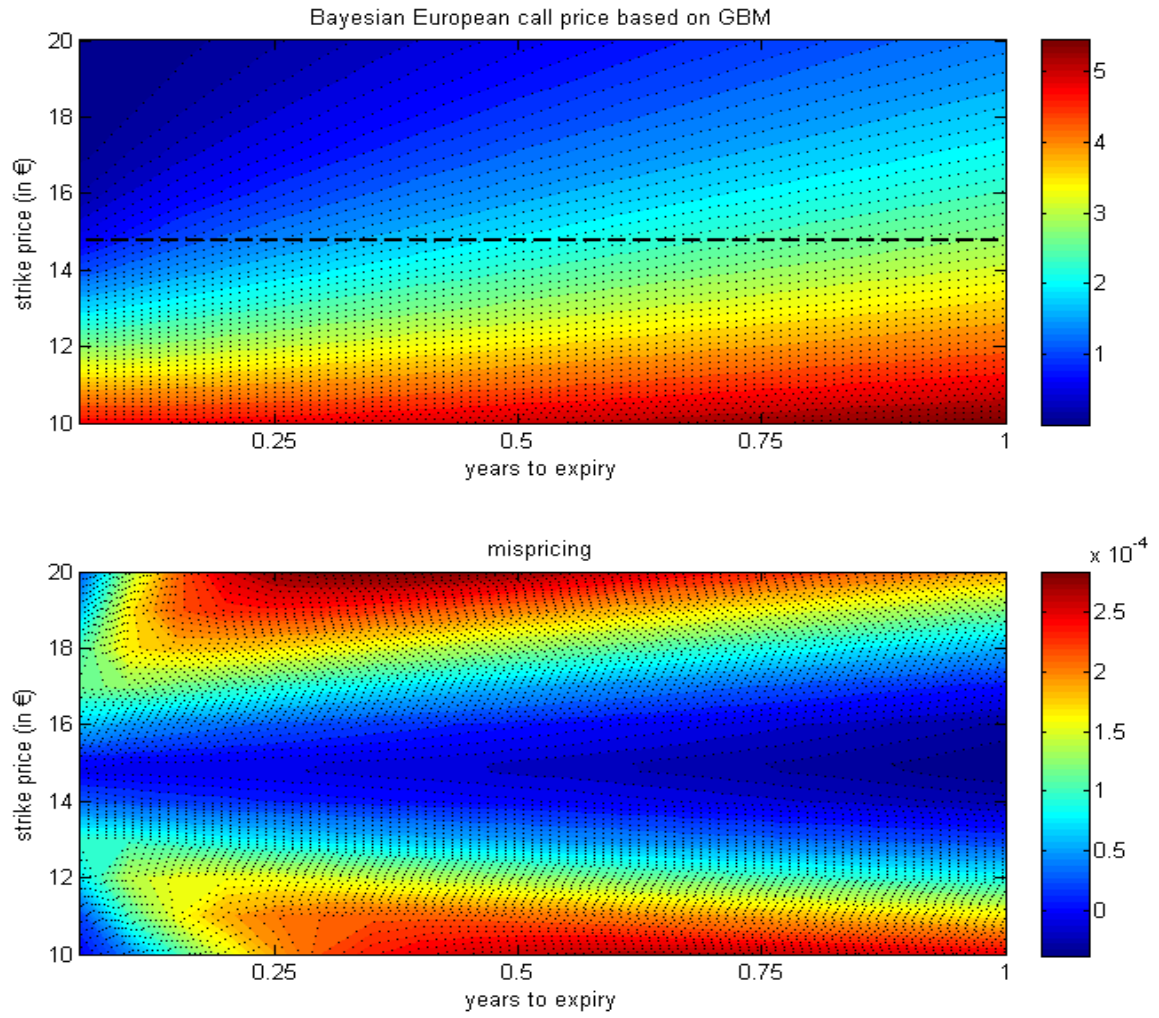


Figure 69: Bayesian option pricing - GBM



Unfortunately, we cannot construct a risk-free hedging portfolio  $dC(S_{t+dt}) - c \cdot dS_t$  by a long position in the option  $C(S_{\mathcal{T}})$  and a short position  $c$  in the underlying stock  $S_{\mathcal{T}}$ . This only allows to eliminate the Brownian risk factor  $W_t^{\mathbb{Q}}$  but not  $J_{\mathcal{T}}^{\mathbb{Q}}$  and  $N_{\mathcal{T}}^{\mathbb{Q}}$ . Option pricing based on equation 18 requires traders not to price the risk of jumps. This is not unrealistic when we assume that the jump risk is independent of the market risk, i.e. the jump risk is unsystematic (Merton 1976). Naturally, there is a price of risk for systematic risks, only. For the special case of  $J_t^{\mathbb{P}} \sim N(\mu_J, \sigma_J)$ , Merton (1976) could even identify a quasi-analytical solution for the price of a plain vanilla European call<sup>183</sup>

$$p_t\left(\theta^{(w)}, r_{\Delta}, \zeta, \Delta, s_t\right) = \sum_{i=0}^{\infty} \frac{\exp(-\lambda_J [\kappa + 1] \Delta) \cdot [\lambda_J [\kappa + 1] \Delta]^i}{i!} BS_t\left(v_i^{(w)}, \rho_i, \zeta, \Delta, s_t\right),$$

where  $BS_t(\cdot)$  is the Black & Scholes formula,  $\rho_i = r - \lambda_J \kappa + i \cdot \ln(1 + k) / \tau$ ,  $v_i = \sqrt{\sigma^2 + i \cdot \sigma_J^2 / \Delta}$ , and  $\kappa = \exp(\mu_J + \sigma_J^2 / 2) - 1$ .

To calculate the Bayesian option prices, we average  $p_t\left(\theta^{(w)}, r_{\Delta}, \zeta, \Delta, s_t\right)$  for our  $W = 10,000$  samples  $\theta^{(w)}$  resulting from the MCMC estimation. In table 33 you can find call prices for some combinations of strike price and maturity.<sup>184</sup>

call prices	$\Delta = 1/4$	$\Delta = 1/2$	$\Delta = 1$
$\zeta = 14\text{€}$	1.7128€ (0.0618€)	2.3703€ (0.0795€)	3.2424€ (0.1047€)
$\zeta = 14.77\text{€}$	1.3984€ (0.0569€)	2.0269€ (0.0789€)	2.9087€ (0.1080€)
$\zeta = 15\text{€}$	1.3084€ (0.0561€)	1.9313€ (0.0787€)	2.8148€ (0.1088€)

Table 33: GBMJ option prices

In the lower subplot of figure 70 you can find a visualization of the Bayesian option prices (upper subplot) and the mispricing ( $\hat{p}_t^{Bayes} - \hat{p}_t^{freq}$ ) of the frequentist prices. The mispricing is modest with a maximum of 1.58€-cent which can however total to a considerable amount in a large portfolio. Generally, the mispricing is most critical for strike prices at-the-money ( $\zeta \approx 14.77\text{€}$ ) and for long maturities.

### 12.3.3 Basic Stochastic Log-Variance (BLV)

The basic log-variance (BLV) model belongs to the class of stochastic volatility models where latent variance  $\exp(h_t)$  is independent of the log-returns  $Y_t = \ln S_{t+1} - \ln s_t$ . In such a case

<sup>183</sup>Hence, we can avoid to sample  $S_{\mathcal{T}}$  according to equation 18.

<sup>184</sup>Please keep in mind that we have approximated  $dN_t^{\mathbb{P}} \sim Pois(\lambda_J \cdot \Delta)$  by  $dN_t^{\mathbb{P}} \sim Ber(\lambda_J \cdot \Delta)$ .

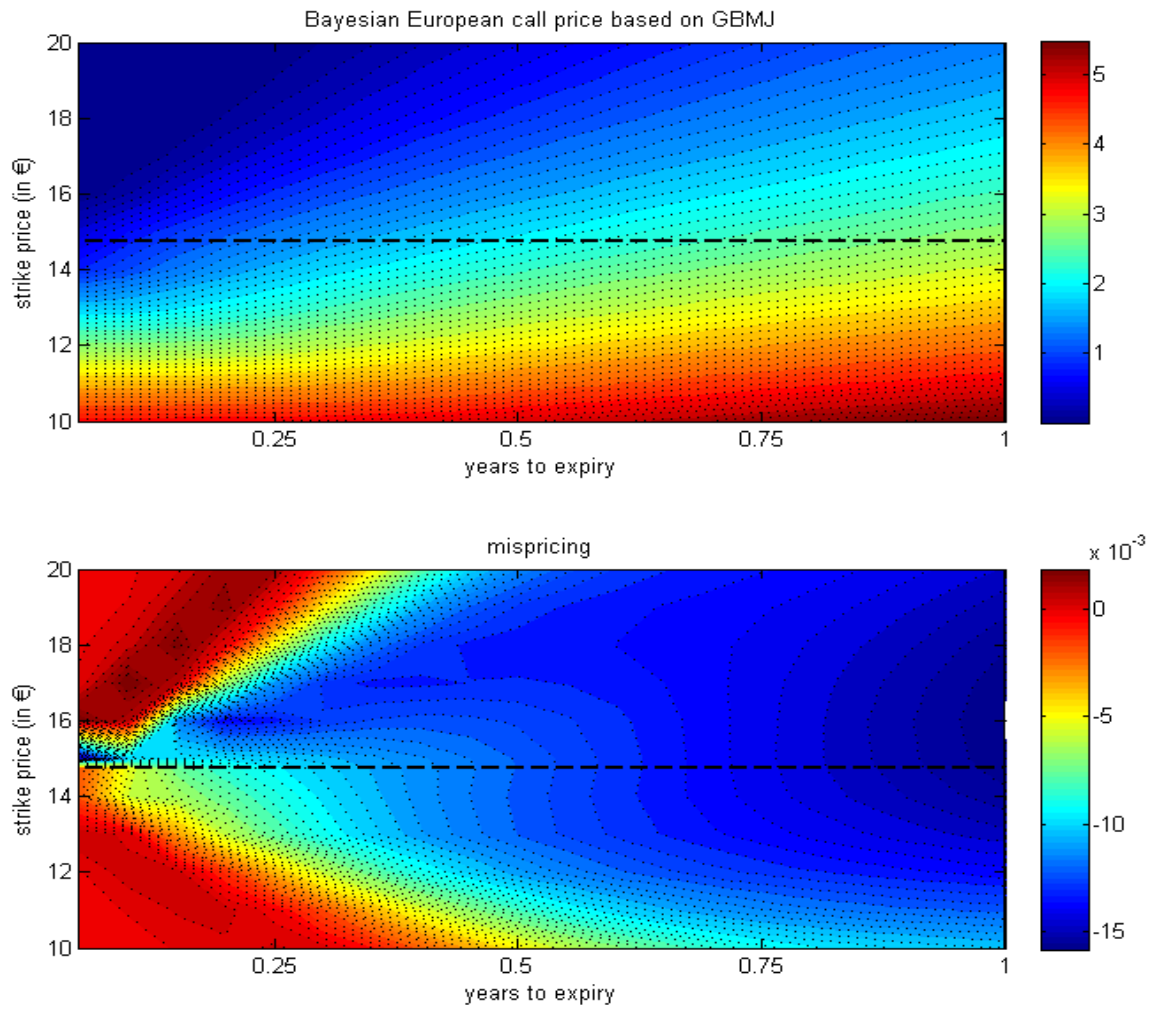


Figure 70: Bayesian option pricing - GBMJ

of independence, Hull & White (1987) show that the knowledge of the average variance  $\overline{\sigma^2}(\mathcal{T})$  up to exercise in  $\mathcal{T}$  is sufficient for option pricing. If we have no analytical solution for  $\overline{\sigma^2}(\mathcal{T})$ , we can run a Monte Carlo simulation. In the case of Wiener increments the call price is simply<sup>185</sup>

$$p_t(\theta) = p_t(\theta, r_\Delta, \zeta, \Delta, s_t) \approx \frac{1}{W_1} \sum_{w_1=1}^{W_1} BS_t \left( \sqrt{\overline{\sigma^2}(\mathcal{T})(w_1)}, r_\Delta, \zeta, \Delta, s_t \right),$$

where

$$\overline{\sigma^2}(\mathcal{T})(w_1) = \frac{1}{\mathcal{T} - t} \sum_{t=t+1}^{\mathcal{T}} \exp \left( h_t^{(w_1)} \right)$$

is the average BLV variance between  $t$  and  $\mathcal{T}$  for sample path  $w_1$ . The frequentist option price is  $p_t(\hat{\theta})$  given the point estimates  $\hat{\theta}$ . From the Bayesian perspective we need additionally reflect our parameter uncertainty

$$p_t^{BLV} = \int p_t^{BLV}(\theta) \pi(\theta|y) d\theta \approx \frac{1}{W_1 \cdot W_2} \sum_{w_1=1}^{W_1} \sum_{w_2=1}^{W_2} BS_t \left( \sqrt{\overline{\sigma^2}(\mathcal{T})(w_1, w_2)}, r_\Delta, \zeta, \Delta, s_t \right),$$

where  $\overline{\sigma^2}(\mathcal{T})(w_1, w_2)$  is the average variance in sample path  $w_1$  based on the MCMC posterior parameter sample  $\theta^{(w_2)}$ ,  $w_2 = 1, \dots, W_2$ .

In the upper subplot of figure 71 we present the Bayesian option prices for several combinations of strike price and maturity. In table 34, we have summarized some of these option prices (standard deviation).

The option prices are below those of the GBM and GBMJ. Nevertheless they are in the same magnitude. In contrast to the other models, the standard deviations of the sampled option prices are large (see brackets in the table). This is why we run a Monte Carlo simulation to sample the mean variance. Fortunately, this is no evidence for instable option price estimators. We generated  $W_1 = 100$  times the average variance for a total of  $W_2 = 10,000$  posterior parameter vectors  $\theta^{(w_2)}$ . The resulting option prices are extremely stable. Repeated sampling according to this procedure results in option prices and standard deviations varying in the range of 0.00 up to 0.03€-cents.

<sup>185</sup>This approach is not applicable for GARCH models as, by definition, the variance is a function of the squared error term of the observation process. Noh et al. (1994) use a GARCH model to estimate the future mean volatility which they plug in the Black & Scholes formula. Because this approach features a mispricing, Bauwens & Lubrano (2002) propose a simulation of the  $\mathbb{Q}$ -measure option payoff  $C(s_{\mathcal{T}})$ . Unfortunately, there is no multi-period risk-neutral process for GARCH models. Alternatively, Bauwens & Lubrano (2002) generate a local (one period) risk-neutral process according to Duan (1995) and sample future log-returns period by period, i.e.  $f_{Y_{\mathcal{T}}}(y_{\mathcal{T}}|\cdot, y_t) = \int f_{Y_{\mathcal{T}}}(y_{\mathcal{T}}|\cdot, y_{\mathcal{T}-1}) \cdot f_{Y_{\mathcal{T}}}(y_{\mathcal{T}-1}|\cdot, y_{\mathcal{T}-2}) \cdot \dots \cdot f_{Y_{\mathcal{T}}}(y_{t+1}|\cdot, y_t) d(y_{\mathcal{T}-1} \dots y_{t+1})'$ .

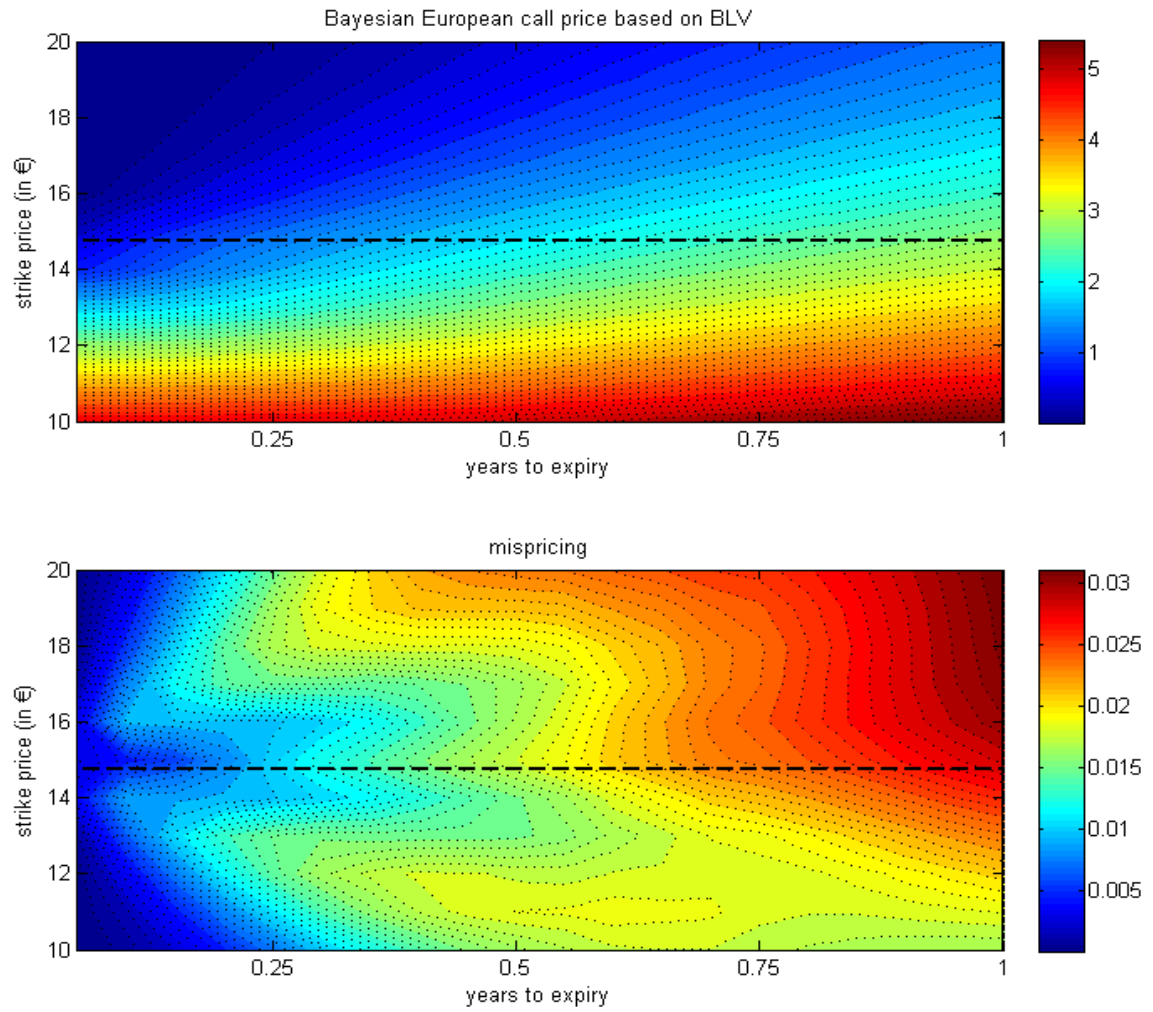


Figure 71: Bayesian option pricing - BLV

call prices	$\Delta = 1/4$	$\Delta = 1/2$	$\Delta = 1$
$\zeta = 14\text{€}$	1.5903€ (0.45447€)	2.2694€ (0.48647€)	3.1517€ (0.51102€)
$\zeta = 14.77\text{€}$	1.3199€ (0.40581€)	1.9404€ (0.47671€)	2.8197€ (0.52719€)
$\zeta = 15\text{€}$	1.2302€ (0.40302€)	1.8467€ (0.47597€)	2.7263€ (0.5311€)

Table 34: BLV option prices

In the lower subplot of figure 71, we visualize the mispricing ( $\hat{p}_t^{Bayes} - \hat{p}_t^{freq}$ ) which is only low for short maturities and grows with increasing maturity. We have calculated a maximum mispricing of  $\approx 3\text{€}$ -cents for in-the-money options with large maturities.

In this section we have exemplarily compared the Bayesian option pricing based on three stochastic processes (GBM, GBMJ, and BLV) we had calibrated by means of MCMC simulations (see section 10). We could show that the option prices based on the workhorse in finance - the GBM - are nearly not affected by parameter uncertainty - although there is a systematic underestimation of the call prices by the frequentist option pricing. Merely the estimator of the volatility is relevant for the pricing under the GBM. However, this parameter features an extremely informative (low standard deviation) and symmetrical posterior resulting in a low variation of nearly symmetrical conditional option prices. For the GBMJ we could identify an overpricing of the frequentist approach with a maximum of  $1.5\text{€}$ -cents for at-the-money options with a maturity of one year. Similar to the GBM, the BLV model features a systematic underpricing of frequentist pricing peaking to  $3\text{€}$ -cents for in-the-money calls with a maturity of one year. Albeit we had to simulate the option prices under the BLV model, our prices are extremely stable with variations (up to  $0.03\text{€}$ -cents) far below the mispricing.

## 13 Softcore-Bayesianism

So far, we have introduced Bayesian statistics as the most important area of HCB. Bayesian statistics transparently reveals parameter and model uncertainty in contrast to frequentist statistics. Hence, a faithful risk management should switch to the efficient Bayesian estimation techniques that are not that complicated but different from frequentist ones. Nevertheless, Bayesian (exactly like frequentist) statistics substantially rests on reliable data for the inference on risk factors.<sup>186</sup> Unfortunately, the quality of data is often lousy in real life.

Alternatively, *Softcore-Bayesianism* (SCB) offers a framework to transparently reveal human expertise in form of continuous or discrete distributions (incl. scenarios) instead of point estimates glossing over personal uncertainty.

*Bayesian Risk Management* (BRM) summarizes elaborated and practicable guidelines, techniques, and tools for a well-thought-out quantification and handling of subjective assessments. In the following, we introduce the main agents and groups involved in BRM (see section 14). Later, we present the *integrated Bayesian Risk Analysis* (iBRA) concept which separates the identification, quantification, and management of critical variables into some well defined modules (see section 15).

Often, expertise of specialists is needed as problems are highly complex. This can substantially improve the quality of the risk management but requires a careful elicitation of in house and external professionals - easily to incorporate into our iBRA concept. We overview the basic ideas of expert elicitation (see section 16) and explain some elicitation techniques (see sections 16.3, 16.5, and 16.6) as well as elicitation tools (see section 17). Because of the exceptional position of prediction markets as elicitation tools we separately dwell on their special features (see section 18).

## 14 Agents in BRM

Generally, up to three groups of people are involved in BRM: risk analysts, decision makers, and experts (see figure 72). The *risk analyst* is the person or group that tries to identify and examine critical menaces and opportunities and has to report to the decision maker in form of studies, models, and personal meetings. Risk analysts are managed by and subject to the supervision of the decision maker. The *decision maker* is a person or group that has the decision-making authority about how a department, cooperation, or an institution has to react to critical menaces and opportunities. We simply define *experts* as all those persons the risk analyst is willing to query hoping to improve her risk analysis. Hence, a person becomes

---

<sup>186</sup> Additionally, mathematical Bayesian updating requires adequate model assumptions and prior distributions (see section 2.1).

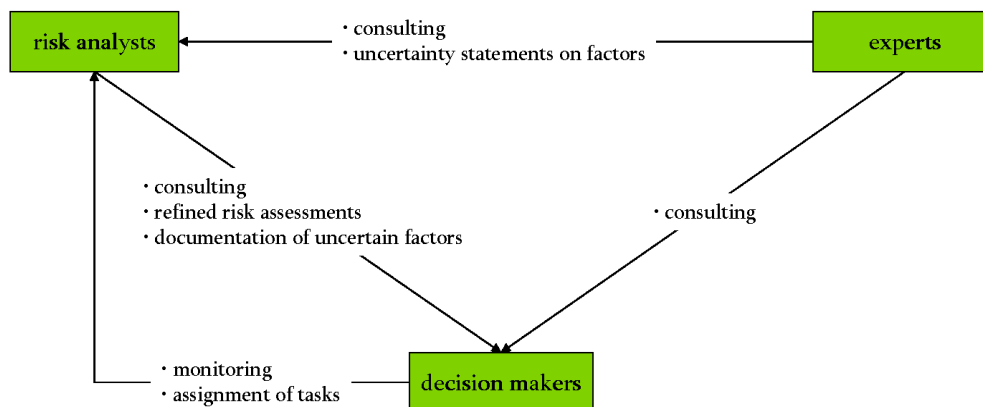


Figure 72: groups involved in Bayesian Risk Management

an expert by selection which can be direct via pre-selection of certain persons (risk analyst, decision maker, or other experts) or indirect via a self-selection of persons that fulfill certain criteria.

Although BRM advocates expert consulting, an integration of external experts from outside the department, corporation, or institution is not always reasonable or necessary. In house experts can often contribute lots of expertise. The expertise of the risk analysts and decision makers should also not be neglected. In fact, the expert, the risk analyst, and decision maker is often the same person.

## 15 Integrated Bayesian Risk Analysis (iBRA)

In the following we introduce a practicable framework allowing an organization to set up an integrated risk management for some critical *target variables*<sup>187</sup> when there is a lack of reliable or informative data. We have sketched the *integrated Bayesian Risk Analysis* (iBRA) concept that summarizes guidelines, technique, and tools for the *subjective* identification, quantification, and evaluation of critical factors affecting the target variable.<sup>188</sup> For practicability, iBRA highlights the risk analyst as invaluable expert. Nevertheless, an integration of in house and external experts is easily realized for strategic factors.

The Bayesian roots of the iBRA propagate a permanent updating of the results and a constant learning to improve the risk management. Consequently, iBRA is a cycle separated

<sup>187</sup>e.g. value-at-risk of a credit/ asset portfolio, output of a company, or cash flow

<sup>188</sup>Although the iBRA concept concentrates on subjective risk management, the identification step seems highly relevant in the model building of pure data-based risk management like credit risk.

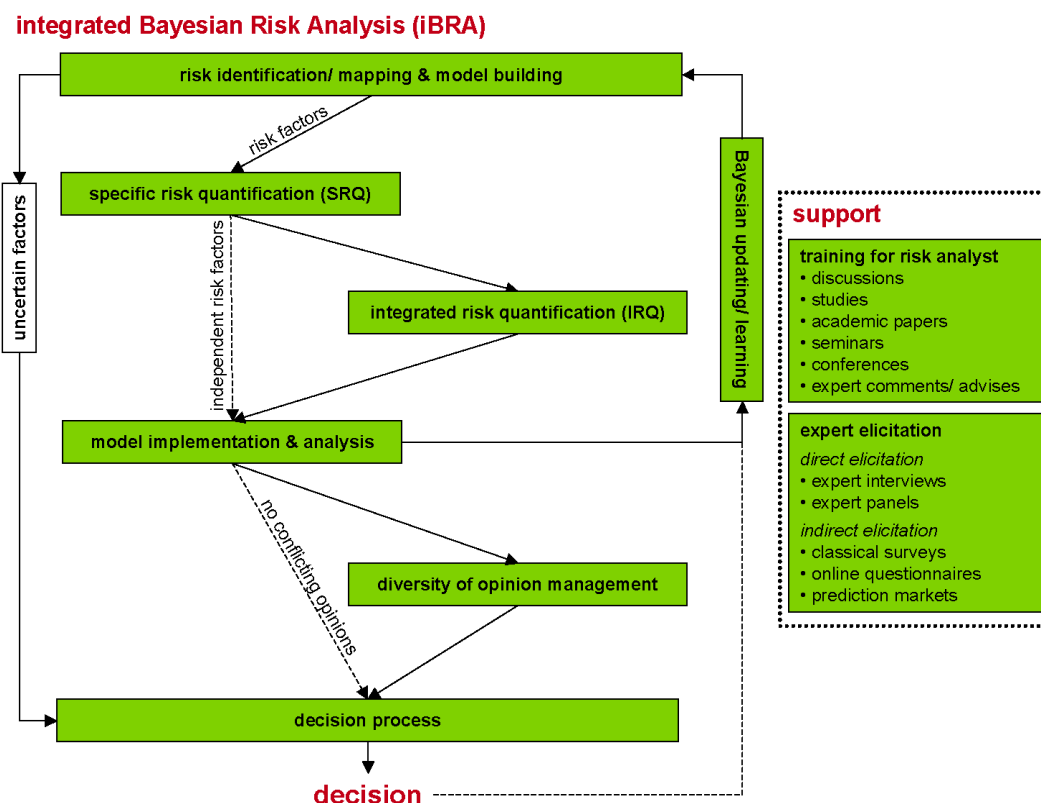


Figure 73: integrated Bayesian Risk Analysis cycle

into several sub-steps (see figure 73), we introduce in the next sections.

## 15.1 Identification, Risk Mapping, & Model Building

The task of a risk analyst is to quantify a (some) target variable(s), e.g. a company's cash flow. Often a direct quantification is not possible because it is influenced by numerous factors that are often highly interrelated. Models can normally help the analyst to grasp the scope of the risk management issue.

In the following, we concentrate on the risk analysis of discretionary non day-to-day business problems. They normally require a great deal of analytical skills and a sufficient level of practicable experience. A reasonable starting point for the identification of critical factors is a brainstorming of potential causes for problems and their negative as well as positive consequences. The task can be simplified by resolving the task into sub-problems. In the early stage, a simple list of all potential critical factors (= dragon kings) is enough.

On principle, there are two basic groups of factors:



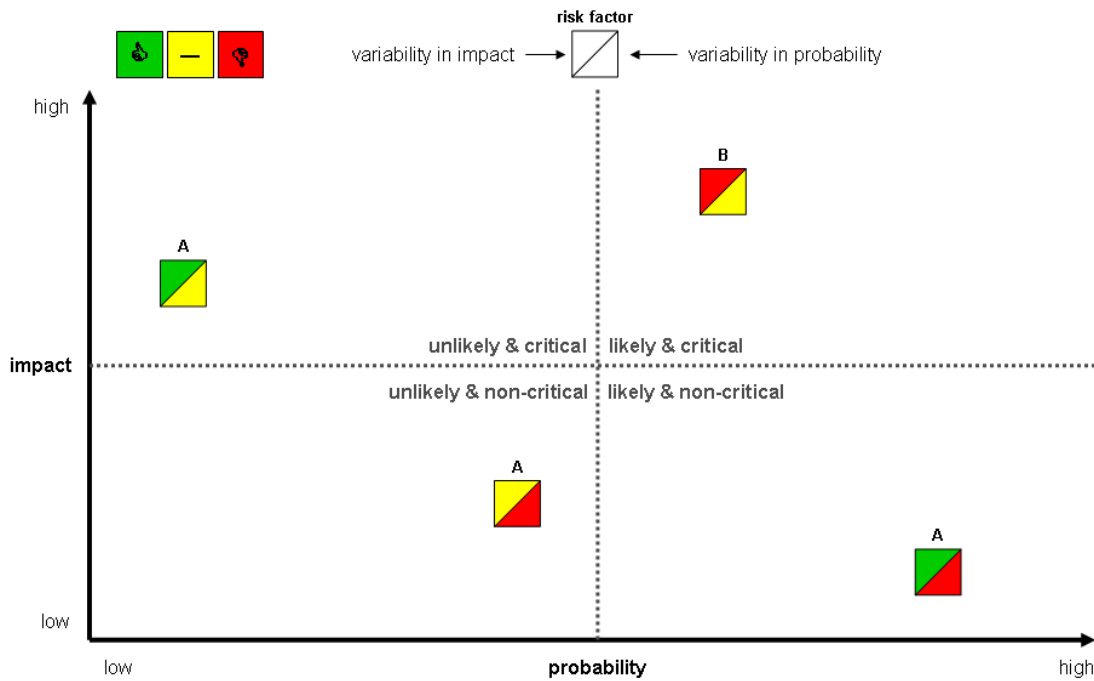


Figure 74: probability impact map

- *Occurrence variables* (e.g. regulation, natural disaster, or electrical power outage) are predominantly uncertain whether they will occur or not. Their impact on the target variable might be substantial but only if the variable has realized.
- *Level variables* (e.g. oil/ CO2 price or demand for goods) are mainly uncertain because of their actual level.

After potential factors have been identified, they should be ranked. A *probability-impact map* can support the ranking of occurrence variables (see figure 74). Most important factors are in quadrant "likely & critical" and less important ones in "unlikely & non-critical". A simple rule could decide according to the expected impact which is simply the map area left and below the factor. The analysis could be additionally improved by accounting for the uncertainty in the probability and impact assessments.

Level variables should be ordered according to their variability and their expected impact on the target variable. A *variability-impact table* could help to visualize the task (see figure 75). Variables with a high variability and a high impact on the target variable are more important than factors with a high variability but a low impact, or vice versa.

Probably, the number of identified factors is too large. The risk analyst can only quantify

	variable 1	variable 2	variable 3	variable 4	variable 5
expected value	50 €	12 €	35 €	3h	100t
variability level					
impact on target variable					

low	medium	high
-----	--------	------

Figure 75: variability-impact table

discrete or continuous distributions for the most important factors (= *risk factors*). Less critical or non-quantifiable factors (= *uncertain factors*) can be estimated by point estimates or need to be dumped and documented. Probability-impact maps and variability-impact tables could complement the analysis.

Because of the subjectivity of the assessments, a risk analyst needs further criteria to classify the identified risk factors (*Bayesian due diligence*): It is important to deliberate on the question whether the risk analyst (or the experts) has sufficient information to form useful discrete or continuous distributions for all critical variables. Besides, the risk analyst should always regard an external assessment (in house as well as external experts) with suspicion. It could be heavily affected by arrogance, pride (fear of loss of face), vengeance, or envy. Risk factor distributions might be too informative (concentrated) in respect of the available information.

Consequently, it might be better to transform some critical factors to constants. A sensitivity analysis allows to reveal the overall impact of the dumped uncertainty on the target variable. If impact is notable, a more extensive expert elicitation could be triggered. Otherwise, there is a good chance that a point estimate is not problematic.

After the determination of risk factors and uncertain factors, the risk analyst should set up a crude *risk map* (see figure 76) which consists of all risk factors and some important uncertain factors in form of point estimates (constants):

- The analysis of a problem can be reduced in complexity if independent groups of risk factors can be identified.
- Non-natural constants are factors we have decided to be not uncertain about. Generally, they correspond to the mean or most likely value of some uncertain factors.
- Decision variables (e.g. number of produced items) are variables which are not uncertain for the decision maker because she can decide on them.
- Intermediate variables are (deterministic) functions depending on decisions, constants, and risk variables.

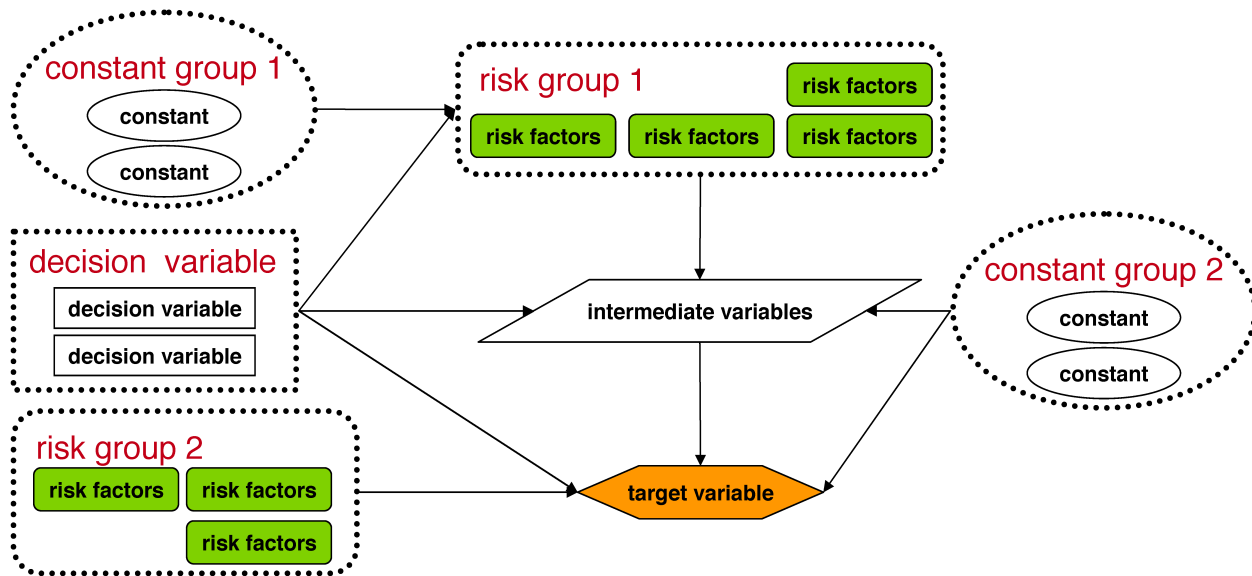


Figure 76: risk map

The designing of a crude risk map normally helps to understand the basic structure of the problem and to reduce the complexity of the model. Sometimes, such risk maps are that simple that they can be used for short discussions with experts and for presentations to the decision makers within the decision process.

Unfortunately, such crude risk maps are not sufficient to set up quantitative models that can calculate the target variable. An explicit description of all critical interrelations between all risk factors, constants, decision, intermediate, and target variables is required.

An *influence matrix* allows to identify the cause-and-effect chain of the risk factors (see figure 77 and Krahl & Wagner 2009). It simply rates the influence (not correlation) of the risk factor in the row on the factor in the respective column.<sup>189</sup> Krahl & Wagner (2009), pp. 130, define the levels 0... "no or very low influence", 1... "low or time-lagged influence", 2... "medium influence", and 3... "strong or very strong influence". The active (row) sums reflect the influence of the risk factor on other factors while the dependency of the risk factor on other factors is summarized by the passive (column) sum. Risk factors with a large *dynamic index* - product of active and passive sum - seem highly interrelated to other factors and should not be assumed as independent.

Once critical interrelations have been identified, the risk analyst can set up a more sophis-

<sup>189</sup>The influence matrix is to reveal the normally non-symmetrical cause-and-effect chain instead of the symmetrical correlation describing parallel or reverse behavior.

risk factors	A	B	C	D	E	active sum
A		2	3	1	0	6
B	1		2	3	3	9
C	0	2		1	0	3
D	2	1	0		1	4
E	3	1	3	2		9
passive sum	6	6	8	7	4	
dynamic index	36	54	24	28	36	

Figure 77: influence matrix

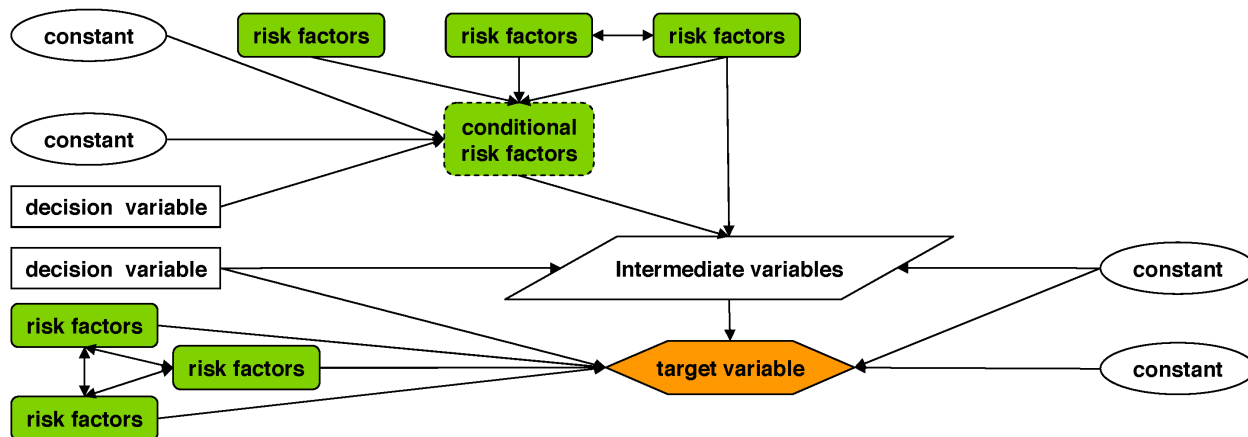


Figure 78: influence diagram

licated risk map, called *influence diagram* (see figure 78). In contrast to crude risk maps, influence diagrams can reach extreme complexities as they explicitly describe the model structure.<sup>190</sup> Moreover, the risk analyst has to distinguish between simple risk factors and conditional risk factors.

The uncertainty about independent risk factors is defined by a discrete or continuous probability distribution. Dependent factors need to be described by a multivariate (or meta-) distribution. To reduce complexity, it is advisable to find independent basic risk factors. E.g., it might be much more easy to model kerosine and gas price as functions of the crude oil price than to quantify a multivariate distribution for kerosine and gas price.

Conditional risk factors are described by statistical distributions depending on decision vari-

<sup>190</sup>The calculations of the WestLB/ Potsdam Institute for Climate Impact Research study "German power utilities - caught in the CO2 trap?" (Garz et al. 2009) are based on such a complex influence diagram implemented in the software Analytica.

ables, constants, intermediate variables, or other conditional risk factors. As long as their influence is only restricted to some parameters of the distribution, there is no big problem. If however they influence many parameters the so called *conditional probability tables* can easily explode in size. Hence, a well-thought-out structured influence diagram can substantially reduce complexity of model calibration.

In the next step, the risk analyst should implement the model according to the influence diagram. In this early stage, there is no requirement for an exact calibration. A preliminary implementation allows to identify logical errors in reasoning and counter-intuitive model results.

Although we have mainly stressed the risk analyst as expert, in house or external experts can often contribute a great deal of expertise. If the gain in insight justifies the additional efforts, the risk analyst should elicit them. Experts can help to identify critical factors and their interrelations. Expert comments on the probability-impact map and risk map as well as on the ranking of the variables seem extremely beneficial. Such a brainstorming can help to reduce imminence of black swans as the black swan of the risk analyst might be an uncertain factor or even a risk factor for an expert.

In the coming sections we introduce concepts how to calibrate a model by separately quantifying the risk factors (see section 15.2) and by determining their interrelations (see section 15.3).

## 15.2 Specific Risk Quantification (SRQ)

BRM is based on the belief that a quantification of uncertainty in form of risk factors is superior to point estimates.<sup>191</sup> In real life, an extensive analysis usually leads to a complex model that is hardly to handle when the uncertainty and all dependency structures of all factors are modelled in distributional form. We should therefore concentrate on the most critical factors. Less important factors (or factors we have no reliable information on) should be treated as constants or be dumped. Furthermore, we have propagated to identify independent basic risk factors (e.g. oil price instead of gas and kerosine price) or at least in-

---

<sup>191</sup>A prime example is the attempt of the Energy Model Forum (EMF) in 1980 to predict the oil price up to 2020 (see Ötsch 2008, pp. 31). The combination of several models and scenarios resulted in 120 projections. From a classical point of view there exist no clear guidelines to deal with more or less plausible projections which all could turn out to be appropriate. The EMF simply calibrated all models to a most-likely reference scenario and presented the results without any weighting. Such an approach entails a fundamental non-communicated uncertainty. Alternatively, Draper (1995) weighted the 120 projections according to the plausibility of the respective scenarios while the models were assumed to be equivalent. In contrast to the Energy Model Forum (EMF), Draper (1995) could clearly identify a considerable risk (although in the left tail of the oil price distribution) for an oil price equal or below the realized price (\$13) in 1986.

dependent groups of dependent risk factors (oil and CO2 price vs. regulation on automotive and aviation industry).

When the risk analyst could set up a model following the principles above, there are several (nearly) independent risk factors that allow a separate quantification. This step — we call *specific risk quantification* (SRQ) — enables the risk analyst to concentrate on the risk factors one at a time. This concept reduces the burdens for expert elicitation fundamentally as in house or external experts with some specialized expertise can be of use.<sup>192</sup>

Generally, we recommend two basic elicitation techniques for a separate quantification of risk factors:

- *fixed interval method* (see section 16.3.5): The risk analyst sub-divides all possible outcomes of a risk factor into a fixed number of intervals. Subsequently, she (or the expert) allots weights to each of them. Interval probabilities can be inferred from those weights. This allows to set up a discrete probability distribution or to calibrate parametric distributions.
- *variable interval method* (see section 16.3.6): The risk analyst directly defines parameters of special elicitation distributions (e.g. triangular or modified PERT distribution).

### 15.3 Integrated Risk Quantification (IRQ)

Unfortunately, a separate quantification of risk factors can cause trouble if some of the risk factors feature strong correlations. E.g., the CO2 price and the coal (lignite) price seem not to be independent. A traditional approach is to run some empirical estimations of Pearson or other (rank) correlation coefficients which depend on historical data. A risk analyst might (realistically) expect a fundamental change in the future dependence structure because of coming regulations to fight climate change. Empirical data analysis can generally not deal with such problems. Subjective assessments are needed.

The *integrated risk quantification* (IRQ) module offers techniques to cope with such situations. Actually, IRQ deals with two different kinds of dependence structures:<sup>193</sup>

- *correlations between risk factors*: To some extent, subjective reasoning propagates some combination of outcomes of several risk factors while it rules out (to some extent) some other combinations.

---

<sup>192</sup>If the risk analyst wants to consult experts, special elicitation tools are needed that expose the experts to the elicitation techniques. We prefer to distinguish between expert panels (e.g. focus groups or Delphi studies), expert interviews, classical surveys (e.g. questionnaires), and automated surveys (e.g. online questionnaires or prediction markets). See section 17 for more details.

<sup>193</sup>The most tricky situation arises when there is correlation between auto-correlated risk factors.

- *auto-correlation of risk factors*: According to subjective reasoning, the outcome of a time-series risk factor is somehow related to its outcomes in previous periods.

### 15.3.1 Correlated Risk Factors

First and foremost, in the case of correlated risk factors, the BRM advises all risk analysts to leave no stone unturned to avoid the IRQ module - without wrongly assuming independence. A nearly all purpose tool is to fathom the cause-and-effect chain. Often, the risk factors depend on the same independent basic factors which cause the correlations. Consequently, we can evade correlated risk factors. If there is no way out, there are several methods to deal with dependencies.

A very simple but often ignored approach to eliminate correlation is the *transformation of risk factors*. Garz et al. (2009) set up a multi risk factor Monte Carlo simulation to appraise the uncertain cash flows of German electric utility portfolios consisting of different energy sources (coal, lignite, natural gas, renewable). The results critically depend on the electricity wholesale prices and the costs of generating electricity. A blue-eyed modelling would quantify all commodity prices resulting in a multi-dimensional distribution or in a highly complex conditional probability table. Usually, both approaches are too complex to deal with. A diligent analysis of the problem finds relative - not absolute - commodity prices important for cash flow (see Garz et al. 2009). The calibration can be simplified by defining a distribution for the price of a numeraire (e.g. coal) and subsequently the relative prices in respect of that numeraire. In the special case of two dependent risk factors, the uncertainty on the (price) ratio could directly be quantified.

Sometimes, it is not possible to eliminate the correlation between different risk factors but there are some techniques that allow to approximate them. Primarily, the risk analyst should structure the crude risk map to reach as many independent groups of risk factors as possible. The complexity of the model is by far more affected by the number of correlated risk factors within independent groups than by the number of such groups. If there are groups of two or three correlated factors the *visual copula* approach could be applied (see section 16.5.1 for more details). It simply separates the analysis into two steps. First, the specific uncertainty on each risk factor is quantified (see SRQ in section 15.2). Finally, the risk analyst or the experts need to choose or rank scatter plots produced by one or several copula functions for different levels of (rank) correlations.

An alternative is the *envelope method* allowing to model correlations for a few risk factors (see section 16.5.2 for more details). A risk factor is chosen as a numeraire and separately quantified. Subsequently, all other dependent risk factors are assumed to follow a simple elicitation distribution (e.g. triangular or modified PERT distribution). Finally, the risk analyst defines the parameters of the elicitation distribution given the numeraire. E.g.,

risk factor scenarios	A1	A2	B1	B2	B3	C1	C2	D1	D2
A1									
A2									
B1	2	4							
B2	2	5							
B3	4	3							
C1	2	4	5	4	2				
C2	4	3	2	3	5				
D1	3	3	4	4	2	4	1		
D2	3	3	2	2	4	2	4		

Figure 79: consistency matrix

given a coal price  $x$ , the most likely CO2 price is  $y$  and the CO2 price will never be below  $y_{\min}$  and above  $y_{\max}$ . As the specification is done separately for each risk factor except the numeraire, the envelope can be simply visualized by a two-dimensional plot. Then the risk analyst can simply draw lines for the most likely, minimum, and maximum value of the CO2 price given different oil prices.

The techniques we have presented are only feasible for a low number of interrelated risk factors. In more complex situations there is no alternative to stress tests or scenario analyses. *Stress tests* are used to describe the tail characteristics of the target variable (e.g. credit portfolio loss) - often the unfavorable tail, only. The critical task is to identify plausible and consistent stress scenarios inducing a relevant "worst case" (and "best case"). The adjectives "plausible", "consistent", and "relevant" simply mean that (1) the assumed outcomes of the risk factors in the respective scenario are based on sound judgement, (2) the simultaneous outcomes of different risk factors are not mutually exclusive,<sup>194</sup> and (3) the probabilities for even more extreme stress scenarios are negligible.

The *scenario analysis* is a generalization of the stress test approach. The designing of at least "worst case" and "best case" scenarios as well as some "medium" reference scenarios can illustrate the full range of outcomes possible for the target variable. Once the scenarios are enhanced by probabilities (*Bayesian scenario analysis*), this approach allows to approximate the full distribution of the target variable.

The identification of adequate reference scenarios is a challenging task. The scenarios should roughly cover all possible and relevant basic storylines reflecting the simultaneous uncertainty on all risk factors. Optimally, each reference scenario is a prime example of an important storyline. A *consistency matrix* can help to simplify the problem (see figure 79 and Krahl & Wagner 2009). First of all, it requires the risk analyst to rank the risk factors according to their importance as the designing of the reference scenarios is a cascade starting from the

<sup>194</sup>E.g., stable or rising tax revenues do not seem realistic in times of an economic depression in Germany.



most important risk factor (e.g. A is more important than B and B is more important than C and C is more important than D). In the next step, the risk analyst defines a number of scenario outcomes for each risk factor (e.g. A1 & A2 for A, B1 & B2 & B3 for B, ...). Then for every outcome of a specific risk factor the risk analyst (or an expert) rates the consistency with the scenario outcomes of the more important risk factors. Krahl & Wagner (2009), pp. 130, propose the following ratings: 1... "totally inconsistent", 2... "partially inconsistent", 3... "independent", 4... "mutually promotive", and 5... "highly mutually promotive".

The scenario generation starts with risk factor A where the risk analyst has to decide whether it has outcome A1 or A2. Assume the risk analyst chooses A1. Then she has to define a scenario outcome for risk factor B (B1, B2, or B3). Except for the construction of a "worst case" scenario, an outcome for B should be chosen which has the highest score in column A1 - here B3. If all scores are 1 or 2 a relevant outcome for B has been ignored and needs to be included. Subsequently, the risk analyst should find a suitable outcome for risk factor C (C1 or C2). Now, the risk analyst could follow different rules, e.g., take the outcome of C which maximizes the sums or products in column A1 and B3 (here C2). An exclusion of an outcome which has at least one rating of 1 seems reasonable. In this way, the scenario outcomes can be defined for all risk factors. As risk factor A is the most important factor, the next scenario should start with A2 and would imply B2, etc.

Without any weights or probabilities, the scenario analysis can only help to understand the range of uncertainty on the target variable. The scenarios could be misinterpreted as equally likely, which can be counterproductive in the case of "worst case" and "best case" scenarios. E.g., the authors of the IPCC Special Report on Emissions Scenarios (SRES) designed six families of emission projections to guarantee comparable results of global circulation models (see IPCC 2000). The scenarios were formulated to roughly cover all plausible and possible emission paths. Probability statements or at least weights were not offered. Consequently, the scenarios are not beneficial for a non-expert decision maker.

From the BRM perspective, probabilities or weights need to be attached to the scenarios. Subsequently, the risk analyst should plot the marginal risk factor distributions whether they agree with her assessment.<sup>195</sup> If not, the scenarios weighting might be inconsistent or the analysis lacks in scenarios describing some key storylines. This approach can even help to identify critical dragon kings.

---

<sup>195</sup> Assume, the target variable is defined by three risk factors A, B, and C. The risk analyst constructs the scenarios to fully describe A and B while C is only adjusted to be in line with these scenarios. Then highly relevant or likely outcomes of C could have been ignored.

### 15.3.2 Auto-Correlated Risk Factors

A special case of dependency structures is the auto-correlation of time-series risk factors which results in a two-dimensional problem. On the one hand, the risk analyst wants to quantify the *marginal distributions* of the risk factor for different points in time. On the other hand, the realization of the risk factor in one period should be interrelated to previous realization. E.g., if logical reasoning propagates that a high (low) CO<sub>2</sub> price in 2015 tends to be followed by a high (low) price in 2016, the CO<sub>2</sub> price is assumed to be positively auto-correlated.

Within the IRQ module we propose three basic approaches: stochastic processes, weighted projections, and time-series copula. A rather simple approach is to calibrate a mean-reverting *stochastic process* (e.g. Ornstein-Uhlenbeck, Cox-Ingersoll-Ross, or constant elasticity of variance process). Mean-reverting processes allow to quantify a long-term equilibrium for the risk factor. Other parameters are normally a mean-reversion rate and a variance level/factor which are not that intuitive. Alternatively, a stochastic process can be calibrated to statements on means and quantiles. Unfortunately, the modeling of assessments via stochastic processes is rather inflexible. Mean-reverting processes are generally defined by three up to four parameters. This number is mostly too low for an adequate approximation of the auto-correlation and the time-dependent risk factor distribution. Either, the number of (consistent) statements is reduced to the number of parameters or a (quadratic) loss function is needed to find a parameterization.

An alternative is the definition of *weighted projections* for the risk factor. Similar to the scenario analysis, the projections should reflect all relevant storylines. The pro of this approach is its nearly unlimited flexibility. Similar to the scenario analysis, this approach suffers from the risk that the weighted reference scenarios can result in marginal risk factor distributions contradicting the risk analyst's assessments.

Another method is the *time-series copula* which is a combination of the envelope method and a copula for all points in time (see section 16.5.4 for more details). In contrast to the standard envelope method the numeraire is no other risk factor but simply the time. In a first step, the risk analyst calibrates elicitation distributions to the risk factor for all periods. Then, by means of a Gaussian copula, the realization of the risk factors is simultaneously simulated for all periods. The time-series copula method is extremely flexible in modeling the marginal distributions but restricted in respect of the auto-correlation. We apply a correlation matrix equivalent to auto-correlations with (symmetrical) exponential decay in the time-difference. Such an auto-correlation structure seems adequate for most time-series risk factors.

## 15.4 Model Implementation and Analysis

In the modules before, the risk factors and their interrelations have been identified and quantified while uncertain factors have been transformed to point estimates or have been dumped. The essential task of this module is (1) the analysis of the quantified risk factors (marginal distributions) and the uncertain factors (point estimates), (2) the calibration of the model, and (3) the documentation of the model results. These steps should always be accompanied by a permanent evaluation whether a model updating or a new quantification is necessary.

In a first step, all risk factor distributions should be critically analyzed according to their information content, plausibility, and their credibility.<sup>196</sup> Theoretically, a very flat distribution reflects the high uncertainty of the assessor. The major difficulty for the risk analyst is to decide whether such a distribution results from a competent, information based assessment or from a pure ignorance of the topic.<sup>197</sup> In the first case, this is an alarm signal. The risk factor and its interrelations could be too complex for people to offer more precise assessments. In the latter case, the risk analyst should wonder whether she has addressed the adequate person to quantify the risk factor.

Another, maybe even harder, problem is to evaluate the information level of concentrated distributions. There is the danger of some overoptimistic assessments. The risk might be underestimated when the assessor has no expertise on the risk factor or no access to (new) relevant information.

If the statements are rated uninformative, the risk analyst should think about the elicitation of (more) competent experts. If the risk analyst believes that there is no such (affordable) expertise allowing to quantify the risk factors, she should try to modify the model by replacing the intractable risk factor by another much easier to handle one.

After the evaluation of the risk factor distributions, the model can be calibrated to these results. The risk factor distributions substantially increase the complexity of the model. Mostly, the distribution of the target variable can only be calculated by means of Monte Carlo techniques. Consequently, the distribution is the result of a large number of simulations given the generated realizations of the risk factors and the point estimates of the uncertain factors. The target variable distribution covers all uncertainty in the risk factors. Some statistics (e.g. mean or standard deviation) can help to summarize basic characteristics for a better evaluation.

Unfortunately, there is always some unconsidered uncertainty in point estimates of uncertain

---

<sup>196</sup>Of course, the point estimates of the uncertain factors need also to be checked. In the following, we will concentrate on risk factors for a clearer presentation.

<sup>197</sup>This seems challenging when the risk analyst herself stated the assessments. Then she needs to judge on her own previous work.

factors. Sensitivity analyses can help to identify the most critical factors and their impact on the target variable. Unexpected results should be explainable otherwise a more in-depth analysis of the factor and its interrelations is needed.

For the decision process, the target distribution with its statistics should be enhanced by a documentation of the sensitivity results for the most important factors. To roughly describe the range of ignored uncertainty, a "worst case" and a "best case" target variable distribution should be simulated. This requires the risk analyst to find some reasonable and consistent "worst case" and "best case" storylines for *all* relevant uncertain factors. A consultation of in house and external experts could improve the quality of the estimates. If the target variable distribution is only slightly sensitive to extreme "worst case" and "best case" scenarios of the constants, the treatment of uncertain factors as constants seems reasonable.

## 15.5 Diversity of Opinion Analysis

By now, we have described the iBRA concept under the assumption that the risk analyst sets up exactly one model and calibrates it by risk factor distributions and point estimates of the uncertain factors. Although the iBRA concept advises to permanently reconsider the model (informal Bayesian updating), we have ignored a potential *indecisiveness* and *diversity of opinions*. The risk analyst might often favor several contradicting or competing assessments on the critical factors and their interrelations. The consulting of experts increases the intricacy of the risk analysis even more.

However, competing assessments sometimes are no bug but a feature. They reveal the uncertainty (a person with competing assessments) or disagreement (several persons with competing assessments) on the risk factors. Makridakis (1986), pp. 16 & table 1, compared several studies on different information sources producing (point) forecasts. He could not identify a generally superior information source. Hence, it would be grossly negligent solely to rely on one source (Armstrong 2001). According to a rule of thumb, combining information sources (in contrast to picking one of them) improves forecast accuracy the more the assessments of the sources are negatively correlated.

In SCB we distinguish between three basic paradigms for the handling of competing assessments (see figure 80 for expert aggregation): behavioral aggregation, no aggregation, and mathematical aggregation. The aggregation of disagreeing expert assessments is supported by several approaches following the paradigm of *behavioral aggregation*. Generally, behavioral tools require a certain level of interaction between experts. Ideally, experts reach a consensus on their assessments. The most dominant tools are deliberation panels and Delphi studies. A deliberation panel (see section 17.1.2) is a formally unstructured discussion group of several experts which is more or less managed by a coordinator. The task for the

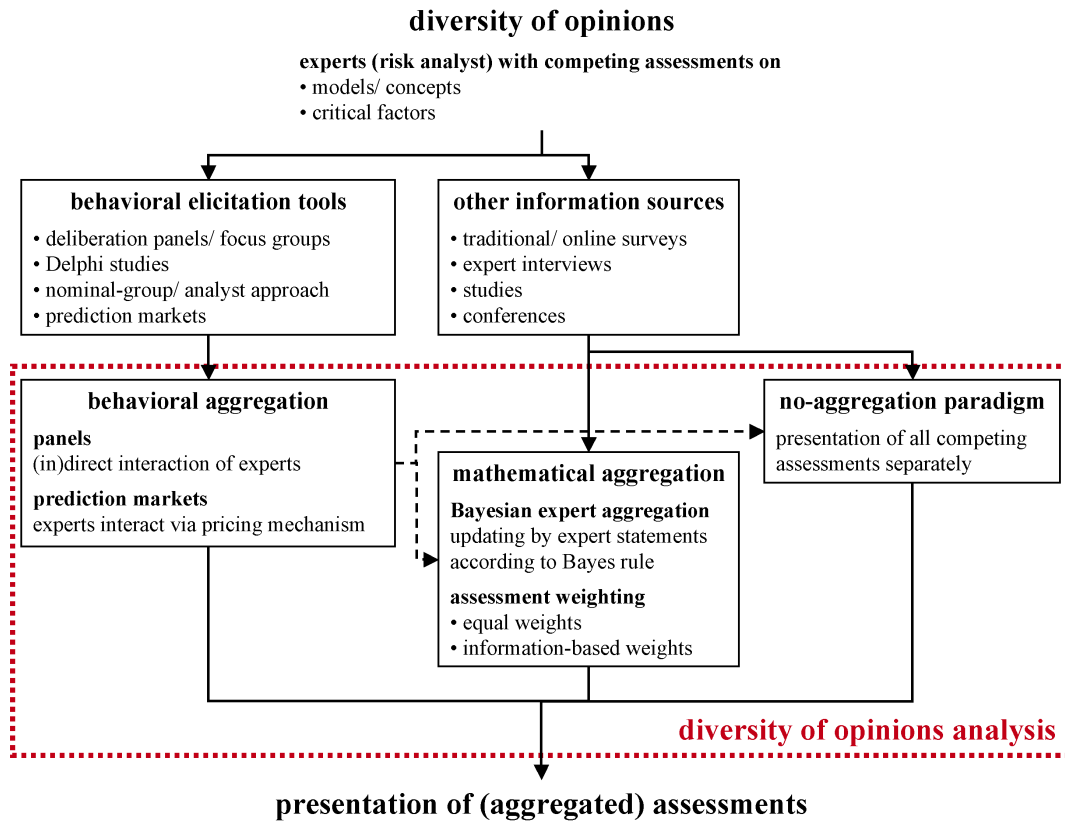


Figure 80: diversity of opinions

experts is to find a consensus.

Unfortunately, such expert panels do not always reach a consensus and there is a real risk for some undesired group behavior like group pressure. Hence, the success of a deliberation panel critically depends on the coordinator. To avoid such flaws, Delphi studies (see section 17.1.2) do not allow for a direct contact of the experts. Instead, experts are separately asked for their assessments, subsequently presented to the other experts. In an iterative (laborious) process the Delphi method is hoped to result in converging expert assessments.

Nominal-group-techniques like the SANDIA-NRC protocol (see section 17.1.2) try to profit from the direct interaction of experts but want to avoid undesired effects. Hence, initial deliberation panels are followed by separate interviews where the experts can state their individual assessments.

The behavioral elicitation tools, just sketched above, mostly entail a direct contact between the experts and a coordinator (mostly the risk analyst). This mostly entails an expensive and time-consuming coordination which is not economical in daily business. Alternatively,

there is the concept of prediction markets (see section 18) that profit from the automatic aggregation via the pricing mechanism - justified by Hayek's theorem (Hayek 1945). Experts can trade on the realization of risk factors as well as on competing models and concepts. The market prices can be interpreted as a market consensus on the topic. Ideally, the market prices contain all information any of the trader has gathered. Prediction markets are well-understood for uncertainties that will resolve in the near future. Unfortunately, many important uncertainties will stay unclear for a long time. Therefore, we propose the concept of the prediction market PCXtrade which runs short term markets on long-term events (see section 18.9.2).

When direct behavioral elicitation tools could not reach consensus or surveys have been conducted, the risk analyst is confronted with competing assessments. Consequently, she can simultaneously present them according to the no-aggregation paradigm or apply some form of mathematical aggregation

Supporters of the *no aggregation* paradigm (e.g. Morgan & Keith 2008) fear that an aggregation of different propagated models or risk factor distributions could result in an inadmissible loss of information. Alternatively, they favor tools (e.g. simple pdf plots or box plots<sup>198</sup>) that reveal all conflicting assessments to the decision maker who informally aggregates by her expertise (see figure 81).

There are two different forms of *mathematical aggregation*: assessment weighting and Bayesian expert aggregation. *Assessments weighting* summarizes the weighting of competing assessments stemming from one person<sup>199</sup> (risk analyst) as well as different experts<sup>200</sup> or (e.g. econometric, physical) models. The most dominant weighting tool is the linear opinion pool which is a weighted arithmetic average of the competing assessments (see section 16.6.1). The weights can either be uninformative or informative. In the first case, assessments on a risk factor are equally weighted. In the latter case, weights should base on some reasoning of the risk analyst, the decision maker, or the assessor herself.

Fischer & Harvey (1999) showed in a laboratory experiment that a subjective weighting by persons without any idea about the accuracy of different information sources performed worse than equal weighing but better when these persons had such information. Armstrong (2001)

---

<sup>198</sup>A box plot is an easy tool to visualize the features of some observations without assuming any distribution. The box reflects the range of the middle 50% of all observations while the horizontal line inside the box is the median separating the lower from the upper 50% of all observations. A non-symmetrical box means skewness in the data. The whiskers have a maximum of 1.5 times the length of the box. Observations below and above the respective whiskers are marked as outliers.

<sup>199</sup>People are often undetermined about the "correct" model. The iBRA cycle can easily be applied to this situation. The concept of imprecise probabilities (e.g. see Levi 2000) offers the theoretical background. Nevertheless, we do not believe that it is reasonable and economically justifiable to allow one person (either risk analyst or expert) to state several risk factor distributions.

<sup>200</sup>See section 16.6.1 where we describe assessments weighting for conflicting expert assessments.

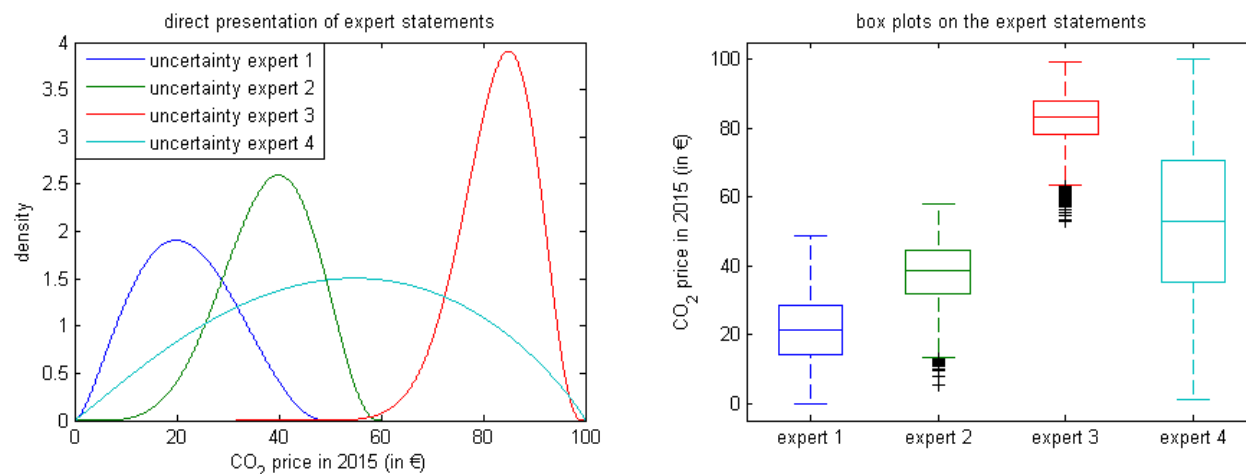


Figure 81: no aggregation paradigm - Continuous and discrete distributions as well as box plots can illustrate the full range of disagreement.

presents an overview of 30 empirical studies. By just equally weighting different information sources they achieve, on average, a 12.5% reduction (ranging from 2% up to 24%) of the forecasting errors. Under some ideal conditions, aggregated assessments were more accurate than the best single information source. As the error of the aggregated assessment is never greater than the worst single assessment, this approach is also very interesting in situations where wrong predictions result in intolerable losses.

Generally, the determination of informative weights is situational. The risk analyst can be confronted with (1) competing assessments of non-human information sources like statistical or econometric models, (2) competing ideas or models, or (3) competing pure subjective assessments of different experts.

The future performance of models could be weighted by posterior model probabilities available in Bayesian statistics (see sections 8 and 10.3). However, this approach presumes that the data - used to calculate the probabilities - is informative for future developments. Often, there is no such reliable data. Then, there is no way out to define subjective weights in form of peer weights, self weights, or group weights.

There are different approaches to find adequate weights for expert (and model) assessments: the past calibration, the analysis of key criteria, or again the use of pure subjective weights. The weighting of experts (or models) according to their *past calibration* simply reflects their past performance in forecasting the outcome of the risk factor.<sup>201</sup> Unfortunately, the calibra-

<sup>201</sup>Based on many years of experiences with psephology the German election research institute Allensbach weights the interviewees in polls according to their intentions to go to the election. A statement "sure" is weighted with 1.0, "likely" with 0.7, and "surely not" with 0 (Berlemann & Schmidt 2001, p. 16). Here, the

tion approach requires the assumption of a repeatable risk factor which is a critical assumption. Alternatively, calibration could be measured according to the forecasting performance on comparable risk factors. Is the past performance of a weather forecaster comparable to her future performance? This seems quite reasonable. However, is the past performance of an investment banker a good indication for her future profits?

In the case of unparalleled risk factors or in situations where the past forecasting performance is unserviceable to future success, alternatives are needed. Clemen & Winkler (1986), p. 45, prefer to assume exchangeable experts. Then with experts from the same "league", experts should equally weighted.

We do not favor such an equal weighting. It just tries to simplify the problem and ignores important additional information. The iBRA concept propagates a criteria based weighting of experts (*Bayesian due diligence*) following the request "weigh evidence, not experts!" (Kaplan 1992, p. 61). We do not keep faith with a belief of objectivity and call for *inter-subjective* rather than "objective" criteria for the evaluation of experts. Ideally, there exist such inter-subjective criteria - any decision maker or risk analyst would come to the same appraisal. In reality, we can only hope to find some key criteria - any rational person would judge in the same direction. Such criteria could be the practical and theoretical experience of the expert,<sup>202</sup> no-stakes condition,<sup>203</sup> interdisciplinary skills,<sup>204</sup> anti-mainstream opinion,<sup>205</sup> or deliberate opinion.<sup>206</sup> Within the iBRA concept, we favor a simple *reputation map* for the experts which is based on a traffic lights categorization (see figure 82). The risk analyst (or decision maker) simply rates different experts or expert groups by red (=bad), yellow (=medium), or green (=good). Subsequently, she decides on the (mis)information content of the expert statements. In fact, we believe that it is not reasonable and too cumbersome to find a weighting scheme for the key criteria. Instead, the risk analyst should subjectively weight the experts in a manner that she can justify by the criteria. A rigorous documentation reduces the danger of arbitrary coloring by the risk analyst.

Regrettably, there are some delicate intricacies linked to subjective expert weighting. As long as the expert assessments originate from studies or conferences, there are no interpersonal problems. The weighting of experts directly invited by the risk analyst seems more

---

performance in forecasting is the person's activity to go to the polls.

<sup>202</sup>We assume consensus on the fact that people prefer experts with more theoretical (or practical) experience as long as there is no other difference.

<sup>203</sup>Rational, utility-maximizing people with stakes in competing projects are prone to false testimonies and disinformation (see Kadane & Winkler 1988 and section 16.3.1).

<sup>204</sup>A higher level of interdisciplinary skills seems correlated with more problem solving competence.

<sup>205</sup>An expert having a competing assessment compared to the mainstream does not need to be right. However, the consideration of a broad scope of opinions seems reasonable.

<sup>206</sup>An expert who has formed an own view on the problem should be preferred compared to persons that have not.



risk factor: A	expert 1	expert 2	expert 3	expert group I	expert group II
practical experience					
theoretical experience					
no-stakes condition					
interdisciplinary skill					
anti-mainstream opinion					
deliberate opinion					
information content	medium	low/ critical	medium	risk of disinformation	high
weight	medium	0	medium	0	high

Figure 82: reputation map

problematic. A disproportionately low weighted expert could feel to suffer a snub. Therefore, we favor *self-assessments* of experts on the quality of their forecasts. Within the iBRA concept, we propose the wording "no idea", "extremely uncertain", "uncertain", "moderately uncertain", "confident", and "very confident". All experts who state "no idea" are ignored. This avoids to account for experts pushed to their statement. The crux of the matter is an adequate transformation of the wording into numerical weights. The risk analyst (or decision maker) could define a personal weighting scheme. Alternatively, we have carried out a survey on the common understanding of these statements. We come to the following weighting scheme (see table 35 and figure 83):<sup>207</sup>

no idea	extremely uncertain	uncertain	moderately uncertain	confident	very confident
0%	6.0%	13.2%	20.3%	28.9%	31.6%

Table 35: expert weighting scheme

An alternative to the weighting approach for competing expert assessments is the concept of *Bayesian expert aggregation* (see section 16.6.2). Its basic idea is the *rational updating* of the decision maker's prior uncertainty on a risk factor by expert statements which are used like traditional data. The resulting posterior is a consensus any rational decision maker would agree on given the prior and the expert statements. To reach an inter-subjective posterior, the prior is often modelled uninformative and the decision maker is treated as an additional

<sup>207</sup>We used our online questionnaire PCXquest for a short survey on people's perception of the competence of experts who self-rate their assessments by "extremely uncertain", "uncertain", "moderately uncertain", "confident", and "very confident". A total of 57 persons could rate the self-assessments with 0... "no trust at all" up to 10... "full trust". We transformed these absolute weights (upper subplot in figure 83) to relative weights summing up to one (lower subplot). We used the resulting medians for our weighting scheme.

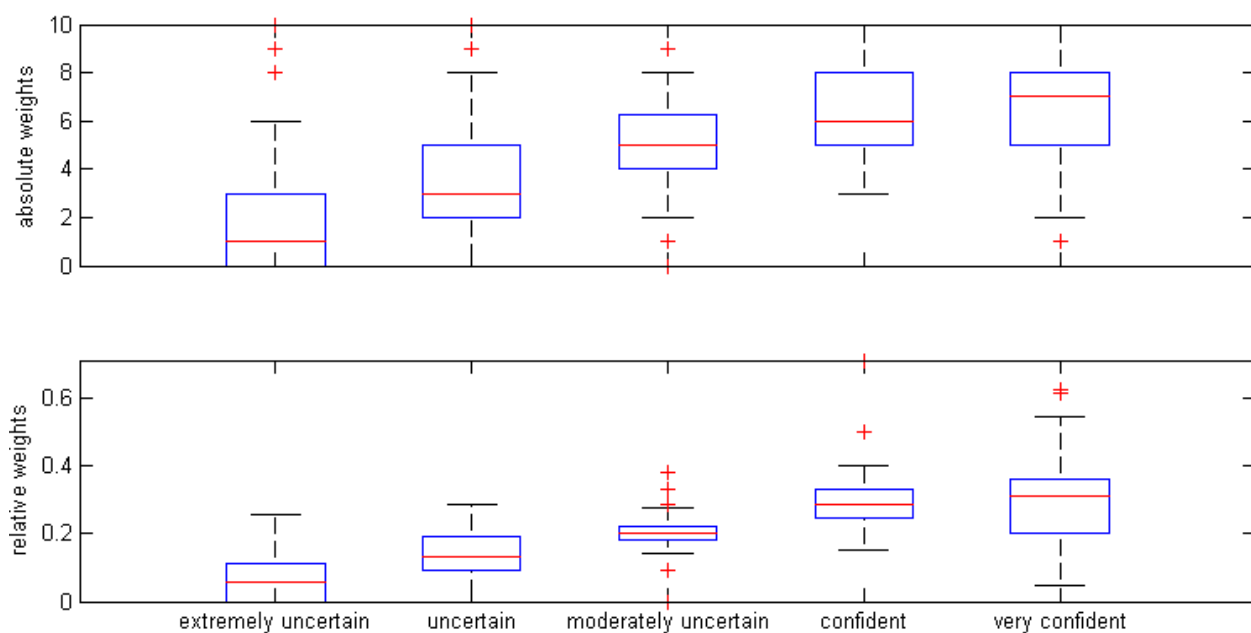


Figure 83: confidence in expert's self-rating - The box plots (whiskers set to a maximum of 1.0 times the interquartile range) show: The more confident the expert in her assessment the more confident people in the expert assessment. Nevertheless, 1/3 (1/5) of all respondents expressed a lower level of faith for assessments rated by "very confident" than by "confident" ("moderately uncertain").

expert. Then the updating does only depend on the form of the joint distribution of the expert statements. Generally, Bayesian expert aggregation accounts for correlated expert statements. Such correlations can fundamentally affect the quality of expert statements. Highly correlated statements provide some indication of a single information source most experts refer to. Metaphorically speaking, the Bayesian approach aggregates information sources instead experts. An against mainstream expert is given more weight than mainstream experts. This Bayesian idea reduces the danger of ignoring Cassandras warning against a bubble.

There is no clear answer to the question which mathematical aggregation scheme (assessment weighting or Bayesian expert aggregation) should be preferred when there are competing expert assessments. Both approaches use fundamentally different concepts. The weighted uncertainty on a risk factor is an average of competing opinions. The aggregated Bayesian density (= posterior) is the result of a rational (inter-subjective) process of learning (mathematical Bayesian updating). Although the Bayesian approach is theoretically more appealing, it is practically hard to implement.

## 15.6 Decision Management

Once the risk analyst has identified, quantified, and aggregated all uncertainty and all relevant interrelations, the results need to be evaluated in respect of their implications on management decisions in the cooperation. As this dissertation concentrates on the risk analysis instead of optimal decision-making, we make this section short.<sup>208</sup>

A crucial point in the iBRA design is the exact definition of the decision maker who can be the risk analyst herself or another person or group within the cooperation. The decision maker needs to be briefed on the risk factors and their interrelations. Moreover, a concentrated documentation should inform the decision maker on uncertain factors and interrelation which could be crucial but have not been quantified. If the results critically base on expert assessments, the presentation of the respective reputation maps seems advisable. We avoid to present exact algorithms offering the "optimal" decision. We are rather in favor of a decision maker acting on her gut feeling. In his remarkable book "Bauchentscheidungen" (English "gut decisions"), Gigerenzer (2008) presents an unbelievable number of examples where gut feeling can outperform an explicit decision process. Although Gigerenzer (2008) highlights the success of spontaneously acting according to someone's gut feeling, we rather propagate to base the decision on the analysis and documentation of the risk analyst. This might allow the decision maker to become aware of unexpected value drivers and scenarios critical for the cooperation.

---

<sup>208</sup>See Peterson (2009) for a full introduction to decision theory.

Nevertheless, the decision maker should keep in mind that notwithstanding an exemplary risk analysis there is always a myriad number of black swans (see section 2.2 and Taleb 2007) potentially foiling any projections decisions are based on. The decision maker and the risk analyst are not at the mercy of such black swans. Bayesian estimates can better reveal uncertainty on the risk assessments than classical point estimates. Moreover, a black swan for them might be an uncertain factor or even a quantifiable risk factor for some experts. Last but not least, no decision should critically rely on the assessments of the risk analysis. A safety cushion should be able to compensate unexpected minor up to medium variations. The inadequateness of human planning has coined the famous sentence of the chief of staff of the Prussian Army Helmuth Graf von Moltke the Elder (1800 – 1891): "Kein Plan überlebt die erste Feindberührung" (English "No campaign plan survives first contact with the enemy") summarizing the idea of uncertainty (in war) of the Prussian military theorist Carl von Clausewitz (1780 – 1831).<sup>209</sup> This is closely related to the *ludic fallacy* of Taleb (2007) describing people's tendency to believe that their models can really describe reality which allows planning in a quasi deterministic world. However, risk analysis, planning, and decision making - in our Bayesian understanding - should rather raise the awareness of the critical factors and their interrelations. This enables persons to quickly react and to be prepared for unpredictable events.

## 16 Expert Elicitation

The basic idea of BRM is to explicitly quantify subjective uncertainty on risk factors. So far, we have not concentrated on the fact who is a suitable source for the assessments. For simplicity, we have mostly assumed that the risk analyst herself just expresses her uncertainty. Of course, the risk analyst or even a group of risk analysts are mostly professionals but might not have expertise on all topics they deal with. Consequently, a consultation of respective experts seems appropriate as long as the additional invested time and money can be justified by the expert assessments.

There are different levels of expert consultation. Experts can be asked (1) for some advice on the identification of risk factors as well as on the construction and on the evaluation of models, (2) for some point estimates, or (3) for probabilities expressing their uncertainty on risk factors and their possible outcomes. The latter is called *expert elicitation*.

We do not want to discuss adequate general strategies for cooperation and integration of in house and external experts. This could be better performed by management consultancies

---

<sup>209</sup>See the Boston Consulting book "Clausewitz on strategy: inspiration and insight from a master strategist" interpreting Clausewitz's ideas of the unpredictability of war for the today's management (von Ghyczy et al. 2002).

like McKinsey, Roland Berger, Boston Consulting, or Deloitte Consulting. Although the expert consulting of *point estimate* (normally most likely or expected value of a risk factor) is still going strong (Morgan & Keith 2007), we do not concentrate on this topic for two reasons: First, point estimates conceal the uncertainty of an expert. Second, there is a bad habit that the dispersion of point forecasts of different experts is normally interpreted as the uncertainty of a group. Instead, it is just the observed *disagreement of the experts* (Nesvisky 2006).<sup>210</sup> Theoretically, all experts can be very uncertain although they do not disagree about the forecasted value. Gürkaynak & Wolfers (2005), p. 17, found that uncertainty and disagreement comove with a low correlation. Hence, they conclude that disagreement is not a good proxy for uncertainty.

In this and the next sections, we introduce the concept of expert elicitation which is an important instrument in different areas like decision analysis, psychology, risk analysis, Bayesian statistics, mathematics, and philosophy (Hora & Jensen 2002, p. 2). Even if the risk analyst abstains from expert elicitation, its understanding can help to improve the quantification of her own uncertainties.

In expert elicitation, the risk analyst is directly or indirectly confronted with one or several experts whose assessments might vary between experts and over time. Consequently, the risk assessments are afflicted with two kinds of uncertainty:

- There is the *communicated uncertainty* of the expert which depends on the expert's substantive and normative expertise.<sup>211</sup> In the simplest form, the expert estimates the most plausible outcome of the risk factor enhanced by a verbal statement about her personal degree of belief that the forecast becomes true. In the most complex form the expert communicates a complete probability or density distribution for the risk factor. A point forecast is only sufficient when experts have complete information and they feel certain.
- The *analyst's uncertainty* reflects the risk analyst's degree of belief in the expert assessment whether it is informative or misleading. The analyst has to decide whether she deals with a risk factor or an uncertain factor (see section 2.2).

Actually, nobody should expect *perfect foresight*. Experts are invited as professionals on a topic to state their uncertainty about a risk factor. It is important to understand that there is no classification whether a subjective assessment is "true" or "wrong" (Hora & Jensen

---

<sup>210</sup>Mankiw et al. (2003) approximate the disagreement of the experts by the standard deviation of the point forecasts.

<sup>211</sup>Again, the substantive expertise is an expert's level of adequate and useful information on the risk factor, while normative expertise just describes the expert's ability to communicate this information in a probabilistic form. See section 2.2 for more details.

2002, p. 4). There are no "objective" probabilities intrinsic to an object or event but there are subjective probabilities intrinsic to the assessors.<sup>212</sup> Accordingly, it is more important to reproduce an expert's subjective probability assessments as exactly as possible because the communicated as well as the analyst's uncertainty is critically affected by the normative expertise.

## 16.1 Heuristics

Tversky & Kahneman (1974) stated in the mid of the 1970ies that (non-expert) people have problems to express their assessments in form of subjective probabilities. Therefore, they would use *heuristics* (= rules of thumb) to reduce complexity, that lead to systematic predictable biases like conjunction fallacy, base rate neglect, and miscalibration:

- *representativeness*: The probability that something or someone belongs to a group is judged by its similarity to this group. E.g., a blond-haired person is commonly disproportionately often assumed to be a Swede as there is the opinion that nearly all Swedes are blond. The *base rate* of blond Swedes compared to all blond-haired people worldwide is *neglected*.
- *availability*: People tend to judge the frequency of an event by the ease of remembering past examples. Therefore, rare extreme events appear to occur more frequently than rather personally not so important events.
- *adjustment and anchoring*: People usually anchor their (probability) estimates and adjust other information to be in accord with the anchor. E.g., when you want people to estimate the number of inhabitants of Berlin and you state that London is another important European capital that has 7.5 million inhabitants, the anchoring would bias the estimations to 7.5 million inhabitants although the real number is 3.4 million.

Since Tversky & Kahneman (1974), the literature has mainly related to this article and its biases although Tversky and Kahneman relaxed their statements later (Kynn 2008, p. 239). Kynn (2008) summarizes the research on probability elicitation that helps to elicit "good" probability figures. To do so, it is "not only what we ask experts to assess, but how we ask it" (Kynn 2008, p. 240). The question about the number of inhabitants of Berlin is not univocal. The wording could easily be understood that it is a fact that Berlin has the same number of inhabitants like London. Hence, the biased estimation could be rather because of a wrong wording than of the incapacity of people to express good assessments.

---

<sup>212</sup>See Nau (2001) who argues that de Finetti (1974), p. x, was right with his statement "probability does not exist".

Another problem is that people usually violate the *law of total probability* as the total of all their probabilities sums to more than 100%, called *subadditivity*. However, even Bernoulli suggested that this is not a big problem for subjective probabilities (Kynn 2008, p. 246).

The *conjunction effect* describes the observation that people partly estimate the probability for A is less than the probability for A given B. A prime example was brought forward by Kahneman & Tversky (1982), p. 126:

*"Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations. Which is more probable? (i) Linda is a bank-teller; (ii) Linda is a bank-teller who is active in the feminist movement."*

People tend to answer (ii) to be more probable than (i). From probability theory this is formally wrong. Kynn (2008), pp. 247 argues that the conjunction effect can arise from a misleading question. People maybe seem to conclude in response to (ii) that (i) means that Linda is a bank-teller **and not** in the feminist movement. The interviewer should pose the question more clearly without requiring people to know statistical calculus.

In general, people tend to have no problems to increase the probabilities rather than to decrease the others consistently. To avoid such problems odds ratios can be used as no counterbalancing is needed (Kynn 2008, p. 250). Alternatively, weights instead of probabilities seem adequate, too.

However, over the years, there have been found several more biases which cannot be traced back to the fact of misunderstandings: (1) People tend to be more confident about their prediction skills than their historical record of correct responses would suggest (Kynn 2008, p. 253). This *overconfidence effect* depends on the type of the task and the level of difficulty of the question and is (2) accompanied by the *hard-easy effect* which describes that people are overconfident for hard questions and under-confident for easy questions (Kynn 2008, p. 253). (3) People have the tendency to overestimate events with very low probabilities and underestimate events with high probabilities. This is called *favorite-longshot bias* (Thaler & Ziemba 1988).<sup>213</sup> (4) People trade according to their desires rather than to their own assessments.<sup>214</sup> (5) People seem to have problems to express and interpret complex thoughts (Ostrover 2005, p. 10). (6) People tend to secure certain gains as well as to avoid certain losses which results in insuring future gains but bearing uncertain losses. This is called

<sup>213</sup>Wolfers & Zitzewitz (2004) compared Tradesports ([www.tradesports.com](http://www.tradesports.com)) prices and option prices from Chicago on the S&P 500. They found that extremely high and low outcomes of the S&P 500, which seem to be very unlikely, are relatively overpriced in Tradesports.

<sup>214</sup>Forsythe (1999) found evidence that people's trading in political stock markets is correlated with their party identification. They are motivated by partisanship. However this could be a special problem of political polls and political stock markets.

*reflection effect* (Kahneman & Tversky 1979). (7) People seem to have problems to handle and understand probability assessments. They can rather distinguish between an event that is certain or not certain than between events that have a probability of 60% or 50%. Consequently, they are more concerned about the loss of certainty, 100% probability to 90%, than 60% to 50%, as they are interested to reduce uncertainty to zero. This phenomenon is called *certainty effect*.<sup>215</sup> (8) People often ignore information which is shared by several sources but concentrate on information which is not shared. Moreover, they seek for the differences in the structure of the alternatives. They rather decide for an alternative because it is differently presented. Therefore, in a second presentation they would choose another alternative. This is called the *isolation effect* or *von Restorff effect* (von Restorff 1933).

Other important cognitive biases are:<sup>216</sup> conservatism or endowment effect,<sup>217</sup> bandwagon fallacy,<sup>218</sup> choice-supportive and confirmation bias,<sup>219</sup> congruence bias,<sup>220</sup> status quo bias,<sup>221</sup> *déformation professionnelle effect*,<sup>222</sup> distinction bias,<sup>223</sup> need for closure,<sup>224</sup> neglect of prob-

---

<sup>215</sup>E.g., people taking part in a Russian Roulette are generally willing to pay more of their lifetime income to get rid of one out of one bullet in a six-barreled gun than one bullet out of two. Theoretically, people should pay more in the one out of two scenario as in both cases the probability is reduced by the same amount but there still exists a probability of death in the second scenario in which all money is worth less for them (Zeckenhauser 1996, endnote 16).

<sup>216</sup>An excellent survey of nearly all fallacies you can find in Wikipedia article "List of cognitive biases" and its related links.

<sup>217</sup>People are willing to pay more for things they already possess to keep them instead to pay for the same things when they don't own them. E.g., there is an experiment with students. The half of them are handed beer mugs. Then they are offered money to give back the mugs. The other students were simultaneously offered to take money or a mug. Although both situations should be the same for a rational utility maximizing agent, students first equipped with the mug generally named a higher price for the mug (Daniel Kahneman in a Newsweek; issue of April 10, 1995; [www.newsweek.com/id/110181/output/techbiz/biz](http://www.newsweek.com/id/110181/output/techbiz/biz) (last revised July 1, 2008))

<sup>218</sup>People tend to rely on the mainstream because it is more comfortable to follow a herd instinct instead to oppose the assessment of the majority (see Nadeau et al. 1993).

<sup>219</sup>Ex post, people tend to put lipstick on the pig by stating partly indefensible positive features supporting the option as well as ignoring negative features opposing the option they have chosen before. This is partly because people really remember this way - *choice-supportive bias* (see Mather & Johnson 2000 and Mather et al. 2000) - and partly because people try to find, respectively ignore new information to justify their assessment - *confirmation bias* (see Lord et al. 1979).

<sup>220</sup>People tend to stick to a hypothesis they initially believed to be correct. They prefer to modify the hypothesis rather than to consider alternative ones (see Wason 1960).

<sup>221</sup>People refuse to reconsider an established assessment as long as there is no overwhelming evidence.

<sup>222</sup>Once experts are asked being specialists for a certain topic, there is the danger that she believes the whole problem could be explained by that area of expertise.

<sup>223</sup>The assessment of two different options depends on the fact whether both are analyzed separately or simultaneously. In a simultaneous analysis small differences are seen relatively more distinctive.

<sup>224</sup>People dislike to be uncertain. Consequently, people tend to come to a conclusion.



ability bias,<sup>225</sup> wishful thinking,<sup>226</sup> optimism bias,<sup>227</sup> overconfidence effect,<sup>228</sup> ambiguity effect,<sup>229</sup> attentional bias,<sup>230</sup> gambler's fallacy,<sup>231</sup> hindsight bias,<sup>232</sup> illusory correlation<sup>233</sup> and disregard of regression toward the mean.<sup>234</sup>

Kahneman & Tversky (1979) concluded from such biases that the *expected utility theory*<sup>235</sup> might be not suitable to describe decision making under risk. As an alternative they proposed the *prospect theory* for describing decision making under uncertain future events with known probabilities.

The prospect theory features two steps, *editing* and *evaluation*. In the edition step humans sort all possible outcomes of the future event relative to a *reference point*, which can be the worst case or the status quo. All outcomes which are more (less) preferable than the reference point are seen as gains (losses). In the evaluation step a human being calculates a subjective *overall value of an edited prospect* (Kahneman & Tversky 1979, p. 276)

$$V = \sum_{k=1}^K w(p_k) \cdot v(x_k),$$

where  $p = (p_1, \dots, p_K)'$  and  $x = (x_1, \dots, x_K)'$  are the subjective probabilities and payoffs of the  $K$  possible outcomes. The formula above has two functions which distinguish it from the expected utility formula:

- The function  $w(\cdot)$  is the *subjective probability weighing function* which reproduces the observation that people overreact to small probabilities and underreact to large probabilities. This overweighting property should not be mixed up with the overestimation

---

<sup>225</sup>Sometimes, people deliberately ignore uncertainty.

<sup>226</sup>People state relatively more positive than negative forecasts as they tend to form assessments according to their hopes rather than according to the apparent evidence.

<sup>227</sup>People tend to be over-optimistic about planned actions resulting in too high likelihoods for positive outcomes and too low likelihoods for negative outcomes.

<sup>228</sup>People tend to over-estimate the correctness of their assessments resulting in a miscalibration of subjective probabilities.

<sup>229</sup>People tend to prefer options with known probabilities over options with uncertain but in the mean equal probabilities.

<sup>230</sup>People tend to test for some requirement of an assessment, only. Consequently, people can fail to notice contradicting evidence.

<sup>231</sup>People have problems to understand the idea of stochastic independence. Instead, they believe an outcome (e.g., head of a coin) will realize next because it has not realized for a long time.

<sup>232</sup>People tend to see the realized outcome of an events relatively more predictable after the realization than before.

<sup>233</sup>People tend to find strong correlations (group stereotypes) where there are no or only low level interrelations.

<sup>234</sup>People tend to expect extreme outcomes to repeat.

<sup>235</sup>see Schoemaker (1982) for more details

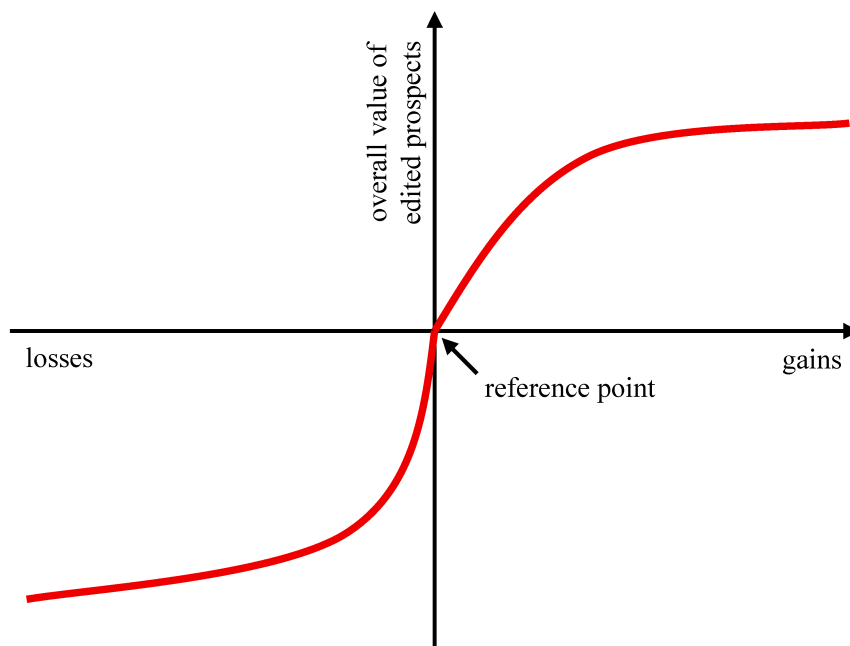


Figure 84: prospect theory - The gains value function is concave in gains and convex in losses (Kahneman & Tversky 1979, figure 3).

of low probabilities as in the Prospect Theory it is assumed that people know the "true" probabilities but state them differently

- The function  $v(\cdot)$  is the *value function* that reflects that people seem to have a concave (convex) appraisal of gains (losses). Moreover, the value function is in general steeper for losses than for gains (see figure 84).

People come to a decision by choosing a strategy, e.g. insurance ( $I$ ) vs. no insurance ( $NI$ ), which maximizes subjective *overall value of an edited prospect*

$$\max(V^I, V^{NI}) = \max\left(\sum_{k=1}^K w(p_k) \cdot v(x_k), v(\psi)\right),$$

with  $\psi$  as the insurance premium for a fixed gain ( $\psi > 0$ ) or loss ( $\psi < 0$ ). Irrespective the weighting function, people prefer to exchange risky gains against a fixed payment which equals the expected value of the risky ones or is even slightly below. Hence, people favor insurances to fix gains as they maximize their overall value of edited prospects. In contrast,

they rather refuse to insure losses except the insurance premium is far below the expected value of the risky payments. Consequently, whether persons buy an insurance or not depends on the reference point, which decides what payoffs are gains and losses. Especially when the reference point is the status quo people tend to accept high insurance premiums to lock future gains, immediately, while they tend to take high risks instead of insuring against future losses (reflection effect).

Although the prospect theory seems to be promising to explain some of the biases, the rest of the paper follows the expected utility theory. This is not only because of convenience. Harrison (1994) reviewed several biases which are supposed to indicate that the expected utility theory is wrong. However, he concluded that modification in the experiment designs would even result in support of the traditional expected utility theory.

## 16.2 Some Theoretical Thoughts on Expert Quality & Selection

So far, we have presented some of the most important biases a risk analyst should keep in mind when she tries to elicit experts. However, are there clear criteria to evaluate the quality of subjective statements? A standard approach is to calculate a person's *calibration* - how close is the mean probability assessment compared to the observed frequency. However, there is some opposition to this concept. Calibration seems not to be suitable to evaluate subjective probability assignments (Aven 2003, p. 64). Subjective probabilities convey totally different information compared to "objective" probabilities:

- The "objective" probabilities are an immanent part of the forecasted object (respectively sequence), e.g. a "fair" dice, while subjective probabilities express the uncertainty of the forecasting subject.
- The "objective" probabilities are an axiomatic value while the subjective probabilities are variables that change with incoming information.
- While an assignment of probability one to the actual realizing outcome and zero to the others would be a miscalibration in classical statistics these are desirable subjective probabilities.<sup>236</sup>

---

<sup>236</sup>The concept of "objective" probabilities is not optimal for the performance measurement of credit portfolio ratings. Banks separate their credit portfolios into different rating classes from low to high probabilities of default. The aim is that the long-term default frequencies match the announced class probabilities. This is measured by the calibration. However, rating classes with probabilities different than zero or one are just a sign for a not perfect *selectivity*, a common measure in credit risk (see Bemann 2005). Such classes are only necessary because of a lack of information about the situation of the debtors. Optimally, all finally defaulted debtors should have been classified in the default class (= no credit granted) and all others in the premium class.

Suppose a "fair" dice with no bias to a specific number. In classical understanding of aleatory uncertainty, the optimal estimate would be to assign an "objective" probability of 1/6 to all numbers. In Bayesian understanding, subjective probabilities of 1/6 just reflect a person with the belief that the dice is not biased. An assignment of probabilities for a "fair" dice different to 1/6 needs not to be a failure but may rest on additional information. An experienced gambler could forecast the outcome of the next die role above average.

Nevertheless, calibration seems to be a suitable quality criterion as long as the risk analyst is interested in a frequency. Patt & Schrag (2003), p. 22, believe that (lay)persons are rather able to give probabilities for events where a frequency observation is at least theoretically possible than for singular events without past data as such probabilities convey a degree of freedom.

Often, but not always, there is more than one expert on a topic. There is the question whom to ask? According to the concept of *concordant beliefs* of Milgrom and Stokey (1982) rational people are assumed to interpret information in the same way (inter-subjective updating) irrespective of their (heterogeneous) prior beliefs. Consequently, it would be sufficient to arbitrarily choose an expert to ask for *expert probabilities* (Hanson 2002, p. 2). In reality, such expert probabilities can only be elicited when it is possible to reach a consensus between all experts<sup>237</sup> or when there is an expert community with a nearly homogeneous knowledge base.<sup>238</sup>

Unfortunately, no consensus can be reached on most topics - even within research communities. The degree of discordance of expert assessments critically depends on how information is spread over several experts. Bruggen et al. (2006), p. 3, distinguish between (1) *low knowledge-heterogeneity* and (2) *high knowledge-heterogeneity*. Information of low knowledge-heterogeneous problems are homogeneously spread over informants. Information is differently dispersed (in type and quality) for high knowledge-heterogeneous problems. It is common in corporations as much of the information is widely spread and undocumented, residing in the minds of the employees (Ostrover 2005, p. 9). Moreover, many employees are professionals on isolated topics. And companies often lack of systematic procedures for consolidating massive amounts of information into quantitative metrics (Ostrover 2005, p. 9).

Consequently, it is relevant which expert or group of experts to ask. In a situation of competing expert assessments, the risk analyst should generally prefer to consult as many experts as possible. A subsequent informal or formal aggregation of the expert statements

---

<sup>237</sup>Deliberation groups or Delphi studies could help to find common prior. Afterwards, the risk analyst needs only to contact one of those experts for further assessments.

<sup>238</sup>Weather forecasters usually make very similar predictions for the weather weeks or month in the future as they refer to some base rates, e.g. long term historic data. The more the forecasted day approaches the more the forecasts diverge because of new private information (Hanson 2002, p. 2).

by the risk analyst (or decision maker) normally results in a better risk assessment.<sup>239</sup>

## 16.3 Basic Elicitation Techniques

In this section we present some techniques that allow for a separate quantification of the risk analyst's or expert's uncertainty on a risk factor  $Y$ .<sup>240</sup> For simplicity, we only use the term expert even though the expert can be the risk analyst herself.

We start with techniques (scoring rules, promissory notes, lotteries, and lottery insurance markets) that can deal with experts not willing to reveal their assessments without any monetary compensation. All approaches are grounded on assumptions of profit or utility maximization of rational experts as well as on the Bayesian understanding of probabilities as a willingness to bet. Regrettably, these techniques are extremely academic and too demanding for real life.

Hence, we favor two basic and simple interval techniques (see sections 16.3.5 and 16.3.6 as well as Garthwaite et al. 2005) that can even account for imprecise probabilities: weighting of intervals (*fixed interval technique*) or calibrating probability distributions (*variable interval technique*).

All elicitation techniques require to some extent that the experts are able to quantify their latent probability assessment (normative expertise). Unfortunately, people can often not meet this requirement in real life. In section 16.4 we present some empirical results on the actual calibration of people and offer a Bayesian quantification of the miscalibration of subjective statements.

### 16.3.1 Scoring Rules

Due to the fear of cheap talk, advanced elicitation techniques have been developed to incentivize experts to report truthfully. One of the most important techniques is the *scoring rules* approach which offers experts monetary incentives when they reveal their "true" probability assessments. The first rule, the logarithmic scoring rule, was presented by Good (1952). Since that time countless other scoring rules have been presented.

---

<sup>239</sup>An intuitive example, showing the gains from aggregation when information is spread, comes from Kalovcova (2007), p. 4: Assume the world will be tomorrow in one of the six states  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$ , or  $F$  and there are five experts that give you an assessment according to their knowledge. The experts may fail to advise us the right state  $A$ . This may be modelled by expert drawing three balls with replacement from an urn with  $10 \times A$  and two balls of  $B$ ,  $C$ ,  $D$ ,  $E$ , and  $F$ . E.g., the experts draw  $AAB$  (expert 1),  $AEE$ ,  $ABF$ ,  $ACD$ , and  $CDF$ . Then expert 1 advises state  $A$ , expert 2 states  $E$  and the others have no clue. If we had an aggregation mechanism, we could observe the experts have drawn five times  $A$ , and two times  $B$ ,  $C$ ,  $D$ ,  $E$ , and  $F$ .

<sup>240</sup>See section 15.2 for the *specific risk quantification* module of the iBRA concept.

The huge advantage of scoring rules is that the operator (e.g. the risk analyst) of a *proper* scoring rule does not need to know the "true" expert probabilities. Instead, the experts are rewarded by payoff rules maximizes their expected payoff for a "true" statement, only. However, some assumptions are needed: (1) The experts know their "true" probability assessments but do not want to reveal them for free (Kynn 2008, p. 255). (2) The experts need to understand the payoff structure of scoring rules. (3) Experts should be utility maximizers and consequently willing to reveal their subjective probabilities for cash. In the following we present some formulas for a better understanding of scoring rules.<sup>241</sup>

Scoring rules are (monetary) rewarding schemes that try to get an expert to reveal her subjective probability assessment  $\pi = (\pi_1, \dots, \pi_K)'$  on the  $K$  outcomes of a discrete or discretized risk factor. The risk analyst can only observe the expert's reports  $r_k$ , for  $k \in \{1, \dots, K\}$ . Consequently, the scoring rule pays  $s_k(r)$  given outcome  $k$  realizes and the report is  $r = (r_1, \dots, r_K)'$ . A *proper* scoring rule maximizes the expected payoff when the expert reveals her subjective assessment ( $r = \pi$ )

$$\pi = \arg \max_r \sum_{k=1}^K \pi_k s_k(r), \quad (19)$$

where  $\sum_{k=1}^K r_k = \sum_{k=1}^K \pi_k = 1$ . Prominent proper scoring rules are summarized in table 36:

	scoring rule
quadratic	$s_k(r) = a_k + b \cdot r_k - b \sum_{k=1}^K r_k^2 / 2$
spherical	$s_k(r) = a_k + b \cdot r_k / \sqrt{\sum_{k=1}^K r_k^2}$
logarithmic	$s_k(r_k) = a_k + b \cdot \ln r_k$
power law	$s_k(r) = a_k + b \cdot \alpha \int_0^{r_k} \varepsilon_k^{\alpha-2} d\varepsilon_k - b \sum_{k=1}^K r_k^\alpha, \quad \alpha \geq 1$

Table 36: prominent proper scoring rules (Hanson Jan. 2002, p. 4)

There is no other proper scoring rule than the logarithmic where the payoff  $s_k(\cdot)$  for outcome  $k$  only depends on the expert's probability report  $r_k$  (Hanson Jan. 2002, p. 4).

It is important to know, that proper scoring rules only hold when experts maximize their expected payoff (see equation 19). However, this assumption is only suitable for risk-neutral experts. Risk-averse experts maximize their expected utility (Hanson 2002, p. 5). This is not equivalent to a maximization of their payoff. An example may provide a better understanding of scoring rules.

Assume, the risk analyst wants to elicit the probabilities of two mutually exclusive outcomes,  $A$  and  $B$  ( $P(A \cap B) = 0$  and  $P(A) + P(B) = 1$ ) of a risk factor. A simple logarithmic

<sup>241</sup>For a closer description of scoring rules see Lindley (1985), pp. 23.

scoring rule seems suitable<sup>242</sup>

$$s_k(r) = \begin{cases} \ln r, & \text{when } k = A \text{ occurs} \\ \ln(1-r), & \text{when } k = B \text{ occurs} \end{cases},$$

where  $r$  ( $1-r$ ) is the expert's probability report of outcome  $A$  ( $B$ ). Furthermore,  $f_W(w|A)$  and  $f_W(w|B)$  are the expert's subjective probability distributions for her wealth in a world given outcome  $A$  or  $B$ . Then, the expert's expected utility - inclusive the payments of the scoring rule - is (Kadane & Winkler 1988, pp. 359)

$$E(U(W)) = \pi \int U(w + \ln r) f_W(w|A) dw + [1 - \pi] \int U(w + \ln(1-r)) f_W(w|B) dw,$$

where  $\pi$  ( $1-\pi$ ) is the expert's subjective probability assessment of outcome  $A$  ( $B$ ). For a utility maximization expert it follows

$$\frac{r}{1-r} = c \frac{\pi}{1-\pi},$$

where

$$c = \frac{\int U'(w + \ln r) f_W(w|A) dw}{\int U'(w + \ln(1-r)) f_W(w|B) dw}.$$

The report  $r$  matches the subjective assessment  $\pi$  when  $c = 1$ . This is true when  $U(\cdot)$  is linear, i.e. the expert is risk-neutral. The *no-stakes condition* of Kadane & Winkler (1988) is not sufficient to guarantee  $c = 1$ . It simply states that the expert's wealth is independent of the outcome of the risk factor, i.e.  $f_W(w|A) = f_W(w|B)$ , because she has no other stakes in a special outcome of the risk factor.<sup>243</sup>

Scoring rules suffer from a *thick market problem*<sup>244</sup> because they are not able to produce a single consensus estimate when experts give different responds (Hanson 2003, p. 108). This is problematic when we assume that the knowledge about a risk factor is dispersed over different experts. In such a situation the different experts might come to different assessments when they judge on the basis of different, only partly overlapping information. Scoring rules do not offer a mechanism to deal with diverse answers (Hanson Jan. 2002). Indeed, it is possible to apply a weighting mechanism for the different answers. Unfortunately, a weighting according to the scoring rule payoffs is not univocal. Even when proper

<sup>242</sup>This corresponds to a logarithmic scoring rule  $s_i(r_i) = a_i + b \cdot \ln r_i$ , where  $i = A, B$ ,  $a_i = 0$ , and  $b = 1$ .

<sup>243</sup>An extreme example is the expert elicitation of the probability to win a nuclear war. The expert could deliberately underestimate that probability in order to prevent the war because her wealth inclusive the payments of the scoring rule would be worthless in such a situation.

<sup>244</sup>In contrast, prediction markets (see section 18) rather suffer from a *thin market problem*. Their results improve the more traders take part. However, they might fail with too less traders. In the case of one trader standard continuous double-auction markets don't work.

scoring rules are used they can lead to different rankings of the experts (Winkler & Murphy 1968, p. 756). Moreover, the resulting aggregated probability distribution does not allow to retrace to the marginal distributions of the single agents. This is called *opinion pooling problem* (see Garg et al. 2004). It is not far-fetched that the resulting weighted distribution could be disapproved by all interviewed experts.

### 16.3.2 Promissory Notes

The previous approach assumes experts with clear probability assessments. Experts just need to be asked for their probability estimates, although monetary incentives are often needed. However, the discussion - initiated by Tversky & Kahneman (1974) - on the human problems to formulate subjective probability statements indicates that a *direct* query of probabilities seems critical. Alternatively, Bayesian theory propagates to *indirectly* derive subjective probabilities (= degrees of belief) from the person's bets on a risk factor (willingness to bet, de Finetti 1976).

A very simple approach is the use of *promissory notes* (Kadane & Winkler 1988) which pay  $x_A > 0$  when outcome  $A$  realizes and otherwise nothing. Then, it is asked for the maximum price  $\lambda$  the expert is willing to pay for one note. The expert's reported probability for outcome  $A$  is  $r_A = \lambda/x_A$ .

The odds ratio of a promissory note is (see Kadane & Winkler 1988 or Kadane & Winkler 1987)

$$\frac{r_A}{1 - r_A} = c \frac{\pi}{1 - \pi},$$

where

$$c = \frac{\lambda}{x_A - \lambda} \frac{\int [U(w - \lambda + x_A) - U(w)] \cdot f_W(w|A) dw}{\int [U(w) - U(w - \lambda)] \cdot f_W(w|\bar{A}) dw}$$

and  $\bar{A}$  is the complement outcome of  $A$ .

The *no-stakes condition* of Kadane & Winkler (1988) is not sufficient for  $c = 1$ . Even when a risk-averse expert has no other stakes outside the elicitation, the promissory note approach underestimates the experts probability assessments for  $A$  because of  $c < 1$ . Consequently, this simple procedure is not suitable to elicit subjective probabilities.

### 16.3.3 Lotteries

Lotteries were another early approach to separate probability assessments and preferences of experts in elicitation. An expert decides between two lotteries with identical payoffs.<sup>245</sup> In a first lottery the expert receives a payoff  $x_A$  when outcome  $A$  occurs and  $x_B$  in the case

<sup>245</sup>see Kadane & Winkler (1988), pp. 357, for more details



of outcome  $B$ . Both outcomes are mutually exclusive. A second *independent* lottery has the same payoff  $x_A$  ( $x_B$ ) with the reported probability of  $r$  ( $1 - r$ ). Without losing generality we presume  $x_B = 0$ . The experts have to set the probability report  $r$  at such a value that they are indifferent between the lotteries.

Assume,  $f_W(w|A)$  and  $f_W(w|B)$  as the probability distributions for the wealth of the expert without any lottery payoff given a world where  $A$  or  $B$  has realized. Then, the expert's expected utility for the first lottery equals

$$E_1(U(W)) = \pi \int U(w + x_A) f_W(w|A) dw + [1 - \pi] \int U(w) f_W(w|B) dw,$$

with  $\pi$  as the expert's probability assessment of outcome  $A$  and  $f_W(w|A)$  is the expert's latent uncertainty on her wealth given outcome  $A$ . Respectively, the expected utility for lottery 2 is

$$\begin{aligned} E_2(U(W)) &= r \left[ \pi \int U(w + x_A) f_W(w|A) dw + [1 - \pi] \int U(w + x_A) f_W(w|B) dw \right] \\ &+ [1 - r] \left[ \pi \int U(w) f_W(w|A) dw + [1 - \pi] \int U(w) f_W(w|B) dw \right]. \end{aligned}$$

The expert is indifferent between the two lotteries when  $E_1(U(W)) \stackrel{!}{=} E_2(U(W))$ . After some arrangements it follows the odds ratio (Kadane & Winkler 1988, pp. 358)

$$\frac{r}{1 - r} = c \frac{\pi}{1 - \pi}, \text{ with}$$

where

$$c = \frac{\int [U(w + x_A) - U(w)] \cdot f_W(w|A) dw}{\int [U(w + x_A) - U(w)] \cdot f_W(w|B) dw}.$$

In the case of  $c = 1$ , the expert's announced value  $r$  is equivalent to her probability assessment  $\pi$ . This is true for linear  $U(\cdot)$ , i.e. the expert is risk-neutral, or when the no-stakes condition holds, i.e.  $f_W(w|A) = f_W(w|B)$ .

#### 16.3.4 Lottery Insurance Market

Apart from lotteries, monetary elicitation techniques fail to elicit the "true" probabilities of risk-averse and rational experts - irrespective of whether they have other stakes or not. Unfortunately, lotteries in their basic form can not be used for risk factors with more than  $n = 2$  possible outcomes. Hanson (2002) proposes the *lottery scoring rule* as an alternative that pays lottery tickets instead of cash.

First, the expert is rewarded with  $K$  different lottery tickets for  $K$  different reports  $r = (r_1, \dots, r_K)'$ , for  $\sum_{k=1}^K r_k = 1$ , on the probability of the outcomes  $k = 1, \dots, K$ . Subsequently,

the lottery on the realized outcome will only be conducted. All other lotteries will be canceled and their tickets will become worthless.

The ticket on the realized outcome  $k$  pays  $x$  with an *outcome-dependent* probability  $\alpha_k \in [0, 1]$  and 0 with a probability of  $1 - \alpha_k$ . The state dependent probability is determined by a scoring rule  $\alpha_k = s_k(r) \in [0, 1]$ ,<sup>246</sup> which depends on the expert report  $r$ . Then a "faithfully" report  $r_k = \pi_k$  maximizes the expert's expected utility.

$$E(U(W)) = \sum_{k=1}^K \pi_k \cdot [[1 - \alpha_k] U(w) + \alpha_k U(w + x)] = U(w) + [U(w + x) - U(w)] \cdot \sum_{k=1}^K \pi_k \alpha_k.$$

Once the no-stakes condition does not hold, the expert's wealth depends on the outcome of the risk factor. The basic lottery scoring rule does not guarantee "faithfully" reports. A solution is a competitive *insurance market* with  $J$  price-taking experts (Hanson 2002, pp. 5). For outcome  $k$ , an expert  $j$  can hedge,  $h_{kj}$ , her outcome-dependent wealth  $w_{kj}$  to get a hedged wealth of  $z_{kj} = w_{kj} + h_{kj}$ . Hedging does only take place between the experts. For each outcome  $k$  the aggregated hedging payments are zero over all experts, i.e.  $\sum_{j=1}^J h_{kj} = 0$ . The price for an insurance payment  $x$  in  $k$  is  $p_k$  which is formed by demand and supply. All rational experts  $j = 1, \dots, J$  maximize their expected utility

$$E_j(U_j(\cdot)) = \sum_{k=1}^K \pi_{kj} U_j(z_{kj}),$$

where  $\sum_{k=1}^K p_k h_{kj} = 0$ . Given the insurance payments  $p = (p_1, \dots, p_K)'$ , expert  $j$  seeks for the outcome-dependent payoffs  $z_{kj}^*$  that satisfy  $\pi_{kj} U_j'(z_{kj}^*) \stackrel{!}{=} \eta_j p_k$  for  $\eta_j = \sum_{k=1}^K \pi_{kj} U_j'(z_{kj}^*)$ . The experts will trade as long as the market has not reached an equilibrium where no expert wants to buy or sell any hedging. After trading run dry, the insurance market is closed. Then the hedging price  $p_k$  for a payment of  $x$  is the market consensus on the probability of outcome  $k$ .

The scoring rules, promissory notes, lotteries, as well as the lottery insurance markets are theoretically and mathematically very elegant. However, they seem not feasible in real life: (1) These techniques are mostly designed for risk-neutral and Bayesian utility maximizing experts. (2) Even when experts would behave in such a way, they would need to understand the payoff mathematics. (3) Experts are assumed to know their probabilities exactly but are not willing to state them for free. In our understanding, it is doubtful whether experts really would deliberately express false statements as long as the elicitation process is well

<sup>246</sup>The payments of the scoring rules need to be between zero and one. This condition can be fulfilled for all reports by the quadratic and spherical scoring rules but not by the logarithmic scoring rule as  $\lim_{r \rightarrow 0} \ln r = -\infty$ .

thought-out and organized. Consequently, we prefer the simple fixed and variable interval techniques.

### 16.3.5 Fixed Interval Technique

The basic idea of the fixed interval technique is to express uncertainty on an uncertain discrete or discretized risk factor  $Y$  by weighting different intervals. Therefore, the risk analyst needs to separate the domain of the risk factor  $Y$  into  $K$  intervals  $y^I = (y_1^I, \dots, y_K^I)'$ , where  $y_k^I \in [y_k^{low}, y_{k+1}^{low})$ , for  $k = 2, \dots, K - 1$ , and  $y_K^I \in [y_K^{low}, y_K^{up}]$ .<sup>247</sup> The intervals need to be disjunct, i.e.  $y_k^I \cap y_l^I = \emptyset$  for  $k \neq l$ , and should cover the whole domain of  $Y$ . Subsequently, the expert has to allocate  $K$  weights  $w = (w_1, \dots, w_K)'$  to the intervals. It is possible to endow the expert with a predefined amount of weights  $\tilde{w}$  that  $\sum_{k=1}^K w_k \stackrel{!}{=} \tilde{w}$  must hold. A free allocation is possible, too. For a more standardized presentation of the experts uncertainty on the risk factor  $Y$  we just need to normalize the weights to get the expert's *discrete prior distribution*  $\pi(y) = (\pi_1, \pi_2, \dots, \pi_K)'$ , where the interval probabilities calculate by  $\pi_k = w_k / \sum_{l=1}^K w_l$ .<sup>248</sup>

A visualization of the weights can support the expert to express her uncertainty. It is preferable to draw the prior distribution on a sheet of paper. We call such plots *expert histograms*. Unfortunately, expert histograms can be counterintuitive when the intervals  $y_k^I$  feature different widths. Then a *modified expert histogram*  $\tilde{\pi}(y) = (\tilde{\pi}_1, \tilde{\pi}_2, \dots, \tilde{\pi}_K)'$  is preferable as it accounts for the different ranges:

$$\tilde{\pi}_k = \frac{w_k \cdot Range_k}{\sum_{l=1}^K w_l \cdot Range_l},$$

where  $Range_k = y_{k+1}^{low} - y_k^{low}$ , for  $k = 1, \dots, K - 1$  and  $Range_K = y_K^{up} - y_K^{low}$ .

A special problem arises when interval 1 or  $K$  possesses infinite limits, i.e.  $y_1^{low} \rightarrow -\infty$  or  $y_K^{up} \rightarrow \infty$ . Then the analyst can decide to ask the expert for an absolute minimum  $y^{\min}$  or maximum  $y^{\max}$ . Alternatively, it is possible to model the tails (1) by a function exhibiting *exponential decay*

$$\begin{aligned} \tilde{\pi}_1^{exp}(y) &= \tilde{\pi}_2 \cdot \exp(-\lambda [y_2^{low} - y]), & \text{for } y \in (-\infty, y_1^{up}] \\ \tilde{\pi}_K^{exp}(y) &= \tilde{\pi}_{K-1} \cdot \exp(-\lambda [y - y_{K-1}^{up}]), & \text{for } y \in [y_K^{low}, \infty) \end{aligned}$$

or (2) by a function featuring *power law decay*<sup>249</sup>

$$\begin{aligned} \tilde{\pi}_1^{pow}(y) &= \tilde{\pi}_2 \cdot [y_2^{low} - y + 1]^{-\lambda}, & \text{for } y \in (-\infty, y_1^{up}] \\ \tilde{\pi}_K^{pow}(y) &= \tilde{\pi}_{K-1} \cdot [y - y_{K-1}^{up} + 1]^{-\lambda}, & \text{for } y \in [y_K^{low}, \infty). \end{aligned}$$

<sup>247</sup>In the case of no finite  $y_1^{low}$  and  $y_K^{up}$  the intervals are open.

<sup>248</sup>E.g., a weight on interval  $k$  twice as large as on interval  $l$  corresponds to a subjective assessment that the probability for  $Y$  to realize in interval  $k$  is twice as high as for realizing in  $l$ .

<sup>249</sup>Heavy-/ fat-tailed distributions possess a power law decay, where  $P(X > x) \sim x^{-[1+\lambda]}$ , for  $x \rightarrow \infty$

Both approaches guarantee a smooth decay starting from the modified probabilities of interval 2 and  $K - 1$ . The areas under these tails equal the *non*-modified expert probabilities  $\pi_1$  and  $\pi_K$ .

A suitable value  $\lambda > 0$  needs to be determined. For the exponential decay the parameter  $\lambda$  needs to guarantee

$$\begin{aligned} \frac{\tilde{\pi}_2}{\lambda} \cdot \exp(-\lambda [y_2^{low} - y]) \Big|_{-\infty}^{y_2^{low}} &\stackrel{!}{=} \pi_1, \text{ for } y < y_2^{low} \\ -\frac{\tilde{\pi}_{K-1}}{\lambda} \cdot \exp(-\lambda [y - y_{K-1}^{up}]) \Big|_{y_{K-1}^{up}}^{\infty} &\stackrel{!}{=} \pi_K, \text{ for } y > y_{K-1}^{up}. \end{aligned}$$

Hence, it is  $\lambda = \tilde{\pi}_2/\pi_1$ , for  $y < y_2^{low}$ , and  $\lambda = \tilde{\pi}_{K-1}/\pi_K$ , for  $y > y_{K-1}^{up}$ . In the case of power law decay we just solve

$$\begin{aligned} -\frac{\tilde{\pi}_2}{1-\lambda} \cdot [y_2^{low} - y + 1]^{1-\lambda} \Big|_{-\infty}^{y_2^{low}} &\stackrel{!}{=} \pi_1, \text{ for } y < y_2^{low} \\ \frac{\tilde{\pi}_{K-1}}{1-\lambda} \cdot [y - y_{K-1}^{up} + 1]^{1-\lambda} \Big|_{y_{K-1}^{up}}^{\infty} &\stackrel{!}{=} \pi_K, \text{ for } y > y_{K-1}^{up}, \end{aligned}$$

Assuming  $\lambda > 1$  then  $\lambda = [\pi_1 + \tilde{\pi}_2]/\pi_1$ , for  $y < y_2^{low}$ , and  $\lambda = [\pi_K + \tilde{\pi}_{K-1}]/\pi_K$ , for  $y > y_{K-1}^{up}$ .

An example might help to understand the differences between the approaches. Assume, a risk analyst has asked an expert for interval weights to express her uncertainty on a risk factor  $Y$ . The hypothetical answers and resulting probabilities are summarized in the table 37.

	$[-5, -2)$	$[-2, 0)$	$[0, 0.5)$	$[0.5, 1)$	$[1, 2)$	$[2, 3.5)$	$[5.5, 6)$	$[6, 10]$
$w_k$	5	4	3	2	3	3	4	6
$\pi_k$	0.16	0.13	0.1	0.06	0.1	0.1	0.16	0.19
$\tilde{\pi}_k$	0.05	0.06	0.19	0.13	0.1	0.06	0.06	0.05
$\tilde{\pi}_k^{exp}$	$\lambda = 0.4$	0.06	0.19	0.13	0.1	0.06	0.06	$\lambda = 0.33$
$\tilde{\pi}_k^{pow}$	$\lambda = 1.4$	0.06	0.19	0.13	0.1	0.06	0.06	$\lambda = 1.33$

Table 37: examples for interval weights

In the upper subplot of figure 85 we visualized the expert weights. For the middle subplot and  $\lambda > 0$ . Some reserve the term "fat tail" for distributions with infinite variance where  $0 < \lambda < 2$  ([http://en.wikipedia.org/wiki/Fat\\_tail](http://en.wikipedia.org/wiki/Fat_tail), last revised January 28, 2009). See Karagiannis et al. (2007) for a discussion on exponential and power law decay.

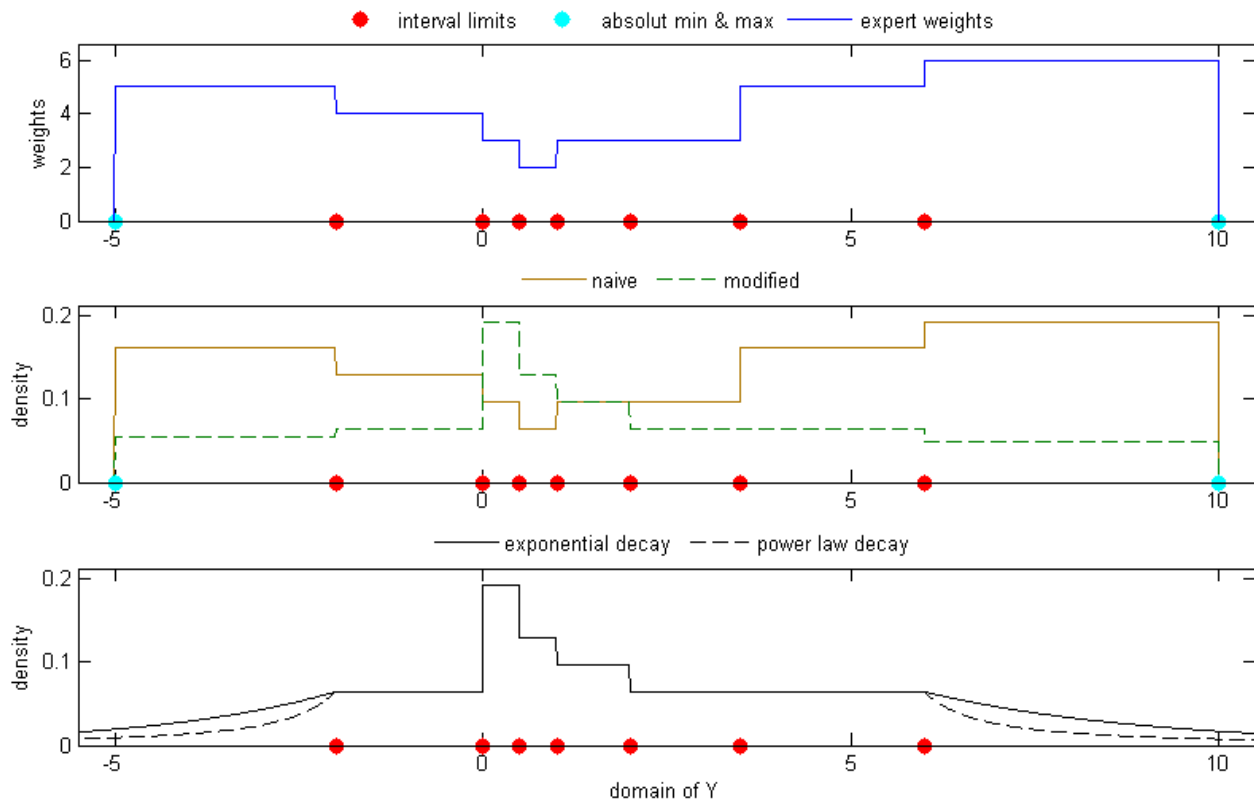


Figure 85: (modified) expert histograms

we assume a natural minimum ( $y_1^{low} = -5$ ) and maximum ( $y_K^{up} = 10$ ) for  $Y$ . It is easy to see that the basic expert histogram is misleading because of the heterogeneous interval widths. Hence, decisions should be based on a modified expert histogram. In the lower subplot we model the tails for the modified expert histogram. The power law decay falls faster as it dominates the exponential decay far in the tails.<sup>250</sup>

### 16.3.6 Variable Interval Technique

An alternative to the fixed interval method is the *variable interval technique* which propagates a subjective calibration of a stochastic distribution. It seems obvious to describe expert statements by distributions normally used in data analysis, like Gaussian, Student's  $t$ , exponential, chi-square, gamma, beta, or binomial distribution. However, a direct inquiry for parameters seems difficult as these distributions mostly have non-intuitive parameters. This

<sup>250</sup> An alternative to the (modified) expert histograms is the implementation of kernel estimators like the Gaussian or the Epanechnikov kernel function.

is even true for assuming a Gaussian distribution for our risk factor  $Y \sim N(\mu, \sigma)$ . The mean value  $\mu$  is not a problem as the mean value, mode, and median collapse. In contrast, most people, even experts, might have a problem to quantify the standard deviation  $\sigma$ . Although such distributions could be used to match them to an elicited modified expert histogram, the dominant approach is to directly elicit special elicitation distributions like the triangular distribution or the (modified) PERT distribution.

### Triangular Distribution

The *triangular distribution* is one of the workhorses in elicitation as it is easy to use even when there is little information on the risk factor  $Y$ . It simply defines  $Y$  by  $a$ ,  $c$ , and  $b$  which are the lower and the upper bounds as well as the most likely value of  $Y$ , which is more intuitive than asking for the mean value. The respective pdf is

$$Tri(y|a, b, c) = \begin{cases} \frac{2[y-a]}{[c-a][b-a]}, & \text{for } a \leq y \leq b \\ \frac{2[c-y]}{[c-a][c-b]}, & \text{for } b < y \leq c \end{cases},$$

where  $E(Y) = [a + b + c]/3$  and  $Std(Y) = \sqrt{a^2 + b^2 + c^2 - ab - ac - bc}/\sqrt{18}$  are both equally sensitive to all three parameters.

Often analysts try to avoid asking experts for the *absolute* minimum or maximum ( $a$  and  $c$ ). This would provoke experts to state extreme values. Alternatively, the risk analyst can elicit a *credible interval*  $\mathcal{CR}(q)$  with credible level  $q$ . It is simply an interval where  $Y$  is presumed to realize with probability  $q \cdot 100\%$ . Hence, the expert is asked for a *practical* minimum and maximum ( $\tilde{a}$  and  $\tilde{c}$ )<sup>251</sup>

$$(a^*, c^*)' = \arg \min_{a, b} \left( \left[ \frac{1-q}{2} - F_{Tri}(\tilde{a}|a, b, c) \right]^2 + \left[ \frac{1+q}{2} - F_{Tri}(\tilde{c}|a, b, c) \right]^2 \right),$$

where  $F_{Tri}(\cdot|a, b, c)$  is the cdf of a triangular distribution and  $a^*$  as well as  $c^*$  are the implied absolute minimum and maximum. In the upper subplot of figure 86 you can find triangular pdfs for fictive expert statement on the CO2 price. All three statements have equal minimum and maximum values but diverge for the most likely value. In the lower subplot we show the variations of the triangular distributions (for fixed most likely price) when the expert is asked for practicable minimum and maximum instead of absolute bounds. Generally, this results in  $a^*$  and  $c^*$  more left and right compared to  $a$  and  $b$ . Unfortunately, this approach can lead to counterintuitive results as negative CO2 prices feature positive probability.

<sup>251</sup>In fact, the credible level  $Q$  is uncertain to the risk analyst. In section 16.4, we present some results of an elicitation where we have tried to quantify the uncertainty on  $Q$ . In section 16.5.2, we introduce an algorithm that accounts for the uncertainty on  $Q$ .

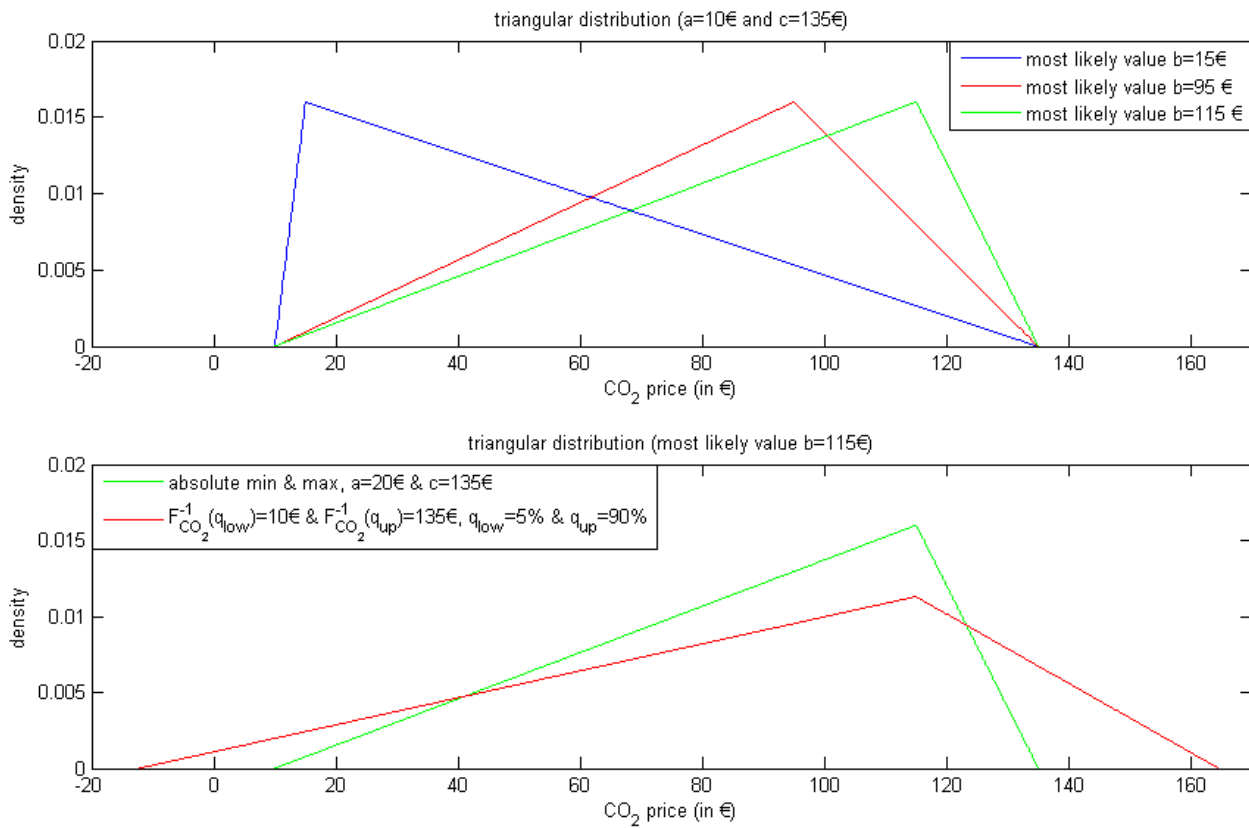


Figure 86: triangular distribution

**(Modified) PERT**

A more flexible distribution compared to the triangular one is the *modified PERT* (Program Evaluation and Review Technique) distribution. It is again defined by an absolute minimum  $a$ , absolute maximum  $c$ , and the most likely value  $b$ , of  $Y$ .<sup>252</sup> Additionally, its form depends on a shape parameter  $\gamma$  controlling the influence of extreme absolute minimum and maximum expert statements. The modified PERT is defined by

$$PERT_{\text{mod}}(y|a, b, c, \gamma) = \text{Beta}\left(\frac{y-a}{c-a} \middle| \alpha, \beta\right),$$

for  $\gamma > 0$  and  $\text{Beta}(\cdot|\alpha, \beta)$  as the pdf of a beta distribution with parameters

$$\alpha = \frac{[\mu - a] \cdot [2b - a - c]}{[b - \mu] \cdot [c - a]} \quad \& \quad \beta = \frac{\alpha \cdot [c - \mu]}{\mu - a} \quad \& \quad \mu = \frac{a + \gamma \cdot b + c}{2 + \gamma}.$$

<sup>252</sup> Again, it is possible to ask experts for the practical minimum and maximum of  $Y$ , beside the most likely value  $b$ .

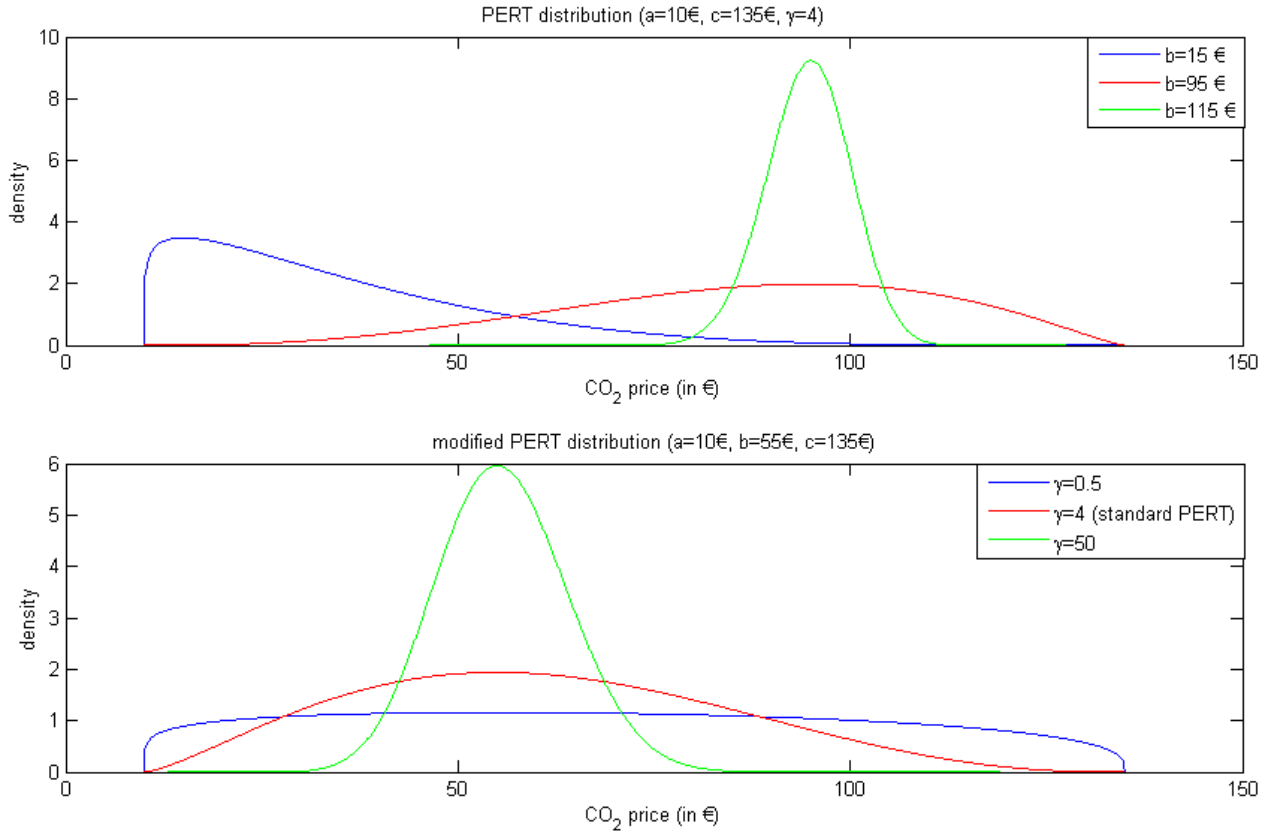


Figure 87: (modified) PERT distribution

The mean  $\mu$  of the modified PERT distribution is  $\gamma$  times more sensitive to the most likely value  $b$  than to the absolute minimum or maximum.<sup>253</sup> By moving  $\gamma$  up and down, the expert can express how likely values of  $Y$  are near  $a$  or  $b$ . In the upper subplot of figure 87 you can see different modified PERT distributions resulting from different most likely values,  $b = (15, 95, 115)'$  for hypothetical CO<sub>2</sub> prices. Given a most likely value  $b = 115\text{€}$  different modified PERT distributions ( $\gamma = (0.5, 4, 50)'$ ) are plotted in the lower subplot of figure 87.

## 16.4 Credibility Intervals

Elicitation distributions like the triangular or modified PERT distribution are easily to introduce to experts but require to state an absolute minimum and maximum. This can result in two problems: (1) Experts state an extreme minimum and maximum which they believe

<sup>253</sup>The basic PERT distribution is defined for  $\gamma = 2$ .



is rather unlikely. Then a triangular distribution puts too much probability to the tails. Although this effect could be compensated by the calibration of a modified PERT distribution, the identification of an adequate shape parameter could be time-consuming. (2) Experts could simply state practicable instead of absolute extrema.

For the latter proposal, there arises the question what "practicable" means? The risk analyst could simply query for the predefined  $q/2$  and  $1 - q/2$  quantiles of a risk factor  $Y$ : "Please state a lower (upper) bound for  $Y$  where you believe that the probability for the final outcome of  $Y$  below (above) this bound is  $q/2 \cdot 100\%$ ".<sup>254</sup>

Unfortunately, such an approach relies on the assumption that experts have no systematic bias in interpreting numerical probabilities. Alternatively, experts could be confronted with verbal probability statements.<sup>255</sup> A well-known probability wording is used by the Intergovernmental Panel on Climate Change (see IPCC 2008, p. 11): The risk analyst could ask for bounds where a lower and higher outcome of  $Y$  is "extremely unlikely" ( $< 0.05$ ), "very unlikely" ( $0.05 - < 0.10$ ), or "unlikely" ( $0.10 - < 0.33$ ).<sup>256</sup> Patt & Schrag (2003) show that people tend to have an inter-subjective understanding of the IPCC probabilities which is however biased. In contrast, Budescu & Wallsten (1986) find people to possess a stable but not inter-subjective rank ordering of such phrases. This can result in a fundamental miscommunication between the risk analyst and the expert (Fillenbaum et al. 1991).

Because of these contradicting research results, we propagate an alternative way which applies verbal statements which do not require a fixed numerical interpretation. We identified three "simple" questions to query experts for some basic *credible intervals*:

1. *high probability interval*: "Please state a range for  $Y$  for which you are confident that the final outcome of  $Y$  is covered with high probability and there is only a low probability that the value is not covered!"
2. *fifty/ fifty interval*: "Please state a smaller range for  $Y$  for which you believe that the interval has a fifty/ fifty chance to cover the final outcome of  $Y$ !"
3. *median interval*: "Please estimate a value for  $Y$  for which you think it is equally probable that the final outcome of  $Y$  is below (exact) or above that cutting point!"

The crux of these credible intervals are their credible levels  $Q$  which are risk factors for the risk analyst. An adequate distribution is needed to reflect the risk analyst's uncertainty. In

<sup>254</sup>see Garthwaite et al. (2005) for comparable approaches

<sup>255</sup>Budescu (1988) present evidence that there is often a small difference between decisions based on numerical or verbal statements.

<sup>256</sup>Other statements are "virtually certain" ( $> 0.99$ ), "extremely likely" ( $> 0.95 - 0.99$ ), "very likely" ( $> 0.90 - 0.95$ ), "likely" ( $> 0.66 - 0.90$ ), "more likely than not" ( $> 0.50$ ), "about as likely as not" ( $0.33 - 0.66$ ) (IPCC 2008, p. 11).

the elementary approach, the risk analyst simply states a prior  $\pi(q)$  by calibrating a beta distribution  $Q \sim \text{Beta}(a_0, b_0)$  according to some basic statistics,

$$E(Q) = \frac{a_0}{a_0 + b_0} \ \& \ Var(Q) = \frac{a_0 b_0}{[a_0 + b_0]^2 [a_0 + b_0 + 1]} \ \& \ Mode(Q) = \frac{a_0 - 1}{a_0 + b_0 - 2}.$$

Alternatively, the risk analyst can calculate a posterior  $\pi(q|Elicitation)$  based on some data helping to interpret the human understanding of credible intervals. Within the BRM research project we run an empirical analysis by asking  $J = 22$  (20 for median interval) arbitrary chosen persons (= experts) to state the credible intervals (see above) for  $n = 10$  almanac questions.<sup>257</sup> For all experts, we count the numbers of "hits"  $x_j \in \{0, 1, \dots, 10\}$ , i.e. the number of intervals that covered the "correct" answers. Given the risk analyst knows the latent credible level  $Q_j = q_j$  and the expert's answers can be assumed to be independent, the number of hits  $X_j$  should be binomially distributed  $X_j|n = 10, q \sim \text{Bin}(10, q)$ . Based on our  $n = 10$  observations we can set up the risk analyst's posterior of expert  $j$ 's credible level  $Q_j$

$$\pi(q_j|x_j) \propto L_{x_j}(q_j) \cdot \pi(q_j) = \text{Bin}(x_j|10, q) \cdot \pi(q).$$

Once the risk analyst assumes a beta-distributed prior  $Q_j \sim \text{Beta}(a_0, b_0)$ , the posterior  $Q_j|x_j \sim \text{Beta}(a_{post}, b_{post})$  is a conjugate prior where  $a_{post} = a_0 + x_j$  and  $b_{post} = b_0 + n - x_j$ . So far, we have shown a possibility how to measure the latent calibration of expert  $j$  beforehand the actual elicitation. This seems rather academic than practical as a risk analyst can count herself lucky to find experts at least taking part in the elicitation. Therefore, we assume experts to be exchangeable for the risk analyst in respect of their understanding of the credible interval questions. Then, given the stereotypical credible level  $Q = q$  and all answers are independent, the whole number of observed hits  $\sum_{j=1}^J X_j$  should follow a binomial distribution  $\sum_{j=1}^J X_j|n \cdot J, q \sim \text{Bin}(10 \cdot J, q)$ . Consequently, the posterior of a

---

<sup>257</sup>The questions were: "What is your estimation of the number of inhabitants of Florence (Italy) in June 2006?", "What is your estimation of the equatorial circumference of Earth?", "What was the official population of the People's Republic of China according to the 2000 census?", "What do you believe is the share of African Americans in the total U.S. population in 2007?", "What do you believe was the popular vote share of US president Barack Obama in the 2008 U.S. presidential election?", "What do you believe was the maximum daytime temperature measured in Palma de Mallorca (Spain) in 2008?", "According to the Forbes magazine, Warren Buffet was the 'World's Billionaires #1' in 2008. What do you believe did Warren Buffett own of a net worth in 2008 (in US-\$)?", "During the 2006 FIFA World Cup in Germany, 64 matches were played. What do you believe was the average number of goals per match?", "The Mount Everest Elevation (Nepal/China, Tibet) is the World's highest mountain. What do you believe is the exact elevation of the Mount Everest (in m/ ft)?", "The Euro (€) is one of the most important currencies in the world. What do you believe was the 2007 worldwide share of the Euro in the official foreign exchange reserves?"

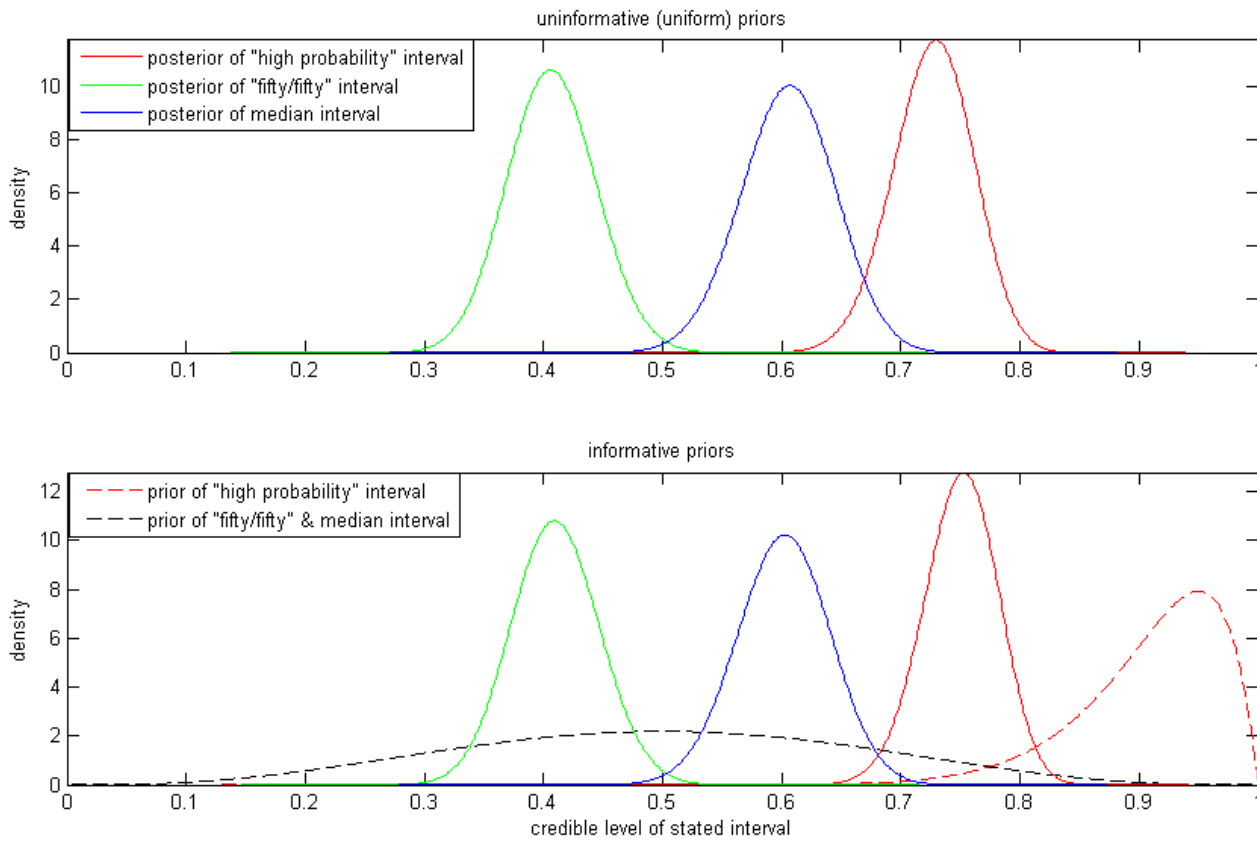


Figure 88: posterior credible level

stereotypical credible interval  $Q$  is

$$\pi(q|x) \propto \text{Bin}\left(\sum_{j=1}^J x_j \mid 10 \cdot J, q\right) \cdot \pi(q).$$

Again, when the risk analyst assumes a beta prior  $Q \sim \text{Beta}(a_0, b_0)$ , the posterior  $Q|x \sim \text{Beta}(a_{post}, b_{post})$  is a conjugate prior where  $a_{post} = a_0 + \sum_{j=1}^J x_j$  and  $b_{post} = b_0 + \sum_{j=1}^J [n - x_j]$ . In table 38, we have summarized the posteriors  $\pi(q|x)$  (see also upper subplot in figure 88) resulting from our empirical analysis. To purely rely on the data, we have assumed uniform priors corresponding to a beta distribution with  $a_0 = b_0 = 1$ .

Our results support the heuristic of overconfidence of subjective statements for the "high probability" (1) and the "fifty/ fifty" interval (2). While the posterior of (1) is far below one, the posterior of (2) is below 0.5. Consequently, the risk analyst should take into consideration that experts offer too narrow intervals. Surprisingly, the posterior of  $Q$  for the median

	$10 \cdot J$	$\sum_{j=1}^J x_j$	$a_{post}$	$b_{post}$	$Mode(Q_{post})$	$Std(Q_{post})$
(1)	220	161	162	60	0.73	0.03
(2)	220	92	93	129	0.42	0.03
(3)	200	117	118	84	0.59	0.04

Table 38: posteriors for credible levels (uniform priors)

interval has most of its probability above 0.5. This is a sign that people tend to overestimate the "true" value.

These empirical results need to be taken with a pinch of salt as they base on the elicitation of almanac questions. However, this is a general problem. Experts are normally elicited on (currently unknowable) future events, while most research is on probability judgments of John Citizens for general knowledge questions. This is why the risk analyst can immediately measure the quality of such statements (Vose 2008, pp. 395). In contrast, experts recognize that the realizations of future events is unknown to everyone and not definitely answerable in advance. As a consequence the risk analyst can expect a lower level of overconfidence for questions on future events (Ronis & Yates 1987, pp. 194). This fact we tried to account for by imposing a prior for the "high probability" interval with a mean close to unity while we have more concentrated the other priors around 0.5 (see table 39 and the lower subplot of figure 88).

	$a_0$	$b_0$	$Mode(Q)$	$Std(Q)$	$a_{post}$	$b_{post}$	$Mode(Q_{post})$	$Std(Q_{post})$
(1)	20	2	0.95	0.06	181	61	0.75	0.03
(2)	4	4	0.5	0.17	96	132	0.42	0.03
(3)	4	4	0.5	0.17	121	87	0.58	0.03

Table 39: posteriors for credible levels (informative priors)

Only the posterior (1) has slightly shifted to unity while the posteriors of the other credible levels are more close to 0.5. However, there is no remarkable difference between the posteriors.

Another curious result of our study is the self-assessment of the interviewees. After they had answered all questions, we asked 17 of them to state their assessments how many of their "high probability", "fifty/fifty" intervals actually cover the "correct" value and in how many cases the true value will lie below or equal the median (see table 40 and figure 89).

Although the interviewees were instructed to state an interval where they are rather certain that the "correct" value will not fall outside, there were several persons that ex post felt

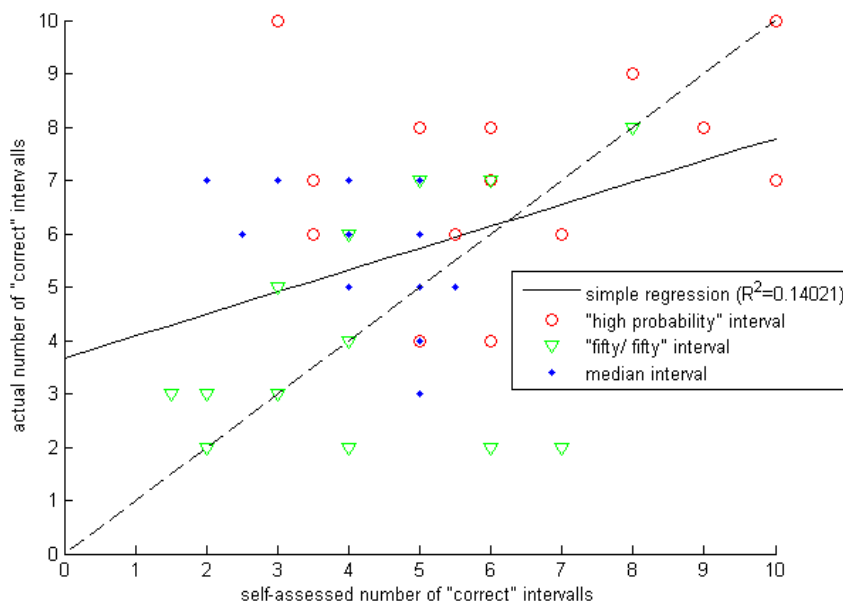


Figure 89: self-assessment

interval	self-assessment	# of covering intervals
(1)	6.32 (2.12)	7.29 (1.76)
(2)	4.44 (1.89)	4.18 (2.10)
(3)	4.35 (1.01)	5.76 (1.20)

mean value (standard deviation)

Table 40: self-assessment results

some overoptimism and feared too narrow intervals. Generally, we optically found no strong relationship between the self-assessment of the experts  $\hat{x}_j$  and their actual performance measured in "correct" answers  $x_j$ .

Additionally, we run (for simplicity) a frequentist least squares estimation of the linear regression  $x_j = \alpha + \beta\hat{x}_j + \varepsilon_j$  where  $E(\varepsilon_j) = 0$ .<sup>258</sup> We estimated  $\hat{\alpha} = 3.675$  and  $\hat{\beta} = 0.411$ .<sup>259</sup> There is some evidence for a positive relationship but it is rather weak (Pearson's  $\rho = 0.374$ ,  $R^2 = 0.140$ ).

<sup>258</sup> Here, we do not distinguish between the different credible intervals.

<sup>259</sup> We do not test the parameters as we want to avoid a distributional assumption for  $\varepsilon_j$ .

## 16.5 Elicitation Techniques for Integrated Risks

In risk management we are interested in describing important target variables  $B$ . E.g.,  $B$  can be a company value depending on  $n$  uncertain risk factors  $Y = (Y_1, \dots, Y_n)'$  (interest rate, consumption, commodity prices).<sup>260</sup> The uncertainty on  $B$  normally results from two sources:

1. A *model* describes the relationship between the quantity  $B$  and the risk factors  $Y$

$$B = h(Y_1, \dots, Y_n) + \varepsilon,$$

where  $h(\cdot)$  is a deterministic (non)-linear function sometimes enhanced by a measurement or misspecification error  $\varepsilon$ .

2. The *joint distribution of the risk factors*

$$P(Y_1 = y_1, \dots, Y_n = y_n) = f_Y(y_1, \dots, y_n),$$

captures the uncertainty on the risk factors and reflects how much information about one factor can be gained from the observation of the others. The joint distribution complicates when  $B$  is a function of risk factors at different point in times.

In the following section we only deal with techniques that allow to approximate the joint distribution in point 2. We start with non-time dependent risk factors, i.e.  $Y_i = Y_{it}$ . In the basic situation of independent risk factors, their joint distribution can be calculated by a factorization of its marginal distributions (see figure 90)

$$P(Y_1 = y_1, \dots, Y_n = y_n) = f_{Y_1}(y_1) \cdot \dots \cdot f_{Y_n}(y_n).$$

Although modelling independence is so straightforward, it can underestimate risk as observed joint extreme events in real life feature a low probability under the independence assumption (see corners in figure 90). Hence, we present techniques allowing to model correlations between different risk factors as well as to handle auto-correlated risk factors.

### 16.5.1 Visual Copula

We start with the presentation of a technique, we call *visual copulas*, that allows to approximate the correlation of two up to three different risk factors. Its basic idea is to confront

---

<sup>260</sup>E.g., Garz et al. 2009 design a model that estimates the discounted cash flows of different portfolios of German power stations.

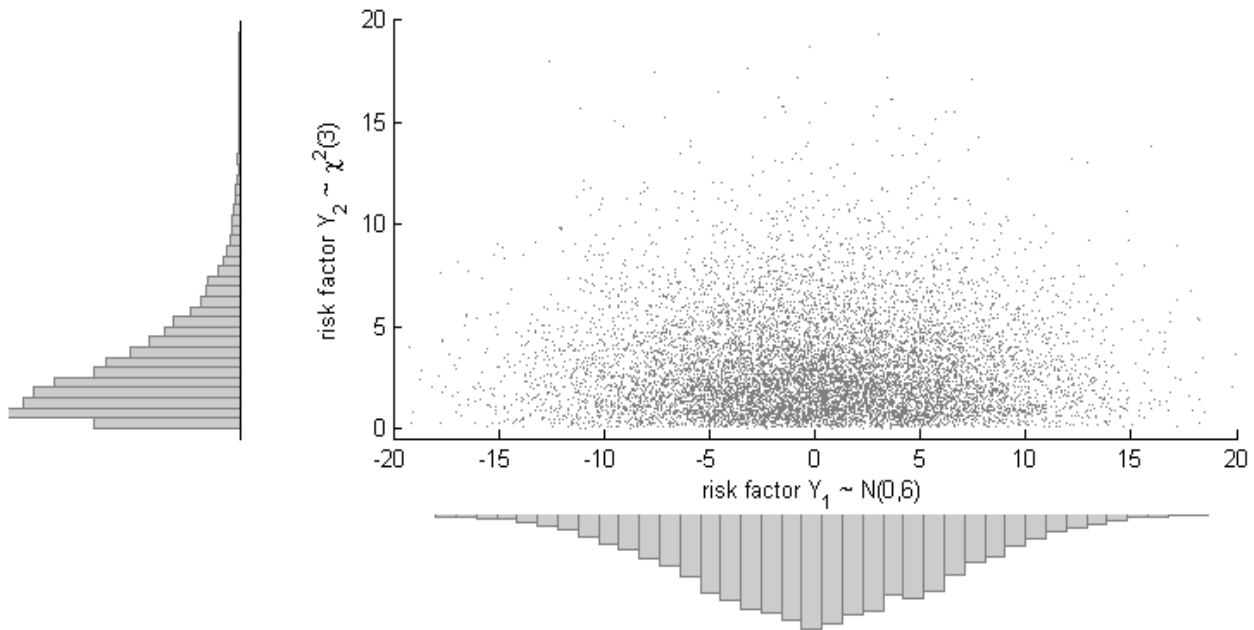


Figure 90: independent risk factors

the expert with scatter plots produced by one or several copulas. In the appendix A.8.1 you can find an introduction to the concept of copulas.

The visual copula technique for two risk factors  $\{Y_1, Y_2\}$  is a three step approach. In a first step, the expert separately quantifies her uncertainty on both risk factors  $Y_1$  and  $Y_2$  in distributional form (see section 15.2 for more details). In the next step, the expert has to choose an adequate correlation structure which is generated by a copula. As the copulas mainly differ in their ability to model tail dependence, the expert should weight the likelihood for joint extreme events of  $\{Y_1, Y_2\}$  (see table 41).<sup>261</sup>

	extreme negative $Y_2$	extreme positive $Y_2$
extreme negative $Y_1$	$w^{--}$	$w^{-+}$
extreme positive $Y_1$	$w^{+-}$	$w^{++}$

Table 41: weighting scheme for the likelihood of joint extreme events

According to the expert’s assessment the risk analyst can chose a suitable copula function to model the joint density of the risk factors (see table 42).<sup>262</sup>

<sup>261</sup> A weight  $w = 0$  would state negligible tail dependence.

<sup>262</sup> See figure 104 in appendix A.8.1 for scatter plots produced by different copulas.

	Gauss	Student's	Clayton*	Gumble*	Frank
$\{Y_1^-, Y_2^-\}$	—	▲	▲	—	—
$\{Y_1^+, Y_2^+\}$	—	▲	—	▲	—
$\{Y_1^-, Y_2^+\}$	—	▼	▼	—	—
$\{Y_1^+, Y_2^-\}$	—	▼	—	▼	—

▲/▼ sampling with positive/ negative (rank) correlation

\* negative dependence only by  $|\tau|$  & sampling  $\{Y_1, 1 - Y_2\}$

Table 42: copulas for modelling of joint extreme events

The risk analyst should prepare several scatter plots (see figure 91) produced by the select copula function and different pre-define levels of rank correlations.<sup>263</sup> The number of presented plots could be reduced by ignoring rank correlations between  $-0.5$  and  $0.5$  apart from zero, as they are hard to recognize in scatter plots. Once the expert has decided for one scatter plot (a weighting would also be possible), Monte Carlo techniques can be used to generate  $W$  correlated pairs  $\{y_1^{(w)}, y_2^{(w)}\}_{w=1, \dots, W}$  of the risk factors (see appendix A.8.2 for the Monte Carlo algorithms).

The visual copula approach is still applicable to model the dependence structure of three risk factors  $Y_1$ ,  $Y_2$ , and  $Y_3$ . For the sake of simplicity, we only use the Gaussian copula which needs an additional step in elicitation to guarantee a positive definite covariance matrix,  $c' \Sigma c > 0$  for  $c \in R^3$ . First, the expert separately expresses her uncertainty on the risk factors. Afterward, scatter plots for the tuples  $\{Y_1, Y_2\}$  and  $\{Y_1, Y_3\}$  are presented to the expert for different levels of rank correlation  $\tau \in [-1, 1]$ . In the third step, the expert can choose between different scatter plots for  $\{Y_2, Y_3\}$ . To guarantee a positive definite covariance matrix, the rank correlation of  $\{Y_2, Y_3\}$  is restricted to  $\tau(Y_2, Y_3) \in [\tau_{\min}, \tau_{\max}]$ .

The boundaries for the third rank correlation can be easily calculated by the following algorithm which bases on a reversed *Cholesky-decomposition* for a three-dimensional covariance matrix of a multivariate Gaussian distribution with zero means and variances equal to one:<sup>264</sup>

1. Set the desired Kendall's rank correlations  $\tau(Y_1, Y_2), \tau(Y_1, Y_3) \in [-1, 1]$  and transform them to Pearson's correlation  $\rho = \sin(\tau \cdot \pi/2)$ .
2. Set the last rank correlation  $\tau(Y_2, Y_3)$  in such a way that Pearson's  $\rho(Y_2, Y_3) \in [\rho_{\min}, \rho_{\max}]$

<sup>263</sup>E.g.,  $\tau(Y_1, Y_2) \in \{-0.85, -0.65, -0.5, -0.25, 0, 0.25, 0.5, 0.65, 0.85\}$ . For the Student's t copula, the risk analyst has additionally to decide in favor of a certain degree of freedom  $v$ .

<sup>264</sup>In this case the covariance matrix  $\Sigma$  and the correlation matrix  $\Gamma$  are equal.



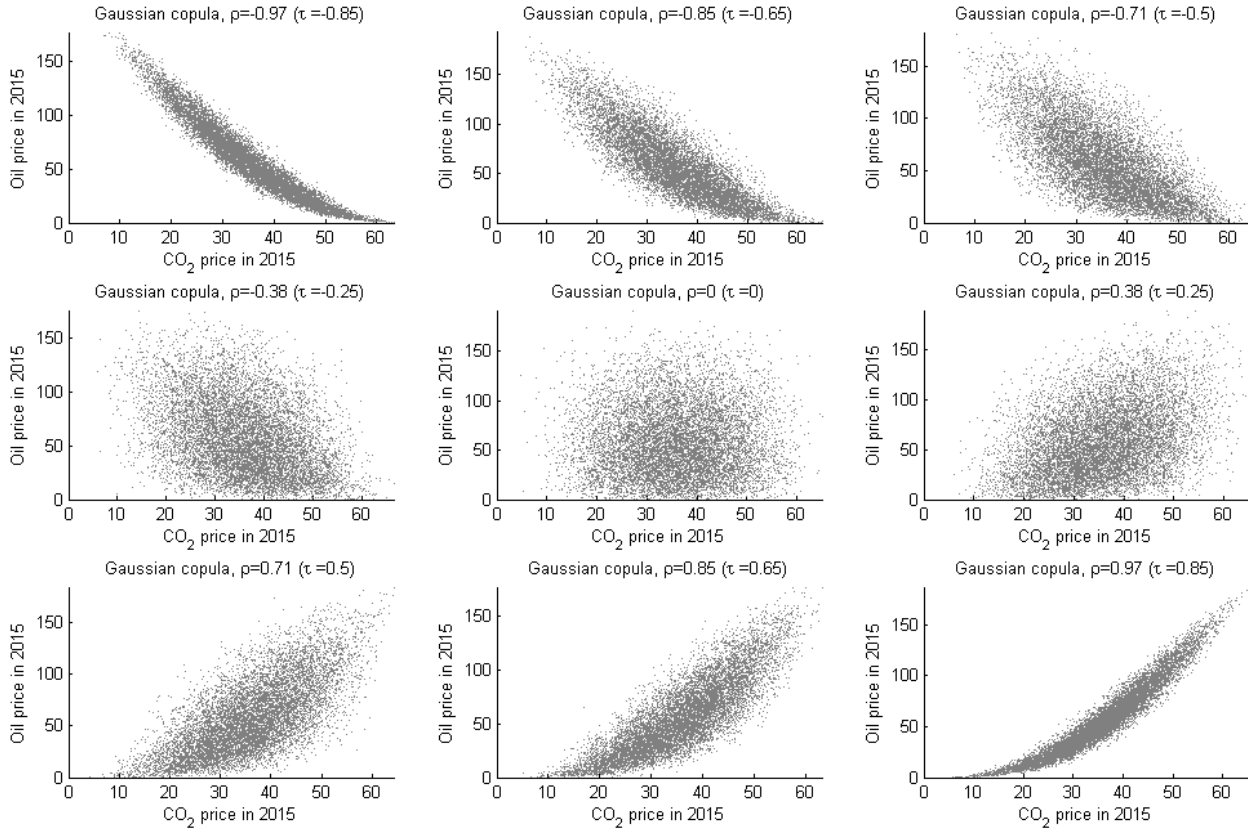


Figure 91: Gaussian copula for different (rank) correlations

where

$$\begin{aligned}
 \begin{pmatrix} \rho_{\max} \\ \rho_{\min} \end{pmatrix} &= \frac{\frac{\Sigma_{12}\Sigma_{13}}{\text{Var}(Z_1)} \pm \sqrt{\left[ \text{Var}(Z_3) - \frac{[\Sigma_{12}]^2}{\text{Var}(Z_1)} \right] \left[ \text{Var}(Z_3) - \frac{[\Sigma_{13}]^2}{\text{Var}(Z_1)} \right]}}{\sqrt{\text{Var}(Z_2) \text{Var}(Z_3)}} \\
 &= \rho_{12}\rho_{13} \pm \sqrt{[1 - [\rho_{12}]^2] [1 - [\rho_{13}]^2]},
 \end{aligned}$$

where  $\text{Var}(Z) = 1$  and  $\rho_{ij} = \rho(Y_i, Y_j)$ .

Unfortunately, visual copulas for higher dimensions ( $n > 3$ ) result in a more laborious procedure to guarantee positive definite covariance matrices. Hence, the risk analyst should not apply the visual copula technique to such situations. Instead, the envelope method or scenario analysis should be preferred.

### 16.5.2 The Envelope Method

Another technique for modeling dependency between a few risk factors  $Y = (Y_1, \dots, Y_n)'$  is the *envelope method* (Vose 2008, pp. 280) which assumes the distributions of conditional risk factors  $Y_{-1}|y_1 = (Y_2|y_1, \dots, Y_n|y_1)'$  as function of a *numeraire* (exogenous risk factor)  $Y_1$ .<sup>265</sup> Hence, this approach runs in two steps.

1. Simulate  $W$  samples  $y_1^{(w)}$  of the numeraire from  $Y_1 \sim Dist(\theta_1)$  with parameter vector  $\theta_1 = (\theta_{11}, \dots, \theta_{1K_1})'$ .
2. Simulate  $W$  samples  $y_i^{(w)}|y_1^{(w)}$  from  $Y_i^{(w)}|y_1^{(w)} \sim Dist(\theta_i^{(w)})$  where  $i = 2, \dots, n$  and the parameter vector  $\theta_i^{(w)} = (g_{i1}(y_1^{(w)}), \dots, g_{iK_i}(y_1^{(w)}))'$  is defined by functions of the numeraire.

Although this concept is rather flexible, the risk analyst should restrict to triangular and modified PERT distribution for the elicitation of  $Y_i|y_1, i = 2, \dots, n$ .<sup>266</sup> For both distributions, the expert is only asked for the (practical) minimum  $a$  and maximum  $c$  as well as the most likely value  $b$  of the conditional factors  $Y_i$  given the numeraire  $Y_1 = y_1$ .<sup>267</sup> This can be done in different ways: (1) The expert draws three lines in a  $Y_1$ - $Y_i$ -plot (see figure 92) or (2) states  $L$  tuples  $\{a_l, b_l, c_l\}_{l=1, \dots, L}$ , given some realizations  $y_1^{(1)} < \dots < y_1^{(L)}$  of the numeraire. The points can easily be connected by straight lines.

As already mentioned (see section 16.3.6), it is possible to ask the expert for the practicable instead of the absolute minimum and maximum ( $\tilde{a}$  and  $\tilde{b}$  vs.  $a$  and  $b$ ) of a risk factor. Unfortunately, the expert's understanding of "practicable" is latent for the risk analyst, and often not explicit to the expert. In such a situation the risk analyst is uncertain about the credible level  $q$  of the credible interval and consequently about the absolute minimum and maximum

$$\int_{\tilde{a}}^{\tilde{c}} Tri(y_i|a, b, c) dy_i = q.$$

In section 16.4, we have analyzed the understanding of people when they are asked for the practicable minimum and maximum and could find a posterior distribution for the credible level  $\pi(q|Elicitation)$  which we can use for the envelope method. Here, we present our approach for the special case of a triangular distribution.

<sup>265</sup>Of course, it is possible to construct a more complex factor model where the distributions of factors  $Y_2, \dots, Y_{\tilde{n}}$  are functions of factor  $Y_1$  while the distributions of the factors  $Y_{\tilde{n}+1}, \dots, Y_n$  are again functions of  $Y_1, \dots, Y_{\tilde{n}}$ .

<sup>266</sup>see section 16.3.6 for more on these distributions

<sup>267</sup>In the case of the modified PERT distribution the shape parameter needs to be determined, too. This could be done by the risk analyst or the expert has to choose between several distributions differing in the shape parameter.

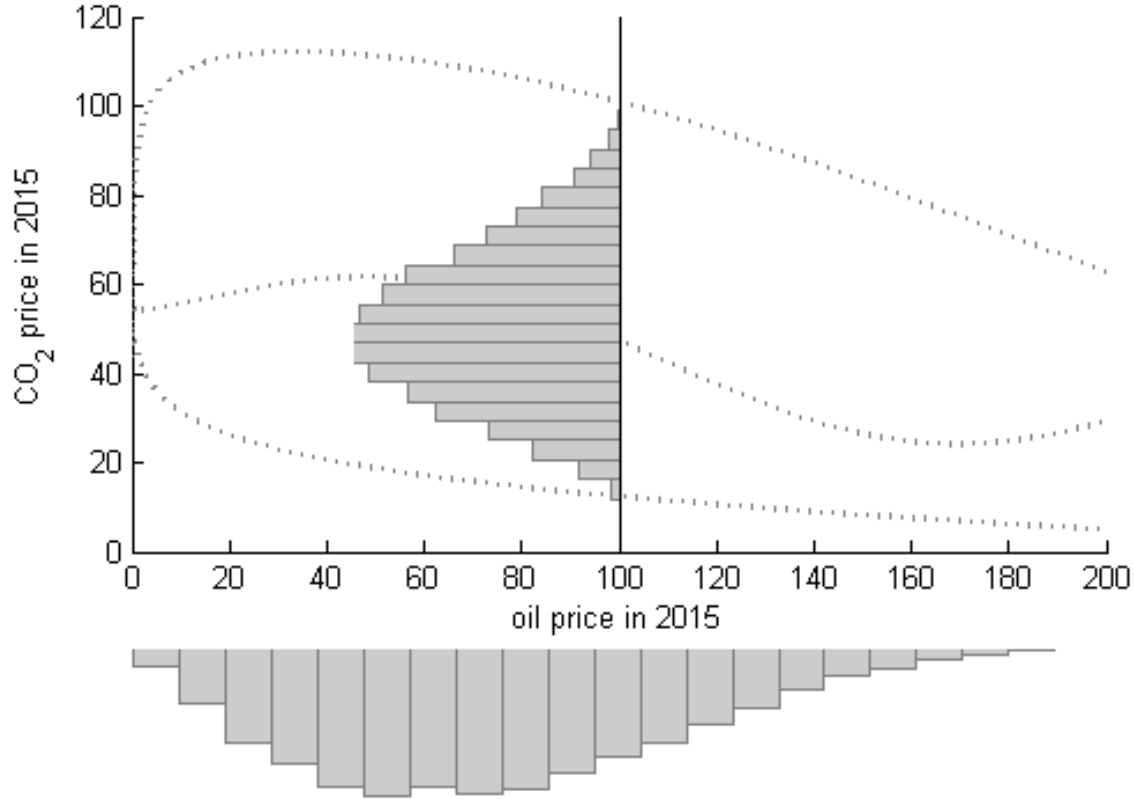


Figure 92: envelope method

1. Simulate  $W$  samples  $y_1^{(w)}$  of the numeraire from  $Y_1 \sim Dist(\theta_1)$  with parameter vector  $\theta_1 = (\theta_{11}, \dots, \theta_{1K_1})'$ .
2. Simulate  $W$  samples  $q^{(w)}$  from a prior or posterior of the credible level  $Q$ .<sup>268</sup>
3. Simulate  $W$  samples  $y_i^{(w)} | y_1^{(w)}$ , for  $i = 2, \dots, n$ , from  $Y_i^{(w)} | y_1^{(w)} \sim Tri(a_i^*, b_i, c_i^*)$  where

$$(a_i^*, c_i^*)' = \arg \min_{a_i, c_i} \left( \left[ \frac{1-q}{2} - F_{Tri}(\tilde{a} | a_i, b_i, c_i) \right]^2 + \left[ \frac{1+q}{2} - F_{Tri}(\tilde{c} | a_i, b_i, c_i) \right]^2 \right),$$

$F_{Tri}(\tilde{a} | a_i, b_i, c_i)$  is the cdf of a triangular distribution,  $a_i = g_{ia}(y_1^{(w)})$ ,  $b_i = g_{ib}(y_1^{(w)})$ , and  $c_i = g_{ic}(y_1^{(w)})$ .

<sup>268</sup>You can find our empirical posterior  $\pi(q | Elicitation) = Beta(q | \alpha_{post}, \beta_{post})$  in section 16.4. Here, we use alpha and beta to avoid confusion with the minimum and most likely value of  $Y$ .

The envelope method can be extended to a situation where a conditional risk factor  $Y_n|y_{-n}$  depends on the  $n-1$  realizations  $y_{-n} = (y_1, \dots, y_{n-1})'$  of the numeraires  $Y_{-n} = (Y_1, \dots, Y_{n-1})'$ . Although it is possible that these numeraires are correlated, e.g. modeled by a copula, we want to keep the presentation straight and simple. Hence, we assume independent numeraires and restrict our presentation to the triangular distribution. Moreover, we disregard uncertainty on the subjective understanding of minimum and maximum and simply ask for the absolute minimum  $a$  and maximum  $b$  as well as the most likely value  $b$ .

Even in such a simplified situation, the challenge is to find a parameter vector  $\theta_n = (a, b, c)' = (g_a(y_{-n}), g_b(y_{-n}), g_c(y_{-n}))'$  that results in a consistent distribution for  $Y_n|y_{-n}$ . Again, a multi-step approach is required.

1. Simulate  $W$  samples of the numeraires  $y_{-n}^{(w)}$  from the distributions  $Y_i \sim Dist(\theta_i)$  where  $i = 1, \dots, n-1$  and  $\theta_i = (\theta_{i1}, \dots, \theta_{iK_i})'$ .
2. The expert needs to state the conditional parameters  $a_i^{(w)} = g_a^{(i)}(y_i^{(w)})$ ,  $b_i^{(w)} = g_b^{(i)}(y_i^{(w)})$ , and  $c_i^{(w)} = g_c^{(i)}(y_i^{(w)})$  for all numeraires  $Y_i$ ,  $i = 1, \dots, n-1$ . This can easily be done by drawing three lines in the respective  $Y_i$ - $Y_{n-1}$ -plots.
3. Optionally, the expert can weight the influence of the different numeraires  $Y_{-n} = (Y_1, \dots, Y_{n-1})'$  on  $Y_n$  by  $w_i \geq 0$ .
4. An aggregation algorithm is needed which produces consistent parameters. We propose  $a^{(w)} = \max(a_1^{(w)}, \dots, a_{n-1}^{(w)})$ ,  $c^{(w)} = \min(c_1^{(w)}, \dots, c_{n-1}^{(w)})$ , and<sup>269</sup>

$$b^{(w)} = \frac{\sum_{i=1}^{n-1} w_i \cdot b_i^{(w)} \cdot \mathbf{1}_{(b_i^{(w)} \in (a, c))}}{\sum_{j=1}^{n-1} w_j \cdot \mathbf{1}_{(b_j^{(w)} \in (a, c))}}.$$

Although people often have an assessment on the range of a risk factor  $Y_n$  given one of the numeraires  $y_i$ , it seems beyond the scope of human beings to state all ranges for  $Y_n$  given all possible combinations of  $y_{-n}$ . Therefore, we believe that the most restrictive minimum and maximum, given the realized  $n-1$  risk factors, should apply. The most likely value  $b$  instead is simply the weighted average of the most likely values. As we do not want to generate a most likely value lower than  $a$  or larger than  $b$  we exclude such  $b_i^{(w)}$  by the means of an indicator function.

Nevertheless, there is the danger that our algorithm ends in a situation where  $c \leq a$ . This might be avoided when  $a$  is a restriction for  $c$ . However, we prefer to accept the risk for

<sup>269</sup>If the expert has not weighted the numeraires they are treated equally important.

two reasons. (1) A situation where  $c \leq a$  reveals an inconsistent quantification of the risk factors' dependence structures. (2) The algorithm above should only be applied in a direct and intensive interaction with the expert which allows for training and modification by trial and error.

Generally, the envelope method can be used to describe dependence between the distribution of a conditional risk factor  $Y_i$  and one up to a few uncertain numeraires  $Y_j$ ,  $i \neq j$ . Unfortunately, this approach does normally not work for a time-series risk factor  $Y_t$  with the deterministic numeraire time  $t$ . Although the envelope method could flexibly model a time-dependent distribution for  $Y_t$ , it neglects auto-correlation of  $Y_t$  on past realizations  $y_{t-\Delta}$ ,  $\Delta = 1, 2, \dots$ . In the following sections, we present some techniques that allow for an approximation of auto-correlated time-series risk factors.

### 16.5.3 Stochastic Processes

A basic approach for modeling auto-correlation of a risk factor  $Y_t$  is to use stochastic processes. Especially basic mean reverting processes seem to be most suitable. They are specified by an intuitive long-term mean reversion level (equilibrium) and feature analytical solutions.

The two most prominent mean-reverting processes are the Ornstein-Uhlenbeck and the Cox-Ingersoll-Ross processes (see sections 10.1.3 and 10.1.4). The Ornstein-Uhlenbeck process (also known as Vasicek process when applied to interest rates) is defined by

$$dY_t = \kappa [\mu - y_t] dt + \sigma dW_t,$$

where  $\kappa$  is the mean reversion rate,<sup>270</sup>  $\mu$  is the long-term mean (or mean reversion level),  $\sigma$  is the volatility, and  $dW_t \sim N(0, \sqrt{dt})$  is the increment of a Wiener process. The analytical solution to this process is

$$Y_{t+\Delta} | y_t \sim N \left( y_t e^{-\kappa\Delta} + \Delta [\mu - \kappa y_t], \sigma \sqrt{[1 - e^{-2\kappa\Delta}] / 2\kappa} \right).$$

Although the Ornstein-Uhlenbeck process features an analytical solution it seems only suitable in expert elicitation when the auto-correlated risk factor  $Y_t$  is allowed to become negative.

Therefore, Cox et al. (1985) extended Vasicek's interest rate version of the Ornstein-Uhlenbeck process by a variable volatility term

$$dY_t = \kappa [\mu - y_t] dt + \sigma \sqrt{y_t} dW_t.$$

---

<sup>270</sup>The mean reversion rate is much easier to understand by the transformation  $\Delta_{1/2} = \ln 2 / \kappa$  termed half-life time. It simply states the time needed to pass half-way from  $y_t$  to the long-term mean  $\mu$ . See section 10.1.3 for more details and a derivation.

The Cox-Ingersoll-Ross process also possesses an analytical solution (Cox et al. 1985, pp. 391)

$$2cY_{t+\Delta}|y_t \sim \chi_{NC}^2(v, \delta),$$

where  $2cY_{t+\Delta}$  is non-central chi-square distribution with  $v = 2q + 2$  degrees of freedom, a non-centrality parameter  $\delta = 2u$ , and

$$c = \frac{2\kappa}{\sigma^2[1 - e^{-\kappa\Delta}]} \ \& \ u = cy_t e^{-\kappa\Delta} \ \& \ q = \frac{2\kappa\mu}{\sigma^2} - 1.$$

The expected value and the variance of  $Y_{t+\Delta}$  equal

$$\begin{aligned} E(Y_{t+\Delta}|y_t) &= y_t e^{-\kappa\Delta} + \mu [1 - e^{-\kappa\Delta}] \\ Var(Y_{t+\Delta}|y_t) &= y_t \frac{\sigma^2}{\kappa} [e^{-\kappa\Delta} - e^{-2\kappa\Delta}] + \mu \frac{\sigma^2}{2\kappa} [1 - e^{-\kappa\Delta}]^2. \end{aligned}$$

In expert elicitation the Ornstein-Uhlenbeck and the Cox-Ingersoll-Ross process can perfectly be calibrated to  $J = 3$  (consistent) expert statements. A calibration to  $J > 3$  assessments generally results in an approximation. The next problem is to find adequate expert statements. A direct elicitation of the long-term mean  $\mu$  seems unproblematic while it is not for  $\kappa$  and  $\sigma$ . Hence, we propose an elicitation for some practicable lower and upper bounds ( $\tilde{a}$  and  $\tilde{c}$ ).<sup>271</sup> Again, there is uncertainty on the expert's understanding of "practicable". The risk analyst does not know the expert's understanding of the credible level  $Q$  that the risk factor  $Y_t$  will realize within the interval  $[\tilde{a}, \tilde{c}]$  in  $t$ . To reflect a rational uncertainty, we can simulate credible levels  $q$  from the posterior  $\pi(q|Elicitation)$  (see sections 16.4 and 16.5.2). Then, a quadratic target function like

$$TF_w(\kappa, \mu, \sigma) = \left[ \frac{1 - q^{(w)}}{2} - F_{Y_{t+\Delta}}(\tilde{a}_\Delta | \kappa, \mu, \sigma) \right]^2 + \left[ \frac{1 + q^{(w)}}{2} - F_{Y_{t+\Delta}}(\tilde{c}_\Delta | \kappa, \mu, \sigma) \right]^2 \quad (20)$$

could be used to minimize  $\sum_{w=1}^W \sum_{\Delta=1}^{J/2} TF_w(\kappa, \mu, \sigma)$  by adequate parameters  $\kappa^*$ ,  $\mu^*$ , and  $\sigma^*$  and  $W$  samples from the prior/posterior of  $Q$ . In equation 20,  $F_{Y_{t+\Delta}}(\cdot | \kappa, \mu, \sigma)$  is the analytical solution of the Ornstein-Uhlenbeck or Cox-Ingersoll-Ross process while  $\tilde{a}_\Delta$  and  $\tilde{c}_\Delta$  are the practicable minimum and maximum of the risk factor  $Y_{t+\Delta}$  in  $t + \Delta$ .

Within our research group, we have discussed the future developments of the average CO2 price within the EU ETS between 2013 and 2020 (after the second trading period). We could reach consensus on (1) a triangular distribution of the average CO2 price in 2013 (in 2009

<sup>271</sup>A direct elicitation for the most likely value  $b$  seems problematic as the Ornstein-Uhlenbeck process produces a symmetrical solution. Although the Cox-Ingersoll-Ross model results in an analytical solution allowing for asymmetry, the non-central chi-square distribution is not able to model all imaginable relations  $\tilde{a} < b < \tilde{c}$ .

prices)<sup>272</sup> and on the practicable lower and upper bounds in 2015, 2017, and 2020 (see table 43).

	2013	2015	2017	2020
$\tilde{a}_\Delta$	10€	15€	20€	30€
$\tilde{c}_\Delta$	70€	100€	125€	150€
$b_\Delta$	45€			

Table 43: research group assessments on future average CO2 prices

Based on these statements and on our uncertainty of the latent credible level  $Q$ , we have applied the following algorithm

1. Simulate  $W_1$  times the credible level  $q^{(w_1)}$  from the prior or posterior of  $Q$ .
2. Find the practicable minimum  $\tilde{a}_{2013}^{(w_1)}$  and maximum  $\tilde{c}_{2013}^{(w_1)}$  given  $q^{(w_1)}$ . For each  $w_1 = 1, \dots, W_1$ , sample  $W_2$  times the CO2 price  $y_{2013}^{(w_1, w_2)}$  given  $\tilde{a}_{2013}^{(w_1)}$ ,  $b_{2013}$ , and  $\tilde{c}_{2013}^{(w_1)}$ .
3. Find adequate parameters  $[\kappa^*]^{(w_1)}$ ,  $[\mu^*]^{(w_1)}$ , and  $[\sigma^*]^{(w_1)}$  minimizing the target function in equation 20 given  $q^{(w_1)}$  as well as  $\tilde{a}_t^{(w_1)}$  and  $\tilde{c}_t^{(w_1)}$ , for  $t = 2013, 2015, 2017, 2020$ . For each  $w_1 = 1, \dots, W_1$ , sample  $W_2$  times the CO2 price paths  $\left\{ y_t^{(w_1, w_2)} \right\}_{t=2014, 2015, \dots, 2020}$  given  $y_{2013}^{(w_1, w_2)}$ ,  $[\kappa^*]^{(w_1)}$ ,  $[\mu^*]^{(w_1)}$ , and  $[\sigma^*]^{(w_1)}$ .<sup>273</sup>

In upper subplot of figure 93 you can find  $W = W_1 \cdot W_2 = 10,000$  simulated paths of the CO2 price and our assessments on the practicable minimum and maximum as well as the most likely value of 2013. Additionally, we compare the histograms of the CO2 price paths to our assessments (lower subplots). We calculate the quality of the approximation by the mean absolute and square error,  $MAE = 0.058909$  and  $MSE = 0.048175$ .<sup>274</sup> The MAE is

<sup>272</sup>Our approach is a little bit cumbersome as there are two EU ETS markets on CO2 - second and third (post 2012) trading period. Unfortunately, there is not yet a future market for emission rights of the third period. Hence, we start our modelling in 2013.

<sup>273</sup>For numerical reasons we approximate the non-central chi-square distribution by the Gaussian distribution.

<sup>274</sup>Here, we calculate the *mean absolute* and *square error* (MAE and MSE) by

$$MAE = \sum_{w_1=1}^{W_1} \sum_{t \in \mathcal{T}} \frac{|\alpha_{w_1 t}| + |\beta_{w_1 t}|}{W_1 2n_{\mathcal{T}}} \quad \& \quad MSE = \sum_{w_1=1}^{W_1} \sum_{t \in \mathcal{T}} \sqrt{\frac{\alpha_{w_1 t}^2 + \beta_{w_1 t}^2}{W_1 2n_{\mathcal{T}}}}$$

where

$$\alpha_{w_1 t} = \frac{\sum_{w_2=1}^{W_2} 1_{(y_t^{(w_1, w_2)} < \tilde{a}_t)}}{W_2} - \frac{1 - q^{(w_1)}}{2} \quad \& \quad \beta_{w_1 t} = \frac{\sum_{w_2=1}^{W_2} 1_{(y_t^{(w_1, w_2)} > \tilde{c}_t)}}{W_2} - \frac{1 + q^{(w_1)}}{2}.$$

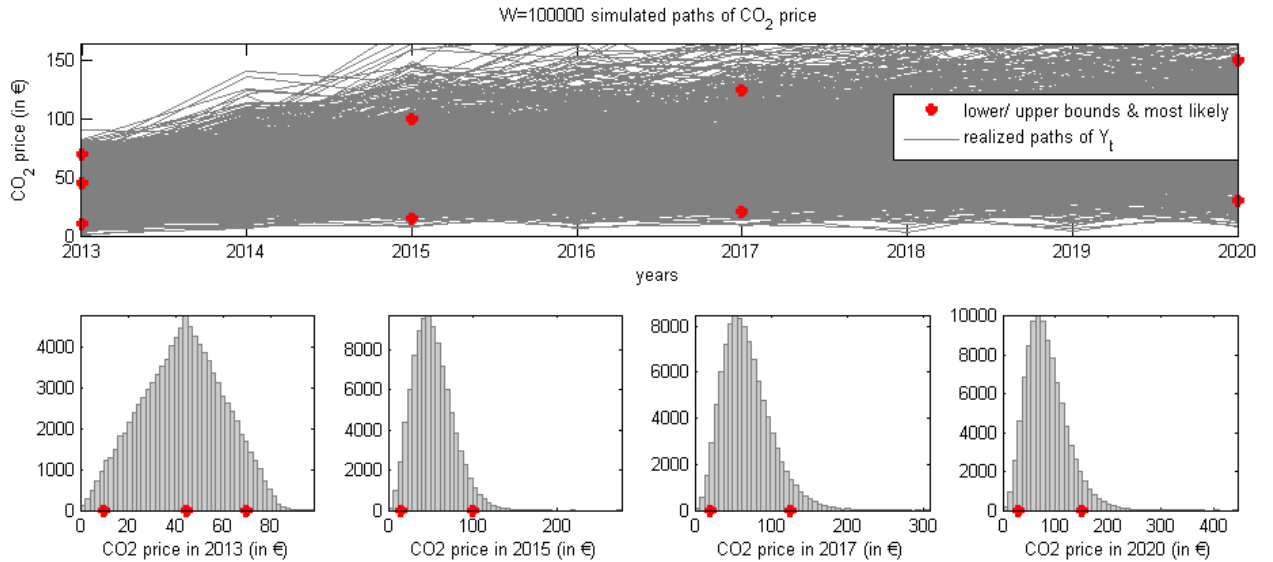


Figure 93: stochastic processes for expert elicitation

simply to interpret. It states the mean deviation between the quantiles ( $\tilde{a}_\Delta$  and  $\tilde{c}_\Delta$  should equal  $[1 - q]/2$  and  $[1 + q]/2$ ) of our assessments and the sampled ones.

#### 16.5.4 Time-Series Copula

Calibrating stochastic processes to reflect expert statements is not a rather flexible approach as stochastic processes are mostly defined by few parameters. Therefore, an approach is needed which can reproduce large fluctuations in the structure of risk factor  $Y_t$  over the years. The envelope method can easily produce samples  $y_t$  from a distribution with time-dependent parameters. Even switching between the distributions can easily be done. However, the envelope method is not able to produce auto-correlation in the time-paths. Hence, we propose a combination of the envelope and copula method we call *time-series copula* which is able to reproduce time-variant expert statements on  $Y_t$  as well as auto-correlation. In our modification the auto-correlation is restricted to positive values which seem adequate for most auto-correlated risk factors. The time-series concept consists of three steps: (1) time-dependent envelope method, (2) definition of auto-correlation structure, and (3) simultaneous sampling of  $Y = (Y_1, \dots, Y_T)'$  via a copula.

---

$W_1$  and  $W_2$  are the number of times we sample a credible level  $q^{(w_1)}$  and a path given  $q^{(w_1)}$ . This means we generate  $W = W_1 \cdot W_2$  sample paths.  $\mathcal{T} = (2013, 2015, 2017, 2020)'$  - hence  $n_{\mathcal{T}} = 4$  - are the years we offer statements on practicable minimum  $\tilde{a}_t$  and maximum  $\tilde{c}_t$  of the CO2 price.





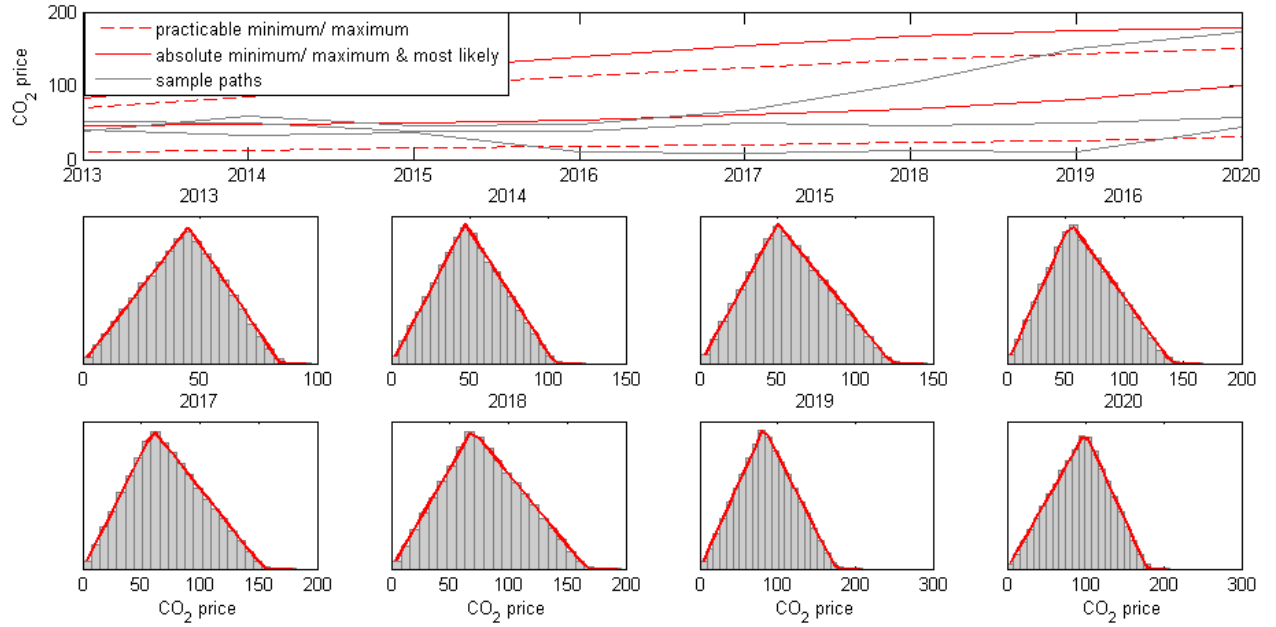


Figure 94: time-series copula

number of samples, the empirical Kendall's rank correlation matrix converges to the defined one and the marginal distributions are sampled exactly according to the expert statements. In the following, we apply the time-series copula to the our research group assessments on the future CO<sub>2</sub> price (see table 43). In contrast to the stochastic process approach, we calibrate a triangular distribution to our assessment. As the time-series copula allows for an arbitrary number of statements, we join the different statements for the practicable minimum  $\tilde{a}$  and maximum  $\tilde{c}$  as well as the most likely value  $b$  by an adequate polynomial. As we need the absolute minimum  $a$  and maximum  $c$  we minimize a quadratic target function defined on the practicable minimum and maximum (see equation 20). Again, we are uncertain about our understanding of "practicable". Hence, we use the posterior  $\pi(q|Elicitation)$  on the credible level to generate  $W$  samples  $q^{(w)}$ . After some discussion, we have decided to set  $\lambda = 0.5$  resulting in a relatively fast decreasing auto-correlation.<sup>275</sup> In the upper subplot of figure 94, you can find some exemplarily sampled CO<sub>2</sub> price paths featuring the desired auto-correlation. As expected, the copula approach allows to exactly sample from the marginal

<sup>275</sup>The auto-correlations depends on the absolute time difference (in years) between the CO<sub>2</sub> prices:

	$\Delta = 0$	$\Delta = 1$	$\Delta = 2$	$\Delta = 3$	$\Delta = 4$	$\Delta = 5$	$\Delta = 6$	$\Delta = 7$
$\tau_{\Delta}$	1	0.607	0.368	0.223	0.135	0.082	0.050	0.030

distributions of  $Y_t$ , for  $t = 2013, \dots, 2020$  (lower subplots in figure 94).<sup>276</sup> We measure the quality of the time-series approach by the mean absolute and square error (see p. 195) which are  $MAE = 0.032339$  and  $MSE = 0.036418$ . The time-series copula approach has a better approximation than the calibration of stochastic processes. However, this is not the full story. The time-series copula approach is by far more flexible as it could even approximate many more expert statements and even more complex structures (e.g. bounds could be  $W$ -shaped).

## 16.6 Mathematical Expert Aggregation

Armstrong (2001) comes to provoking conclusions in a meta-analysis:

1. Organizations should not concentrate to find the best expert but a diversity of assessments.
2. When an organization has access to several experts it should not bring them together in group meeting as the independence of the forecasts vanishes.<sup>277</sup>

In section 15.5, we have generally overviewed the scope of approaches to cope with a diversity of opinions. Especially behavioral tools like panel discussions, focus groups, or prediction markets allow for an elegant aggregation (see sections 17 and 18). Unfortunately, such tools can fail to aggregate because experts could not agree on a group assessment or the prediction market lacks in liquidity. In such a situation, the diversity of opinions could simply be documented or a mathematical expert aggregation technique could concentrate the information. In the following, we introduce mathematical expert aggregation techniques which can be subdivided in weighting techniques and Bayesian expert aggregation. Although we use the phrase "expert", we also mean the output of statistics as well as any kind of model.

### 16.6.1 Expert Weighting

Expert weighting is a relatively simple axiomatic approach of aggregating different expert assessments.<sup>278</sup> Here, we concentrate on the dominant linear and logarithmic opinion pool (Clemen & Winkler 1999, pp. 190). The *linear opinion pool* is a (weighted) arithmetic average of  $J$  different expert assessments on a risk factor  $Y$ . We do not distinguish whether

---

<sup>276</sup>The differences between the histograms (grey) and the theoretical marginal distributions (red) originate from the uncertainty on  $Q$ . We plot the theoretical marginal distributions for the mean posterior credible level  $E(Q)$  while the samples  $y_t^{(w)}$  base on different realizations of  $Q$ .

<sup>277</sup>Huffcutt & Woehr (1999) empirically found job interviews with a panel of interviewers less accurate to select the best job applicant than interviews with a single interviewer.

<sup>278</sup>A more general term is assessments weighting when conflicting assessments only result from one person.

the  $J$  statements originate from  $J$  experts (incl. the risk analyst) or whether some experts state several assessments (e.g. problem of imprecise probabilities).

In the case of continuous expert statements, e.g. in form of calibrated elicitation distributions (see section 16.3.6), the linear opinion pool is defined by<sup>279</sup>

$$\pi(y) = \sum_{j=1}^J w_j \pi_j(y),$$

where  $\pi_j(y)$  is the calibrated pdf of expert  $j$  and  $\sum_{j=1}^J w_j = 1$ .

There are some nice properties of the linear opinion pool: (1) The average distribution is zero for  $Y = y$  which were excluded by all experts and has positive probability for  $Y = y$  when at least one expert holds such an outcome probable.<sup>280</sup> (2) The pool opinion exactly equals each expert's uncertainty when all experts state the same assessment (unanimity property). (3) The linear opinion pool is the only combination scheme which satisfies the marginalization property (Clemen & Winkler 1999, p. 189). It simply means that it is irrelevant whether the risk analyst weights the expert's marginal distributions  $\pi(y_i) = \sum_{j=1}^J w_j \pi_j(y_i)$  of risk factor  $Y_i$  or whether she first weights the joint probability  $\pi(y) = \sum_{j=1}^J w_j \pi_j(y)$  of all  $n$  risk factors  $Y = (Y_1, \dots, Y_n)'$  and then calculates the marginal  $\pi(y_i) = \int_{y_{-i}} \pi(y) dy_{-i}$ , where  $y_{-i}$  is  $y = (y_1, \dots, y_n)'$  without  $y_i$ .

Although Ranjan & Gneiting (2008) theoretically show that the linear opinion pool is not calibrated even when calibrated experts are weighted, we believe in the linear opinion pool as an excellent tool for two reasons. First, Ranjan & Gneiting (2008) propose a beta transformed linear opinion pool which can only be applied on repeatable events. Second, and most important, the linear opinion pool has empirically proved to be a good decision rule which can fundamentally reduce miscalibration, at least for repeatable events (e.g. Sanders 1963; Clemen & Winkler 1986; Vislocky & Fritsch 1995; Graham 1996).

Another axiomatic approach is the *logarithmic opinion pool*<sup>281</sup>

$$\pi(y) \propto \prod_{j=1}^J \pi_j(y)^{w_j}$$

which equals the geometric mean for  $w_j = 1/J$ . The logarithmic opinion pool features the external Bayesianity property (Clemen & Winkler 1999, p. 190) which means that there is no

<sup>279</sup>If the expert statements are in discrete form, the linear opinion pool is  $\pi_k = \sum_{j=1}^J w_j \pi_{jk}$ , where  $\pi_{jk}$  is the (modified) probability assessment of expert  $j$  on interval  $k \in \{1, \dots, K\}$  and  $\sum_{j=1}^J w_j = \sum_{k=1}^K \pi_{jk} = 1$ .

<sup>280</sup>In contrast, the Bayesian aggregation via copulas (see section 16.6.2) results in zero probability for outcomes of  $Y$  at least one expert excludes.

<sup>281</sup>In the discrete case, it is  $\pi_k \propto \prod_{j=1}^J \pi_{jk}^{w_j}$  for intervals  $k = 1, \dots, K$ .

difference whether we update each expert distribution with new information and then weight them or whether we first weight them and then update the aggregated expert distribution. When deciding for an axiomatic approach it is important to keep in mind that no rule satisfies all desirable features (Clemen & Winkler 1999, p. 191). Nevertheless, the most important fact for weighting is not the technique but the expert weights  $w_j$  that need to be determined carefully (see discussion in section 15.5).

### 16.6.2 Bayesian Expert Aggregation

A different mathematical expert aggregation technique - academic attention has shifted to - is *Bayesian expert aggregation* which updates the decision maker's (or risk analyst's) prior uncertainty  $\pi(y)$  on risk factor  $Y$  by  $J$  expert statements  $s = (s_1, \dots, s_J)'$  assumed as data observations. The expert statements can originate from individuals or groups as well as from statistics or any kind of model. Based on priors and expert statements, the decision maker receives an updated posterior (Clemen & Winkler 1999, p. 190)

$$\pi(y | s_1, \dots, s_J) \propto f_S(s_1, \dots, s_J | y) \cdot \pi(y) = L_s(y) \cdot \pi(y),$$

where the risk factor  $Y$  is treated like a Bayesian parameter and  $f_S(s_1, \dots, s_J | y)$  is any distribution reflecting the decision makers uncertainty on the statements.

For simplicity the prior distribution is often modeled without information (improper prior) and possible prior information of the decision maker is implemented by an additional expert statement  $s_{J+1}$ .<sup>282</sup> The likelihood function  $L_s(y)$  is required to account for the precision and the bias of the expert statements. In the case of distributional statements, precision measures the expert's uncertainty while bias is understood as level of calibration (Clemen & Winkler 1999, p. 190). In the case of point estimates precision and bias can be understood the way presented in table 44.<sup>283</sup>

Winkler (1981) proposed a Bayesian aggregation algorithm which allows updating from  $J$  experts expressing their uncertainty on  $Y$  by a Gaussian distribution with mean  $m_j$  and standard deviation  $\sigma_j$ . In the basic setting the assessments of the experts are reduced to the mean values  $m = (m_1, \dots, m_J)'$  as point estimates of  $Y$  while the standard deviation statements  $\sigma$  are disregarded. The basic idea is to understand the point estimate  $m_j$  as a simple realization of the expert's uncertainty on the mean  $M_j$ . Consequently, the expert's "true" estimation error  $\varepsilon_j^* = y^* - M_j$  is random where  $y^*$  is the final outcome of risk factor  $Y$ . Additionally, the experts are assumed to be unbiased, i.e.  $E(\varepsilon_j^*) = 0$ . As the

---

<sup>282</sup>The treatment of the decision maker as expert has a huge advantage. A basic prior of the decision maker is conditionally independent of her uncertainty on the expert statements. In contrast, in the formulation above, the uncertainty of the decision maker is correlated with those of the experts.

<sup>283</sup>In the case of repeatable risk factors precision can be labeled as repeatability or reproducibility.

	low bias	large bias
low precision	- estimates close to "true" value - repeated estimates relatively heterogenous	- estimates far away from "true" value - repeated estimates relatively heterogenous
high precision	- estimates close to "true" value - repeated estimates close to each other	- estimates far away from "true" value - repeated estimates close to each other

Table 44: precision &amp; bias of an expert assessment

decision maker is uncertain on  $Y$  we cannot model  $\varepsilon_j^*$  directly but the conditional expert error  $\varepsilon_j = [\varepsilon_j | y] = y - M_j$  instead. Then, the resulting posterior is

$$\pi(y | m_1, \dots, m_J) \propto f_\varepsilon(\varepsilon_1, \dots, \varepsilon_J | y) \cdot \pi(y) = L_\varepsilon(y) \cdot \pi(y),$$

where  $L_\varepsilon(y) = f_\varepsilon(\varepsilon_1, \dots, \varepsilon_J | y)$  models the *joint* conditional expert errors. For simplicity, the errors  $\varepsilon \sim N\left((0, \dots, 0)', \Sigma\right)$  are assumed Gaussian.<sup>284</sup>

A speciality of Bayesian aggregation is the consideration of dependent expert statements which can result in a simultaneous bias in the same direction (positive correlation) and means a lower "net-number" of experts we can use for updating. The covariance matrix  $\Sigma$  allows to calibrate the model to the decision maker's uncertainty on the expert statements and their correlation.

The posterior of the Winkler model (see Winkler 1981, p. 483, or Clemen & Winkler 1985, p. 430)

$$Y | \mu_1, \dots, \mu_J \sim N\left(\frac{\sigma_0^{-2}\mu_0 + e'\Sigma^{-1}m}{\sigma_0^{-2} + e'\Sigma^{-1}e}, [\sigma_0^{-2} + e'\Sigma^{-1}e]^{-1/2}\right), \text{ for } e = (1, \dots, 1)',$$

is a conjugate prior for a Gaussian prior  $Y \sim N(\mu_0, \sigma_0)$ .

The covariance matrix  $\Sigma$  can "freely" be chosen to express confidence in the experts.<sup>285</sup> Winkler (1981), p. 484, simply sets the main diagonal to  $\sigma_1, \dots, \sigma_J$ . Nevertheless, it is hard to state a positive definite covariance matrix reflecting the decision makers assessment on the dependencies between each of the  $J[J-1]/2$  pairs of experts. Clemen & Reilly (1999), pp. 213, empirically analyze different methods to elicit correlation levels. They find a combination of different approaches seems more promising. E.g., expert could be asked for correlations and subsequently scatter plots could visualize these statements.

<sup>284</sup>Lindley (1983) presents a Student's t distribution approach.

<sup>285</sup>In Winkler (1981), pp. 485, you can find an extension of this approach, when the covariance matrix is not known.

Clemen & Winkler (1986), p. 45, propagate the assumption of exchangeability for experts when they are from the same "league".<sup>286</sup> Consequently, a simplification is to assume equal covariances. Then the resulting correlation is restricted to  $\rho \in (-1/[J-1], 1]$ . The more experts we use for updating the less negative correlation is possible. However, it seems that experts' forecast errors  $\varepsilon$  are mostly positively correlated because of interconnections, same sets of information, and equal structures of models (Clemen 1989). Especially, in the situations when experts ride a bubble, we should expect positive correlations.

All things considered, the (Gaussian) Bayesian aggregation approach of Winkler (1981) is remarkable as it accounts for dependent information sources of different experts. It makes aware that expert elicitation should not simply be an acclamation and doubling the number of experts will not always result in a doubling of forecasting quality. In the Gaussian model increasing positive correlation between experts (forecasting errors) results in a different Gaussian posterior, as the "net-information" available for updating reduces. Assuming a homogeneous expert uncertainty  $\sigma$ , Clemen & Winkler (1985), pp. 430, present the *equivalent number of independent experts* which transforms the number of  $J$  dependent experts to the number of independent experts  $N^*$  resulting in the same posterior uncertainty. Under normality, the joint standard deviation for independent expert statements  $S_j \sim N(\mu_j, \sigma)$  is  $\sigma/\sqrt{J}$ . Assume the decision maker consults  $J$  independent experts while her prior information equals  $J_0$  expert statements. Then, the posterior uncertainty of the decision maker is  $\sigma_{post} = [J_0/\sigma^2 + J/\sigma^2]^{-1/2} = \sigma [J_0 + J]^{-1/2}$ , for  $\sigma_0 = \sigma/\sqrt{J_0}$ . The equivalent number of independent experts

$$J^* = \sigma^2 e' \Sigma^{-1} e$$

calculates by comparing  $\sigma_{post}$  to  $[\sigma_0^{-2} + e' \Sigma^{-1} e]^{-1/2}$  from the Gaussian model (for correlated expert statements) (see Clemen & Winkler 1985, p. 431).

Nevertheless, the Gaussian approach has several shortcomings (Clemen & Winkler 1999, p. 192): (1) Expert errors need to be adjusted to the unimodal and symmetrical Gaussian distribution, (2) the posterior is always Gaussian which (3) may put positive probability on realizations of  $Y$  which are excluded by all experts. These arguments might be better understood with some examples. In figure 95, the Bayesian aggregation is shown for the case of weighted and unweighted experts. In the upper subplot you can find the different (Gaussian) expert statements and the decision maker's posteriors for different levels of assumed expert error correlations if experts are *equally* weighted. The higher the assumed correlation the more uncertain (broader) becomes the posterior reflecting a decreasing level of independent information. The case of different weighted experts is more complicated and shown in the lower subplot. The experts are weighted according to their variance statements. A lower

<sup>286</sup>In the case of different types of experts (e.g. econometric models, judgmental forecasts, extrapolation methods) exchangeability should hold for subsets of experts.

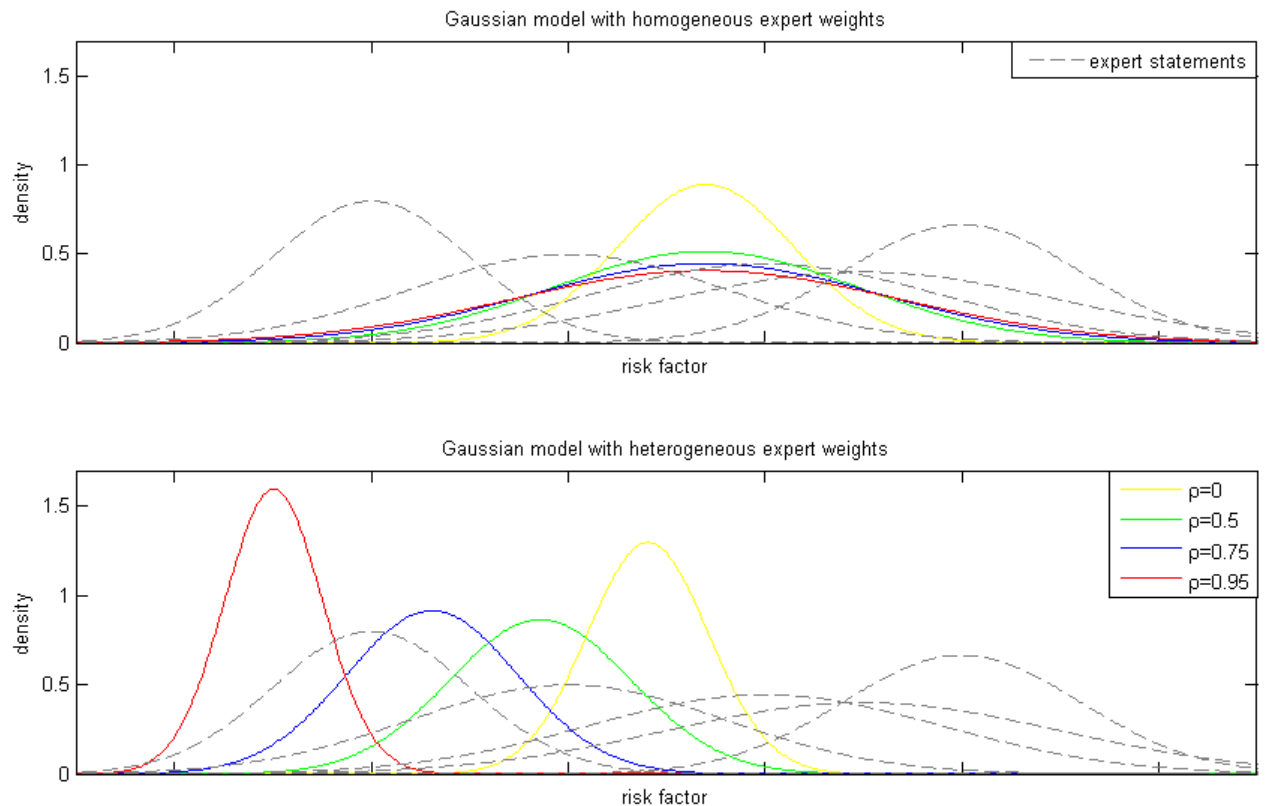


Figure 95: Bayesian expert aggregation (Gaussian model)

variance reflects a higher level of reliance. When the expert correlation rises, the Gaussian model assumes a higher level of common information. Different abilities in interpretation of the common information result in different expert statements. As we prefer experts with good interpretation skills, reflected by a low variance statement, the posterior tends to the expert statement with the lowest variance.

When the expert correlation approaches unity, a simultaneous bias of all experts should not be ruled out. However, the Gaussian model is not able to fully reflect this idea because its posterior is unimodal. Consequently, it is not possible to simultaneously model fears of an over- and underestimation. In the lower subplot it can be seen that the Gaussian model tries to highly weight the expert with the lowest variance statement and to account for a simultaneous bias of all experts. As a consequence, the Gaussian model only offers the risk of a simultaneous overestimation. A simultaneous underestimation would assume that the expert with the lowest variance statement would have the largest prediction error.

A remarkable transition to a more flexible Bayesian aggregation technique offers the cop-



ula proposal of Jouini & Clemen (1996) which allows for arbitrary marginal distributions although a homogeneous dependence structure is assumed between the experts. In detail, the decision maker's posterior on  $Y$  is

$$\begin{aligned}\pi(y|s_1, \dots, s_J) &\propto f_S(s_1, \dots, s_J|y) \cdot \pi(y) \\ &= c(F_{S_1}(s_1|y), \dots, F_{S_J}(s_J|y)|\vartheta) \cdot \prod_{j=1}^J f_{S_j}(s_j|y) \cdot \pi(y).\end{aligned}$$

The copula approach allows the decision maker to separately specify her uncertainty on the  $J$  expert statements  $s = (s_1, \dots, s_J)'$  by  $f_{S_j}(s_j|y)$  and on their correlations where

$$c(u_1, \dots, u_J|\vartheta) = \frac{\partial C(u_1, \dots, u_J|\vartheta)}{\partial u_1, \dots, \partial u_J}$$

is the copula density with respective parameter vector  $\vartheta$  (see appendix A.8.1).

Jouini & Clemen (1996), pp. 452, propose the following specification of  $f_{S_j}(s_j|y)$  and  $c(u_1, \dots, u_J|\vartheta)$ : Assume the decision maker and  $J$  experts state their uncertainty on  $Y$  by  $\pi_j(y)$ , where  $j = 1, \dots, J$  and  $m_j$  are the respective medians of these distributions. Although in real life, the full uncertainty of the experts, expressed by  $\pi_j(y)$ , might be correlated, the approach is only manageable if correlation is restricted to some key features. Jouini & Clemen (1996), pp. 452, assume that the median estimate  $m_j$  is only a realization of the expert's uncertainty on the median  $M_j$ . Hence, there is a random conditional expert error  $\varepsilon_j = [\varepsilon_j|y] = y - M_j$ . An appropriate specification of  $\varepsilon_j$  is  $g(\varepsilon_j) = \pi_j(\varepsilon_j + m_j)$  which has the same shape like the expert statement  $\pi_j(y)$  but a median at zero. As Jouini & Clemen (1996) presume  $M_j$  to have a median equal to  $y$  ( $M_j$  is unbiased) its conditional marginal can be described by  $f_{M_j}(\tilde{m}_j|y) = g(y - \tilde{m}_j) = \pi_j(y - \tilde{m}_j + m_j)$ . The respective cdf is simply  $F_{M_j}(\tilde{m}_j|y) = 1 - F_{\pi_j}(y - \tilde{m}_j + m_j)$ . Here  $\pi_j(\cdot)$  and  $F_{\pi_j}(\cdot)$  are the expert's uncertainty on  $Y$  while  $f_{M_j}(\tilde{m}_j|y)$  and  $F_{M_j}(\tilde{m}_j|y)$  reflect the conditional uncertainty of the decision maker on the expert statements. This results in the following posterior

$$\pi(y|m_1, \dots, m_{J+1}) \propto \prod_{j=1}^{J+1} f_{M_j}(m_j|y) \cdot c(F_{M_1}(m_1|y), \dots, F_{M_{J+1}}(m_{J+1}|y)|\theta),$$

where Jouini & Clemen (1996) assume a Frank copula and an uninformative prior. The  $J$  experts and the decision maker state their uncertainty by  $J + 1$  priors  $\pi_j(y)$  which are transformed to  $J + 1$  distributions  $f_{M_j}(\tilde{m}_j|y)$  with respective median statements  $\tilde{m} = m = (m_1, \dots, m_{J+1})'$ .

A Frank copula seems reasonable in this context as it produces a remarkable posterior  $\pi(y|\cdot)$  for the risk factor  $Y$ . In the case of a forecast error correlation close to unity there is a good

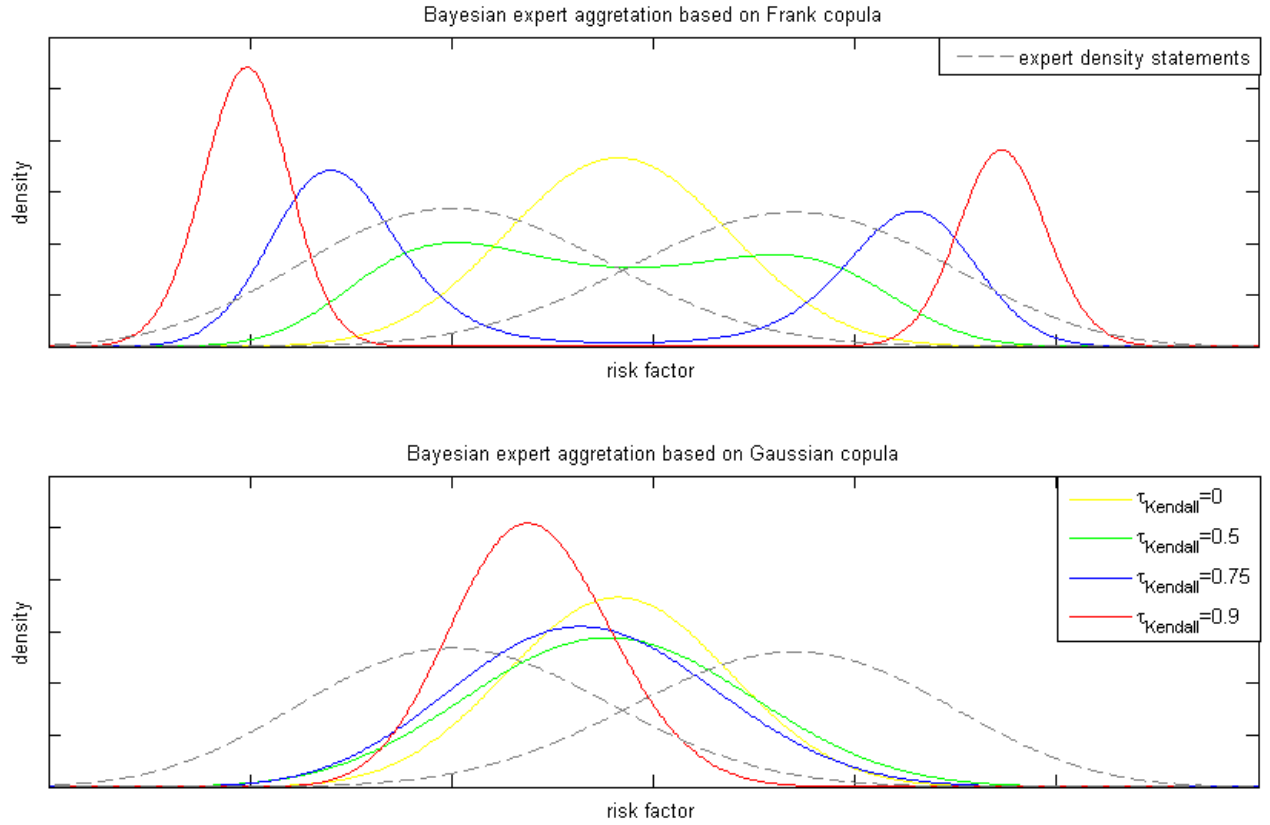


Figure 96: Bayesian expert aggregation (copula approach)

chance that all experts have simultaneously under- or overestimated the "true" outcome of  $Y$ . A Frank copula accounts for this fact (see upper subplot in figure 96). Unfortunately, a Frank copula becomes hard to set up for  $J > 2$  experts. A multi-step approach could be used where coupled subgroups of experts are again coupled by a Frank copula.<sup>287</sup> An alternative is the Gaussian copula with a density

$$c(F_{M_1}(m_1|y), \dots, F_{M_{J+1}}(m_{J+1}|y)|\theta) = |\Gamma|^{-1/2} \exp\left(-\frac{1}{2}z'[\Gamma^{-1} - I]z\right),$$

where  $\Gamma$  is the correlation matrix and  $z = (\Phi^{-1}(F_{M_1}(m_1|y)), \dots, \Phi^{-1}(F_{M_{J+1}}(m_{J+1}|y)))'$ . This approach allows easily to aggregate a nearly unrestricted number of expert distributions but does generally not result in a bimodal posterior (see lower subplot in figure 96).

<sup>287</sup>In the case of a large number of expert statements (= observations), the copula density  $c(\cdot|\vartheta)$  could be estimated (see e.g. Autin et al. 2010).

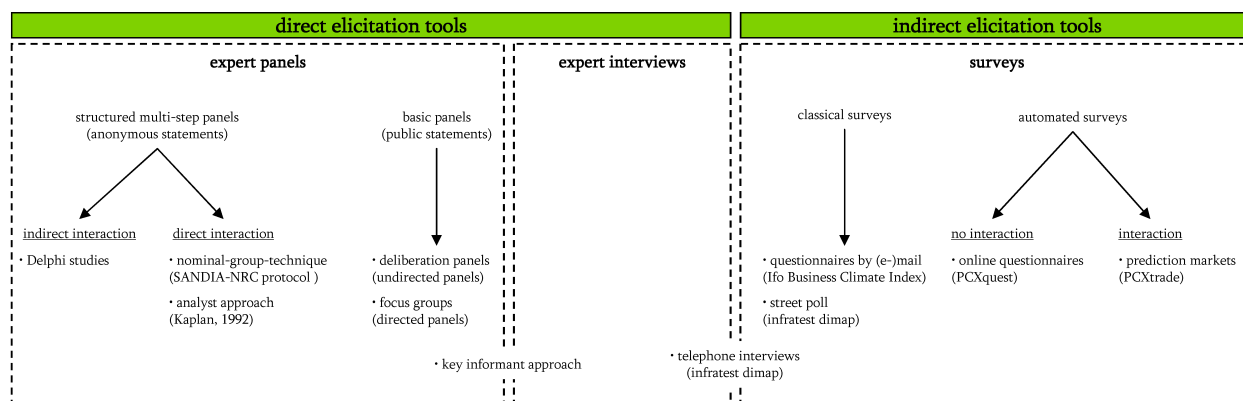


Figure 97: expert elicitation tools

## 17 Expert Elicitation Tools

In the section above, we have presented techniques that help to quantify and manage competing expert assessments. However, this also requires to expose experts to those techniques. This is done by *expert elicitation tools* that should be compared according to their quality of information elicitation, information exchange, and information aggregation (Bruggen et al. 2006, pp. 7):

- *Information elicitation* is the process of extracting the individual assessment of an expert who should not be influenced by the risk analyst. The fact whether the expert was right or wrong is not a criterion for the quality of the elicitation tool.
- *Information exchange* quality is determined by the possibility for the experts to exchange their knowledge and to benefit as well as to learn from others' knowledge.
- *Information aggregation* is the mechanism to aggregate the knowledge dispersed over the experts.

In the following, we present different elicitation tools. We basically distinguish between direct elicitation tools (expert interviews and expert panels) where there is a direct contact between risk analyst and experts and indirect elicitation tools (surveys) where there is mostly no direct contact and experts can often decide when and where to state their assessments (see figure 97).

## 17.1 Direct Elicitation Tools

### 17.1.1 Expert Interviews

The most dominant tool in expert elicitation is the *expert interview* which can have different levels of complexity. In the simplest version, the risk analyst informally interviews the expert via a phone call or during lunch. Normally, this only results in some point estimates and uncertainty statements. To elicit more elaborated expert assessments, the risk analyst needs to prepare an in-depth guideline with all questions. Sometimes, the risk analyst needs to customize the guideline because of the expert statements during the interview. Generally, the risk analyst conducts the interview but should avoid to force her point.

The advantage of expert interviews is the simple integration of different elicitation techniques (e.g. calibrate a distribution and weight these intervals) and its flexibility to personalize the elicitation. Moreover, within a short time, it is possible to gather lots of different expert statements which a well-versed risk analyst can relatively easily appraise. Unfortunately, professional and sophisticated expert interviews are extremely time-consuming in preparation, realization, and reporting.<sup>288</sup> Furthermore, there is no interaction between the experts. This might be favorable because there are no flaws like cascades and group pressure (see section 17.1.2). However, this can also be unfavorable as soon as there emerge new points of view. Then the risk analyst has to decide whether she ignores that point in the following interviews to guarantee comparability. An alternative to expert interviews are expert panels which normally allow for an interaction between the experts.

### 17.1.2 Expert Panels

(Multi-)Expert panels gather persons supposed to possess superior knowledge about some risk factors or uncertain factors. A border case of an expert panel is the *key informant approach* (KIA) which is widespread in practice because of its simplicity (Bruggen et al. 2006, p. 3). The panel only consists of one expert, called *key informant*. The key informant is picked as representative because of her knowledge and willingness to communicate. The KIA is accompanied by problems like individual biases, random errors, and the inability to aggregate information spread over different experts. In fact, the key informant approach can also be interpreted as expert interview which theoretically requires no second opinion.<sup>289</sup>

(Multi-)Expert panels are behavioral aggregation approaches for combining judgemental forecasts of different experts (Bruggen et al. 2006, p. 4). There are panels with direct and

---

<sup>288</sup>A prime example is Krause (2010) who interviewed 15 German automotive experts (representatives from car manufacturers, investors, non-governmental organizations, associations, and science) to collect and compare their assessments of CO2 emission reduction options for the German automotive industry.

<sup>289</sup>In real life, the renaming to key informant approach is often just a smokescreen to avoid more interviews.

indirect interactions between experts. The aggregation mechanism is normally performed by a simple or (knowledge/ competence/ confidence) weighted average when experts do not interact or reach no consensus.

The basic form of such expert panels are undirected *deliberation panels* (or *group discussions*) which allow for an in-depth elicitation. Although deliberation panels bear no risk of manipulation and bubbles, they have the potential to fail because of social pressure, groupthink,<sup>290</sup> and informational influence. Panels tend to amplify rather to correct individual errors. Additionally, they usually reduce variance in the assessments when panel members come to one group statement. In the end, they tend to be more confident with their judgement, even when the group is totally wrong and group polarizations has taken place (Sunstein 2006, pp. 70 & pp. 82). Moreover, deliberations often suffer from *hidden profiles*, a consequence from the *common-knowledge effect* (Sunstein 2006, p. 75): Deliberations do not perform well in disclosing information held by only few group members. Cascades are another problem where group members are influenced by others and neglect their own knowledge (Sunstein 2006, pp. 78):

- *informational cascades*: Single group members, uncertain with their assessment, join the predominant group opinion increasing its predominance. Hence, more group members could be convinced and join, too.
- *reputational cascades*: Group members are convinced that they are right. But they do not express their own opinion as they do not want to contradict the predominant group opinion. The more group members do so the more group pressure builds up.

There are different approaches to avoid or reduce such problems. On the one hand there is the idea of a directed deliberation group, called *focus groups*.<sup>291</sup> The moderator (e.g. the risk analyst) of a focus group should be impartial without formulating own assessments. Her task is rather to guarantee every group member to get a word in edgeways and has to prevent the discussion to wander from the subject. Nevertheless, it is hard to dispel group pressure. Alternatively to focus groups, there have been proposed *structured (multi-step) expert panels* which can substantially reduce the danger of social pressure and cascades because experts do not need to express their assessments in front of other experts. The widespread *Delphi studies*, developed by the RAND Corporation in the 1950ies, let all experts separately reveal their assessments which are anonymously presented to the whole group. This is repeated until the group reaches consensus. The disadvantages of Delphi methods are the equal weighting of

---

<sup>290</sup> Janis (1972) shaped the term *groupthink* to describe consensual false conclusion caused by peer-group influences in the context of U.S. foreign policy.

<sup>291</sup> see Kitzinger (1995) or Morgan (1996) for more details

even less informed experts and logistical problems which do not allow for a frequent updating (Bruggen et al. 2006, p. 4).

Another class of structured expert panels bases on the *nominal-group-technique* (first proposed by Delbecq & van de Ven 1971) which allows for a direct contact of experts but tries to reduce the need of experts to openly commit to their assessments. It guarantees a relatively equal integration of all experts as all assessments are considered (van de Ven & Delbecq 1974). Normally, all experts separately state their assessments (in written form) which are anonymously presented to the group and subsequently discussed. Finally, experts need to separately rank or weight these assessments. A special version of such panels for the elicitation of physical risks is defined in the SANDIA-NRC protocol,<sup>292</sup> developed by the US SANDIA National Laboratories and the US Nuclear Regulatory Commission (NRC). It advises to gather experts to panels twice (Hora & Jensen 2002, pp. 2):<sup>293</sup> In a first meeting, the experts freely discuss the issue and approaches to find critical factors. Additionally, experts are trained in probability elicitation<sup>294</sup> to give them some hints about potential biases in elicitation. In a second meeting, experts present their models and data they use for quantification. Then, they are separately asked to state their assessments. The results support the expert panel to form a probability distribution.

Another rather distinct type of expert panels could be labeled as *analyst approach*. Kaplan (1992) propagates panels where experts present their competing arguments and evidence to a non-expert risk analyst. Based on this information, the risk analyst solely determines the group assessment. This guarantees that the panel always produces an aggregated assessment. This approach seems extremely demanding. It is important that the risk analyst explicitly identifies such facts all experts agree on as well as those facts there is some disagreement. Additionally, she should enable all experts to state all their arguments. However, the most important challenge for the (non-expert) risk analyst is to separate evidence from good presentation.

## 17.2 Indirect Elicitation Tools

Direct elicitation tools, presented above, are mostly extremely expensive and time consuming for experts and the risk analyst. The availability of experts and the exact timing of a direct elicitation procedures seems however to be the main restriction.

From the perspective of a researcher, in house experts can be motivated if they research on

---

<sup>292</sup>Slight variations are allowed like inviting experts of one or several fields (Hora & Jensen 2002, p. 2).

<sup>293</sup>see Morgan & Keith (2008) for a similar approach

<sup>294</sup>To assess the experts' ability to formulate probabilities, experts are queried by almanac questions about quantities outside their area of expertise. The project leader knows the quantities as the questions are from the encyclopedia.

a complementary topic or there is a good personal relationship. Normally, external experts can only be accessed by strategic cooperations (with other (research) institutions, private business, and politics) which are mostly initialized by long-time personal relationships. In contrast, medium and large scale businesses can motivate experts much more easily. A clever intern charging system to clear activities of departments makes in house expertise accessible to risk analysts. External expertise can be retrieved from a better use of memberships in associations, e.g. Federation of German Industries (BDI), or political pressure groups. Moreover, scientists are highly interested in cooperations with the private sector because this is a general claim from public financiers and students.

However, if the risk analyst is not able to raise the resources for a direct elicitation or cannot manage an adequate timing, *indirect elicitation tools* are an alternative which avoid to overstrain the expert's willingness of cooperation. They mostly allow experts to decide when and where to state their assessments but cannot produce such informative elicitation results like direct tools. Generally, experts are asked for short statements (e.g. by means of fixed or variable interval techniques) on some risk factors which normally requires a transmitter between the risk analyst and the experts.

The most widespread indirect elicitation tools are *classical surveys*. Questionnaires on the risk factors can be distributed in form a hardcopy by mail or attached to e-mails. There are two special versions of classical surveys: telephone interviews and street polls. Traditional telephone interviews can be conducted by the risk analyst herself or by several assistants. In the first case, a telephone interview is an expert interview we have already discussed in section 17.1.1. Nevertheless, we prefer to understand telephone interviews as surveys when the number of interviews is large (maybe more than 20). The term "survey" highlights the overall picture of the assessments of a broad range of experts. Generally, telephone interviews should only be conducted by assistants when the number of experts is large and they are no authority but rather John Q. Publics. The same applies to street pools where assistants arbitrarily (or based on some basic criteria) address passers-by.

Another class of indirect elicitation tools are *automated elicitation tools* which offer an interactive automated transmitter that has normally marginal maintenance costs after it has been set up. We favor web-based tools like prediction markets and online questionnaires.

In prediction markets,<sup>295</sup> experts can trade on the outcome of risk factors for several minutes up to months or years (fixed interval technique). Online questionnaires are web-based versions of traditional questionnaires (fixed and variable interval technique) that are often more flexible and allow for a visualization of the expert statements.

The most crucial difference between both concepts is the interaction of the experts. While

---

<sup>295</sup> An alternative term for prediction market is *online market*. However, this can result in confusions with the concept of online based market platforms where subcontractors compete for orders.

online questionnaires exclude any information exchange between them, prediction markets fundamentally rely on the dynamic interaction of experts. Hence, questionnaires can even be run with a single expert while markets need at least two experts. Consequently, a questionnaire seems preferable when the risk analyst fears some negative interactions between the experts (bubbles) while prediction markets are suitable when interaction is hoped to improve elicitation.<sup>296</sup> Additionally, online questionnaires are far less time consuming than prediction markets which are again relatively low time consuming compared to direct elicitation tools like extensive expert interviews or expert panels.

Within our research project, we have set up the internet page *Potsdam Climate Exchange* (PCX) hosting our online questionnaire platform PCXquest and our prediction market platform PCXtrade. In the following we shortly present the PCXquest. The concept of prediction markets is more complex and beyond the scope of a subsection. Hence, section 18 exclusively deals with it and our prediction market platform PCXquest.

### Online Questionnaire PCXquest

In our perspective, an online questionnaire should allow elicitation of experts who are somewhere in the world and are not willing to invest too much time. According to these constraints we have developed the online questionnaire platform PCXquest hosting several questionnaires consisting of an arbitrary number of questions.

On the main page of the PCXquest the name and the runtime of the questionnaire is displayed. After clicking on the name the questionnaire starts with an introduction text. Subsequently, the questions are presented one by one. Each question belongs to one of six basic types:

- *question type I (most likely realization)*: Questions of type I simply offer up to 15 intervals of an uncertain variable. The expert is allowed to chose one interval, only. Of course, this type can also be applied for questions beside probability elicitation (e.g. Yes/No-questions).
- *question type II (weighting)*: Question type II is our workhorse in elicitation. The risk analyst can break an uncertain variable in up to ten intervals. The expert can weight intervals by up to 30 weight levels predefined by the risk analyst. It is possible to visualize the expert statements by a bar chart.
- *question type IIIa (allocate probabilities)*: Question type IIIa requires the expert to spread up to 100 percentage points on up to 15 intervals. An additional correction field guarantees that statements always sum up to 100%. Similar to question type II a

---

<sup>296</sup>Indeed, the interaction of experts in a prediction market is indirect - via the pricing mechanism. This is close to the interaction in Delphi studies where there is no direct contact between the experts, too.



bar chart can visualize the probabilities. Because of its design, we should only prefer type IIIa to type II if an interval of the uncertain variable can be understood as "rest interval" (e.g. "CO2 prices > 200€") and an explicit probability statement is needed.

- *question type IVa (variable intervals)*: Question type IVa is designed to ask for up to ten characteristics (mostly minimum, maximum, and most likely value) of an elicitation distribution like the triangular and modified PERT distribution.
- *question type IIIb (allocate shares)*: In fact, question type IIIb has the same structure as type IIIa. Again, the experts are asked to allocate exactly 100 percentage points. However, IIIb asks for the shares of different elements of a domain instead of probabilities. E.g., if we ask for the German energy mix in 2020, the statements are not probabilities but shares.
- *question type IVb (verbal statements)*: Question type IVb allows for short statements on up to 10 different questions. Similar to question type IIIb this is no probability elicitation but can sometimes be useful.

After the last question, experts are asked for their professional background. As the PCXquest is freely accessible via internet and we did not implement a login procedure - to reduce the time burden for the experts - we have developed the concept of *expert codes*. On the background page we have created a field where experts can enter a predefined code. This allows us to match the statements and the experts.

We are convinced that the PCXquest is one of the elicitation tools which asks experts for the lowest level of cooperation. Moreover, creation of additional questionnaires is easily done by the risk analyst and the expert statements are stored in a text/ Excel file. Consequently, the PCXquest, or a similar platform, can easily be implemented in the day-to-day business to quantify uncertainty of in house and external experts. Online questionnaires could be run on a regular basis (monthly, quarterly, etc.) for a constant group of in house experts (one person of each key department) and external experts to monitor crucial changes in important risk factors.

## 18 Prediction Markets

The *Hayek Hypothesis* (Hayek 1945) assumes the market mechanism is excellent in collecting and aggregating diverse information which is relevant for the value of commodities and assets. This was formalized by the *rational expectation theory* of Muth (1961). The three central facts of the Hayek Hypothesis are (Berlemann 2003, pp. 176):

- Competitive markets lead to Pareto efficient allocations where all possible gains are realized.
- Market participants behave as if they possess all available information even when it is initially asymmetrically distributed.
- The market mechanism offers incentives that market participants seek an efficient amount of new information.

According to the *efficient markets hypothesis* (EMH) all information is reflected in prices of *efficient* markets (Fama 1970) which holds irrespective whether the traders are fully or partly informed (Dahan et al. 2007, p. 10). This is shown in classical rational expectation models (e.g. Grossman 1981; Plott & Sunder 1982) when traders are assumed to update their beliefs by observing market prices and quantities (Dahan et al. 2007, pp. 9). Plott & Sunder (1982) presume that the market price in the equilibrium may reflect even more information than the sum of the individual information of the traders. This is called *crystal ball hypothesis*.

These findings have actually been observed for *financial* and *betting markets*<sup>297</sup> showing a good prediction ability.<sup>298</sup> Therefore, a new class of *prediction markets*,<sup>299</sup> that are conducted on *prediction market platforms*,<sup>300</sup> has been developed. Prediction markets are strongly related to the idea of measuring market expectations. The payoffs in such prediction markets solely depend on the realization of an uncertain risk factor (e.g. CO2 price in 2013) or the outcome of the elements of an uncertain future event (e.g. election results of parties or concepts of products<sup>301</sup> in marketing).<sup>302</sup>

<sup>297</sup>Very often stock, bond, currency, and commodities markets are called *speculative markets* as they allow traders to bet on future prices. A special case of speculative markets are *betting markets* that trade securities only designed for traders to bet on particular matters of fact, like horse races or sport events (Hanson Summer 2007, p. 5).

<sup>298</sup>The current financial crisis (since 2007/8) has challenged these claims. However, one of the most prominent Cassandras - Robert Shiller famous for the Case-Shiller index - sees the problems in a bad regulation not in the markets themselves. Indeed, Robert Shiller even believes that markets on all possible risk factors can be the solution to the crisis (see section 18.1 as well as Shiller 2008 and Shiller 1993).

<sup>299</sup>According to Tziralis & Tatsiopoulos (2007), who analyzed 152 articles dealing with prediction markets, there are five highly frequent names for prediction markets: (1) *information markets*, (2) *prediction markets*, (3) *electronic (stock) markets*, (4) *virtual (stock) markets*, and (5) *decision markets*. Beside them, names are used like *idea futures markets* (see Hanson 1992; Passmore & Cebeci 2003; Hahn & Tetlock 2006), *forecasting markets* (see Berlemann et al. 2005) and *artificial markets* (see Pennock et al. 2001).

<sup>300</sup>In this paper the name *prediction market platform* is used to distinguish a prediction market on a special topic from its platform. E.g., PCXtrade is a platform for several prediction markets on climate related topics.

<sup>301</sup>Dahan et al. (2007) call this *concept trading*.

<sup>302</sup>see section 18.3 for more on winner-takes-all and index markets

There is no universally accepted terminology and definition of prediction markets. Berg et al. (2003), p. 1, state that prediction markets "... are designed and conducted for the primary purpose of aggregating information so that market prices forecast future events. These markets differ from typical, naturally occurring markets in their primary role as a forecasting tool instead of a resource allocation mechanism". Ottaviani & Sørensen (2007), p. 555, interpret prediction markets as "... simple financial markets that are created with the purpose of collecting information, but serve no liquidity purposes".

Of course, prediction markets do not fully equal real financial markets which is rather an advantage (Berlemann 2008):

- Although there is a large number of real financial markets, a market on the topic of interest (1) does often not exist at all, (2) does not exist in the respective region, or (3) does not exist for the respective time or maturity. In contrast, a prediction market can easily be tailored to a specific topic of interest. Berlemann et al. (2005) conducted several prediction markets on the Bulgarian inflation and US dollar - Bulgarian leva exchange rate to get additional prediction instruments as Bulgarian financial markets were underdeveloped and regular surveys were not available. Leigh et al. (2003) analyzed "Saddam Securities" from September 2002 to February 2003 at TradeSports that only paid if Saddam Hussein is ousted. They used this prediction market as a proxy for a "war in Iraq" market and achieved the results that a 10% rise in probability of war increased the oil price by about \$1 and the S&P 500 by 1.5%.
- Even when a suitable real financial market exists, very strong assumptions on a stochastic process including the market's risk premium is needed to estimate the market expectation. In contrast, prediction markets offer a full probability distribution since different *pure/ Arrow-Debreu securities*<sup>303</sup> are traded.
- Trading at financial markets often relies on an assessment of the expected future stock price. Traders in prediction markets, instead, are explicitly asked to quantify their uncertainty on the respective risk factor. Consequently, prediction markets could deliver an extension to common future markets especially in context of the current financial crisis. While future markets offer traders (e.g. farmers) a possibility for hedging against bad weather conditions, a prediction market could deliver likelihoods for the weather in addition (Hahn & Tetlock 2006, p. 274).

For the purpose of elicitation, prediction markets do a good job for several reasons (Ostrover 2005, p. 12): *participation*: Everyone with access to relevant information can contribute

---

<sup>303</sup> Arrow-Debreu securities pay exactly one unit of money when a certain outcome of an event realizes and otherwise nothing.

to *public mass (prediction) markets*<sup>304</sup> (1) wherever they want, because participation in prediction markets is not restricted on a special place, and (2) whenever they want, because trading is possible 24 hours per day, 7 days per week during the life-span of the market. *motivation*: The reward mechanism sets incentives for informants to reveal their information. *anonymity*: The fear of reprisal for revealing unpopular beliefs does not exist. *interaction*: The pricing mechanism lets traders interact. *computation*: The market prices can easily be observed and employed for forecasts. *costs*: After a prediction market platform has been set up, new prediction markets are nearly without any additionally costs for the platform operator.

A *no hierarchies* argument is stated by Meirowitz & Tucker (2004), p. 334: Even in large organizations with very complex hierarchies, where bottom-up information has to pass several supervisors, prediction market prices can directly be observed by the top management. Moreover, prediction markets have inherited the feature of *real-time* predictions from real financial markets when the liquidity is sufficient. Traders are allowed to buy or sell shares immediately when they get new information. In contrast, online questionnaires or questionnaires sent by mail or e-mail do not reveal changing assessments of the interviewees during their life-spans. There is the risk that important information becomes public during the questionnaire life-span which results in a bias between the late and early answers.

## 18.1 From Past up to Now

The first forerunners of today's prediction markets were betting markets on the U.S. presidential elections starting nearly 150 years ago. In organized markets, contracts were traded that paid off a fixed dollar amount in the case of the winning of a certain candidate. A 5% commission was common. The amounts invested in the betting markets were enormous. The maximum was reached in 1916 with 165 million \$ (in 2002 \$) which was twice the total spending on the election campaign (Rhode & Strumpf 2004, p. 128).

Rhode & Strumpf (2004) analyzed US presidential elections between 1868 and 1940 and found very successful betting markets with a remarkable forecasting performance although the information via media was relatively sparse compared to today.<sup>305</sup> In one case, only, the favored candidate - one month before election - was not the winner. Even state-specific forecasts were quite accurate although there were no scientific polls before the mid of the 1930ies (Rhode & Strumpf 2004, p. 129). From 1916 on, the largest non-scientific survey

---

<sup>304</sup>E.g., the market platform PCXtrade hosts public mass markets on climate change related topics. Everybody, who is interested in, can trade in the mass markets via internet. Beside them the PCXtrade also runs *expert (prediction) markets* which restrict access to select groups of experts.

<sup>305</sup>However, Rhode & Strumpf (2004), p 128, report that the New York Times nearly published quotes on daily basis during the elections 1896, 1900, 1904, 1916, and 1924.

was the *Literary Digest* poll which sent millions of postcard ballots via mail. In 1916 the *Literary Digest* failed to forecast the landslide victory of F. Roosevelt while the betting markets did. Nevertheless, they disappeared in the late 1930ies - not because of low forecast accuracy but other forms of gambling<sup>306</sup> attracted a wide range of people and scientific polls were introduced.

Hanson (1992), calling the prediction market stocks "idea futures", offered the first introductory article to the concept of prediction markets. A few years before the first application of such a modern prediction market was conducted on the 1988 U.S. presidential elections (Forsythe et al. 1992). From that time on especially so called *political stock markets* have shown good results compared to traditional political forecasts (Berlemann & Schmidt 2001) which might explain the nearly uncountable number of efforts to use prediction markets on other topics.<sup>307</sup>

Although Hanson (1992) was one of the first who proposed prediction markets as a social institution that creates visible consensus on topics other than pure election markets, the research was focused on political stock markets in the 1990ies. Up to 1998 the number of publications in the area of prediction markets was low with all together less than 20. From 1999 on the number of publications sharply rose to 14 in 2002, 22 in 2004 and 34 in the first 8 months of 2006 (Tziralis & Tatsiopoulos 2007). The majority of prediction market publications (72%) deal with applications of prediction markets, 22% focused on descriptions, and 18% on theoretical research (Tziralis & Tatsiopoulos 2007).

In 2004 the concept of prediction markets were popularized by James Surowiecki's popular science book "The Wisdom of Crowds" (Surowiecki 2004) as one part of a *wise-crowd vision* that also incorporates Wikipedia, MySpace, and open source software. This is related to the *cheap-crowd vision* highlighting the possibility to substitute expensive experts by the cheaper crowd.

There are two additional concepts in the focus of academic discussion on prediction markets: decision markets and macro markets. Hanson (Oct. 2007), p. 75, defines prediction markets as *decision markets* designed to directly inform a decision maker (e.g. market operator) who can influence the market's underlying. E.g., a prediction market on the ratification of the Kyoto protocol by the USA is a decision market for the U.S. government while a market on the number of hurricanes is not.

In decision markets traders could be discouraged to reveal their information honestly since the decision maker is expected to change the settings for the underlying (*intervention problem*). When the market prices indicate to an adverse outcome the decision maker can change

---

<sup>306</sup>New York legalized pari-mutuel markets on horse races in 1939 (Rhode & Strumpf 2004, p 139).

<sup>307</sup>E.g., Berlemann (2008) concludes that it might be useful to extend political market methodologies to monetary policy instruments.

the underlying. The traders detecting the undesirable development are not rewarded for their information as subsequently the stock prices on the adverse outcome will fall. Therefore, the traders will incorporate this danger and should not be willing to fully reveal their knowledge and beliefs (Ledyard 2006, p. 46).

A special version of decision markets are terrorist markets where a subgroup of criminal traders can influence the underlying instead of the decision maker. The *Policy Analysis Market* (PAM) platform, part of a program called *FutureMap*, was feared to be such a terrorist market. The PAM was planned to forecast military and political instability around the world that affects US interests. Because of practical reasons PAM should be focused on eight nations in the Mideast and five parameters: military activity, political instability, economic growth, U.S. military activity, and U.S. financial involvement. Forecasts were planned for each quarter of the following two years (Hanson July 2006, pp. 259). In addition, PAM traders should give predictions on the U.S. GDP, world trade, total U.S. military casualties, and total western terrorist casualties. Finally, there was planned to offer traders the possibility to propose some miscellaneous markets (Hanson Oct. 2007, p. 77).<sup>308</sup> The FutureMap project including the PAM was cancelled by the US Senate in 2003 after really bad press coverage and apprehensions by politicians that terrorists could manipulate prices to mislead the authorities or try to make money by trading at the PAM beforehand they make a terrorist attack.<sup>309</sup> Hanson (Oct. 2007), p. 83, one of the developers of PAM, disagrees with these arguments because the PAM was designed as a decision market on geopolitical trends — not on terrorist attacks. Moreover Meirowitz & Tucker (2004), p. 335, wonder why the hierarchy mechanism of intelligence apparatus is less objectionable than prediction markets.

Beside the proposal of decision markets there is the discussion on *macro markets*. Shiller (1993) argues that financial markets can only be used for hedging relatively small risks. Hence, he proposes *macro markets* on main national and international economic variables, like national incomes. Athanasoulis et al. (1999) believe that such macro markets can help persons as well as huge pension funds to eliminate fluctuations in country-specific growth performance.

---

<sup>308</sup>As the outcomes of the different markets would not be independent, the PAM was planned as a *combinatorial prediction market* platform where the participants can trade conditional stocks of the form: "payment of 1€ when a certain parameter of a certain country in a certain time interval has realized given that another parameter of a certain country in a certain time interval has occurred". Because this leads to an exponentially increasing number of stocks the PAM is based on Hanson's market scoring rule (MSR, see section 18.6.7) offering infinite liquidity. As the MSR needs an initial subsidy \$50,000 were bankrolled (Hanson Oct. 2007, p. 78). Berg & Rietz (2003) show how conditional political prediction markets can be used for decision support.

<sup>309</sup>Meirowitz & Tucker (2004) give an overview about the press coverage on the PAM as "terrorist market".

The *Economic Derivatives* platform, set up by Goldman Sachs and Deutsche Bank, can be seen as a prime example of the application of prediction markets as a large scale hedging instrument against macroeconomic risks.<sup>310</sup> Economic Derivatives allows institutional traders to hedge against U.S. non-farm payrolls, retail sales, business confidence, initial unemployment claims, and Euro-area harmonized CPI (Wolfers 2006) by using pari-mutuel markets (see section 18.6.4) - usually open for about one hour one morning before the data release.<sup>311</sup> As the markets, partly preceded by another market one or two days before, are conducted very shortly before the data release, Gürkaynak & Wolfers (2005), p. 5, believe that the Economic Derivatives markets can be used to hedge event risk rather than business cycle risk. Besides hedging, Wolfers (2006) recommends such markets for (at least) short-time economic forecasts instead of complicated models.<sup>312</sup>

## 18.2 Basic Structure of Prediction Markets

Prediction markets normally take place via internet or in laboratories (computer pools) on prediction market platforms. The participants (= traders, experts) have to register and get a registration code.

In *real money markets* the participant can invest a certain sum of money that is mostly restricted to a maximum at academic prediction market platforms. According to the amount invested the participants get trader accounts which they can use for trading. Apart from commercial real money prediction markets that earn money via fees on the transactions (e.g. TradeSports or Betfair) academic prediction markets are generally *zero sum games*. This means all invested money is finally redistributed depending on the traders' success. In *virtual money markets*, Wolfers & Zitzewitz (2004) call them *pseudo-markets*, each trader is initially endowed by the same amount of virtual money.

After registration (and remittance in real money markets) the traders are allowed to trade on some or all prediction markets of the prediction market platform. The maturity of the markets varies from minutes (mainly markets in laboratories) to several months or years

---

<sup>310</sup>Of course, Economic Derivatives is not a classical macro market in the sense of Shiller (1993).

<sup>311</sup>Other macroeconomic markets for GDP, international trade balance, and U.S. CPI have been set up or are planned. E.g., the Chicago Board of Trade offers federal funds rate futures and options and the Chicago Mercantile Exchange CPI futures contracts (Gürkaynak & Wolfers 2005, footnote 1).

<sup>312</sup>Gürkaynak & Wolfers (2005) compared 153 Economic Derivatives auctions within 2.5 years to the *survey forecasts* - consensus forecasts of usually about 30 forecasters - released by Money Market Services (MMS) on the Friday before data release. The result was that market prices contained all information of the surveys and were better able to explain behavior of stock prices in respect of economic news. Indeed, market prices seemed to mirror the information level of the financial markets (Gürkaynak & Wolfers 2005, p. 12). The average survey tends to stick too long with bad forecasts instead to incorporate new information. Market prices did not. This is partly because the markets took usually place days after the surveys.

(internet markets).<sup>313</sup> At each market there are contracts, called *stocks* or *bonds*, on *all* possible outcomes of a risk factor or on all elements of an event (e.g. parties in an election). On each stock several *shares* are issued and can be traded. E.g., assume a "hurricane market 2009" which is on the number of hurricanes in the North Atlantic in 2009. At this market there are  $K$  stocks: stock 1 (0 - 1 hurricane), stock 2 (2 - 3 hurricanes), ..., stock  $K$  (15 -  $\infty$  hurricanes) which cover *all* possible outcomes (0 -  $\infty$  hurricanes). Each stock consists of  $n_k$  (mostly  $n_k = n$ ) tradable shares.

*Classical prediction markets* have some characteristics in common. Traders have the possibility to trade shares on the different stocks in a so called *secondary market* where only the participants can buy and sell. When there are several bids (buy order) or asks (sell order), the trades are normally ordered by time of entering the market. The prediction market platform usually displays the highest bid price and the lowest ask price for each stock. The market price of a stock only reflects the last price at least one share was traded. This is called *marginal pricing*.

A further possibility to buy and sell stocks is the *bank* (or *primary market*). At the bank the participants can only buy and sell complete *unit portfolios*, also called *bundles*. Unit portfolios comprise one share of each stock and can be bought or sold during the whole duration of the prediction market for a fixed price. Generally, traders are not allowed to go short or to take out a loan at the bank. The fixed price of the unit portfolio, often set at 1 (€, \$, etc.), determines the rational sum of prices at the secondary market.<sup>314</sup>

For the participants it is only possible to buy shares or unit portfolios when they possess the adequate amount of money. For selling them it is necessary to possess the appropriate number of shares on the respective stocks. Buying or selling orders can be limited by time.

### 18.3 Market Liquidation

At maturity, liquidation is normally determined by the observed realization of the event under consideration while the payoff of the unit portfolios equals its fixed price to guarantee a *zero sum game* for all traders as a whole. According to the liquidation mechanism prediction markets are called *winner-takes-all markets*, *index markets* and *spread markets* (see table 45).

In *winner-takes-all markets* only the shares on the realized outcome get a payoff equal to the fixed price of the unit portfolio. All other shares of the remaining stocks have no payoff. Consequently, such shares are *pure/ Arrow-Debreu securities* (Berlemann & Nelson 2005,

---

<sup>313</sup>At the Foresight exchange there are prediction markets that last for several years. E.g., trading on the claim "Humanly mobile robot, by 2036" started in 1997 and is planned to mature in 2036, at the latest.

<sup>314</sup>see section 18.5. for more on arbitrage strategies



mechanisms	details	market expectation of
winner-takes-all	shares on the realized outcome pay 1€, other shares nothing	probabilities for the different outcomes
index	shares pay according to a relative portion	expected value of different elements of an event
spread	payoff if realization of the outcome is above the share price	expected quantiles of a risk factor

Table 45: liquidation mechanisms of prediction markets (Wolfers &amp; Zitzewitz 204, table 1)

p. 8). Such a prediction market is complete as the risk originating from the respective risk factor can fully be hedged by holding a unit portfolio (Berleemann 2008).

In *index markets* the payoff of each stock is proportional to the relative share of different elements of an uncertain event. In contrast to winner-takes-all markets, index markets reveal point estimates instead of uncertainty on the elements. Index markets are popular in political stock markets where each stock represents a party or a politician. The payoff of each stock is determined by the vote share (= element of the event "election"). Prediction markets could also assess the energy mix in Europe by 2020.

In the rather academic *spread markets* traders bid on expected quantiles of a risk factor. Suppose a market "the average global temperature rise up to 2050 will be more than  $y^\circ\text{C}$ ". The price of the stock is fixed and the payoff too. The trader solely bids on the cutoff point  $y$ . When the global temperature will be above  $y$ , the trader wins the fixed payoff. Otherwise, they get nothing. The relationship between fixed stock price and payoff determines the quantile the traders reveal. E.g., a price of 1€ and a payoff of 2€ will lead to the 50% quantile (median) while a price of 4€ and a payoff of 5€ reveal the market assessment of the  $4/5=80\%$  quantile.

Finally there exists the highly theoretical concept of *non-linear index* stocks with payoffs depending on the squared outcome  $y^2$  of the risk factor  $Y$ . Prediction markets on such stocks reveal the market expectation about the squared outcome  $E(Y^2)$ . Combining a linear and a non-linear market, the implied market volatility of the risk factor  $Y$  can be calculated by

$$\text{Std}(Y) = \sqrt{E(Y^2) - E(Y)^2}.$$

However, it is doubtful, whether prediction market traders are able to rationally trade on such non-linear index stocks.

## 18.4 Real Money vs. Virtual Money Markets

At least in *real money markets*, traders have a monetary incentive to reveal their true assessments about the outcome of a risk factor. Even in *virtual money markets* (no compensation or prizes) traders can earn reputation in a community by succeeding in the markets. In general prediction markets with real money are assumed to achieve better results. Monetary incentives are presumed to let traders take part and reveal their subjective assessments as this strategy leads to a maximization of the traders' expected profits and utilities<sup>315</sup> while trading with virtual money is just *cheap talk* with no consequences (Wolfers & Zitzewitz 2006, p. 20).

Nevertheless, the perception of the superiority of real money markets is more based on theoretical reasoning than empirical results. Rosenbloom & Notz (2006) found real money markets on non-sports events significantly more accurate than virtual money markets. Nevertheless, there are some counter-examples: Empirically, Span & Skiera (2003), p. 1317, could not find a difference in the prediction quality compared to real money markets. Luckner (2006) observed non-public real money markets on the FIFA World Cup 2006 to perform worse than virtual money markets with prizes. Servan-Schreiber et al. (2004) found that real and virtual money markets on American Football outcomes at TradeSports.com (real money) and NewsFutures (virtual money) during the 2003-2004 NFL season had a comparable prediction ability. They suppose that this is a result of the higher motivation at real money markets and the more efficient information aggregation at virtual money markets. Maybe, this result could be misleading as a general finding because sport markets are closely related to a huge sports fan community with motivation patterns others than money.

A commonly used version of virtual money markets is to reward traders with money or prizes depending on their virtual gains. Traders do not need to invest their own money which could be less discouraging. This could attract a large number of new traders.

When the best performing traders are rewarded with prizes, traders might only be motivated that hope to be one of these best. All others have at least no monetary incentive to reveal their expertise as it makes no difference whether one is the worst of all traders or the best under the non-rewarded traders. When an active trader comes to the conclusion that she will most likely not win a prize, the trader could start a more risky strategy to become one of the best which than results in biased market prices in form of overvalued prices for subjectively very unlikely events (Berlemann et al. 2005, p. 21). A possibility to ease the problem is to publicly rank people to give them non-monetary incentives. This could however keep away some potential traders fearing disgrace.

Although the prediction performance is, at least theoretically, in favor of real money markets,

---

<sup>315</sup>In fact, the utility of traders may not be maximized by a faithfully revelation of their expectations when the *no-stakes condition* is violated (see section 16.3.1 and Kadane & Winkler 1988).

a market operator should keep in mind other issues conveyed by real money markets:

(1) *time-lag*: Even with online banking, traders in online markets cannot trade immediately after they have registered. The market operator will not activate the trader accounts before the trader's money is transferred on the market operator's bank account. Some people may not be willing to take part in the market when they are not allowed to start trading immediately or will have lost their interest up to the activation of their accounts.

(2) *transaction costs*: Potential investors could be deterred by transaction costs for transferring and re-transferring the money. Within the Single Euro Payments Area (SEPA),<sup>316</sup> starting in 2010, this will be no problem even for international payment transactions all over European. Global prediction markets should also allow payment systems like PayPal. The transaction cost argument in the Bulgarian financial sector was one of the reasons for Berlemann et al. (2005) to run virtual money markets.

(3) *non-disposable investments*: The longer the life-span of a prediction market the harder it is to motivate people to invest real money as they usually don't get back their investment until the maturity of the market. Berlemann et al. (2005), p. 8, propose to pay an interest on the investment. This could at least reduce the opportunity costs. In the first political stock market in the Czech Republic in 2002, Chalik et al. (2005) subsidized all investments with additional 10% of the invested amount if at least 20% of the their initial value were traded.

(4) *credibility*: A market operator needs credibility as she/ it collects the invested money. The longer a market lasts the more important the credibility. Risk-averse people may shrink away from investing money in real money markets because they fear they could lose it when they are wrong or when the market operator goes bankrupt or is a criminal. Luckner (2006) argues that in public prediction markets on sports events many traders could be risk-seeking and therefore the real money markets could perform better.

## 18.5 Traders

Most academic real money prediction markets are *zero sum games* for the traders as a whole. At first glance no trader should be willing to take part in such a market with an average expected payment of zero. Every risk a trader bears in the market in form of not holding the same number of shares of each stock (= unit portfolio) is *unsystematic risk* that is not compensated because of arbitrage arguments. Forsythe et al. (1992) present five motives for trading in a prediction market: *Novelty factor*: For most people prediction markets are new and they are curious to trade. *Confidence factor on information*: People believe they

---

<sup>316</sup>Transaction within SEPA, comprising the 27 EU member states, Liechtenstein, Iceland, and Switzerland, will be usually zero for private persons.

possess an above average knowledge in the topic. Consequently, they presume not to take part in a zero sum game. *Confidence factor on interpretation ability*: People believe that they are superior to others in transforming information, e.g. from news papers, that will arise during the market. *Confidence factor on trading talent*: People believe they are better traders than others. *Risk-seeking factor*: People enjoy to gamble. Morris (1995) argues that especially for events that can only be observed with insufficient frequency it is very likely that traders have uncommon priors which results in *heterogeneous beliefs* and are a motive for trading.

Theoretically, traders can be distinguished by different trading strategies, namely arbitrage strategy, expectation strategy, modified expectation strategy and speculation strategy (Berle-  
mann 1999, pp. 8; Berle-  
mann & Nelson 2002, pp. 13):

*Arbitrage strategy* means that participants try to make risk-free profits. When the market price for an unit portfolio is lower than its fixed bank price the participants should buy stocks at the secondary market (synthetical unit portfolio) for the lower price and sell them at the bank. When the market price is higher than the fixed bank price, it is rational to buy unit portfolios at the bank and sell the stocks separately at the market. Both strategies lead to a risk-free profit. However, this is only true when there is enough liquidity in the market for all stocks. Otherwise it is not possible to buy all stocks at the low prices to create a synthetical unit portfolio or there is the risk that the trader is not able to sell all stocks, at all, or only for a lower price. As long as there is no problem of liquidity in a good functioning market these arbitrage possibilities should disappear very fast.

Traders who follow the simple *expectation strategy* decide to buy or sell stocks on the basis of their expectations of the final market outcome. When traders expect a rising (falling) market prices they buy (sell) shares.

Traders following the *modified expectation strategy* reflect the possibility that their expectations could be wrong. Consequently, traders are more and more willing to sell their stocks the closer the market prices narrow the expected values.

The *speculation strategy* is not based on expectation of the final market result. Instead, traders act according to their expectations about the trading behavior of other traders. The effects of speculative trading can be speculative bubbles and irrational price shifts.

There are also some empirical findings on the behavior of traders at prediction markets. Generally, in academic prediction markets, not all registered traders trade at least one time. E.g., in a CPI inflation market at the TU Dresden, only 31 out of 51 registered participants traded at least one time (Berle-  
mann & Nelson 2005; Berle-  
mann 2008). In their 1994 Dutch political stock market, Jacobsen et al. (2000), p. 210, observed 40% of the active buyers and 20% of the active sellers responsible for 90% of the traded shares.

Empirically, Oliven & Rietz (2004) could categorize traders. They used the unit portfolio

and market structure to identify whether traders are market makers or price-takers. They found a self-selection of *market-makers*<sup>317</sup> and *price-takers*. Market makers were less mistake prone and appeared to be more rational. Traders on the buy side produced less violations than traders on the sell-side and markets that were more competitive, measured by bid-ask spread, produced less violations, whereas markets with high total daily dollar volumes were prone to more mistakes (Oliven & Rietz 2004, p. 342). Altogether, market makers had more market-specific experience and general education on average (Oliven & Rietz 2004, p. 349). Do these findings mean that a market operator should only attract market-makers? Berlemann et al. (2005) restricted their virtual money markets to handpicked experts (= expert markets) to avoid attracting uninformed traders. In contrast, there is a widespread literature (e.g. Spiegel & Subrahmanyam 1992; Hanson 2006; Bloomfield et al. 2009) arguing that prediction markets need uninformed traders, called *noise traders*, offering incentives for informed traders, called *insiders*. When the market with solely rational traders and common knowledge is in an efficient equilibrium the *no trade theorem* will apply and no trade will take place (Milgrom & Stokey 1982). This is because everybody knows a trader has private information when she wants to trade. Hence, rational traders are not willing to accept the bid or ask offer.

Beside these theoretical considerations an important decision for a market operator is whether a *public mass (prediction) market* or an *expert market* should be conducted. Public mass markets are prediction markets via internet with no restriction on the access. Everybody is allowed to sign in and trade. There is a self-selection of the traders without any requirement to fulfil certain requirements. In contrast, trading at expert markets is restricted to preselect experts or to persons that fulfill predefined criteria.

The adequate market form mainly depends on the topic and on the purpose of the market. Expert markets are preferable (1) when a professional prediction is needed on a specific topic (e.g. Potsdam Climate Exchange, corporate prediction market for HP, Google, etc.), or (2) when legislation requires to restrict access (Economic Derivatives).

However, prediction markets are often public mass markets. This seems reasonable (1) when the topics are not too special, (2) when the public directly influences the outcome (e.g. political stock markets, Hollywood Stock Exchange, Securities Trading of Concepts (STOC)),<sup>318</sup> (3) when prediction markets are set up for entertainment purposes (e.g. TradeSports and Foresight Exchange), and (4) when prediction market shall measure sentiments in the public which concern politics, business, NGOs, and research (e.g. Potsdam Climate Exchange). Additionally, a market operator can be forced to run public mass markets when there is not

---

<sup>317</sup>Here, the term *market maker* describes traders that mainly set bid and ask prices instead of just accepting offered market prices.

<sup>318</sup>The public elects the candidates or parties, goes to the movies, and purchases products.

a sufficient number of experts available but a prediction is needed.<sup>319</sup>

## 18.6 Pricing Mechanisms

A key characteristic of prediction markets is their pricing mechanism that exist in extremely different forms. In the following, we present the most important as well as the most promising and sophisticated mechanisms. Finally, we compare their pros and cons.

### 18.6.1 Continuous Double Auction (CDA)

Classical prediction markets possess a *continuous double auction mechanism* (CDA) just matching the bids and asks of the market participants. There is only trading, when any trader is willing to buy at least one share for a price (bid order) that must be equal or larger the price (ask order) another trader is willing to sell at least the same number of shares of the same stock. Thus, the market operator bears no financial risk.

CDA markets can differ in small features. Some markets allow partially filling of orders. The remainder can be placed as a book order or be dropped. Often, traders are allowed to impose their offers a time limit. Limit orders, common for CDA markets, are also used in different ways. Normally, the limit price for, e.g. a bid (ask) order is defined as the maximum (minimum) price, a trader is willing to buy (sell) at least one share. Sometimes, the limit prices is understood as the maximum *average* price of all bought shares. Hence, when parts of a bid order are fulfilled at a lower than the maximum price, this mechanism also allows to buy single share above that maximum.

New information in such CDA markets is incorporated very fast. However, the CDA can suffer from illiquidity meaning that there are huge bid-ask spreads possible or empty bid-ask queues. The most successful measures to overcome illiquidity is to increase the number of traders or to motivate them to trade with higher frequency. Unfortunately this is very often hardly possible.

### 18.6.2 CDA with (Non Market Maker) Artificial Traders

One approach to eliminate or reduce human shortcomings in markets is the usage of automated traders. Artificial traders called electronic snipers like *eSnipe* and *AuctionBlitz* have established at eBay. In eBay auction markets it is not a reasonable strategy to reveal the maximum amount of money one is willing to pay. This would give other bidders the chance

---

<sup>319</sup>This is mainly the case in research. As there is often no access to experts, students are used to predict inflation rates, exchange rates, the success of product concepts, etc. (Berlemann et al. 2005, Berlemann 2008, Dahan et al. 2007).

to raise their maximum bids. This inflates the final price. Electronic snipers automatically bid at the last minute of the auction.

Grossklags & Schmidt (2003) have compared the efficiency of CDA prediction markets with and without artificial traders. The automated traders were programmed as arbitrageurs scanning for purchasing or selling arbitrage possibilities. Markets with automated and human traders, informed of the artificial arbitrageurs, seem to possess an improved efficiency.<sup>320</sup> However, in the case that the humans were not informed about the automated traders, the efficiency of the market prices was worse than in the pure humans' market. A crowding out of the human by automated traders could not be found.

### 18.6.3 CDA with Market Maker

At markets with a CDA there can trade one or more (human or automated) market makers. We call such markets *CDA with Market Maker* (CDAMM). In general, market makers need an accreditation of the market operator and often get access to the order books of the market. They are obliged to permanently offer bid and ask rates to raise liquidity in thin markets. Their profit comes from the bid-ask spread. Market makers have the incentive to set fair prices because they profit from a high trading volume. To get optimal narrow bid-ask spreads the market operators can admit several market makers that compete with each other.

In prediction markets the market maker is usually the market operator and the spread is often set to zero. The risk in fulfilling the role of a market maker is the danger of large monetary losses that could lie at least beyond the financial capability of at least most of the academic prediction market operators. That is the reason why professional market makers are mainly financial institutions with profound experience and expert-knowledge.

Automated market makers set prices according to an algorithm determining to what extent the market price increases (decreases) when a bid (ask) order is processed. The Hollywood Stock Exchange (HSX) runs an automated CDA market-maker matching bid and ask orders by generating a market price by a Virtual Specialist software. This bears no risk for the HSX as virtual money (Hollywood\$) is traded only. Other prediction market platforms with automatic market makers are the Iowa Stock Exchange, NewsFutures, ForesightExchange, TradeSports, Inkling Markets, Washington Stock Exchange, BizPredict, YooNew, Net Exchange and Yahoo! Tech Buzz Game. A market maker was planned at the PAM, too (see section 18.1 for PAM).

Exactly like the regular CDA, the CDAMM leads the market participants to reveal their private information continuously as soon as new information arrives. Additionally, a market

---

<sup>320</sup>Grossklags & Schmidt (2003) conducted 18 sessions (six markets of each type: no automated traders, human traders informed about automated arbitrageurs, and human traders not informed about automated arbitrageurs) with 108 humans and 12 programmed traders.

maker offers the advantage to aggregate even more information because traders do not need to wait for a matching offer to make the deal (Ondrus et al. 2007). Bruggen et al. (2006) performed thin prediction markets with market makers and as little as six traders (which is hardly possible with basic CDA markets). Compared to a simple equal weighting of experts, the performance of these markets were significantly better when traders had highly heterogeneous expertise. In the case of rather homogeneous experts, the weighting approach performed slightly better.

#### 18.6.4 Pari-Mutuel Mechanism

*Pari-mutuel (betting) markets* are familiar to a lot of people by horse race wagers. Plott et al. (2003) have no clear theory explaining why pari-mutuel markets should be able to aggregate information. They find empirical evidence in favor of an excellent aggregation - apart from an overestimation of rare events.

In such markets, gamblers can put their money on mutually exclusive and exhaustive outcomes of a risk factor, i.e. bets on different horses to win. After the market has closed and the outcomes can be observed the shares on the "winning" outcomes equally receive all staked money  $M = \sum_{k=1}^K M_k$  reduced by the fees  $F$ . Assume that there are  $K$  mutually exclusive and exhaustive outcomes and  $M_k$  is the amount of money invested in the "winning" stock  $k \in \{1, \dots, K\}$  paying

$$p_k = \frac{M - F \cdot M}{M_k} = \frac{M \cdot (1 - F)}{M_k},$$

per invested unit of money (Pennock 2004, p. 171). The payoff is zero for investments in the other  $K - 1$ .

Normally, pari-mutuel markets quote no prices but odds. At TradebetX there are either *moneyline odds* or *digital odds*. Digital odds differ from traditional odds as they include the traders own stake money as part of the total return. If a trader places a bet of \$20 at digital odds of 5.0 and wins, her total payoff is \$100 (winnings of \$80 plus original stake of \$20). In fractional odds this would be quoted as 4/1 or +400 moneyline. Digital odds are simpler to use than traditional odds and are the most common form of odds quoted outside the U.S. Besides, digital odds relate more closely to probabilities: in a race with four equally matched horses, the probability of each horse winning is 25%. Each horse will have traditional odds of 3/1 or digital odds of 4.0.

Without any fee - common in academic pari-mutuel markets - it is irrelevant whether the total stake is redistributed to the winning stocks or only the "losing" stakes and the investment is repaid

$$p_{win} = \frac{M_{win} + M_{lose}}{M_{win}} \stackrel{!}{=} 1 + \frac{M_{lose}}{M_{win}}.$$



In contrast to continuous double auction markets, pari-mutuel markets are theoretically not able to continuously reflect the aggregated information of its participants as there is an incentive to place stakes shortly before market close. All investors in the "winning" stock will get the same payoff per invested unit of money irrespective the time of investment. Thus, market participants should wait for possible new information and the latest market prices and then decide where to invest. Consequently, the market prices don't reflect the market assessment before market close, partly because market participants cannot cash out. The return on each invested unit of money can only rise when relatively more money is invested in the other stocks. It is possible to allow bettors to change or to cancel their bets. However, they cannot profit from price changes as every invested unit of money has an exact value of one up to the market closes (Pennock 2004, p. 171).

Kalovcova (2007), pp. 5, has summarized the main differences between classical prediction markets and pari-mutuel markets:

- In a prediction market traders with information, also called insiders, will enter the market as early as possible to profit from market price changes. In pari-mutuel markets such insiders will attend as late as possible because the payoff depends on the amount of money invested on the same stock.
- In prediction markets, traders can immediately calculate their expected payoffs. In pari-mutuel markets, they need additional assumptions about invested sum in the stocks (see also Gürkaynak & Wolfers 2005, p. 3).
- When traders split their money equally on all stocks, their return on investment is zero. In a pari-mutuel markets such a portfolio is not risk-free.
- In traditional pari-mutuel markets there is no continuous trading as there is no possibility to resell the tickets.

### 18.6.5 Wagering Markets

Las Vegas style *wagering markets* are similar to pari-mutuel markets with bookmakers or oddsmakers acting as market maker and are called *book* or *house*. They determine the odds in the beginning by (their own) expert judgement and later in response to the different stakes (Pennock 2004, p. 172). In contrast to pari-mutuel markets the odds are fixed at the moment the gamblers put in their wagers while odds may change over time. This is similar to CDA with market makers. The bookmakers earn their money by the bid-ask spread but can incur huge losses. In contrast, pari-mutuel market operators make their profits by fees.

### 18.6.6 Combined-Value Trading (CVT)

From asset pricing theory investors are not interested in single assets but in portfolios to maximize utility coming from returns and risk (Bossaerts et al. 2002). This is why non-diversifiable (systematic) risk is only compensated. Bossaerts et al. (2002) propose a portfolio trading mechanism called *combined-value trading* (CVT) that induces liquidity in thin financial markets (see section 18.6.6). Investors can submit portfolios of securities they want to buy or sell. The CVT mechanism computes prices by optimally combining portfolio orders.

The CVT mechanism is an *intermittent call market* working in the following way: Assume, a prediction market has  $K$  different stocks and  $n_{jk}$  stands for the number of shares of stock  $k$  that trader  $j \in \{1, \dots, J\}$  wants to buy ( $n_{jk} > 0$ ) or sell ( $n_{jk} < 0$ ). The budget constraint  $b_j$  for trader  $j$  states the maximum amount of money this trader is willing to pay for the portfolio  $n_j = (n_{j1}, \dots, n_{jK})'$ . Traders need to enter prices  $p_j = (p_{j1}, \dots, p_{jK})'$  beside  $n_j$  and  $b_j$  in an electronic open book. After a predetermined time, the book is closed and prices and quantities are calculated. Subsequently, the book is opened again.

The CVT algorithm maximizes the market turnover  $\sum_{j=1}^J \alpha_j b_j$  by finding adequate fill fractions  $\alpha_j \in [0, 1]$  for all  $J$  traders.<sup>321</sup> The maximization is not allowed to violate the  $K$  conditions  $\sum_{j=1}^J \alpha_j n_{jk} \leq 0$  guaranteeing that not more shares of a stock are sold than offered. A market clearing is not required. Once the orders are matched, the optimal prices  $p^* = (p_1^*, \dots, p_K^*)'$  are chosen such that Walras' law obtains  $\sum_{k=1}^K p_k^* \sum_{j=1}^J \alpha_j n_{jk} \stackrel{!}{=} 0$ . A solution to  $p^*$  always exists but is not unique as no numeraire is assigned (Bossaerts et al. 2002, p. 1676).

### 18.6.7 Market Scoring Rule (MSR)

Hanson (2002) developed a *market scoring rule* (MSR) which can be used as two-side automated market maker offering continuously bid and ask prices. The advantage of a MSR compared to a CDA with a market maker (CDAMM) is a bounded maximum loss for the market maker.

Traders can buy or sell stocks at any time from or to the market maker at prices that are defined by a cost function. The inherent scoring rule  $s(\cdot)$  offers incentives that rational traders change the market prices in accordance with their subjective beliefs as they can always expect to profit when they disagree with the market assessment. The payoff for the traders is determined by the way the agent has changed the prices on the  $K$  possible outcomes.

---

<sup>321</sup>It is also possible to interpret  $J$  as the number of orders when traders submit several orders.

Assume that the current market prices for the different outcomes of a risk factor are  $p = (p_1, \dots, p_K)'$  and trader  $j$  changes those prices to her reports  $r_j = (r_{j1}, \dots, r_{jK})'$  by buying and selling shares of the  $K$  different stocks. Now suppose outcome  $k$  will finally realize. Hence, shares on stock  $k$  will pay off (e.g. 1€) and all others nothing. Then trader  $j$ 's virtual number of shares of stock  $k$  is

$$n_{jk} = \Delta s_k(r_j, p) = s_k(r_j) - s_k(p) = \begin{cases} > 0, & r_{jk} > p_k \\ = 0 & r_{jk} = p_k \\ < 0 & r_{jk} < p_k \end{cases}, \quad (21)$$

where the scoring rule  $s_k(\cdot)$  equals to the whole number of shares on stock  $k$  in the market. The portfolio value of trader  $j$  calculates by  $n_{jk} \cdot 1$ .

Equation 21 can be interpreted in the following way: When the realized outcome is  $k$  then the optimal forecast would have been  $r_{jk} = 1$  and  $r_{jl} = 0$ ,  $l \neq k$ . Hence, the market scoring rule rewards traders increasing the market price for stock  $k$  and imposes a fine on those who have lowered it. The exact monetary rewards and fines depend on the scoring rule  $s(\cdot)$ . The total turnover (= number of emitted shares on  $k$  times one unit of money)

$$n_k(T) = \sum_{t=1}^T [s_k(r^{(t)}) - s_k(r^{(t-1)})] = s_k(r^{(T)}) - s_k(r^{(0)})$$

for  $T$  reports on the "winning" outcome  $k$  depends on the initial and the final report, only. In the case of a *logarithmic market scoring rule* (LMSR),  $s_k(r_k) = a_k + b \cdot \ln r_k$ , the automated market maker calculates the market price according to (see Hanson Jan. 2002, p. 6; Hanson 2003, p. 111)

$$p_k = \frac{\exp([s_k - a_k]/b)}{\sum_{l=1}^K \exp([s_l - a_l]/b)}.$$

The corresponding cost function<sup>322</sup>

$$C(s_1, \dots, s_K) = b \cdot \ln \left( \sum_{k=1}^K \exp([s_k - a_k]/b) \right)$$

describes the amount of money so far invested in all stocks.

The maximum expected payment for the market maker is limited to  $-b \sum_{k=1}^K \pi_k \ln(p_k^{(0)})$  where  $p^{(0)} = (p_1^{(0)}, \dots, p_K^{(0)})'$  are the initial market prices set by the market operator who has the subjective probability assessments  $\pi = (\pi_1, \dots, \pi_K)'$ . Hence, the maximum loss is  $-b \ln(\max_k p_k^{(0)})$ . A market maker who intends to minimize risk should set the initial prices to her expectations,  $p^{(0)} = \pi$ . Then the maximum expected payment equals the entropy of

<sup>322</sup>see Pennock (2006) for the two stocks case

the initial distribution  $-b \sum_{k=1}^K \pi_k \ln(\pi_k)$ . The larger  $b$  the more the market maker can lose but it makes the market more liquid. Thus, the purchase and selling of shares do change the market prices by a lower magnitude and lead to lower price swings (Pennock 2006).

In the following we adopt a numeric example presented by Pennock (2006). Assume that the scoring function for the two stocks  $k = 1, 2$  is  $s_k(r_k) = 100 \cdot \ln r_k$ . The initial investment of the market operator is

$$C(s_1 = 0, s_2 = 0) = C(0, 0) = 100 \cdot \ln(e^0 + e^0) = 69.31\text{€}.$$

For 10 shares of stock 1, a trader has to pay

$$C(10, 0) - C(0, 0) = 100 \cdot \ln(e^{10/100} + e^0) - 100 \cdot \ln(e^0 + e^0) = 5.12\text{€}.$$

Later, when  $s_1 = 400$  and  $s_2 = 300$  shares circulate, the costs for selling 10 shares of stock 1 are

$$\begin{aligned} C(390, 300) - C(400, 300) &= 100 \cdot \ln(e^{390/100} + e^{300/100}) - 100 \cdot \ln(e^{390/100} + e^{300/100}) \\ &= -7.21\text{€}. \end{aligned}$$

Meaning that the seller receives this amount of money. The gain for the trader is  $7.21\text{€} - 5.12\text{€} = 2.09\text{€}$ , a return of 41%.

The current market prices for stocks  $k = 1, 2$  are just the prices for an infinitesimal amount of shares. Mathematically, this is nothing else than the deviation of the cost function

$$p_k = \frac{\exp([s_k - 100]/10)}{\exp([s_1 - 100]/10) + \exp([s_2 - 100]/10)}.$$

In the example above, the market prices for stock 1 are

$$p_1(0, 0) = \frac{e^0}{e^0 + e^0} = 0.5\text{€}, \quad p_1(400, 300) = \frac{e^{400/100}}{e^{400/100} + e^{300/100}} = 0.62\text{€}.$$

Hanson's market scoring rule is not purely theoretic. It was planned to be used in the Policy Analysis Market (PAM) which failed for political reasons. Ledyard (2006) still proposes the MSR for policy markets. And indeed, the logarithmic market maker has been currently implemented at Inkling, the Washington Stock Exchange, BizPredict, Net Exchange, and YooNew.

### 18.6.8 Dynamic Pari-Mutuel Market (DPM)

The *dynamic pari-mutuel market* (DPM) was developed by Pennock (2004) and has been already implemented by Yahoo! Tech Buzz Game. The DPM can be conceptualized as an

one-side market maker that continuously offers ask prices, only (Pennock 2004). However, there is no guarantee that market participants can sell their shares as selling is accomplished by a CDA mechanism. Compared to other mechanisms, even the MSR, the market maker in a DPM bears no risk but a constant pre-determined subsidy is required to start the market. The DPM theoretically features infinitive liquidity like a pari-mutuel market but continuously incorporates new information. In contrast to pari-mutuel markets, the return of investments in the "winning" DPM stock depends on the state of wagering at the time of purchasing. This means the prices of stocks are variable and depend on the number of bets on the stocks so far. The more bets, the more expensive the stock. Similar to pari-mutuel and wagering markets the payoff of a share on the realized outcome in the DPM market is unknown in advance.

### DPM I: Losing Money Redistributed

Pennock (2004), pp. 174, shows how market probabilities can be extracted in a DPM market where the stakes of the "losing" shares are redistributed while the holders of "winning" shares get their investments back. In the case of only two different stocks  $A$  and  $B$ , the payoff per share  $k = A, B$  is

$$\mathcal{P}_k = \begin{cases} M_{\bar{k}}/n_k, & k \text{ "wins"} \\ 0, & k \text{ "loses" } (\bar{k} \text{ "wins"}) \end{cases}$$

where  $n_k$ , describes the number of shares sold of stock  $k$  while  $M_{\bar{k}}$  is the amount of money invested in the stock which is not  $k$ . The initial investment is paid back.

The market prices  $p_A$  and  $p_B$  only hold for purchases and sales of an infinitesimal amount  $\varepsilon \rightarrow 0$  of those shares as each trading influences their prices. Hence, trader  $j$ 's expected gain or loss  $\mathcal{E}_{jA}$  resulting from buying  $\varepsilon \rightarrow 0$  shares of stock  $A$  is

$$\mathcal{E}_{jA} = \pi_j(A) \cdot E_j(\mathcal{P}_A|A) \cdot \varepsilon - \pi_j(B) \cdot p_A \cdot \varepsilon,$$

where  $\pi_j(A) = 1 - \pi_j(B)$  is the subjective probability of trader  $j$  that stock  $A$  will win and  $E_j(\mathcal{P}_A|A)$  describes the expected payoff of one share of stock  $A$  given  $A$  wins. Normalized to exact one share of stock  $A$ , the expected gain or loss is

$$\mathcal{E}_{jA}/\varepsilon = \pi_j(A) \cdot E_j(\mathcal{P}_A|A) - \pi_j(B) \cdot p_A.$$

A risk-neutral market trader  $j$  should purchase shares of  $A$  as long as  $\mathcal{E}_{jA}/\varepsilon > 0$ . Risk-averse traders would stop purchasing above this value.

The individual gain or loss expectations of trader  $j$  for purchasing  $\tilde{n}_{jA}$  shares of  $A$  calculates by

$$\mathcal{E}_{jA}(\tilde{n}_{jA}) = \pi_j(A) \cdot \tilde{n}_{jA} \cdot E_j(\mathcal{P}_A|A) - \pi_j(B) \cdot \int_0^{\tilde{n}_{jA}} p_A(\varepsilon) d\varepsilon. \quad (22)$$

A profit maximizing trader would buy tickets up to  $\mathcal{E}_{j_A}(\tilde{n}_{j_A}) = 0$ .

In an efficient market, the aggregated expected gains or losses for additional emitted  $\tilde{n}_A$  shares should be zero

$$\mathcal{E}_A(\tilde{n}_A) = \pi(A) \cdot E(\mathcal{P}_A|A) - \pi(B) \cdot p_A(\tilde{n}_A) \stackrel{!}{=} 0,$$

as the prediction market is a zero sum game. The variable  $\pi(A)$  describes the market consensus on the probability of  $A$  and on the market expectation  $E(\mathcal{P}_A|A)$  of the payoff per share of stock  $A$  given  $A$  is the "winning" outcome. The market probability can be derived by:

$$\pi(A) = \frac{p_A}{p_A + E(\mathcal{P}_A|A)}.$$

The critical problem here is to determine the conditional market expectation of the payoff. Pennock (2004) assumes a random walk without trend for  $\mathcal{P}_A|A$ , resulting in  $E(\mathcal{P}_A|A) = \mathcal{P}_A$ , and proposes two pricing functions for the market prices that can be used for the integral in equation 22.

*price function I:* The payoffs for tickets on  $k = A, B$  equal the prices for  $\bar{k} = B, A$

$$p_{\bar{k}} = \mathcal{P}_k = \begin{cases} M_{\bar{k}}/n_k, & k \text{ "wins"} \\ 0, & k \text{ "loses" } (\bar{k} \text{ "wins"}) \end{cases} \quad (23)$$

Then the price of a share on stock  $A$  rises when c.p. more money is invested in such tickets. Meanwhile, the price of tickets on  $B$  falls with the rising number of tickets on  $A$ . The resulting market probability for outcome  $A$  is

$$\pi(A) = \frac{M_A n_A}{M_A n_A + M_B n_B}.$$

When there already circulate  $n_A$  shares of stock  $A$ , the price for an additional infinitesimal number of shares  $\varepsilon \rightarrow 0$  is

$$p_A(\varepsilon) = \frac{dC_A(\varepsilon)}{dn_A} = \frac{M_A}{n_B} \exp(\varepsilon/n_B)$$

where

$$C_A(\tilde{n}_A) = \int_0^{\tilde{n}_A} p_A(\varepsilon) d\varepsilon = M_A \cdot (\exp(\tilde{n}_A/n_B) - 1)$$

are the costs to purchase  $\tilde{n}_A$  tickets of  $A$  additionally to  $n_A$ . In a prediction market the market maker should offer a price list for complete shares whose price is calculated by  $C_A(1)$  for the first share,  $C_A(2) - C_A(1)$  for the second share, and so on. Or the market maker offers packages like  $C_A(10)$ ,  $C_A(20) - C_A(10)$ , etc. With these equations the prediction

market is for the operator a zero sum game although an arbitrarily small subsidy for the initial shares is needed

*price function II:* A second price function of Pennock (2004), pp. 175, defines the price ratio by the ratio of wagered money in both stocks

$$\frac{p_A}{p_B} = \frac{M_A}{M_B}. \quad (24)$$

The more money is invested in a stock the more expensive it becomes. The resulting market probability is

$$P_m(A) = \frac{M_A \sqrt{n_A}}{M_A \sqrt{n_A} + M_B \sqrt{n_B}}.$$

Then with already  $n_A$  circulating shares, the price for additionally  $\varepsilon \rightarrow 0$  shares is

$$p_A(\varepsilon) = \frac{dC_A(\varepsilon)}{dn_A} = \frac{M_A}{\sqrt{n_A + \varepsilon} n_B} \exp\left(2\sqrt{\frac{n_A + \varepsilon}{n_B}} - 2\sqrt{\frac{n_A}{n_B}}\right)$$

where

$$C_A(\tilde{n}_A) = \int_0^{\tilde{n}_A} p_A(\varepsilon) d\varepsilon = M_A \left( \exp\left(2\sqrt{\frac{n_A + \tilde{n}_A}{n_B}} - 2\sqrt{\frac{n_A}{n_B}}\right) - 1 \right)$$

are the costs to purchase  $\tilde{n}_A$  shares.

### DPM II: All Money Redistributed

In the case that all invested money is redistributed the respective payoffs are

$$\mathcal{P}_k = \begin{cases} M/n_k, & k \text{ "wins"} \\ 0, & k \text{ "loses"} (\bar{k} \text{ "wins"}) \end{cases},$$

where  $k = A, B$  and  $M = M_A + M_B$  is the whole investment in the market.

The individual expected gain or loss  $\mathcal{E}_{jA}(\tilde{n}_{jA})$  for purchasing  $\tilde{n}_{jA}$  shares calculates by

$$\mathcal{E}_{jA}(\tilde{n}_{jA}) = \pi_j(A) \cdot \tilde{n}_{jA} \cdot E_j(\mathcal{P}_A|A) - \int_0^{\tilde{n}_{jA}} p_A(\varepsilon) d\varepsilon,$$

while the aggregated expected gains or losses for additional emitted  $\tilde{n}_A$  shares should be zero

$$\mathcal{E}_A(\tilde{n}_A) = \pi(A) \cdot E(\mathcal{P}_A|A) - p_A(\tilde{n}_A) \stackrel{!}{=} 0$$

in an efficient market. Then the market probability calculates by

$$\pi(A) = \frac{p_A}{E(\mathcal{P}_A|A)} \stackrel{\text{random walk}}{=} \frac{p_A}{\mathcal{P}_A}.$$

Unfortunately, there results no coherent price function from the price condition in equation 23. Pennock (2004), pp. 176, applies the price condition  $p_A/p_B = M_A/M_B$  of equation 24. Then the market probability is

$$\pi(A) = \frac{M_A n_A}{M_A n_A + M_B n_B}.$$

Even this approach does not lead to a closed-form solution of  $C_A(\tilde{n}_A)$ . Pennock (2004), pp. 177, proposes to solve the equation

$$n_A(C_A) = \frac{C_A \cdot (n_A - n_B)}{M} + \frac{n_B(M + C_A)}{M_B} \ln \frac{M(M_A + C_A)}{M_A(M + C_A)}$$

numerically. It simply states the number of shares that can be bought for a budget (= costs)  $C_A$ .

### Comparing DPM I and II

Using DPM I has the advantage that an investment in the "winning" stock will never lead to a monetary loss (ignoring discounting) as the stakes in the "losing" stocks are redistributed, only. However, DPM I involves the problem that shares on the stocks  $k = A, B$  are not homogeneous. The value of one share of stock  $k$  consists of (1) the equal gains of the redistributed "losing" money  $\mathcal{P}_k$  and (2) the respective market price  $p_A$  which varies over time. Hence, trading such heterogeneous at a secondary market seems difficult.

A DPM II market deals with homogeneous shares on the same stock  $k$  that allow an easy implementation of a secondary market. Since the initial price paid for the shares of the "winning" stock is not directly refunded, all shares on  $k$  have the same payoff.

### 18.6.9 Comparison of Market Types

In the previous sections different kinds of market pricing were presented which mainly differed whether they have a market maker or not. Markets without a market maker, i.e. continuous double auction markets (CDAs), feature a *marginal pricing*. The current market price corresponds to the price at least one share of the stock has been traded for the last time. However, when running a prediction market we are interested in the market assessment instead of the assessment of one trader, which might be a manipulation effort. Hence, the new market price should be immediately "reviewed" by the market. This requires sufficient liquidity which is often not the case in prediction markets. A trader with new information is often not able to immediately buy or sell shares. Moreover, there is the danger of a pricing failure when the market liquidity mainly stems from *limit orders* (see section 18.7.1) —



automatically executed orders. When the trading due to the limit orders eventually takes place traders may prefer different decisions based on their new information (Tetlock 2008).

An alternative is to enhance CDAs with *competing* market makers.<sup>323</sup> Then, the market prices immediately reflect a *professional* appraisal of the latest market orders. This market structure seems superior in low liquid markets. However, there is no chance to find professional market makers, e.g. banks, for (academic) prediction markets. At least the market operator could act as market maker. In virtual money markets there is no financial risk for the operator but often the market operator is no professional in the topic the prediction market is about.

Therefore, several approaches with automated market makers were analyzed. Such mechanisms always offer a market price irrespective of whether the market is liquid or not. Consequently, traders can immediately incorporate their new information in the market, there is no limit order problem and such markets could attract more active traders which enhances liquidity.

However, automated market makers do not guarantee efficient market prices but liquidity, only. When there is no reaction of other traders to the last trader, the new market price simply results from a pricing function without any evaluation of the new information. Another flaw of automated market makers is their inability to adjust to shifts in the trading volume.<sup>324</sup> The impact of a trader on the market price is a (non-)linear function which may additionally depend on the amounts of money invested in different stocks. The market operator has to decide on a function and its parameters before the market starts. To our knowledge the automatic market maker of Xpree is the simplest algorithm. The operator decides (by a factor or amount) how much the market assessment is changed by an investment of virtual \$1,000. The total number and distribution of the shares in the market is irrelevant. In Hanson's MSR the operator has to decide on the maximum loss she is willing to bear. The higher the maximum the less a trader is able to influence the market price.

Altogether, a market operator should simultaneously base a decision in favor of a pricing mechanism according to three or four conditions: (I) *guaranteed liquidity*, (II) *no risk for the market operator* in real money markets, and (III) *continuous incorporation of information*. An additional criterion (IV) is whether the payoff of a share of the "winning" stock is known in advance. Otherwise, a trader has to assess both, the probability for the stock to be the winner and the payoff in this situation. In table 46 there is an overview how some of the presented pricing mechanisms match these requirements.

---

<sup>323</sup>Bookmakers can be seen as a special form of market makers.

<sup>324</sup>see the blog of Chris Hibbert at <http://blog.commerce.net/?p=251> (last revised August 26, 2008).

	I	II	III	IV
CDA	-	+	-/+	+
CDA with MM	+	-	+	+
pari-mutuel market	+	+	-	-
wagering market	+	-	+	+
MSR	+	+	+	+
DPM	-/+	++	-/+	-

The sign "+" ("-") means that the requirement is (not) fulfilled.

Table 46: comparison of market types (based on Pennock (2004), p. 173)

## 18.7 Efficiency of Prediction Markets

Elicitation tools should provide incentives to seek and reveal truthful information as well as offer an algorithm for aggregating diverse opinions (Wolfers & Zitzewitz 2004). Consequently, prediction markets require that the speculations of Hayek (1945), the rational expectations theory (Muth 1961), and the efficient markets hypothesis (Fama 1970) hold. Unfortunately, these theories are no incontrovertible truths but hypotheses which need to be tested (Passmore & Cebeci 2003, p. 4).

In an *efficient financial market* the market prices are assumed to incorporate all available information influencing the value of the underlying. The prices reflect everything, more than a few people know about the stock. Hence, new information of a trader on the stock should already be discounted in the prices. Buying and selling of stocks on the basis of special information or interpretation "is not likely to be profitable" (Black 1975, p. 324). During the 1970ies the efficient markets theory was the dominant paradigm in finance. This was confirmed by empirical research describing the practicable impossibility to beat the market (Dimson & Mussavian 1998).<sup>325</sup>

Some years later, in financial economics, behavioral critiques came up against the efficient markets hypothesis. According to Grossmann & Stieglitz (1980), persistent informational efficient markets are impossible as in the market equilibrium, where all arbitrage possibilities are eliminated, informed traders had no profit from trading and therefore no incentives for costly arbitrage activities or to acquire information.

To a broad public, bursting financial bubbles like the dotcom or the subprime crises in the late 1990ies and since 2007/8 have raised doubts about efficient markets. This reasoning is not made up out of thin air. Brunnermeier & Nagel (2004) found evidence that hedge funds rode the technology bubble in the late 1990ies for a while because the investor sentiments

<sup>325</sup>Dimson & Mussavian 1998 overview the pro and con research on the efficient markets hypothesis.

were predictable. Shortly before the burst the hedge funds deinvested and realized profits. This was an example where rational behavior did not trade against the mispricing but boosted the bubble resulting in inefficient market prices. The same result can be observed when arbitrageurs are not able to enter and exit their trading positions easily or when their capital is insufficient compared to naive traders.

Surprisingly, since the mid 1990ies, more and more evidence has been found in favour of the efficient markets hypothesis as it is hard to profit from even extreme violations of market efficiency (Dimson & Mussavian 1998). Fama (1998) argues that most market anomalies are fragile in the long-term and tend to disappear when they are discovered. Finally, after examining lots of studies attacking efficient-market hypothesis, Malkiel (2003) concludes that the markets are more efficient and less predictable than the discussion let us believe.<sup>326</sup> Especially for prediction markets Oliven & Rietz (2004) show that despite theoretical arguments against the possibility of market efficiency and the possibility of non-rational maximizing traders, markets can be efficient. This is why classical prediction markets are zero-sum games. When an informed trader would not set efficient prices, in a competitive market, there is the danger that another informed trader would do so. This is in the line with Berg et al. (1997) who found in an empirical analysis of 16 U.S. political stock markets that "active" markets predict better.

Despite these findings, there is a huge literature dealing with important market failures in prediction markets like: information traps, informational cascades, herd behavior, overconfidence, winner's curse, informational flooding, manipulations, and illiquidity. The extent of all these stated failures of prediction markets is strongly affected by the traders' level of information on the underlying (risk factor/ event) of the market (Chen et al. 2001, p. 58). Normally, it is hard or nearly impossible at least to distinguish between informed and uninformed traders. In *information traps*, traders even interpret actions of uninformed traders as information driven (Camerer & Weigelt 1991; Nöth et al. 1999). Fortunately, Camerer & Weigelt (1991) observed in laboratory experiments that such traps are only temporary. Very close to information traps are *informational cascades* when single traders join the main market assessments (= stocks with the highest prices) even when they have contradictory information (Nöth & Weber 2003; Sunstein 2006). This leads to even higher prices which might amplify the cascade and a *herd behavior* could begin when traders mimic the actions of other traders (Scharfstein & Stein 1990). *Overconfidence* results in biases when traders negatively affect prices as they stick to their forecast even if better informed traders give other signals (Bruggen et al. 2006, p. 6). Traders put too much weight on their information

---

<sup>326</sup>Of course, the current crash of the financial markets will again marginalize this understanding of finance. However, it does not seem that a significant group of financial investors could have had profited from the current financial crisis since 2007/8.

and do not accept low market prices as an indication that the outcome of the risk factor is very unlikely. In addition with cash constraints this can lead to the *winner's curse* phenomenon (Berlemann & Schmidt 2001, p. 27). The buyer pays too much for a stock according to the unknown fair value.

Another problem for prediction markets is *informational flooding*. Especially in illiquid markets there are only a few active traders who might be overstrained in collecting and incorporating all relevant information into the market. Hopman (2007), p. 132, found that the Intel prediction markets on demand risks got worse in the final month before the actual result. He speculates that the amount of information shortly before the event explodes and the time for assessment shrinks. Hence, humans seemed overburdened.

As *illiquidity* and *manipulations* are supposed as being among the most crucial hazards for the efficiency of prediction markets they are presented separately.

### 18.7.1 Illiquidity

Standard models of market microstructures show that noise traders increase accuracy of market prices as long as informed traders (= insiders) have the financial endowments to profit from the mispricing (Kyle 1985/ 1989,<sup>327</sup> Milgrom & Stokey 1982; Spiegel & Subrahmanyam 1992). Liquidity stemming from noise trading - arbitrary trading based on no information<sup>328</sup> - attracts additional capital by informed traders as it does not counteract informed trading but allows such traders to realize profits (Tetlock 2008, p. 5). Indeed, Jacobsen et al. (2000), p. 227, observed evidence for a better performance of liquid markets, which come along with low bid-ask spreads, a low price impact of a trade, and a high turnover volume (Baker & Stein 2004).

Nevertheless, the market efficiency can only be achieved by informed traders and might fail when their number is too small (Spann & Skiera 2003). Berlemann et al. (2005), p. 20, speculate that there exists a minimum number of informed and active traders that are needed to achieve reasonable forecasts, although the gain in forecast accuracy quickly diminishes with an increasing number of traders above this minimum.

Berlemann & Schmidt (2001), p. 23, propose the number of registered traders as a good proxy for the number of active traders. Ledyard (2006), p. 51, sees a connection between the relationship *number of traders/ number of stocks* and the market efficiency. He cites the

---

<sup>327</sup>Kyle (1985) set up a theoretical base model to analyze information aggregation in prediction markets. The *Kyle model* describes three types of traders: (1) one risk-neutral trader with information, (2) several noise traders, and (3) one risk-neutral market maker. The informed trader makes profits by her private information. However, the market maker has problems to incorporate all information of the informed traders in the price setting because of the dominance of the noise traders.

<sup>328</sup>Nevertheless, noise traders are rational and can expect a zero profit.

prediction market of Bossaerts & Plott (2004) ( $\approx 13$  traders/ stock) which better equilibrated than the market of Bossaerts & Plott (2002) with  $\approx 4$  traders/ stock.

However, the number of registered and active traders can significantly differ. Out of 6,425 employees registered at the Google prediction market platform just about 1,463 placed at least one order (Cowgill et al. 2008, p. 4). Even the number of active traders can be delusive. E.g., roughly 20% of the traders were responsible for 80% of all trades in the influenza market of the Iowa Electronic Markets (Polgreen et al. 2007, p. 278). This ratio might depend on the fact whether the relevant information is homogeneously or heterogeneously spread over the traders.

An efficient market equilibrium heavily rests on the assumption of price-taking traders that cannot influence prices. Indeed, this does not hold in thin markets. Kyle (1989) theoretically analyzes the situation in which traders, with private information, consider their influence on the market prices. This results in non-efficient market prices that do not reflect all information spread over the traders. As a consequence a market operator should raise market liquidity.

Unfortunately, (academic) prediction markets have mostly a low turnover, only. Berlemann & Nelson (2005) run markets on German CPI inflation rates with 32 up to 47 traders and turnovers between 288€ and 1,022€. There have only been some markets attracting a lot of attention and hence a lot of capital. The "Saddam Securities" of TradeSports had up to about 32,000 active traders and a monthly turnover up to more than \$11,000 (Leigh et al. 2003, table 1). At the winner-takes-all political stock market of the Iowa Electronic Markets on the 1996 U.S. presidential election 1,151 participants traded with a monthly volume of \$5,800 and a total turnover of more than \$137,000. The Economic Derivatives market platform of Deutsche Bank and Goldman Sachs is the most professional prediction market. Institutional investors are allowed, only. The market volume is several hundreds of millions of dollars (Wolfers & Zitzewitz 2006, p. 20).

There are several proposals to improve the liquidity of thin prediction markets. One possibility are *call markets*. Transactions take place at predetermined intervals only. All bid and ask orders are aggregated and transacted at once. The clearing price is determined by the market operator based on the number of bid and ask orders. An example for a call market is the combined-value trading (CVT) mechanism of Bossaerts et al. (2002) (see section 18.6.6). An alternative is to run *combinatorial matching markets* which search for combinations of market offers that can be matched (Hanson 2003). On the Google corporate prediction market platform it was tried to motivate traders by conducting so-called "fun" markets which comprised 30% of all markets. Cowgill et al. (2008), p. 5, could successfully test for a positive correlation of the liquidity of "fun" and "serious" markets. Hence, they conclude this approach creates liquidity rather than crowding out traders.

The statement that more liquidity improves market efficiency has been challenged in recent years. Tetlock (2008), p. 33, sees the dotcom bubble during the late 1990ies as a prominent example for a situation when liquid markets are not superior to liquid ones. Baker & Stein (2004) built a model which can describe a situation where increased liquidity in a market with short-sales constraints is rather an indicator of a market sentiment than of efficiency as the market is dominated by irrational investors which underreact to information.

Linnainmaa (2007) ascribes irrational market behavior to the possibility of *limit orders*. A limit order allows a trader to determine for which prices the market operator automatically sells and buys shares. When the trader does not immediately react according to new information, the automatism can trade against the traders information. Based on this insight Tetlock (2008) analyzed short-term TradeSports markets on financial and sporting event outcomes. He concludes *naive liquidity provision* (liquidity triggered by limit orders) does not improve but sometimes reduces market efficiency as prices respond more slowly to new information (Tetlock 2008, pp. 5):

- As limit orders, based on old information, are executed automatically when new information enters the market, an efficient market price response is retarded.
- Limit orders typically buy at low prices and sell at high prices which can result in a favorite-longshot bias where low probabilities are overpriced and high probabilities are underpriced.

An example might help to better understand the argumentation of Tetlock (2008). E.g., the current market assessment for an outcome is 0.10€. Because of new information *all* traders believe the fair price for the outcome is 0.30€. Tetlock (2008) analyzes the situation where the liquidity mainly rests on the limit orders not on a high number of active traders. Consequently, only a few active traders try to buy as many shares as possible offered for less than 0.30€. When these active traders have no deep enough pockets they are not able to raise the market price to 0.30€ although all traders think this is the fair price. This is because limit orders, based on old information, are automatically triggered. Of course, when new information turns out to be true these active traders profit but the market efficiency suffers.

This subsection shortly overviewed the research on a sufficient liquidity level for prediction markets. It can be summarized that a market operator should raise liquidity to improve market efficiency. This is especially needed in CDA markets as often in thin markets traders cannot find a counterpart to immediately execute their buy or sell orders. Generally, such markets allow limit orders which automatically match orders when possible. However, Tetlock (2008) empirically showed that liquidity induced by limit orders can reduce market

efficiency. A possibility to bypass this flaw is to run prediction markets with automatic market makers as they possess three benefits: (1) Markets with market makers should attract more traders as traders are not frustrated by queued orders. (2) There is no *naive liquidity provision* in the sense of Tetlock (2008) since no limit orders are necessary. (3) Even when the number of traders is low, the liquidity is fundamentally improved. By means of a market maker Bruggen et al. (2006) could achieve good results in prediction markets with not more than six traders.

### 18.7.2 Foul Play

The market efficiency can be damaged by different forms of foul play: lying, sabotage, embezzlement, retribution, and manipulation (Hanson 2006, pp. 128). In this subsection these forms are discussed in detail.

#### Lying

Prediction market prices could be influenced by *lying* or *misleading* advisors who are paid by third parties with an interest in a certain market result. Moreover, traders with insider information could launch misleading signals to maintain a low profile about their information level.<sup>329</sup> This is an issue all forecasting methods are affected by (Hanson July 2006, p. 266).

#### Sabotage

When traders cause harm to achieve a profit this is called *sabotage* (Hanson 2006, p. 132). E.g., explaining the attacks of 09/11 by a sabotage motive would presume that the terrorists crashed in the World Trade Center to profit from subsequently falling stock prices. Hanson 2006, pp. 132, claims that sabotage is not a big problem as prediction markets are thin and the underlyings are often events, like presidential elections or inflation, a few traders cannot fundamentally influence.

#### Embezzlement

A motive for foul play in real money prediction markets on company-related events is embezzlement. In markets where employees can win large amounts of money they may neglect their company duties. Employees could withhold relevant information within their department hoping to make profits in the market.

Therefore, corporate prediction markets should not offer direct financial rewards. Sometimes workmates of the traders are allowed to veto single offers. This however is not possible with anonymous trading (Hanson 2006, pp. 132). Consequently, the market operator has to deal with a trade-off between the danger of embezzlement and social pressure.

---

<sup>329</sup>Other traders might be reluctant to trade when they fear some market participants to possess insider information.

### Retribution

Predictions can be deliberately inaccurate when traders hope to profit but fraud cannot be unmasked. When the cheater is in a predominant position, especially in cooperations, other insiders are often not willing to show her up as they fear *retribution*. In this context, such social pressure can be avoided by creating prediction markets with anonymous accounts.

### Manipulation

The most prominent form of foul play is *manipulation* which describes actions that are in contradiction with the information of the manipulators since they intend to fool other traders. In prediction markets, the target of the swindle could also be the market operator or the final consumer of a market forecast. Allen & Gale (1992), p. 505, present a reasonable categorization in action-based, information-based, and trade-based manipulation.<sup>330</sup>

*Action-based manipulations* intend to change the perceived value as well as the actual value which is also called *outcome manipulation* (Ottaviani & Sørensen 2007, p. 556). Especially in corporate prediction markets traders could have incentives to influence the outcome of the underlying. Ottaviani & Sørensen (2007) are the first ones who set up a formal model that describes action-based manipulation in cooperations. Their model incorporates optimists and pessimists on the outcome of an event. Both have the incentive to manipulate the outcome in their favored direction. Although the different manipulations typically do not perfectly off-set, the costly manipulations are partly wasted (Ottaviani & Sørensen 2007, p. 556).

*Information-based manipulations* arise from dissemination of false or misleading information around the market. To cut off manipulations a *disclosure rule* has been proposed which would force traders to reveal their trading activities after executing their orders. However this approach may even create incentives for insiders to trade against their signals since announcements of good and bad company news have often been preceded by insider trading (John & Narayanan 1997).

Most research has been on *trade-based manipulations* (Rhode & Strumpf 2007, p. 4), which can be understood as a trading in the "wrong" direction by selling at good news and buying at bad news (John & Narayanan 1997, p. 218). Jarrow (1992) shows that trade-based manipulation is not profitable in efficient markets as manipulators raise the market price when buying the shares and reducing the price by the same amount when they want to sell them. Allen & Gale (1992) proofed that trade-based manipulation is profitable with rational investors when other traders have a positive probability that the manipulator bases her actions on insider information.

---

<sup>330</sup> All forecasting methods are effected by action- and information-based manipulations while trade-based manipulation is a special failure of prediction markets.



There are two main motives for traders to manipulate. A trader can possess monopolistic information or external stakes. *Information monopolists* have the incentive not to reveal all their private information, at once. Especially in thin markets it is not reasonable to offer e.g. relatively high bid quotes for the preferred stocks. In illiquid markets such orders would change the market prices too fast. The information monopolist cannot profit from purchasing stocks at low prices very long. Consequently, insiders will only gradually disclose their information which can be theoretically understood as a trade-based manipulation.

Manipulators with *external stakes* are traders that try to interfere with the market because of (monetary) interests outside the market which outweigh potential losses and consequences within the market. This means that the *no-stakes condition* of Kadane & Winkler (1988) is violated.

In the previous paragraphs several forms of manipulations have been presented which could have been, if at all, observed at real financial markets, only. Indeed, discussions on prediction market manipulations are mainly hypothetical<sup>331</sup> and little based on empirical data. Rhode & Strumpf (2007) analyzed a large dataset of betting odds on presidential, gubernatorial, and mayoral races over the 1880ies to the 1940ies. There were 46 charges of manipulation/wash-sale/ bluffing events with an average manipulation occurring 7.8 days before the election (median 4 days). They find in nearly all prediction markets speculative attacks but no patterns that indicate manipulation events led to large, irreversible changes in prices. Camerer (1998) tried to manipulate betting odds in horse races by investing large sums in horses to send signals to the other gamblers. Later, he withdraw the money. Although he could visibly interfere the odds he could not realize significant profits by this strategy.

Hanson et al. (2006) studied the effects of manipulation on prediction markets prices in the laboratory. They conducted markets with heterogeneously informed traders in a normal design and a manipulation design that offered incentives to manipulators to increase median price even above the fundamental price. They found that manipulators submit higher bids than non-manipulators. This is the expected result. The bid behavior was affected by manipulation incentives but the presence of manipulation did not harm the relationship between price and fundamental value. Moreover, manipulation had no significant effect on the accuracy of prices measured as the squared difference between market prices and fundamental values. In fact, Hanson et al. (2006) recognized a slight improvement of accuracy in the case of manipulation.

Astonishingly, the average price traded in the markets was higher in the normal design than in the manipulation design which offered incentives to inflate prices. However, the results of Hanson et al. (2006) should be reviewed with care as in these special prediction market designs the participants were aware of the presence of manipulators and their incentives to

---

<sup>331</sup>Hanson (2006) believes that manipulators would reduce average price errors in terrorist markets.

inflate the prices. Moreover, the manipulation gains were a public good all manipulators could participate even manipulators that did not manipulate at all.

Until now, not more than one documented manipulation has been observed in a non-laboratory prediction market. This was at the German political stock market *WahlStreet* conducted on the 1999 Berlin state election (Hansen et al. 2004). Supporters of the German liberal party FDP tried to push FDP shares above 5% for more than a week to give the illusion that the FDP could pass the 5%-barrier to be represented in the state parliament. After the last media coverage they withdraw and prices fell to 3% immediately. As the official vote share of the FDP on election day was 2.2% the manipulators suffered monetary losses. Hansen et al. (2004), p. 462, calculated that for pushing the FDP to 5% seven manipulators would need a total endowments of at least 305.26€ to counterbalance the 200 traders on *WahlStreet* with an average endowment of 29€. The total costs of manipulation would rise up to 134.31€.

Theoretically, it should not be very surprising that, with the exception above, no other prediction market manipulations have been noticed. A broad research community argues that manipulations can make prediction markets even more accurate (Kyle 1985; Milgrom & Stokey 1982; Spiegel & Subrahmanyam 1992). Hanson & Oprea (2004) theoretically show that under risk-neutrality and quadratic utility functions, manipulators with unknown price targets are a special form of noise traders. They define manipulators like noise traders as market participants that trade on "considerations other than their best estimate of the asset value" (Hanson & Oprea 2004, p. 16). Hence, the discussion about the consequences of noise traders in the market holds for manipulators, too (see section 18.7.1). The more noise traders act in a market the more profits an informed trader can realize and consequently the more informed traders are attracted or increase their investments (Hanson July 2006, p. 267).

## 18.8 Market Prices vs. Market Assessments

### 18.8.1 Real Financial Markets

Markets have been often assumed to disseminate aggregate information efficiently. However no commonly accepted theoretical model exists explaining why the *Hayek Hypotheses* should hold under realistic assumptions (Berlemann 2003, p. 189). The relatively good prediction abilities of capital markets - especially commodity stocks like orange juice futures<sup>332</sup> and pari-

---

<sup>332</sup>In fact, there has been a long lasting disagreement about the future market on frozen concentrated orange juice (FCOJ). The discussion mainly has based on Roll (1984) who stated weather in central Florida as the most identifiable factor influencing FCOJ returns. Nevertheless, he could only empirically explain a small fraction of the observed variability in futures prices by weather (Roll 1984, pp. 876/ 879). Consequently, this

mutuel betting markets - have been seen as a verification by reality. There exist two directions of research on information aggregation of markets that deal with iterative aggregation and equilibrium aggregation, respectively (Ledyard 2006, pp. 40).

*Iterative aggregation:* When people start with common prior beliefs, subsequently get new private information, and their posterior beliefs become common knowledge then the final posterior beliefs are the same for all persons (Aumann 1976). This also holds (under some conditions) for people that update their information in a Bayesian style (Geanakoplos & Polemarchakis 1982). Consequently, the *opinion pooling problem* - experts do not agree with the aggregated assessment - should not exist in efficient prediction markets given a liquid market and common priors (Geanakoplos & Polemarchakis 1982; Hanson Nov. 2002). As long as people share their information *honestly* and update in a Bayesian style, the iterative process can result in an aggregation of all information (Ledyard 2006, p. 41).

*Equilibrium aggregation:* This field is based on the *rational expectations theory* where traders' demands (and supplies) reveal their assessments of the future. Market prices are formed by the individuals' demands. Traders incorporate the public information reflected in the prices and adjust their expectations as well as their demands. This again effects market prices that are not in equilibrium as long as not all publicly available information is incorporated. Otherwise behavior of market participants will always change in accordance with new market prices (Plott 2000, p. 3). The equilibrium prices equal the common-knowledge posterior probabilities (Ledyard 2006, p. 41). Nevertheless, this requires Arrow-Debreu securities, price-taking and risk-neutral agents as well as further assumptions that may not be fulfilled in thin markets (Ledyard 2006, p. 43).

Indeed, there are also counter-arguments against the theory that the market prices reflect all information. In a thin market where a few traders know the future for sure and others are very uncertain, revealing important information would result in very high bids for the "winning" stock and very low asks for the "losing" stocks. Uninformed traders could interpret this as superior information and refuse to trade (Ledyard 2006, p. 43). Thus, informed traders have no incentives to reveal their full information.

### 18.8.2 Prediction Markets

The use of prediction markets for the elicitation of market probability assessments has mainly been justified by references to traditional financial markets and by the great success of political stock markets and laboratory experiments. Actually, probability assessments can be derived from traditional financial markets but they differ from those of prediction markets.

---

was cited as a prominent example of the failure of prices to reflect fundamentals (e.g. Shleifer 2000; Hirshleifer 2001). In contrast to this conclusion Boudoukh et al. (2007) could empirically show that unexpected freeze is a crucial factor for future volatility.

A huge theoretical and empirical literature on financial market prices suggests that market prices result from risk-adjusted weighted and discounted future payoffs and are normally not equal or proportional to probabilities (e.g. Björk 2004). Although probability distributions are widely calculated from market prices they are just *risk-neutral*, and therefore partly far away from the "true" market assessments.<sup>333</sup>

In contrast to real financial markets, prediction markets are Arrow-Debreu markets that possess the same number of stocks as outcomes of the risk factor.<sup>334</sup> Such stocks only pay when the respective outcome realizes. As a consequence prediction markets price the payoff for each possible outcome. However, it is not clear immediately whether the prices reflect the market's (probability) assessments.<sup>335</sup> This should not be mixed with the question whether the market's probability assessments are based on reasonable judgements.

There is indeed field and scientific evidence for prediction markets to be good mechanisms for information aggregation (Ledyard 2006, pp. 39). Berlemann & Schmidt (2001) find in a meta study 25 political stock markets to perform better on average than traditional polls. However, take care, political stock markets are mostly index (vote-share) markets which do not predict probabilities but the expected share of an element of an event. Plott (2000) reviews laboratory prediction markets showing that market prices converge to theoretical market equilibriums of very complex non-linear systems by dynamic adjustments.

In contrast to empirical proofs, there was no clear theoretical proof for the quality of market prices and market probability assessments up to the beginning of this decade. In the previous years, there have been several papers by Berlemann (2003), Berlemann & Nelson (2005), Gjerstad (2005), Gürkaynak & Wolfers (2005), Manski (2006), and Wolfers & Zitzewitz (2007) offering proofs that partly seem to contradict each other, at first glance. Actually all approaches have in common the assumption of liquid markets that allow to buy or sell whenever a trader wants to do so. In our opinion this is the most critical assumption and should be kept at the back of the reader's mind in the next subsections.

### Traders with a Second Order Probability

Wolfers & Zitzewitz (2007) and Manski (2006) analyze the relationship between market assessments and market prices in the case of prediction markets with two stocks, only. The

---

<sup>333</sup>For pricing derivatives the real market probability distribution is normally not known as the risk-adequate discounting rate is unknown. In fact, the true distribution is not needed. There exists an infinite number of distributions that lead to the same pricing. For the sake of convenience the risk-neutral distribution is normally used as the discounting rate is just the risk-free interest rate.

<sup>334</sup>In this argumentation, we ignore index markets on events like elections.

<sup>335</sup>There is one exception: *Hanson's market scoring rule* (MSR) is a combination of a traditional prediction market and a *scoring rule* approach (see sections 18.6.7 and 16.3.1). Scoring rules have been proved to reveal the subjective probabilities for risk-neutral people.

special characteristic of their approach is the assumption of a *second order probability* of the traders on the outcome probability.

Manski (2006) shocked prediction market experts that even with risk-neutral price-taking traders the market price and the average belief of the market can fundamentally diverge. He assumes a winner-takes-all market on a dichotomous risk factor  $Y \in \{A, B\}$  where the outcomes  $A$  and  $B$  can be disjunct intervals. There are traded shares on the two stocks  $k = A, B$  paying 1€ when the respective state will realize. The prior distribution of trader  $j \in \{1, \dots, J\}$  on  $Y$  can be described by a Bernoulli distribution  $\pi_j = (\pi_j(A), \pi_j(B))'$  where  $\pi_j(B) = 1 - \pi_j(A)$ . However, Manski (2006) and Wolfers & Zitzewitz (2007) account for traders' uncertain on their probability assessments (imprecise probabilities). Hence, the subjective probability  $Q_k$  of outcome  $k = A, B$  is an uncertain risk factor for the traders. This we formulize by a second order pdf  $\tilde{\pi}_{jk}(q_k)$  and the corresponding cdf  $F_{jk}(q_k)$ .

In the two stocks prediction market of Manski (2006), the risk-neutral trader  $j$  invests *all* her money (wealth)  $w_j$  in that stock  $k = A, B$  where the market price  $p_k$  is below her subjective probability belief. Thus, trader  $j$  holds  $n_{jA} = w_j/p_A$  shares of stock  $A$  and nothing of  $B$  when  $q_{jA} > p_A$  ( $= q_{jB} < p_B$ ), and vice versa. The resulting aggregated *market demand* for shares of stock  $k$  is

$$n_k = \frac{1}{p_k} E(w \cdot 1_{(Q_k > p_k)}) \stackrel{\text{uncorr.}}{=} \frac{w}{p_k} [1 - F_k(p_k)],$$

where  $E(\cdot)$  is the market expectation,  $F_k(\cdot)$  is the cdf of the market on the probability of outcome  $k$ , and  $1_{(\cdot)}$  is an indicator function. The equal sign only holds in the case that the market expectations on the total market investment  $w$  and the probability of outcome  $k$  are not correlated.

In classical prediction markets there exists a bank where an unit portfolio can be bought or sold for  $c\text{€}$ . Without loss of generality, we assume  $c = 1$ . Then, a total investment of  $w\text{€}$  will result in a number of  $n_k = w$  shares of each stock.<sup>336</sup>

As risk-neutral traders will invest all their money in one of both stocks, the prices need to adjust to guarantee that the aggregated supply of both stocks is  $w$  and equals the respective demand in the equilibrium

$$\begin{aligned} n_A &= \frac{w}{p_A} [1 - F_A(p_A)] = w = \frac{w}{p_B} [1 - F_B(p_B)] = n_B \\ &\rightarrow \frac{1}{p_A} [1 - F_A(p_A)] = \frac{1}{p_B} [1 - F_B(p_B)] \end{aligned}$$

After a simple rearrangement, the market price  $p_k$  for stock  $k$  is defined by

$$p_k = P(q_k > p_k) = 1 - F_k(p_k), \quad (25)$$

<sup>336</sup>This rules out arbitrage but only holds in a liquid market.

corresponding to the  $(1 - p_k)$ -quantile of the market assessment  $P(\cdot)$  (respectively  $\tilde{\pi}_k(q_k)$ ) of  $Q_k$ .

The mean market belief  $E(Q_k)$  is an interval as  $\tilde{\pi}_k(q_k)$  is not univocal. The upper bound for  $E(Q_k)$  is  $2p_k - p_k^2$ . It guarantees equation 25 and is the mean of a continuous distribution<sup>337</sup> limiting to the bimodal discrete distribution  $\tilde{\pi}_k(p_k) = 1 - p_k$  and  $\tilde{\pi}_k(1) = p_k$ . The lower bound for  $E(Q_k)$  is  $p_k^2$  and can be derived from a continuous distribution limiting towards the Bernoulli distribution  $\tilde{\pi}_k(p_k + \varepsilon) \xrightarrow{\varepsilon \rightarrow 0} p_k$  and  $\tilde{\pi}_k(0) = 1 - p_k$ . Hence,  $E(Q_k)$  can take any value in the open interval  $(p_k^2, 2p_k - p_k^2)$  and  $\tilde{\pi}_k(q_k)$  can be any distribution bounded by the two distributions presented above.

In the Manski (2006) model with the strong assumption of a *risk-neutral all-or-nothing* investor the market only incorporates the information of investing traders placing *all* their wealth  $w_j$  in one of both stocks. Hence, the market prices solely reflect the beliefs of self-selected traders.

Wolfers & Zitzewitz (2007), p. 3, relax the assumption of all-or-nothing investors by endogenizing the investment decision. In fact, they work with an one stock prediction market allowing short selling where shares pay 1€ when  $A$  occurs and otherwise nothing. This is equal to a market with two stocks on  $A$  and  $B$  and no short selling.

Assume all  $J$  *price-taking* traders in the market are utility maximizers given their probability assessments  $\pi_j(A) = q_j$  and  $\pi_j(B) = 1 - q_j$  for the outcomes  $A$  and  $B$ . Then the utility maximizing demand ( $n_j^* > 0$ ) or supply ( $n_j^* < 0$ ) for stock  $A$  is

$$n_j^* = \arg \max_{n_j} E_j(U_j) = q_j U_j(w_j + n_j[1 - p]) + [1 - q_j] U_j(w_j - n_j p),$$

where  $p$  is the price per share and  $w_j$  is trader  $j$ 's future wealth which is assumed to be known for simplicity. Once  $A$  realizes, trader  $j$  gains  $n_j[1 - p]$  as each share pays 1€ but she has bought them at a price of  $p$ . In contrast, when  $B$  realizes, the loss  $n_j p$  is the purchasing price times the number of bought shares. Maximization results in the first order condition

$$\frac{U'_j(w_j + n_j[1 - p])}{U'_j(w_j - n_j p)} = \frac{p}{1 - p} \frac{1 - q_j}{q_j},$$

which implicates that

- *risk-neutral* traders ( $U'(a) = U'(b)$ ,  $a \neq b$ ) invest *all* their money in the stock ( $n_j > 0$ ) when their probability belief of the event is greater than the market price as the expected profit is positive. When the beliefs are lower, they will short-sell ( $n_j < 0$ ) as many stocks as possible. The same is true for *risk loving* traders ( $U'(a) < U'(b)$ ,  $a < b$ ).

<sup>337</sup>This continuous distribution has probability on the intervals  $(p_k - \varepsilon, p_k)$  and  $(1 - \varepsilon, 1)$ .

- *risk-averse* traders ( $U'(a) > U'(b)$ ,  $a < b$ ) will not invest all their wealth in the stock even in the case of an expected profit.

This is better to understand by using an example. Assume the special case when traders follow a *log-utility*. The optimal demand or supply endogenously calculates by

$$\begin{aligned} n_j^*(q_j) &= \arg \max_{n_j} E_j(U) = n_j \ln(w_j + n_j[1-p]) + [1-n_j] \ln(w_j - n_j p) \\ &= w_j \frac{q_j - p}{p[1-p]}. \end{aligned}$$

The most important special characteristic of the demand under log-utility is that it is *linear* increasing in the individual belief  $q_j$  and there exists a unique solution for  $p \in (0, 1)$ .

Just like Manski (2006), Wolfers & Zitzewitz (2007) believe that traders are uncertain on the probability  $Q$  of outcome  $A$ . Hence, they also model a second order distribution  $\tilde{\pi}_j(q)$ . Then, in the equilibrium, the aggregated demand ( $n^* > 0$ ) and supply ( $n^* < 0$ ) in the market must equal

$$-\int_{-\infty}^p n^*(q) \tilde{\pi}(q) dq = \int_p^{\infty} n^*(q) \tilde{\pi}(q) dq$$

where

$$n^*(q) = w \frac{q - p}{p[1-p]}.$$

and  $\tilde{\pi}(q)$  is the market assessment on  $Q$ . After some rearrangements the market price calculates by

$$p = E(Q) = \int_{-\infty}^{\infty} q \tilde{\pi}(q) dq.$$

Wolfers & Zitzewitz (2007), p. 9, conclude:

- In the case of logarithmic utility functions, the market price  $p$  for a share on outcome  $A$  equals the mean market expectation on the uncertain probability  $Q$  regardless the second order probability  $\tilde{\pi}(q)$ .
- The differences between market price  $p$  and mean beliefs  $E(Q)$  are very low for other reasonable utility functions.
- The difference is zero when  $\tilde{\pi}(q)$  is symmetric in 0.5.

### Prediction Markets as Hedging Instrument

So far, we have analyzed prediction markets with  $K = 2$  stocks and with traders having a second order probability. Henceforth, we follow the derivation of Gürkaynak & Wolfers (2005) where the number of stocks  $K$  is not restricted but there is no second order probability. The basic idea, similar to Wolfers & Zitzewitz (2007), is that traders try to smooth their uncertain future income by hedging their expected wealth by investing in a prediction market.

Assume that the current (known) wealth of investor  $j$  is  $w_j$ . The future wealth depends on the uncertain outcome  $k \in \{1, \dots, K\}$  of the risk factor  $Y$  and can be modelled by  $\beta_{jk}w_j$  where  $\beta_{jk} \in \mathbb{R}$  is a factor. The traders can hold shares of stock  $k$  paying 1€ when outcome  $k$  realizes and nothing otherwise. Then, the optimal numbers of shares  $n_j^* = (n_{j1}^*, \dots, n_{jK}^*)'$  maximizing the expected utility of trader  $j$  calculate by

$$n_j^* = \arg \max_{n_j} E_j (U_j (w_j)) = \sum_{k=1}^K \pi_{jk} \cdot U_j \left( \beta_{jk}w_j + n_{jk} - \sum_{l=1}^K p_l n_{jl} \right),$$

where  $\pi_{jk}$  is the assessment of trader  $j$  on the probability of outcome  $k$ . The term  $n_{jk} - \sum_{l=1}^K p_l n_{jl}$  simply formalizes the fact that trader  $j$  has purchased the shares for the market prices  $p = (p_1, \dots, p_K)'$  and only gets a payment of 1€ on each share she holds of the winning stock  $k$ .

On an aggregated market perspective and under the zero sum game assumption  $n_k = \sum_{k=1}^K p_k n_k$ , the first order condition is<sup>338</sup>

$$\frac{p_k}{\pi_k} = \frac{U'(\beta_k w)}{\sum_{s=1}^K \pi_s U'(\beta_s w)}, \quad (26)$$

where  $\pi_k$  is the aggregated market assessment of the probability of outcome  $k$ .

The interpretation of equation 26 depends on the people trading at the market:

- *risk-neutral traders*: Risk-neutral traders possess a constant marginal utility ( $U'(a) = U'(b)$ ,  $a \neq b$ ). Consequently, the market prices exactly match the market probabilities

$$p_k = \frac{\pi_k U'(\beta_k w)}{\sum_{s=1}^K \pi_s U'(\beta_s w)} = \frac{\pi_k}{\sum_{s=1}^K \pi_s} = \pi_k. \quad (27)$$

<sup>338</sup>This condition results from

$$\frac{\partial E(U)}{\partial n_k} = \pi_k [1 - p_k] \cdot U' \left( \beta_k w + n_k - \sum_{l=1}^K p_l n_l \right) - p_k \sum_{s=-k} \pi_s \cdot U' \left( \beta_s w + n_s - \sum_{l=1}^K p_l n_l \right) \stackrel{!}{=} 0$$

and  $n_k - \sum_{l=1}^K p_l n_l = 0$ , where  $\sum_{s=-k}$  simply means that summation ignores  $k$ .



- *risk-averse traders*: Since risk-averse traders suffer from a loss in wealth more than they profit from the same amount of gains, the marginal utility increases with decreasing wealth ( $U'(a) > U'(b)$ ,  $a < b$ ). As a rule of thumb, the market prices  $p_k$  overestimate the market probability  $\pi_k$  of stocks on "bad" outcomes ( $\beta_k < 1$ ) while they underestimate probabilities of "good" outcomes ( $\beta_k > 1$ ). Generally, risk-averse traders prefer an insurance that pays out in the case of "bad" instead of "good" states.

The previous statements should become clearer with an example. Assume the "true" market assessment of the future wealth is Gaussian distributed with expected value  $w$  (for simplicity  $w = 1$ ) and standard deviation 0.01.<sup>339</sup> The utility function might be isoelastic with a constant relative risk aversion (CRRA)

$$U_\gamma(x) = \frac{1}{1-\gamma} x^{1-\gamma}.$$

In the situation with widely accepted risk aversions of  $\gamma = 5$  or  $\gamma = 20$ , the differences between the "true" market assessments and those derived from the market prices are quite low (see figure 98).

Based on this theoretical reasoning and empirical analyses, Gürkaynak & Wolfers (2005), pp. 30 and pp. 35/ 38, found that the markets on the *Economic Derivatives* market platform<sup>340</sup> feature very low risk premiums and market prices are reasonable approximations of the market probabilities.

### Prediction Markets as an Entertainment Instrument

The previous discussion to compare market prices with market probability assessments presumes a relationship between wealth of a trader and the outcome of the risk factor. Moreover, it is supposed that it is possible to hedge the resulting wealth risk with stocks on that factor. This hedging argument may be correct for the *Economic Derivatives* market platform, but it is not for normal (academic) prediction market platforms with minor monetary stakes (often restricted to a maximum investment). The Iowa Electronic Markets (IEM) only allow a maximum investment of \$500. Traders may not evaluate their stakes in a prediction markets with their effects on their total wealth. Participation is rather because of entertainment reasons (Wolfers & Zitzewitz 2004, p. 113; Wolfers & Zitzewitz 2007, p. 3).

In the following, we present elaborated approaches of Kou & Sobel (2004), pp. 284, and Berlemann (2003), pp. 226/ Berlemann & Nelson (2005), pp. 12, who show that market prices equal (or are proportional to) the market assessment when (1) prediction markets are not part of traders' overall investment strategy and (2) their invested money is locked in.

<sup>339</sup>In a macroeconomic setting a standard deviation of 1% of the wealth might be acceptable.

<sup>340</sup>business confidence, initial unemployment claims, non-farm payrolls, and retail sales (ex autos)

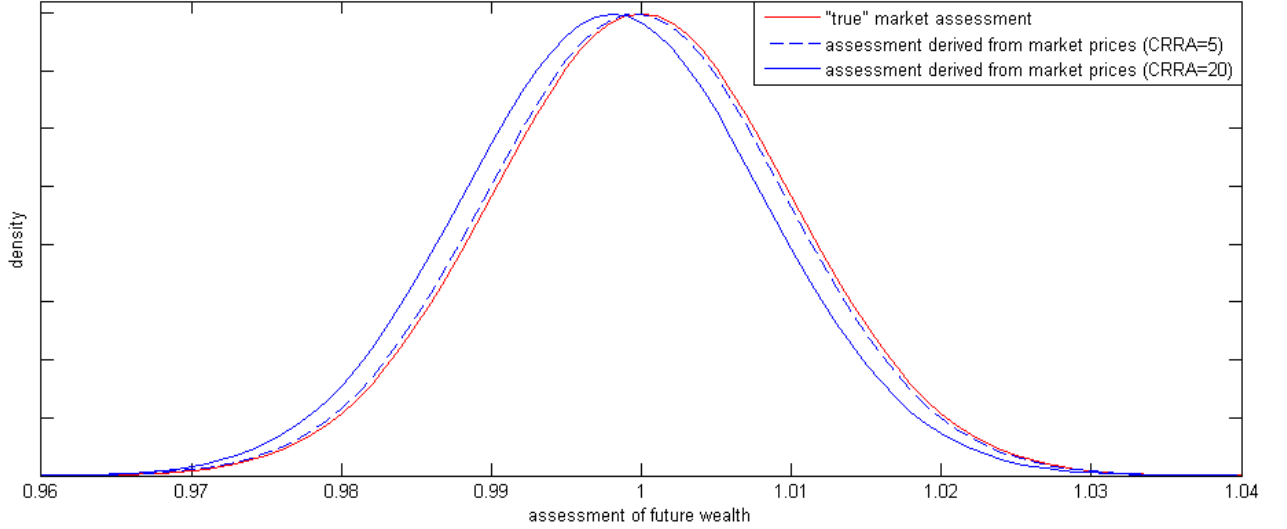


Figure 98: risk aversion - It is assumed that the market's "true" uncertainty on the future wealth is known and can be described by  $N(1, 0.01)$ . In reality, the risk analyst could only deduce the market's uncertainty from the market prices. Given the traders follow an isoelastic utility function with a constant relative risk aversion of  $\gamma = 5$  or  $\gamma = 20$ , the resulting approximations are good proxies for the "true" distribution.

Kou & Sobel (2004), pp. 284, assume each trader  $j \in \{1, \dots, J\}$  decides at time  $t = 0$  to use some "play money"  $\tilde{w}_j(0)$  which is only a part of their much larger total wealth  $w_j(0)$ . This "play money" marks the maximum amount of money, trader  $j$  is willing to spend for entertainment purposes. As soon as traders have separated or reserved their "play money" for entertainment they maximize their "play money" profit. The development of the total wealth  $w_j$  is unimportant, as  $\tilde{w}_j(0)$  is supposed to be negligible small to hedge the total wealth.

In  $t = 0$ , trader  $j$  invests money  $\sum_{k=1}^K \delta_{jk} \leq \tilde{w}_j(0)$  in some or all stocks. The uncertain wealth in  $T$  calculates by

$$\tilde{W}_j(T) = \left[ \tilde{w}_j(0) - \sum_{k=1}^K \delta_{jk} \right] + \sum_{k=1}^K \frac{\delta_{jk}}{p_k(0)} X_k(T),$$

where the term in brackets is the "play money" not invested and  $\delta_{jk}/p_k(0)$  is the number of shares on stock  $k$  trader  $j$  has purchased for  $p_k(0)$  each. The exact payoff  $X_k(T)$  of shares on stock  $k$  are uncertain in  $t = 0$ .

Utility maximizing traders will try to maximize  $E_j \left( U_j \left( \tilde{W}_j(T) \right) \right)$ . As individual traders

are assumed to be price-takers, the first order condition is simply the right-hand derivative

$$\frac{\partial^+ E_j \left( U_j \left( \tilde{W}_j(T) \right) \right)}{\partial \delta_{jk}} = E_j \left( U_j' \left( \tilde{w}_j(0)_i - \sum_{k=1}^K \delta_{jk} + \sum_{k=1}^K \frac{\delta_{jk}}{p_k(0)} X_k(T) \right) \left[ \frac{X_k(T)}{p_k(0)} - 1 \right] \right) \stackrel{!}{=} 0.$$

Traders will buy and sell stocks until the market price will stabilize in an equilibrium. Then, traders have no incentive to trade, meaning  $\delta_j = (\delta_{j1}, \dots, \delta_{jK})' = 0$ . Consequently, the equilibrium market price of stock  $k$  reflects the market expectation of the uncertain future payoff

$$p_k^*(0) = E_j(X_k(T)).$$

When all traders have homogeneous expectations, the market price is the equilibrium price for all traders. This conclusion can be extended to  $t \in [0, T)$  (Kou & Sobel 2004, p. 285).

The approach of Berleemann (2003), pp. 226, and Berleemann & Nelson (2005), pp. 12, is more intuitive and seems to be closer to reality. According to the *arbitrage pricing theory* (APT) the equilibrium price in  $t$  for *pure* or *Arrow-Debreu securities* on outcome  $k$  is defined by

$$p_k(t) = \exp(-[r_f + r_r] \Delta) \cdot E_t(X_k(T)),$$

where  $\Delta = T - t$ ,  $r_f$  is the risk-free return, and  $r_r$  is the risk premium. The index  $t$  indicates that the market expectation  $E_t(X_k(T))$  on the uncertain payoff is based on the information available in  $t$ .

Generally, an arbitrage free and risk-free portfolio should be discounted by the risk-free return  $r_f$ , only. A different discounting rate would create arbitrage possibilities. In a prediction market, there is only the unit portfolio (equal number of shares for all  $K$  stocks) that bears no risk, at all, since it can be bought and sold at the bank for exactly 1€ (in the classical CDA prediction markets) and will pay off 1€ in  $T$  for sure. Hence, the return of an unit portfolio is zero and consequently the risk-free return is  $r_f = 0$ .<sup>341</sup> Here the assumption of "play money" used for the trader's entertainment purposes is crucial. Otherwise the trader would not have any incentive to invest in a prediction market as the (nominal) risk-free rate is generally positive,  $r_f > 0$ , in real financial markets.

Risk-averse traders may wish a risk premium to compensate state-dependent wealth resulting from trading in the prediction market. However, "play money" wealth is constant irrespective the final outcome of the risk factor in a classical zero sum game. At least within the prediction market, there is only diversifiable risk. According to arbitrage arguments, diversifiable risks are not compensated as they are accepted voluntarily. The resulting risk premium is  $r_r = 0$ .

---

<sup>341</sup>Rietz (2005) found in lab experiments that a great deal of arbitrage activity with low profits is needed to eliminate behavioral mispricing. Thus, it is possible but unlikely that a profit maximizing arbitrageur will completely drive out biases. Consequently, behavioral mispricing is extremely robust.

Consequently, the expected payoff of a pure or Arrow-Debreu security on state  $k$  is easy to calculate. It is one in state  $k$  and zero else, i.e.  $E_t(X_k(T)) = \pi_k \cdot 1 + (1 - \pi_k) \cdot 0 = \pi_k$ , where  $\pi_k$ , provided that  $\sum_{k=1}^K \pi_k = 1$ , is the market assessment on the probability that outcome  $k$  will realize.

Altogether, the market price for shares of the stock  $k$  exactly equals the market's probability assessment of outcome  $k$  in the equilibrium

$$p_k(t) = \exp(-[r_f + r_r] \Delta t) \cdot E_t(X_k(T)) = \exp(-[0 + 0] \Delta t) \cdot \pi_k = \pi_k.$$

## 18.9 No "Objective" Liquidation

Our discussion of prediction markets has rest on the assumption that shortly after the market close the market will be liquidated according to the finally realized outcome of the risk factor. There are two areas where prediction markets cannot hold these assumptions for different reason: Prediction markets (1) on the success of different concepts and (2) on the outcome of an event far in the future. Additional, there could be no "objective" liquidation when the *Lucas critique* (Lucas 1976) applies - human behavior reacts on prediction market forecasts. E.g., it is hard to judge the forecasts of the limits of growth of the Club of Rome (Meadows et al. 1972) as "wrong" because they initiated a debate on sustainability.

### 18.9.1 Concept Markets

Dahan et al. (2007) call prediction markets on competing concepts *concept markets* that can only measure the success of the "winning" concept. Chan et al. 2001, p. 21, understand them as *beauty contests* in the sense of Keynes (2007). The market value of the concepts is solely derived from the expectations of the traders as it would be too expensive for firms to enter the market with all concepts. In fact, the actual market success of most concepts will never be observed.

Chan et al. 2001, p. 21 & Dahan et al. (2007) conducted such concept markets with students in marketing to assess the potential market shares for bike pumps and crossover vehicles. The markets lasted 10 up to 60 minutes. Traders had got material informing them about the concepts 10 minutes before trading started. As proxy for the market success of the concepts, Dahan et al. (2007) proposed the volume-weighted average market price.

The crucial point for the concept market approach is whether it can yield comparable results to well-established, but expensive marketing methods. Dahan et al. (2007) calculated statistics like the coefficient of determination  $R^2$  and the average absolute difference between the market shares measured by survey studies and those predicted in the concept markets. Although the markets were only run with students, Dahan et al. (2007) found a

high consistency among different markets on the same concepts and significant correlation with independent survey studies. Consequently, the results should improve the closer the traders are to the target group of the potential products.

There have also been other attempts to run concept markets. To estimate the future performance of IT technologies, namely mobile payment solutions in Switzerland, Ondrus et al. (2007) pooled mobile decision-makers from 20 companies. Beside *multi-criteria decision-making* (MCDM) methods, which try to find a "closest to optimal solution" by analyzing expert assessments of quantitative (e.g. costs, weight) and qualitative (quality of service, beauty) criteria, Ondrus et al. (2007) used prediction markets.<sup>342</sup> They found both methods to be complements. While MCDM methods offer an extremely detailed snapshot of a certain point in time, prediction markets present rather a movie over a time period.

### 18.9.2 Long-Term Prediction Markets

Prediction markets are originally designed for short up to medium term (< 2 years) maturities. Although there is certainly a motivation problem, virtual-money markets have also been conducted for several years at the Foresight Exchange. Markets like "Self-driving cars by 2010" and "CO2 level of 2030" were set up in May 1996 and May 2002 and have been actively traded since then.<sup>343</sup> In real money markets with a long-term life-span, motivation is a more critical problem (Berlemann & Nelson 2005, p. 26):

- The payoff in classical prediction markets cannot take place before the event of the underlying will have realized. When this is some years in the future there arises a motivation problem with the traders.
- The safety of the investments and the objectively executed liquidation of the market far in the future needs to be guaranteed by a commonly accepted institution. Hence, there exists a *counterparty risk*.

### Previous Approaches

There are several proposals to solve the mentioned problems. Berlemann et al. (2005) offered a *staggered system* of prediction markets to motivate traders to take part in medium term Bulgarian inflation markets. They accompanied an one year inflation and an one year

---

<sup>342</sup>Ondrus et al. (2007) called their prediction market platform *MarMix*. All prediction markets were CDA markets enhanced with an automatic market maker. They only run with play-money and consisted of Yes/No winner-takes-all stocks. An initial public offer (IPO) mechanism for proposing new technologies was also implemented.

<sup>343</sup>see [www.ideosphere.com/fx-bin/Claim?claim=Cars](http://www.ideosphere.com/fx-bin/Claim?claim=Cars) & [www.foresightexchange.com/fx-bin/Claim?claim=CO2LVL](http://www.foresightexchange.com/fx-bin/Claim?claim=CO2LVL) (last revised August 28, 2008)

exchange rate market by several markets on the same topic but with shorter maturities. Once traders had registered for shorter markets they could immediately trade in the longer term markets. Although this approach is promising, it seems rather suitable for medium term than for long term markets.

Hanson (2008), p. 7, proposes *judging fees* which are fixed amounts of money or a reduction of the stocks' face values. This money is used to pay a judging group that decides which event realized. This method creates incentives to cash out before the market matures to avoid the judgement group to meet. Therefore, judging fees are reasonable for markets that ask whether an event will have realized before the end of the market (an invention, a political regulation, ...). Prediction markets with judgment fees are usually *audit lotteries*. The winning stock will only pay off if the total market investments can finance the judgment. Otherwise the market is closed without any payoff. The judgment fees idea is sensible to reduce the burden for the market operator as it helps to shorten the market life-span. Nevertheless, it does not offer incentives for traders to invest in prediction markets for a long time.

In normal financial markets it is not unusual that considerable amounts are invested for decades, e.g. by pension funds. The main difference between prediction and financial markets is that the latter ones often allow traders to withdraw their investments and gains at any time which is not a preferable feature for CDA prediction markets. The proof of the identity of market prices and market probabilities requires the *lock-in* of the traders' investments up to market close (see section 18.8.2). Moreover, allowing withdrawing bears the risk of thin markets. In mass markets a steady trader switch could be accepted whereas the exit of some traders would be hard to compensate in expert markets with only a fistful traders. Consequently, it is not advisable to allow deinvestment as the market prices are not anymore directly interpretable as probabilities and (expert) markets could dry out.

Despite of all these problems and uncertainties with long term real money prediction markets, there has been one extremely successful market (Hanson 2008, p. 6). In 1992, Pat Thgisni, a physics student, wrote a paper about superscattering.<sup>344</sup> His professor found the idea not promising but Pat Thgisni had not the funds to run an experiment. Hence, he set up a prediction market on a platform and invested \$200 as initial market turnover, 20\$ for market operator fees and \$100 for a judging group. This judging group has been intended to decide if his superscattering proposal is proved until 2013. In 2008 his idea seemed to be proved and Pat Thgisni reaped a total profit of \$700,000 after investing \$20,000 over the years.

### Short-Term Markets on Long-Term Events

---

<sup>344</sup>"If a neutrino could scatter off all the nuclei in a crystal, at once, the interaction could be a billion billion times more powerful, perhaps allowing neutrino telescopes" (Hanson 2008, p. 4).

A new approach has been proposed by the *Bayesian Risk Management* (BRM) research group operating the *Potsdam Climate Exchange* (PCX).<sup>345</sup> The PCX with its prediction market platform PCXtrade<sup>346</sup> conducts academic CDA prediction markets on climate change related topics which will often realize far in the future ( $> 10$  years). As the PCX is an academic project, financed by the German government, there is no institutional framework which can guarantee the "objective" liquidation of such long-term markets. Moreover, in such specific topics, *expert markets* of the PCXtrade are normally conducted with a fistful of experts, only. It seems unrealistic to motivate experts for trading regularly, once a week, a month, or a quarter, for the next several years when not subsidizing the market with substantial amounts of money.

Therefore, the PCX has decided to run *short-term markets on long-term events* which can last between some minutes, some hours, some days, or some weeks - although the risk factor will realize lots of years in the future. The markets are liquidated shortly after the market close according to the last market quote or the (weighted) average of the last market quotes.<sup>347</sup> To avoid foul play prices manipulations shortly before trading stops, the life-span of the markets is stochastic or is not disclosed to the traders.

A special characteristic of such short-term markets is that there generally comes no further substantial external information into the market during the trading time. Consequently, the price shifts are driven by the following pattern: (1) When the market starts, traders set their bid and ask orders according to their initial information level or wait for the placements of other traders. (2) After this first round, traders can observe the aggregated assessments of all traders. This increases their knowledge level which may lead them to change their assessments on the topic. Then, they may change their previous bid and ask orders or place new ones. (3) After the second round the traders get informed how the aggregated market assessments of all traders have changed. This again increases the level of knowledge of the traders and may lead to further trading. Theoretically, this process should iterate with ongoing time. If there actually enters no further external information and the life-span is long enough, a market equilibrium should be reached and trading stops. A challenge for research is to study how a market equilibrium can be distinguished from an illiquid market and

---

<sup>345</sup>see [www.potsdamclimateexchange.org](http://www.potsdamclimateexchange.org)

<sup>346</sup>see [www.forecastmarkets.org/pik/?action=setlanguage&language=EN](http://www.forecastmarkets.org/pik/?action=setlanguage&language=EN)

<sup>347</sup>The Hayek theorem assumes that the last traded price of a stock bases on more (or at least the same) information than the prices before. Chalikh et al. (2005), p. 9, argue that the chance for successful speculation reduces when maturity of the market approaches. Then, it becomes harder to liquidate for bubble prices. Therefore, they advise not to use interim prices as market assessments but the *last traded price* (LTP). Unfortunately, interim prices often transmit important information in longitudinal studies (see Ondrus et al. 2007) and LTPs entail the danger that single traders could be interested to influence the market price in the last seconds as no correction can take place.

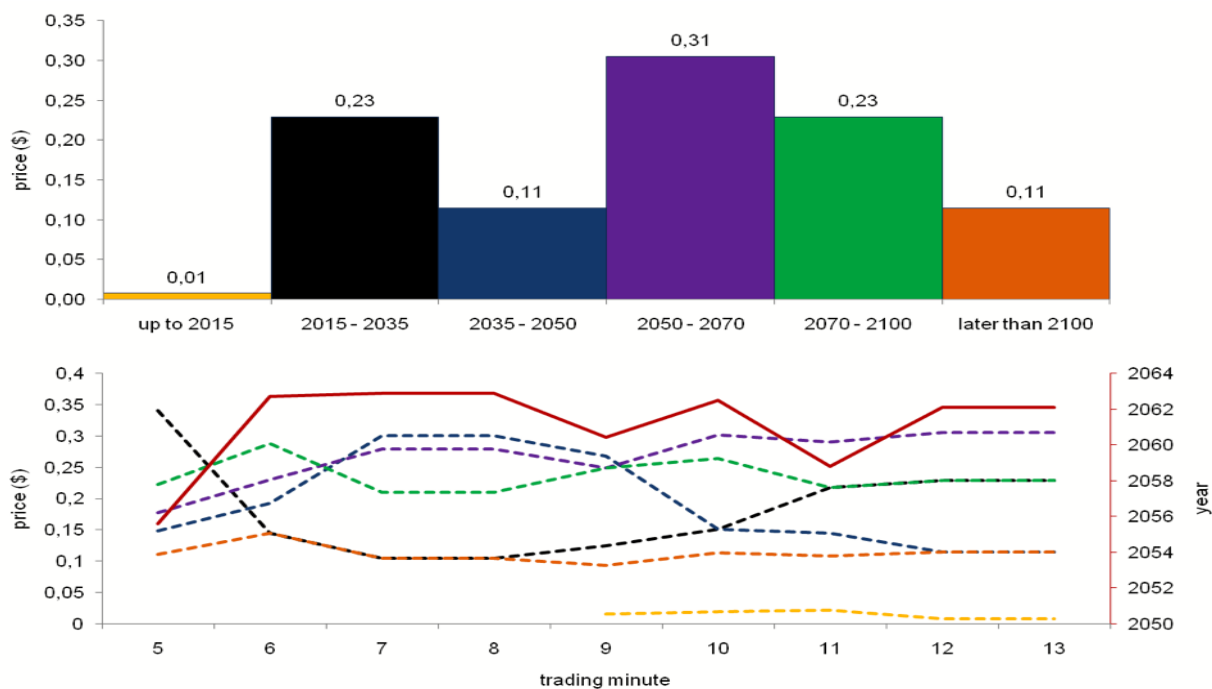


Figure 99: Pittsburgh prediction market - In the upper subplot, the final market prices are presented for the U.S. emission peak. In the lower subplot, you can find the price paths for all stocks (dashed lines) and for the market forecast on the peak year (solid line, right axis). All price paths start when the respective stock has been traded for the first time.

whether it seems reasonable to stop trading in such a situation without offering additional manipulation possibilities. Nevertheless, the main research objective for these markets is to theoretically and empirically prove whether the market prices in such short-term markets can be interpreted as market probabilities.

In the following, we present three prediction markets we have run between June 2008 and June 2009. During the Bayesian Risk Management workshop at Carnegie Mellon University (Pittsburgh, Pennsylvania) on June 11/13, 2008, we run a prediction market on the time when U.S. green house gas emissions will start to fall long-lasting. We could motivate six workshop participants to trade in our market that was closed after 13 minutes. In the upper subplot of figure 99 you can find the final (normalized) market prices at market close.

The surprising result is that traders either believed in a medium-term peak in the period 2015 up to 2035 (23%) or in a long-term peak not before 2050 (65%). There is no exact specification for "later than 2100". From logical reasoning it does not mean "never". In the lower subplot of figure 99, we present the price paths of the different stocks and the price-



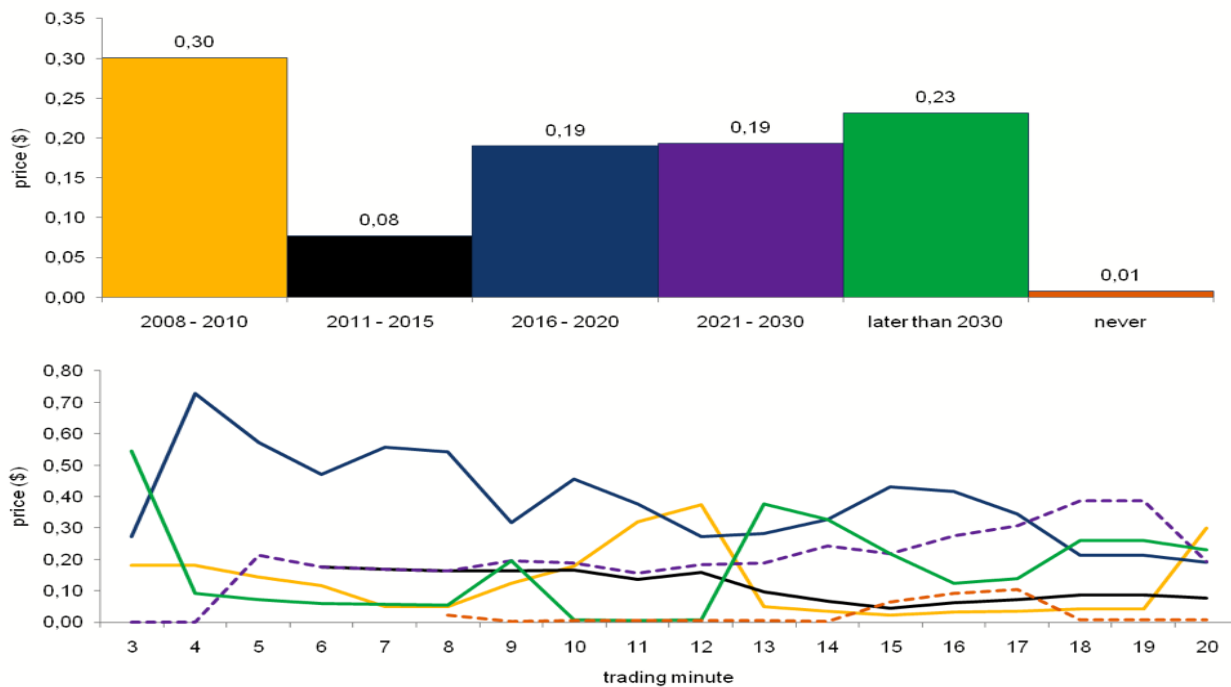


Figure 100: CIRCE prediction market - In the upper subplot, the final market prices are presented for the first year when there is sustainable regulation on water in Spain. In the lower subplot, you can find the price paths for all stocks. All price paths start when the respective stock has been traded for the first time.

weighted mean peak year (2062).<sup>348</sup> The price paths stabilized in the end of the market and trading dried out. Hence, the stop of the market after 13 minutes seems justifiable.

On September 18/19, 2008 we conducted another market during a Bayesian Risk Management training workshop for researchers of the international CIRCE project.<sup>349</sup> Within this workshop we run a prediction market with twelve researcher, mainly from the Mediterranean on the question "When will water prices for Spanish farmers be determined by demand and supply?". The final (normalized) prices (after 20 minutes trading) are presented in the upper subplot of figure 100. Similar to the market on the U.S. emission peak, the market assessment is bimodal. There is high probability on a short-term regulation "2008-2010" - 10% p.a. (three years) — while a regulation in the periods 2011-2015 (five years) and 2021-2030

<sup>348</sup>The stock "later than 2100" is interpreted to stand for the year 2101.

<sup>349</sup>The CIRCE Integrated Project, funded under the European Commission's Sixth Framework Programme, aims to highlight impacts and possible adaptation actions of the climate change in the Mediterranean region including Europe, North Africa and Middle East. See for more details: [www.circeproject.eu](http://www.circeproject.eu) (last revised August 31, 2009).

(ten years) has only low probabilities of 1.6% and 1.9% p.a. The period 2016-2020 (five years) has a slightly higher probability of 3.8% p.a. The results should be taken with a pinch of salt as the market price for the short-term stock "2008-2010" had featured a price jump shortly before the market closed. In retrospect, it is hardly to ascertain whether this jump is in accordance with the general market assessment or whether there were no time for the other traders to correct. The (normalized) market prices before the jump result in an unimodal distribution: "2008-2010" 4% (1.33% p.a.), "2011-2015" 9% (1.8% p.a.), "2016-2020" 21% (4.2% p.a.), "2021-2030" 39% (3.9% p.a.), "later than 2030" 26%, and "never" 1%. A regulation is seen most probable during "2016-2020" while there is nearly no support for the assessment that there will never be such a regulation.

At an advanced training on Bayesian risk management, we ran a prediction market (16 minutes) on the average oil price level in 2020 (in 2009 US-\$). We could gain nine traders — graduate students (global change management) from the University of Applied Science Eberswalde (Germany) on June 12, 2009. The final market price distribution (see upper subplot of figure 101) is unimodal and the interval "\$60.00 – \$99.99" - corresponding to the Brent oil spot price at that time - is seen as most likely (33%). Prices below \$60 (at least \$100) were assumed to have a probability of 29% (37%). Surprisingly, students seemed to think that an oil price in 2020 higher than the peak in mid 2008 is not more likely than 21%. As the price paths seem to have converged, we believe the market life-span of 16 minutes was adequate. Since those students were no experts on commodities, we do not want to overstate the results.

A general assessment of the quality of our market results is hard to find. A widespread method is to calculate the *accuracy* measuring the distance between market forecasts and

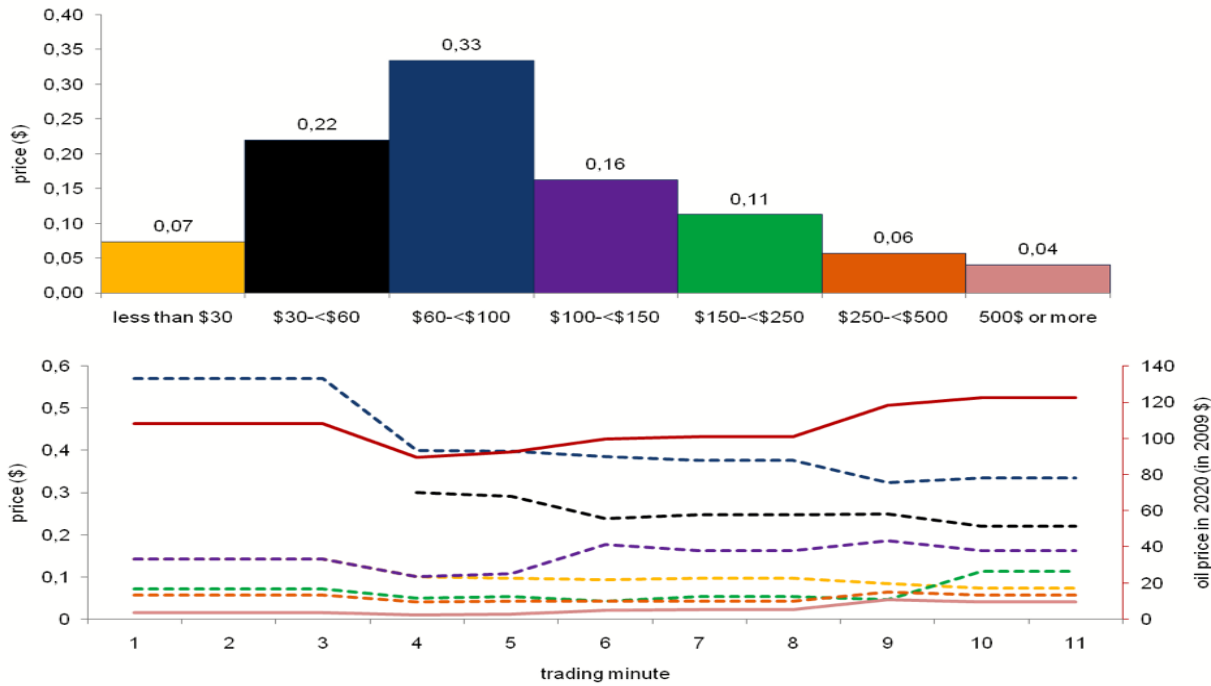


Figure 101: Eberswalde prediction market - In the upper subplot, the final prediction market prices are presented for the average oil price level in 2020 (in 2009 \$). In the lower subplot, you can find the price paths for all stocks (dashed lines) and for the market forecast of the oil price level in 2020 (solid line, right axis). All price paths start when the respective stock has been traded for the first time.

realizations.<sup>350</sup> However, the accuracy is a problematic measure in the context of non-

<sup>350</sup>There exist two basic measures in the evaluation of prediction markets accuracy: *mean absolute error* (MAE) and the *mean squared error* (MSE)

$$MAE = \frac{1}{K} \sum_{k=1}^K |y_k - p_k|, \text{ \& } MSE = \frac{1}{K} \sum_{k=1}^K [y_k - p_k]^2$$

measuring the difference between the (final or weighted) market prices  $p_k$ , for  $k = 1, \dots, K$ , and the final outcome of the risk factor  $Y$  (winner-takes-all markets, e.g. CO2 price in 2013)

$$y_k = \begin{cases} 1, & k \text{ is the realized outcome of } Y \\ 0, & k \text{ is not the realized outcome of } Y \end{cases}$$

or the realized elements  $y_k \in [0, 1]$ , for  $k = 1, \dots, K$  and  $\sum_{k=1}^K y_k = 1$ , of the event  $Y$  (index markets, e.g. election results). Especially in political prediction markets, a forecast is usually preferred that shows small prediction errors for all parties in contrast to a forecast that is very accurate for all but two parties which are totally misspecified. Hence, low MSE could give better indications for possible party coalitions rather than

repeatable events and subjective probabilities which are no frequencies but degrees of beliefs. Consequently, prices of prediction markets only need to (1) continuously respond to new information (Passmore & Cebeci 2003, p. 2) and (2) reflect the aggregated beliefs of the market.

The first claim does not apply to our short-term markets as during the trading minutes no external information on the topic could reach the traders. To assess the aggregation ability of the market, we would need to measure the traders' information levels which is only possible in laboratory experiments where the market operator can spread information by providing traders with different information (Hanson et al. 2006).

Nevertheless, it is possible to analyze the initial and final disagreement of traders.<sup>351</sup> Berlemann & Nelson (2002), pp. 15, propose the *Gini-coefficient*

$$Gini(t) = 1 - \frac{2}{J-1} \left[ J - \frac{\sum_{j=1}^J j \cdot \tilde{w}_j(t)}{\sum_{j=1}^J \tilde{w}_j(t)} \right] \in [0, 1],$$

to calculate the initial ( $t = 0$ ) and final ( $t = T$ ) dispersion of "play money" between the  $J$  traders. The term  $\tilde{w}_j(t)$  is the  $j$ th largest "play money" value. The Gini-coefficient lies between zero and one, while zero means equal distributed funds. The closer  $Gini$  is to one, the more the funds are concentrated to a single trader. In all three markets we have presented above, we cannot measure the initial disagreement by the Gini-coefficient as we have equally endowed all traders with 100 (virtual) units of play money. Hence,  $Gini(0) = 0$ . Nevertheless, we can calculate the diversity of the final trader portfolios:  $Gini_{peak}(T) = 0.047$ ,  $Gini_{water}(T) = 0.104$ ,  $Gini_{oil}(T) = 0.025$ . The coefficients are extremely low. This means that the final portfolio values are rather equal. This we interpret as a sign that there was a low level of disagreement in the markets. All three price distributions are rather flat which indicates high uncertainty while the low Gini-coefficients support the assumption of a low disagreement.

---

a low MAE . In such situations, the value of the MSE, punishing larger errors disproportional compared to small errors, is more informative (Berlemann & Schmidt 2001, p 10).

<sup>351</sup>Disagreement is the traders' diversity of opinions while uncertainty reflects their individual confidence.

## 19 Results and Conclusions

In this dissertation we have discussed the idea of a Bayesian Risk Management (BRM) proposed by the research group Bayesian Risk Solutions (BRS). In the following, we first review our results in respect of our four fundamental goals formulized in the introduction (see section 1). Subsequently, we translate our findings into proposals for a risk management that is better prepared to anticipate upcoming menaces comparable to the current financial crisis since 2007/8.

### 19.1 BRM Structure

Our first goal was to condense the wide scope of BRM into a clear structure based on the concepts Hardcore-Bayesianism (HCB) and Softcore-Bayesianism (SCB). We separated HCB and SCB according to their updating processes that critically depend on the information available (see section 2.1). HCB summarizes all data analysis profiting from the *normative* mathematical Bayesian *updating* via the famous Bayes rule while SCB covers all *informal* (*Bayesian*) *updating* that transparently and conscientiously mobilizes human expertise.

Bayesianism is intrinsically tied to subjective probabilities - a measure for the personal degree of belief. Nevertheless, for a more comprehensive and differentiated overview, we include all efforts that apply the Bayes rule but avoid an incorporation of prior information (see figure 1).

We could identify a clear structure for HCB and SCB and their interrelations (see figure 2). We found SCB as an extremely flexible and stand-alone concept since it accounts for miscellaneous kinds of information. In contrast, HCB is rather restricted because it requires (a) an adequate form of data to apply the Bayes rule, (b) data and model assumptions conjectured to be reliable for inference, as well as (c) advanced information in form of prior distributions. Although uninformative or even improper priors can water down the last claim, HCB always depends on personal appraisals of the reliability of the data and model assumptions. Consequently, there is no HCB without SCB.

### 19.2 Improved Data Analysis

Our second goal was an in-depth discussion on a Bayesian alternative to classical data analysis. We identified several reasons supporting our preference for Bayesian over frequentist statistics:

- The Bayesian understanding does not believe in physical randomness but in uncertainty solely emanating from a lack of information (see section 2).

- Bayesian statistics abandons the traditional idea of maximum likelihood or least square point estimates in favor of a mathematical updating of parameter as well as model uncertainty via the Bayes rule (see sections 6 up to 8). It is rational to account for both parameter and model uncertainty and to make transparent how these propagate into projections and derived risk measures like value-at-risks or option prices. This is highly relevant in the case of a low number of observations.
- The consideration of parameter and model uncertainty results in a convenient side effect. Bayesian statistics does not require any classical hypothesis testing of estimates and of preconditions for those tests. Direct model probabilities can be calculated instead (see sections 8 and 10.3).
- The central Bayesian (parameter) estimation technique is the *Markov Chain Monte Carlo* (MCMC) *simulation* (see section 7). After convergence, this powerful technique efficiently generates (parameter) samples that can describe multivariate uncertainty (on several parameters). Although MCMC samples offer far more information than classical point estimates, this estimation technique is most often not much more time-consuming than classical estimation. In fact, MCMC simulation is often faster for complex models with latent variables - the main reason for the recent popularity of Bayesian statistics even amongst opponents of the Bayesian paradigm. Moreover, the MCMC technique is a universal remedy for an amazing number of heterogeneous estimation problems.

To complement our theoretical discussion we aimed to illustrate the performance of Bayesian statistics on the basis of an empirical study about the European CO2 price risk. We exemplarily fit twelve stochastic processes to four years (April 2005 - May 2009) of European Union Allowance (EUA) price observations. Such a long time-series is unique in empirical work on EUA pricing. Because of a restriction on banking between the first (2005-2007) and second (2008-2012) trading period, we had to deal with two separated spot markets. Our proposal was to profit from a much more liquid future market starting in 2005. By using the cost-of-carry relationship we generated synthetic spot rates from the European Climate Exchange Dec09 future (see section 9.2).

Our time-series covers the turbulent period April/ May 2006 when an oversupply of EUAs was officially announced by the authorities. Therefore, we ran all parameter estimations twice - for the time-series with and without April/ May 2006.

In general, the MCMC algorithms converged fairly quickly and produced impressing estimation results in form of marginal and multivariate posterior parameter distributions (see section 10.2). The variations in the posteriors between the full and short (without April/ May 2006) time-series were reasonable and consistent between all models.

In contrast to frequentist statistics, Bayesian statistics allows for a cardinal performance rating by model probabilities. We could identify the log-variance models with leverage effect as superior for the full and short time-series (see section 10.3). As these models are highly complex we do not believe in their practicability for day-to-day business. Hence, we have analyzed some plain vanilla models (Geometric Brownian motion (GBM), Ornstein-Uhlenbeck, Cox-Ingersoll-Ross, and constant elasticity of variance process). Unfortunately, none of them could systematically outperform.

Another sub-goal of this dissertation was to empirically compare the differences in the risk appraisals between Bayesian and classical statistics. Bayesian statistics generally provided more conservative value-at-risks propagating into higher capital charges (see section 11). Therefore, we embrace Pollard's (Nov. 2007) Bayesian explanation for the capital charge puzzle: Observed capital charges exceed regulatory claims according to Basel II because risk managers unconsciously or off the records account for parameter and model uncertainty.

We additionally analyzed the impact of parameter uncertainty on Bayesian option prices. We found strong evidence that option prices based on the assumption of a GBM - the workhorse in finance - are virtually not affected by parameter uncertainty. This is why option prices only depend on the volatility parameter for which we estimated a highly informative (= concentrated) and symmetrical posterior distribution. However, we could exemplarily identify differences between frequentist and Bayesian option prices for the more complex Geometric Brownian motion with jumps, and the basic log-variance process (see section 12.3).

Beside all pros, we see some obstacles that need to be overcome for a better mainstreaming of Bayesian statistics: (a) There is a lack of user-optimized application packages. At the moment, the most prevalent Bayesian statistics software WinBugs - that we partly used for our estimations - is far from an Apple-like user interface. (b) Lectures in statistics often cover Bayesian statistics as a special subarea of classical statistics whereas it is an alternative understanding of statistics. (c) Bayesian statistics is often discredited because of its subjectivity. In fact, we believe this is not a bug but a feature. Nevertheless, in most cases subjectivity can be reduced to a negligible level (see section 6.2).

### 19.3 Risk Management Beyond Data Analysis

Our third goal was to find *practicable* guidelines and options to *transparently* and *rationaly* incorporate human expertise in risk management *beyond* pure *data analysis*.

First, we identified risk analysts, experts, and decision makers as the most important persons (or groups) in risk management (see section 14).

Based on this taxonomy, we designed the *integrated Bayesian Risk Analysis* (iBRA) concept

(see section 15) proposing a basic framework for an expertise-driven risk quantification and management accompanied by an *informal Bayesian updating* - a permanent reflection and challenge of the findings.

Although iBRA highlights the risk analyst as expert, it allows for a simple integration of in house and external experts when the additional efforts are well justified. Therefore, we overviewed the scope of expert elicitation (see sections 16 up to 18).

We designed iBRA as a multi-step guideline for the risk analysis of a target variable (e.g. company value) mostly depending on several critical factors (e.g. production costs, demand). It starts with the identification of the critical factors that can be prioritized by means of probability-impact maps and variability-impact tables (see figures 74 and 75). Interrelations between the factors are often complex. Risk maps (see figure 76) or influence diagrams (see figure 78) are helpful to structure the problem in a model.

For simplicity, iBRA propagates a separate quantification of the identified factors (see section 15.2 for SRQ step). In section 16.3 we discussed dominant elicitation techniques that can be used for quantification. Mostly, they are well grounded on theory but lack practical feasibility. Therefore, we can only recommend the fixed interval technique and variable interval technique by which experts (or the risk analyst) express their uncertainty in form of interval weights or by the calibration of elicitation distributions (see sections 16.3.5 and 16.3.6).

Unfortunately, people tend to biases when they have to formulize their uncertainty (see section 16.1). That is why we conducted a study on the projection quality of human (credible) interval statements (see section 16.4). By the means of Bayesian statistics, we could quantify a posterior distribution for the credible level of different kinds of credible intervals. We found people tend to overconfidence when they are asked for intervals covering the actual outcome of a factor with high or 50% probability - i.e. they stated too narrow intervals. People also showed evidence for overestimation.

An ignorance of critical dependencies between factors or auto-correlation of factors can fundamentally underestimate relevant risks. Although iBRA summarizes guidelines and techniques to deal with such problems (see section 15.3 for the IRQ step), we advised to focus on those factor interrelations that the analyst thinks to be most important. Analyses of cause-and-effect chains can help to identify independent factors or to define transformations like factor ratios.

If no elimination of dependencies is possible, we proposed the envelope method as well as our visual copula technique for situations with a few correlated factors. The envelope method models correlation by forming a range for dependent factors given the outcome of a numeraire factor (see section 16.5.2). If there is no hierarchic dependence structure, we conceived our visual copula technique for discussing the correlation of the factors on the basis of scatter



plots produced by a copula function for different correlation levels (see section 16.5.1).

When the number of correlated factors is too large, we see no alternatives to stress tests and Bayesian scenario analysis including probability weights for scenarios (see section 15.3). The concept of consistency matrices can support the creation of scenarios (see figure 79).

The approximation of auto-correlations is another tricky task for an expertise-driven risk management. We identified stochastic processes, weighted projections and time-series copulas as promising approaches (see section 15.3.2): The calibration of stochastic processes is easy to implement but their flexibility is modest (see section 16.5.3). Weighted projections in form of scenarios are extremely flexible but can result in contradictory marginal distributions for the factors. Alternatively, we introduced the time-series copula approach (see section 16.5.4). It combines a modified envelope method with the copula concept and allows for a flexible calibration of positive auto-correlation.

Although data analysis is also prone to ambiguous results, an expertise-driven risk management in particular has to deal with a diversity of opinions. Within iBRA we discussed options to present competing assessments to decision makers (see section 15.5).

Optimally, there is a self-aggregation of competing assessments by the direct or indirect interaction of experts (see below for behavioral aggregation tools). Unfortunately, there is often no such behavioral aggregation. Then disagreement can fully be displayed. Alternatively, competing expert assessments can be summarized by mathematical aggregation in form of a basic assessment weighting or a Bayesian expert aggregation.

We found basic assessment (or expert) weighting as most flexible and practicable (see 16.6.1). It is a simple arithmetic or geometric averaging of different personal uncertainties - weighting of different continuous or discrete distributions instead of point estimates. Since there is no "objective" weighting scheme, we suggested experts to self-rate their assessments. In a survey we elicited people's confidence in such expert self-ratings (see section 15.5). Interestingly, people tend to rely in an expert statement the more certain the expert is but partly tend to mistrust "very confident" experts.

Based on expert statements, Bayesian expert aggregation is a normative updating of a decision maker's advanced uncertainty (see 16.6.2). It is rather academic but useful for some theoretical reasoning because it propagates to judge information according to its newsworthiness instead of the pure number of supporters.

Beside elicitation techniques, a Bayesian Risk Management requires some elicitation tools bringing together experts and elicitation techniques (see section 17 and figure 97).

Expert panels feature a direct (e.g. focus groups) or indirect interaction (Delphi groups) of experts while expert interviews are more or less complex interrogations of experts. We call panels and interviews direct elicitation tools. They require a direct contact (incl. telephone) of experts and an inquirer. This is extremely expensive and time consuming.

We instead favored surveys (= indirect elicitation tools) that do not need an arrangement of appointments between experts and inquirer. Although classical surveys like traditional questionnaires are more dominant, we focused on automated elicitation tools like online questionnaires and prediction markets.

Online questionnaires minimize the level of cooperation for the experts improving acceptance for integration into day-to-day business (see section 17.2). We set up the PCXquest - our online questionnaire platform. It summarizes four basic question types corresponding to the constant and variable interval elicitation technique (see above). The PCXquest allowed us to run parallel surveys each comprising a number of questions where experts can additionally self-rate their assessments.

Prediction markets are an alternative. Their excellent prediction performance and their theoretical fundamentals have been well documented by research (see section 18). Prediction markets profit from an interaction between experts via a pricing mechanism. They measure probability assessments in an unconstrained form of "willingness to bet".

Classical prediction markets need to be liquidated according to the final outcome of a risk factor. This restricts them to short-term or medium-term problems. Unfortunately, climate related risks are mostly medium-term up to long-term problems. Therefore, we refined the idea of concept markets, successfully applied in marketing, to our idea of *short-term markets on long-term events* (see section 18.9.2) - a kind of "beauty contest" in the sense of Keynes (2007).

So far, we have run three test markets with different kinds of traders (students, climate scientists, risk experts) on our prediction market platform PCXtrade. We could achieve encouraging results. However, we have seen that experts without experience in trading need to acquaint themselves with this tool. Consequently, we have concluded that we must train the experts more intensively beforehand. Moreover, we plan to implement Hanson's market maker to reduce market complexity (see section 18.6.7).

## 19.4 Non-Arbitrary Risk Management

Our fourth goal was to discuss options to curb the risk of arbitrariness of an expertise-driven risk management. Unfortunately, the basic Bayesian theory does not provide any guidance. Except for the normative mathematical Bayesian updating, it is purely descriptive. In fact, the concept of validation or falsification of human uncertainty is meaningless for a Bayesian thinker since Bayesianism understands probabilistic statements as an expression of the personal level of uncertainty (= degree of belief).

For addressing the risk of arbitrariness, we moot the proposal of a Bayesian due diligence. Bayesian due diligence offers a framework to judge subjective statements according to their

usability for risk management.

Since there is no chance for undisputed "objective" assessments in real live, we compiled some basic inter-subjective criteria rational people should agree on (see figure 82 in section 15.5 for our reputation map). We hope they allow to rate the level of reliable information an assessment is based on.

Our proposal is outside the traditional Bayesian and of course frequentist paradigm. Hence we embedded our Bayesian due diligence concept into a new risk classification referencing to the familiar terms "risk" and "uncertainty" (see section 2.2): We found

- A risk factor is described by probabilistic assessments judged as reliable.
- Although an uncertain factor has been identified as relevant it cannot be described by probabilistic assessments judged as reliable.

We believe a probabilistic quantification according to our iBRA concept (see above) is only advisable for risk factors whereas we oppose an arbitrary quantification of uncertain factors. Instead, we favor a transparent documentation and stress tests for such factors.

Since our Bayesian due diligence propagates a conservative risk management we enhanced our risk classification by black swans - critical but unforeseeable factors. Consequently, a risk management should mobilize risk cushions that can absorb unexpected adverse events of minor up to medium size. Besides, it should seek for new information sources that might provide a better appraisal of possible risks and opportunities. Indeed, a black swan for the risk analyst might be a clear risk for an expert.

## 19.5 Conclusions for the Current Financial Crisis

This dissertation was deliberately designed for providing a Bayesian toolbox for universal purpose. In consideration of the financial crisis since 2007/9 we want to complete our work with a short discussion of potential Bayesian contributions to an improved risk management without arguing that Bayesian Risk Management could have had saved the world from the devastating crisis. In particular, we focus on some catchwords: model failure, insufficient data, and frivolous subjectivism.

In their groundbreaking paper, Colander et al. 2009 believe in a *model failure* explanation for the financial crisis. Economists have developed oversimplified risk and derivatives models propagating unjustifiable stability. In fact, these models could not anticipate the crisis. Maybe, they have even contributed to it.

Bayesian statistics does not deliver new kinds of models but can contribute to a more prudent analysis of existing models (that, of course, have had often proven inadequate during the crisis). It accounts for the generally limited number of observations by a transparent

disclosure of data-induced parameter uncertainty hitherto ignored in quantitative risk models (see section 6). This generally results in more conservative risk assessments (see section 11). Additionally, Bayesian statistics can cardinally measure the relative performance of competing models by posterior model probabilities (see section 8). This allows to identify inferior models more clearly than by classical information criteria. In fact, we were amazed at the distinct model ratings resulting from our empirical analysis of the CO<sub>2</sub> price risk (see section 10.3).

Another important explanation suspects the financial crisis to originate from a risk management relying on *insufficient data* that does not reflect coming menaces. Roughly speaking, more complex (empirical) models would have not anticipated the crisis because the available data from the last decades does not cover comparable turbulent phases. The project "Mainstreaming of Climate Risks and Opportunities in the Financial Sector" has been pioneering in propagating the basic idea of insufficient or devalued historical data (see the project publications at [www.climate-mainstreaming.net](http://www.climate-mainstreaming.net) and especially Onischka 2008). Because of the infrequency or singularity of extreme (financial) crises, we favor unorthodox approaches.

For instance, the Supervisory Capital Assessment Program (SCAP) was a stress test, conducted by the Federal Reserve System, for the 19 largest U.S. financial institutions (see Fed Apr. 2009 and Fed May 2009 for design and results). SCAP analyzed whether these institutions could withstand a worsening recession and further market turmoil. Amongst others, the adequacy of their capitalization was calculated for a pessimistic scenario based on some key macroeconomic factors (GDP:  $-3.3\%/ 0.5\%$ , unemployment rate:  $8.9\%/ 10.3\%$ , house prices:  $-22\%/ -7\%$  for 2009/10).

In the first half of 2010 there has been conducted a similar stress test by the Committee of European Banking Supervisors (CEBS) for European banks. Only seven out of 91 banks (e.g. Hypo Real Estate Holding) failed the test.<sup>352</sup> In retrospective of the subsequent turbulences of the Irish banks the risk assessments of the stress test have been widely criticized. Nevertheless scenario based stress tests are strongly supported from the Bayesian point of view. However they need to be propagated transparently to prevent pressure groups of banks to erode stress scenarios and to advocate legal loopholes in underhand dealings (e.g. only government bonds in the trading book were stressed).

In this dissertation, we have summarized guidelines supporting the risk management to quantify information not contained in historical data (see section 15). Since the designing of scenarios requires statements on diverse factors, an incorporation of in house and external experts seems advisable.

Fortunately, we have summarized elicitation techniques (see section 16) and elicitation tools (see sections 17 and 18) for mobilizing experts' expertise. Unfortunately, an expertise-driven

---

<sup>352</sup>see <http://stress-test.c-eps.org/documents/Summaryreport.pdf>

risk management without any strict supervision would provoke some opposition because of the threat of human insufficiency (see section 16.1 for pitfalls in expert elicitation) and, maybe even more relevant, moral hazard.

In fact, there are more and more people localizing the causes for the crisis in a *frivolous subjectivism*, i.e. (unconsciously practiced) Bayesianism. Exemplarily, we cite Prof. Richard A. Posner (judge on the United States Court of Appeals for the Seventh Circuit in Chicago) and the financial juggler and billionaire George Soros who lament unsupervised subjective risk management in the financial sector where quantitative models only serve as a fig leaf.<sup>353</sup> In personal conversations, (credit) risk managers from different large European banks acknowledged this perception.

In answer to this, we have started the discussion on a criteria based subjective risk management where assessments need to pass some minimum standards. We call this Bayesian due diligence (see sections 2.2 and 15.5). Nevertheless, this is not more than a first step in - we hope - the right direction to nip coming crises in the bud.

---

<sup>353</sup> At iTunes, you can download the interviews: "Posner Says U.S. Will Spend \$13 Trillion to Fight a Depression", Bloomberg on the Economy (April 27, 2009) and "Soros Says Economic Crisis Was Self-Generated by System, Policy", Bloomberg on the Economy (July 8, 2009). See the popular science book Posner (2009)[258].

## A Appendix

### A.1 Distributions

In the following we describe some important distributions often used in this dissertation:

**beta distribution:**  $P_k \sim \text{Dirichlet}(\alpha, \beta)$  (see Dirichlet distribution)

**Dirichlet distribution:**  $P = (P_1, \dots, P_K) \sim \text{Dirichlet}(a_1, \dots, a_K)$  with pdf

$$\text{Dirichlet}(p|a) = \frac{1}{B(a)} \prod_{k=1}^K p_k^{a_k-1},$$

is a multivariate distribution for the random vector  $P = (P_1, \dots, P_K)' \in [0, 1]^K$  under the condition  $\sum_{k=1}^K p_k = 1$ . The term  $B(a) = \prod_{k=1}^K \Gamma(a_k) / \Gamma(\sum_{k=1}^K a_k)$  is the beta function, where  $\Gamma(a) = \int_0^\infty \exp(-\tau) \tau^{a-1} d\tau$  is the gamma function, and  $a = (a_1, \dots, a_K)'$  is the respective parameter vector. The marginal mean and variance are  $E(P_k) = a_k/\tilde{a}$  and

$$\text{Var}(P_k) = \frac{a_k [\tilde{a} - a_k]}{\tilde{a}^2 [\tilde{a} + 1]},$$

where  $\tilde{a} = \sum_{k=1}^K a_k$ . Dirichlet distributed variables  $P_k$ ,  $k = 1, \dots, K$ , can be sampled by  $p_k = z_k / \sum_{k=1}^K z_k$ , where  $Z_k \sim \chi_{2 \cdot a_k}^2$ . In the special case of  $P = (P_1, P_2)' = (P_1, 1 - P_1)'$  the Dirichlet distribution equals a *beta distribution*

$$\text{Dirichlet}(p|a) = \frac{1}{B(\alpha, \beta)} p_k^{\alpha-1} p_k^{\beta-1},$$

where  $\alpha = a_1$ ,  $\beta = a_2$ ,  $E(P_k) = \alpha / [\alpha + \beta]$  and

$$\text{Var}(P_k) = \frac{\alpha \beta}{[\alpha + \beta]^2 [\alpha + \beta + 1]}.$$

**gamma distribution:**  $X \sim \text{Gam}(a, b)$  with pdf

$$f_X(x) = \frac{b^{-a}}{\Gamma(a)} x^{a-1} \exp(-x/b),$$

where  $\Gamma(a) = \int_0^\infty \exp(-\tau) \tau^{a-1} d\tau$  is the gamma function.  $E(X) = ab$  and  $\text{Var}(X) = ab^2$ . *Take care for the specification of the gamma distribution in your software! There is no general binding formulation.*

**inverse gamma distribution:**  $X \sim IG(a, b)$  with pdf

$$f_X(x) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp(-b/x),$$

where  $\Gamma(a) = \int_0^\infty \exp(-\tau) \tau^{a-1} d\tau$  is the gamma function.  $E(X) = b/[a-1]$ , for  $a > 1$ , and  $Var(X) = b^2/[a-1]^2[a-2]$ , for  $a > 2$ . Unfortunately, the inverse gamma distribution is often not implemented in statistical software. Fortunately, there is the link  $X = g(Y) = 1/Y \sim IG(a, b)$  with  $Y \sim Gam(a, 1/b)$ . Hence, we need just to sample  $y$  and then transform it to  $x = 1/y$ .

**multivariate Gaussian (or normal) distribution:**  $X = (X_1, \dots, X_K) \sim N_K(\mu, \Sigma)$  with pdf

$$f_X(x) = [2\pi]^{-K/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} [x - \mu]' \Sigma^{-1} [x - \mu]\right),$$

where  $\mu = (\mu_1, \dots, \mu_K)'$  is the vector of means and  $\Sigma$  is the covariance matrix. In the univariate case ( $K = 1$ ) this simplifies to

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left[\frac{x - \mu}{\sigma}\right]^2\right),$$

where  $E(X) = \mu$  and  $Std(X) = \sigma$ .

**log-Gaussian (or normal) distribution:**  $X \sim \log N(\mu, \sigma)$  with pdf

$$f_X(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \left[\frac{\ln x - \mu}{\sigma}\right]^2\right),$$

where  $E(\ln X) = \mu$  and  $Std(\ln X) = \sigma$ . Consequently, it follows  $E(X) = \exp(\mu - \sigma^2/2)$  and  $Var(X) = [\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$ .

**non-central Student's t distribution:**  $X \sim t_v(\mu, \sigma)$  with pdf

$$f_X(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)\sigma} \left[1 + \frac{1}{v} \left[\frac{x - \mu}{\sigma}\right]^2\right]^{-[v+1]/2}$$

where  $v$  is the degree of freedom,  $E(X) = \mu$ , and  $Var(X) = \sigma v/[v-2]$ , for  $v > 2$ . The Student's t distribution is the special case for  $\mu = 0$  and  $\sigma = 1$ .

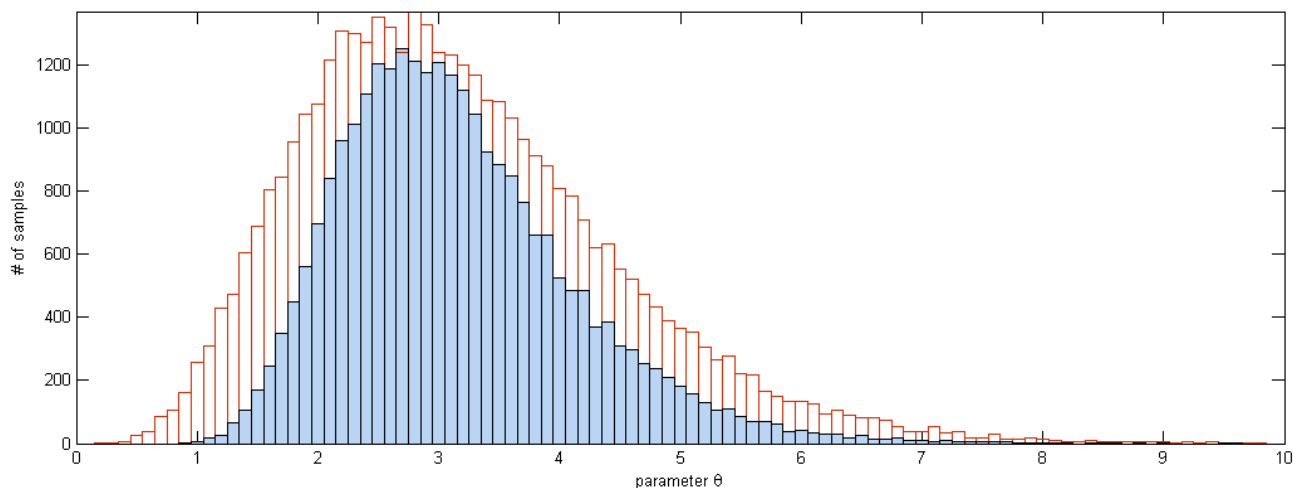


Figure 102: rejection &amp; importance sampling

## A.2 Rejection & Importance Sampling

Assume, we are interested in a target density  $\tau(\theta)$  which is too complex to sample from.<sup>354</sup> However, there is a standard distribution  $\rho(\theta)$  which guarantees  $\tau(\theta) \leq c \cdot \rho(\theta)$  for the whole domain of  $\theta$ , where  $c > 1$  is a scaling factor. The distribution  $\rho(\theta)$  is called *envelope density* as  $c \cdot \rho(\theta)$  needs to envelope the target density everywhere.

This simulation algorithm is called *rejection sampling*:

1. Set  $w = 1$ .
2. Sample a proposal value  $\tilde{\theta}$  from the envelope density  $\rho(\theta)$ .
3. Accept the proposal value,  $\theta^{(w)} = \tilde{\theta}$ , as  $w$ th draw from the target distribution  $\tau(\theta)$  with acceptance probability  $\alpha(\tilde{\theta}) = \tau(\tilde{\theta}) / [c \cdot \rho(\tilde{\theta})]$  and set  $w = w + 1$ . If  $w < W$  go back to 2.

An example for rejection sampling is plotted in figure 102. We generated a total of 40,000 samples (red & blue boxes) from the proposal density  $\rho(\theta)$  (here  $\tilde{\theta} \sim \text{Gam}(6, 0.52)$ ). Finally, just 26,534 samples (blue boxes, only) were accepted to stem from the target density  $\tau(\theta)$  (here  $\theta \sim \log N(1.1, 0.3)$ ). The acceptance probability was calculated with with a scaling factor  $c = 1.5$ .

<sup>354</sup>In the following we base our discussion on Rachev et al. (2008), pp. 63, and Chan & Wong (2006), pp. 110.



Generally, every standard distribution can be used as envelope density by increasing  $c$ . It is only required that  $\rho(\theta)$  has positive density for the full domain of the target density. However, the speed of the rejection sampling critically depends on  $c$ . For an efficient envelope density a  $c \approx 1$  is sufficient to guarantee dominance.

A related concept is *importance sampling* which incorporates all samples generated from the proposal distribution, meaning  $\theta^{(w)} = \tilde{\theta}$  with probability one. The expected value of a real value function  $g(\theta)$ ,  $E(g(\theta)) = \int g(\theta) \tau(\theta) d\theta$ , can be approximated by the arithmetic mean of  $W$  sampled parameters  $\theta^{(w)}$ ,  $w = 1, \dots, W$ . Actually, the samples need to be generated by  $\tau(\theta)$ , which might be intractable. Fortunately, a simple extension of the formula

$$E(g(\theta)) = \int g(\theta) \frac{\tau(\theta)}{\rho(\theta)} \rho(\theta) d\theta \approx \sum_{w=1}^W g(\theta^{(w)}) w(\theta^{(w)})$$

allows to sample i.i.d.  $\theta^{(w)}$ ,  $w = 1, \dots, W$ , from a standard distribution  $\rho(\theta)$ . The samples originating from the wrong distribution are simply reweighted by  $w(\theta) = \tau(\theta) / \rho(\theta)$ .

The approximation quality of the importance sampling estimator critically depends on the variance of the weights. Consequently, the mode and shape of proposal and target density should be matched.

### A.3 Sampling-Importance-Resampling (SIR)

The *sampling-importance-resampling* (SIR) approach of Rubin (1988) is the most important alternative to the MCMC samplers. The SIR bases on the idea of sequential updating where the *updated prior* corresponds to the old posterior

$$\pi(\theta | y_0, \dots, y_{t-1}, y_t) \propto f_Y(y_t | \theta) \cdot \pi(\theta | y_0, \dots, y_{t-1}).$$

Then the SIR algorithm<sup>355</sup> allows to sample from  $\pi(\theta | y_0, \dots, y_{t-1}, y_t)$

1. Sample  $\theta^{(w)}$ ,  $w = 1, \dots, W$ , from the updated prior  $\pi(\theta | y_0, \dots, y_{t-1})$ .
2. Calculate the mean likelihood  $\bar{L}_{y_t} = \sum_{w=1}^W f_{Y_t}(y_t | \theta^{(w)}) / W$ .
3. Calculate the importance weight  $\omega_w = f_{Y_t}(y_t | \theta^{(w)}) / [W \cdot \bar{L}_{y_t}]$  for  $\theta^{(w)}$ ,  $w = 1, \dots, W$ .
4. Resample  $\theta^{(\tilde{w})}$ ,  $\tilde{w} = 1, \dots, \tilde{W} < W$ , from  $\pi(\theta | y_0, \dots, y_{t-1}, y_t)$

<sup>355</sup>see Rubin (1988), Pitt & Shephard (1999), p. 591, Bunnin et al. (2002), pp. 39, or Koch (2007) for more details

- (a) Construct  $W$  intervals on  $(0, 1]$  where interval  $w$  is  $(a_w, b_w]$  with  $a_w = \sum_{k=1}^{w-1} \omega_k$  and  $b_w = \sum_{k=1}^w \omega_k$ .
- (b) Generate  $\tilde{W}$  independent samples  $u_{\tilde{w}}$  from  $Uni(0, 1)$ .
- (c) The  $\tilde{w}$ th sample from  $\pi(\theta | y_0, \dots, y_{t-1}, y_t)$  is  $\theta^{(\tilde{w})} = \theta^{(w)}$  such that  $u_{\tilde{w}} \in (a_w, b_w]$ .

## A.4 Ito's Lemma

Assume a stochastic differential process (SDP)

$$dS_t = \mu(t, s_t, \theta) dt + \sigma(t, s_t, \theta) dW_t,$$

where  $dS_t = S_{t+dt} - s_t$  and  $dW_t \sim N(0, \sqrt{dt})$ , for  $W_0 = 0$ , is the differential of a Wiener process. The terms  $\mu(t, s_t, \theta)$  and  $\sigma(t, s_t, \theta)$  are functions of the time  $t$ , the current realization  $S_t = s_t$ , and a vector of parameters  $\theta$ . Then *Ito's lemma* offers the SDP formulation for a function  $F_t = g(t, S_t)$  by

$$dF_t = \left[ \frac{\partial F_t}{\partial t} + \mu_t \frac{\partial F_t}{\partial S_t} + \frac{1}{2} \sigma_t^2 \frac{\partial^2 F_t}{\partial S_t^2} \right] dt + \sigma_t \frac{\partial F_t}{\partial S_t} dW_t,$$

where  $dF_t = F_{t+dt} - f_t$ ,  $\mu_t = \mu(t, s_t, \theta)$  and  $\sigma_t = \sigma(t, s_t, \theta)$ .

Take care whether the function  $F_t = g(t, S_t)$  depends on  $t$  explicitly. In financial applications the function often reduces to  $F_t = g(S_t)$ . Then  $\partial F_t / \partial t = 0$ .

Ito's lemma can also be extended to jump-diffusion processes

$$dS_t = \mu(t, s_t, \theta) dt + \sigma(t, s_t, \theta) dW_t + Z(t, s_t, \theta) dN_t,$$

where  $Z_t = Z(t, s_t, \theta)$ ,  $t = 1, \dots, T$ , are independent jump magnitudes and  $N_t$  is a counting process. In the case of independent  $dW_t$  and  $dN_t$ , Ito's lemma can be formulized for jump-diffusions by

$$dF_t = \left[ \frac{\partial F_t}{\partial t} + \mu_t \frac{\partial F_t}{\partial S_t} + \frac{1}{2} \sigma_t^2 \frac{\partial^2 F_t}{\partial S_t^2} \right] dt + \sigma_t \frac{\partial F_t}{\partial S_t} dW_t + [g(t, s_t + Z_t) - g(t, s_t)] dN_t.$$

For a better illustration we apply Ito's lemma for Merton's (1976) jump-diffusion model

$$dS_t = \left[ \mu + \frac{1}{2} \sigma^2 \right] s_t \cdot dt + \sigma s_t \cdot dW_t + [e^J - 1] s_t dN_t,$$

where  $J$  is a Gaussian variable,  $N_t$  is a Poisson process with independent  $dW_t$  and  $dN_t$ . As the model is generally estimated for log-returns  $F_t = g(S_t) = \ln S_t$ , we apply Ito's lemma

$$\begin{aligned} d \ln S_t &= \left[ \left[ \mu + \frac{1}{2} \sigma^2 \right] s_t \frac{\partial F_t}{\partial S_t} + \frac{1}{2} [\sigma s_t]^2 \frac{\partial^2 F_t}{\partial S_t^2} \right] dt + \sigma s_t \frac{\partial F_t}{\partial S_t} dW_t \\ &\quad + \left[ \ln (s_t + [e^J - 1] s_t) - \ln s_t \right] dN_t \\ &= \left[ \mu + \frac{1}{2} \sigma^2 - \frac{1}{2} \sigma^2 \right] dt + \sigma dW_t + \ln \frac{s_t + [e^J - 1] s_t}{s_t} \cdot dN_t \\ &= \mu dt + \sigma dW_t + J \cdot dN_t, \end{aligned}$$

where  $J$  is the jump magnitude (or size, width) of the log-returns.

## A.5 Univariate Density Transformation

Assume the density of an uncertain univariate variable  $X$  is  $f_X(x)$  and a second uncertain variable  $Y = g(X)$  is just a function of  $X$ . Then the density  $f_Y(y)$  of  $Y$  can be calculated by the density transformation

$$f_Y(y) = \left| \frac{\partial g^{-1}(y)}{\partial y} \right| \cdot f_X(g^{-1}(y)),$$

where  $|\cdot|$  is the Jacobian. For the density transformation the inverse  $X = g^{-1}(Y)$  needs to exist.

## A.6 Truncation of Distributions

A truncated variable  $Y \sim F_Y$  results from  $X \sim F_X$  restricted to  $(a, b)$ . The pdf calculates by

$$f_Y(y) = \frac{f_X(y)}{F_X(b) - F_X(a)}.$$

Samples from  $Y$  can easily be generated as long as the inverse cdf,  $F_X^{-1}(\cdot)$ , is implemented in your statistical software package. The procedure is the following: Sample  $u$  from  $U \sim Uni(0, 1)$  and calculate

$$y = F_X^{-1}([F_X(b) - F_X(a)] \cdot u + F_X(a)).$$

## A.7 MCMC Samplers

### A.7.1 GBM Gibbs Sampler

Our Gibbs sampler for the GBM can be described by the following algorithm:

1. Define  $\tilde{\delta}^{(w)} = [\tilde{\sigma}^2]^{(w)}$  as well as start values  $\tilde{\mu}^{(0)}$  and  $\tilde{\delta}^{(0)}$ . Set  $w = 1$ .
2. Sample  $[\tilde{\mu}^{(w)} | \tilde{\delta}^{(w-1)}, y] \sim N(m_1, s_1)$ , where
 
$$m_1 = s_1^2 \cdot \left[ m_0/s_0^2 + \sum_{t=1}^T y_t/\tilde{\delta}^{(w-1)} \right] \quad \& \quad s_1^2 = \left[ 1/s_0^2 + T/\tilde{\delta}^{(w-1)} \right]^{-1}.$$
3. Sample  $[\tilde{\delta}^{(w)} | \tilde{\mu}^{(w)}, y] \sim IG(a_1, b_1)$ ,<sup>356</sup> where
 
$$a_1 = a_0 + T/2 \quad \& \quad b_1 = b_0 + \frac{1}{2} \sum_{t=1}^T [y_t - \tilde{\mu}^{(w)}]^2.$$
4. Set  $w = w + 1$ . As long as  $w \leq W$  go back to step 2.
5. Discard the pre-convergence samples  $\theta^{(w)} = \left( \tilde{\mu}^{(w)}, \tilde{\delta}^{(w)} \right)'$ ,  $w \leq B < W$ , where  $B$  marks the burn-in phase.

### A.7.2 GBMJ Gibbs Sampler

Our GBMJ Gibbs sampler is defined by the following steps:

1. Define  $\tilde{\delta}^{(w)} = [\tilde{\sigma}^2]^{(w)}$  and  $\tilde{\delta}_J^{(w)} = [\tilde{\sigma}_J^2]^{(w)}$ . Set start values  $\theta^{(0)} = \left( \tilde{\mu}^{(0)}, \tilde{\delta}^{(0)}, \mu_J^{(0)}, \tilde{\delta}_J^{(0)}, \tilde{p}_J^{(0)} \right)'$ ,  $x^{(0)} = (j^{(0)}, n^{(0)})'$ ,<sup>357</sup> and  $w = 1$ .
2. Sample  $[\tilde{\mu}^{(w)} | \theta_{-\tilde{\mu}}^{(w-1)}, x^{(w-1)}, y] \sim N(m_1, s_1)$ , where
 
$$m_1 = s_1^2 \cdot \left[ m_0/s_0^2 + \sum_{t=1}^T \left[ y_t - j_t^{(w-1)} \cdot n_t^{(w-1)} \right] / \tilde{\delta}^{(w-1)} \right]$$

$$s_1^2 = \left[ 1/s_0^2 + T/\tilde{\delta}^{(w-1)} \right]^{-1}.$$
3. Sample  $[\tilde{\delta}^{(w)} | \theta_{-\tilde{\delta}}^{(\sim w)}, x^{(w-1)}, y] \sim IG(a_1, b_1)$ ,<sup>358</sup> where<sup>359</sup>

$$a_1 = a_0 + T/2$$

$$b_1 = b_0 + \frac{1}{2} \sum_{t=1}^T \left[ y_t - j_t^{(w-1)} \cdot n_t^{(w-1)} - \tilde{\mu}^{(w)} \right]^2.$$

<sup>356</sup>See appendix A.1 for the generation of inverse gamma distributed variables and A.5 for the algorithm of an univariate density transformation.

<sup>357</sup>We set  $j_1^{(0)} = y_t/2$ , for  $t = 1, \dots, T$ , and  $n_t^{(0)} = 0$ , if  $y_t^2 \leq \overline{y_t^2}$ , else  $n_t^{(0)} = 1$ .

<sup>358</sup>The formulation  $\theta_{-\tilde{\delta}}^{(\sim w)}$  reflects the sampled parameter vector  $\theta$  without  $\tilde{\delta} = \tilde{\sigma}^2$ . The index states that  $\theta$  comprises parameters sampled in step  $w$ , here  $\tilde{\mu}^{(w)}$ , and parameters sampled in step  $w - 1$ , here  $\mu_J^{(w-1)}$ ,  $\tilde{\delta}_J^{(w-1)}$ , and  $p_J^{(w-1)}$ .

<sup>359</sup>See appendix A.1 for the generation of inverse gamma distributed variables and A.5 for the algorithm of univariate density transformation.

4. Sample  $\left[ p_J^{(w)} \mid \theta_{-p_J}^{(\sim w)}, x^{(w-1)}, y \right] \sim \text{Beta}(\alpha_1, \beta_1)$ , where

$$\alpha_1 = \alpha_0 + \sum_{t=1}^T n_t^{(w-1)} \quad \& \quad \beta_1 = \beta_0 + T + \sum_{t=1}^T n_t^{(w-1)}$$

5. Sample  $\left[ \tilde{\mu}_J^{(w)} \mid \theta_{-\tilde{\mu}_J}^{(w-1)/(w)}, x^{(w-1)}, y \right] \sim N\left(m_1^{(J)}, s_1^{(J)}\right)$ , where

$$\begin{aligned} m_1^{(J)} &= \varrho_1 \cdot \left[ m_0^{(J)} / \varrho_0 + \sum_{t=1}^T \left[ j_t^{(w-1)} \cdot n_t^{(w-1)} \right] / \tilde{\delta}_J^{(w-1)} \right] \\ \varrho_1 &= \left[ 1 / \varrho_0 + \sum_{t=1}^T n_t^{(w-1)} / \tilde{\delta}_J^{(w-1)} \right]^{-1}, \end{aligned}$$

and  $\varrho_k = \left[ s_k^{(J)} \right]^2$ ,  $k = 0, 1$ .

6. Sample  $\left[ \tilde{\delta}_J^{(w)} \mid \theta_{-\tilde{\delta}_J}^{(w)}, x^{(w-1)}, y \right] \sim \text{IG}\left(a_1^{(J)}, b_1^{(J)}\right)$ , where

$$\begin{aligned} a_1^{(J)} &= a_0^{(J)} + \sum_{t=1}^T n_t^{(w-1)} / 2 \\ b_1^{(J)} &= b_0^{(J)} + \frac{1}{2} \sum_{t=1}^T \left[ j_t^{(w-1)} - \tilde{\mu}_J^{(w)} \right] \Big|_{n_t^{(w-1)}=1}^2. \end{aligned}$$

7. Sample  $\left[ N_t^{(w)} \mid \theta^{(w)}, j^{(w-1)}, y \right]$ , for  $t = 1, \dots, T$ , from

$$P\left(N_t^{(w)} = 1 \mid \theta^{(w)}, j^{(w-1)}, y\right) = \frac{f_{Y_t}(y_t \mid \cdot, 0) \cdot \tilde{p}_J^{(w)}}{f_{Y_t}(y_t \mid \cdot, 0) \cdot \tilde{p}_J^{(w)} + f_{Y_t}(y_t \mid \cdot, 1) \cdot \left[ 1 - \tilde{p}_J^{(w)} \right]},$$

where

$$f_{Y_t}(y_t \mid \cdot, k) = \left( y_t \mid \tilde{\mu}^{(w)} + k \cdot \tilde{\mu}_J^{(w)}, \sqrt{\tilde{\delta}^{(w)} + k \cdot \tilde{\delta}_J^{(w)}} \right)$$

and  $P\left(N_t^{(w)} = 0 \mid \cdot, y\right) = 1 - P\left(N_t^{(w)} = 1 \mid \cdot, y\right)$ .

8. Sample  $\left[ J_t^{(w)} \mid \theta^{(w)}, n^{(w)}, y \right] \sim N(d, c)$ , for  $t = 1, \dots, T$ ,

$$d = c^2 \cdot \left[ [y_t - \tilde{\mu}^{(w)}] / \tilde{\delta}^{(w)} + \tilde{\mu}_J^{(w)} / \tilde{\delta}_J^{(w)} \right] \quad \& \quad c^2 = \left[ 1 / \tilde{\delta} + 1 / \tilde{\delta}_J^{(w)} \right]^{-1}.$$

9. Set  $w = w + 1$ . If  $w \leq W$  go back to step 2.

10. Discard the pre-convergence samples  $\theta^{(w)}$ ,  $w \leq B < W$ .

### A.7.3 OU MCMC Sampler

We implement the following hybrid Metropolis-Hastings sampler for the OU process on log-returns:

1. Run a ML estimation on

$$\theta^{ML} = \arg \max_{\theta} \prod_{t=1}^{T-1} \phi \left( \ln s_{t+1} | \tilde{\mu} + [\ln s_t - \tilde{\mu}] \exp(-\kappa \Delta), \sigma \sqrt{\frac{1 - \exp(-2\kappa \Delta)}{2\kappa}} \right),$$

and store the respective covariance matrix  $\Sigma^{ML}$ .

2. Set start values  $\theta^{(0)} = (\kappa^{(0)}, \tilde{\mu}^{(0)}, \sigma^{(0)})' = \theta^{ML}$ . Set hyper parameters  $m_{\kappa} = \kappa^{ML}$ ,  $m_{\tilde{\mu}} = \tilde{\mu}^{ML}$ ,  $m_{\sigma} = \sigma^{ML}$ ,  $\varsigma_{\kappa}^2 = \Sigma_{\kappa}^{ML}$ ,  $\varsigma_{\tilde{\mu}}^2 = \Sigma_{\tilde{\mu}}^{ML} \cdot c$ , and  $\varsigma_{\sigma}^2 = \Sigma_{\sigma}^{ML} \cdot c$  for a factor  $c$ . Set  $w = 1$ .
3. Use a Random Walk Metropolis sampler for  $\kappa$

- (a) Generate a proposal value from  $\kappa^* \sim N(\kappa^{(w-1)}, \varrho_{\kappa})|_{\kappa^* > 0}$ , where  $\varrho_{\kappa}$  is the standard deviation of the proposal distribution.
- (b) Accept proposal value  $\kappa^{(w)} = \kappa^*$  with probability  $\min(\tau(\kappa^*)/\tau(\kappa^{(w-1)}), 1)$ , where

$$\tau(\kappa) = \prod_{t=1}^{T-1} \phi(\ln s_{t+1} | m(t, w-1), \varsigma(w-1)) \cdot \phi(\kappa | m_{\kappa}, \varsigma_{\kappa})|_{\kappa \geq 0},$$

for  $m(t, w-1) = \tilde{\mu}^{(w-1)} + [\ln s_t - \tilde{\mu}^{(w-1)}] \exp(-\kappa \Delta)$  and

$$\varsigma(w-1) = \sigma^{(w-1)} \sqrt{[1 - \exp(-2\kappa \Delta)]/2\kappa}.$$

Otherwise, set  $\kappa^{(w)} = \kappa^{(w-1)}$ .

4. Use a Random Walk Metropolis sampler for  $\tilde{\mu}$

- (a) Generate a proposal value from  $\tilde{\mu}^* \sim N(\tilde{\mu}^{(w-1)}, \varrho_{\tilde{\mu}})|_{\tilde{\mu}^* \in [\tilde{\mu}_{\min}, \tilde{\mu}_{\max}]}$ , where  $\varrho_{\tilde{\mu}}$  is the standard deviation of the proposal distribution.
- (b) Accept proposal value  $\tilde{\mu}^{(w)} = \tilde{\mu}^*$  with probability  $\min(\tau(\tilde{\mu}^*)/\tau(\tilde{\mu}^{(w-1)}), 1)$ , where

$$\tau(\tilde{\mu}) = \prod_{t=1}^{T-1} \phi(\ln s_{t+1} | m(t, w-1), \varsigma(w-1)) \cdot \phi(\tilde{\mu} | m_{\tilde{\mu}}, \varsigma_{\tilde{\mu}})|_{\tilde{\mu} \in [\tilde{\mu}_{\min}, \tilde{\mu}_{\max}]},$$

for  $m(t, w - 1) = \tilde{\mu} + [\ln s_t - \tilde{\mu}^{(w-1)}] \exp(-\kappa^{(w)} \Delta)$  and

$$\varsigma(w - 1) = \sigma^{(w-1)} \sqrt{[1 - \exp(-2\kappa^{(w)} \Delta)] / 2\kappa^{(w)}}.$$

Otherwise, set  $\kappa^{(w)} = \kappa^{(w-1)}$ .

5. Use a Random Walk Metropolis sampler for  $\sigma$

- (a) Generate a proposal value from  $\sigma^* \sim N(\sigma^{(w-1)}, \varrho_\sigma) |_{\sigma^* > 0}$ , where  $\varrho_\sigma$  is the standard deviation of the proposal distribution.
- (b) Accept proposal value  $\sigma^{(w)} = \sigma^*$  with probability  $\min(\tau(\sigma^*) / \tau(\sigma^{(w-1)}), 1)$ , where

$$\tau(\sigma) = \prod_{t=1}^{T-1} \phi(\ln s_{t+1} | m(t, w - 1), \varsigma(w - 1)) \cdot \phi(\sigma | m_\sigma, \varsigma_\sigma) |_{\sigma > 0},$$

for  $m(t, w - 1) = \tilde{\mu}^{(w)} + [\ln s_t - \tilde{\mu}^{(w)}] \exp(-\kappa^{(w)} \Delta)$  and

$$\varsigma(w - 1) = \sigma \sqrt{[1 - \exp(-2\kappa^{(w)} \Delta)] / 2\kappa^{(w)}}.$$

Otherwise, set  $\kappa^{(w)} = \kappa^{(w-1)}$ .

6. Set  $w = w + 1$ . If  $w \leq W$  go back to step 3.

#### A.7.4 Student's t GARCH MCMC Sampler

Our MCMC sampler for the *GARCH* (1, 1) model with the Student's t errors is a hybrid Metropolis-Hastings algorithm:

- 1. Set start values  $\theta^{(0)} = (\gamma_0^{(0)}, \gamma_1^{(0)}, \omega^{(0)}, \alpha^{(0)}, \beta^{(0)}, v^{(0)})'$  and  $w = 1$ .
- 2. Calculate  $\tilde{\sigma}_t^{(w)}$ , based on  $\omega^{(w-1)}, \alpha^{(w-1)}, \beta^{(w-1)}$ , and  $u_t^{(w)} = y_t - \gamma_0^{(w-1)} - \gamma_1^{(w-1)} r_t$  for  $t = 1, \dots, T$ .<sup>360</sup>
- 3. Sample  $\eta_t^{(w)} | v^{(w-1)}, y_t \sim \text{Gam}(a, 1/b_t)$ ,  $t = 1, \dots, T$ , where  $a = [v^{(w-1)} + 1] / 2$  and

$$b_t = \frac{[u_t^{(w)}]^2}{2 [\tilde{\sigma}_t^{(w)}]^2} + \frac{v^{(w-1)}}{2}.$$

---

<sup>360</sup>We set  $\tilde{\sigma}_0^2 = \widehat{\text{Var}}(y)$  and  $u_0 = 0$ .

4. Sample  $v^{(w)}$  via the Griddy Gibbs sampler from the log-kernel

$$\frac{Tv_*}{2} \ln \frac{v_*}{2} - T \ln \left( \Gamma \left( \frac{v_*}{2} \right) \right) + \left[ \frac{v_*}{2} - 1 \right] \cdot \sum_{t=1}^T \ln \eta_t^{(w)} - \frac{v_*}{2} \cdot \sum_{t=1}^T \eta_t^{(w)} - v_* \lambda_0,$$

for the fixed points  $v_* = 2.1, 2.2, \dots$

5. Sample  $(\gamma_0^{(w)}, \gamma_1^{(w)})' \sim N_2(m_\gamma, \Sigma_\gamma)$ , where

$$m_\gamma = \Sigma_\gamma [z'D^{-1}z\gamma^{LS} + \Sigma_0^{-1}m_0]^{-1} \quad \& \quad \Sigma_\gamma = [z'D^{-1}z + \Sigma_0^{-1}]^{-1}.$$

The matrix  $z = (\mathbf{1}, r)$  is a  $T \times 2$  matrix with  $\mathbf{1}$  as a vector of ones and the EURIBOR rates  $r = (r_1, \dots, r_T)'$ . The matrix  $D$  is a  $T \times T$  identity matrix with  $[\tilde{\sigma}_t^2]^{(w)} / \eta_t^{(w)}$ ,  $t = 1, \dots, T$ , on the main diagonal, and  $\gamma^{LS} = [z'D^{-1}z]^{-1} z'y$  is the least square estimate on  $y_t = \gamma_0 + \gamma_1 \cdot r_t$ .

6. Sample the GARCH parameters  $\omega^{(w)}$ ,  $\alpha^{(w)}$ , and  $\beta^{(w)}$ .<sup>361</sup>

- (a) Run a ML estimation

$$\theta^{ML} = \arg \max_{\omega, \alpha, \beta} \prod_{t=1}^T \phi \left( y_t \mid \gamma_0^{(w-1)} + \gamma_1^{(w)} r_t, \tilde{\sigma}_{*t} / \sqrt{\eta_t^{(w)}} \right),$$

where  $\theta^{ML} = (\omega^{ML}, \alpha^{ML}, \beta^{ML})'$  and  $\tilde{\sigma}_{*t}^2 = \omega + \alpha \cdot [u_{t-1}^{(w)}]^2 + \beta \cdot \tilde{\sigma}_{*,t-1}^2$ . Store the covariance matrix,  $\Sigma^{ML}$ , of the ML estimates.

- (b) Generate a proposal vector  $\theta^* \sim t_{v^{(w)}}(\theta^{ML}, \Sigma^{ML})$ . To account for the prior distributions repeat sampling as long as  $\theta^{ML} \notin (\vartheta_{\min}, \vartheta_{\max}]^3$ .

- (c) Set  $(\omega^{(w)}, \alpha^{(w)}, \beta^{(w)})' = \theta^{ML}$  with an acceptance probability

$$\min(w(\theta^{ML}) / w(\theta^{GARCH}), 1),$$

where  $\theta^{GARCH} = (\omega^{(w-1)}, \alpha^{(w-1)}, \beta^{(w-1)})'$  and

$$w(a) = \pi \left( a \mid \gamma_0^{(w)}, \gamma_1^{(w)}, v^{(w)}, \eta^{(w)}, y \right) / t_{v^{(w)}} \left( a \mid \theta^{ML}, \Sigma^{ML} \right).$$

Otherwise, set  $(\omega^{(w)}, \alpha^{(w)}, \beta^{(w)})' = (\omega^{(w-1)}, \alpha^{(w-1)}, \beta^{(w-1)})'$ .

7. Set  $w = w + 1$ . If  $w \leq W$  go back to step 2.

<sup>361</sup>When the GARCH proposals have not been accepted we repeat this step twice at most. This can avoid to repeat all other steps before to get a new GARCH parameter vector.



### A.7.5 MSGARCH MCMC Sampler

Our MCMC algorithm is a hybrid Metropolis-Hastings with the following steps:

1. Define  $\tilde{\delta}_t^{(w)} = [\tilde{\sigma}_t^2]^{(w)}$  and set start values  $\theta^{(0)}, y_0, x^{(0)}, \tilde{\sigma}_0^2$ ,<sup>362</sup> and  $w = 1$ .
2. Sample from  $\mu_k^{(w)} \left| \theta_{-\mu_k}^{(\sim w)}, x^{(w-1)}, y \sim N(m_k, s_k), \right.$ <sup>363</sup> where

$$m_k = s_k^2 \cdot \left[ \sum_{t|k}^T \hat{y}_t^{(k)} / \tilde{\delta}_t^{(w-1)} + m_0 / s_0^2 \right] \quad \& \quad s_k^2 = \left[ \sum_{t|k}^T 1 / \tilde{\delta}_t^{(w-1)} + 1 / s_0^2 \right]^{-1}.$$

The mean log-return under regime  $k$  is marked by  $\hat{y}_t^{(k)}$  and  $\sum_{t|k}^T$  is only the sum over dates in regime  $k$ .

3. Calculate  $u_t^{(w)} = u_t - \mu_k^{(w)}$  and

$$\tilde{\delta}_t^{(w-1)} = \omega_{x_t^{(w-1)}}^{(w-1)} + \alpha_{x_t^{(w-1)}}^{(w-1)} \cdot \left[ u_{t-1}^{(w)} \right]^2 + \beta_{x_t^{(w-1)}}^{(w-1)} \cdot \tilde{\delta}_{t-1}^{(w)},$$

for  $t = 1, \dots, T$ . We assume  $\tilde{\delta}_0^{(w-1)} = \sum_{t=1}^4 \tilde{\delta}_t^{(w-1)} \cdot weight_t$ , where  $weight = (4, 3, 2, 1) / 10$ .

4. Count the regime transitions from regime  $k$  to regime  $l$  by  $\lambda_{kl}^{(w)} = \sum_{t=1}^T 1_{[x_{t-1}^{(w-1)}=k]} \cdot 1_{[x_t^{(w-1)}=l]}$ .
5. Sample  $P_{kk}^{(w)} \left| \theta_{-pk}^{(w)/(w-1)}, y, x^{(w-1)} \sim Beta \left( \alpha_{0k} + \lambda_{k1}^{(w)}, \beta_{0k} + \lambda_{k2}^{(w)} \right), \right.$  for  $k = 1, 2$ .<sup>364</sup>
6. Sample the regimes  $x_t^{(w)}$ , for  $t = 1, \dots, T$ , separately.

- (a) Calculate the (shortened) conditional posterior  $\kappa_t^{(w)} = P \left( X_t^{(w)} = k \left| \theta_t^{(w)}, x_{-t}^{(w)}, y \right. \right) = 1 / [1 + \exp(c_l - c_k)]$  defined in equation 9 where  $c_j = \sum_{i=t}^T \ln \phi \left( y_i \left| \mu_{x_i}^{(w)}, \tilde{\sigma}_i^{(w)} \right. \right) + \ln p_{x_{t-1}, k}^{(w)} + \ln p_{k, x_{t+1}}^{(w)}$ .

<sup>362</sup>(1) For a fast convergence, our algorithm requires heterogeneous start values for the GARCH parameters in the different regimes. (2) We set  $y_0 = \sum_{t=1}^5 y_t / 5$ . (3) For all  $y_t$  where  $y_t^2 > 3/T \cdot \sum_{t=1}^T y_t^2$  we set  $x_t^{(0)} = 2$ , otherwise  $x_t^{(0)} = 1$ . (4) We match  $\tilde{\sigma}_0^2$  to the variance of the log-returns under the regime  $x_0^{(0)}$ .

<sup>363</sup> $\theta_{-\mu_k}^{(\sim w)}$  means all parameters except  $\mu_k$ . Some of the parameter values have been sampled in MCMC step  $w$  others in  $w - 1$ .

<sup>364</sup>For  $K \geq 3$  regimes we can receive conjugate priors when we implement Dirichlet distributed priors (see Dirichlet distribution in section A.1). Then we can sample from  $P_k \left| \theta_{-pk}, x, y \sim Dirichlet(a_{k1} + \lambda_{k1}, \dots, a_{kK} + \lambda_{kK}) \right.$ , where  $P_k = (P_{k1}, \dots, P_{kK})'$  is the  $k$ th row of the transition matrix.

- (b) Sample a variable  $u$  from  $U \sim Uni(0, 1)$ . If  $\kappa_t^{(w)} \leq u$  chose  $x_t^{(w)} = 1$ , otherwise  $x_t^{(w)} = 2$ .
7. Sample  $\pi \left( \vartheta_k | \theta_{-\vartheta_k}, x_{-t}^{(w)}, y \right) \propto \prod_{t=1}^T \phi \left( y_t | \mu_{x_t}^{(w)}, \tilde{\sigma}_t^{(w)} \right)$ , for  $\vartheta_k = (\omega_k, \alpha_k, \beta_k)'$  and  $k = 1, 2$ , in a two step procedure with a multivariate Student's t independence chain Metropolis-Hastings algorithm (Rachev et al. 2008, pp. 70/ pp. 210)

- (a) Use implemented ML tools in your software to find GARCH parameters maximizing the log-likelihood

$$\vartheta_k^{ML} = \arg \max_{\vartheta_k} \prod_{t=1}^T \phi \left( y_t | \mu_{x_t}^{(w)}, \tilde{\sigma}_t^{(w)} \right), \quad k = 1, 2,$$

where  $\tilde{\delta}_t^{(w-1)} = \omega_{x_t^{(w-1)}}^{ML} + \alpha_{x_t^{(w-1)}}^{ML} \cdot \left[ u_{t-1}^{(w)} \right]^2 + \beta_{x_t^{(w-1)}}^{ML} \cdot \tilde{\delta}_{t-1}^{(w)}$ . Generally, ML packages offer the covariance matrix  $\hat{\Sigma}^{(k,w)}$  of the estimated parameters  $\vartheta_k^{ML}$ .

- (b) Sample a proposal vector  $\vartheta_k^* \sim t_{v^*} \left( \vartheta_k^{ML}, \hat{\Sigma}^{(k,w)} \right)$ ,<sup>365</sup> under restriction  $\vartheta_k \in [\vartheta_{\min}, \vartheta_{\max}]$ , and let it compete with the old vector  $\vartheta_k^{(w-1)}$  by the independence chain Metropolis-Hastings algorithm.

8. Set  $w = w + 1$  and go back to 2 as long as  $w < W$ .

## A.8 Copulas

### A.8.1 Basics on Copulas

Dependence structures between different risk factors  $(Y_1, \dots, Y_n)'$  have a huge influence on risk management as extreme outcomes of different risk factors can simultaneously emerge. For simplicity, the multivariate Gaussian distribution is often used to model dependency. It is implemented in most statistical software packages, e.g. Matlab or R. However the multivariate Gaussian distribution belongs to the class of *multivariate distributions* which require marginal distributions  $f_{Y_i}(y_i) = f_{Y_i}(y_i | \theta)$ ,  $i = 1, \dots, n$ , from the same distribution although the parameter values  $\theta$  can be different. Additionally, the Gaussian approach can only model linear dependency via a correlation matrix.

For our purposes these requirements are too strict, especially when we want to elicit expert statements on prices defined on the positive line, only. We need the broader class of *meta-distributions* which generate (non)-linear dependence structures but do not premise equally

<sup>365</sup>A sample from a multivariate non-central Student's t distribution,  $Y \sim t_v(\mu, \Sigma)$ , can be generated via  $Y = \mu - \Sigma^{-1/2} [v - 2] / v \cdot X$ , where  $X \sim t_v(0, 1)$ .

distributed marginal distributions. Such meta-distributions can be generated by the copula approach.

A copula is a multivariate uniform distribution on the  $[0, 1]^n$  hypercube for the quantiles of the risk factors  $Y_1, \dots, Y_n$

$$C(u_1, \dots, u_n) = P(Y_1 \leq y_1, \dots, Y_n \leq y_n) = F_Y(Y_1 \leq y_1, \dots, Y_n \leq y_n),$$

where quantile  $u_i$  is the realization of  $U_i = F_{Y_i}(Y_i) \sim Uni(0, 1)$ . The dependence structure between the quantiles  $U_i$  is controlled by the copula function which allows to indirectly pattern dependence structures between the risk factors.

The *theorem of Sklar* guarantees that every joint distribution can be modeled by its marginal distributions and a copula function. If all marginals are continuous distributions the copula is unique. Consequently, different meta-distributions can result from different copulas given the same marginal distributions. This allows to separate the meta-distribution into the marginal distributions and the dependence structure

$$\begin{aligned} P(Y_1 = y_1, \dots, Y_n = y_n) &= f_Y(Y_1 = y_1, \dots, Y_n = y_n) \\ &= \prod_{i=1}^n f_{Y_i}(y_i) \cdot c(F_{Y_1}(y_1), \dots, F_{Y_n}(y_n)), \end{aligned}$$

where

$$c(u_1, \dots, u_n) = \frac{\partial C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}.$$

The predominant classes of copulas are the elliptical and the Archimedean copulas. The *elliptical copulas* have normally no closed form and are restricted to symmetrical dependence structures. However, they are defined by a  $n$ -dimensional correlation matrix which allows flexible (symmetrical)  $n$ -dimensional dependence structures. Moreover, the elliptical copulas can easily be sampled with statistical software when standard distributions are implemented. Two common elliptical copulas are summarized in table 47.

	$C(u_1, \dots, u_n   \vartheta)$	dependence parameter vector $\vartheta$
Gaussian	$\Phi_{\Sigma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$	$\rho_{ij} \in [-1, 1]$
Student's t	$t_{v, \Sigma}(t_v^{-1}(u_1), \dots, t_v^{-1}(u_n))$	$\rho_{ij} \in [-1, 1], v \in (0, \infty)$

Table 47: elliptical copulas

The class of *Archimedean copulas* features closed (additive) forms

$$C(F_{Y_1}(y_1), \dots, F_{Y_n}(y_n) | \vartheta) = AC(y_1, \dots, y_n | \vartheta) = \Psi^{-1}\left(\sum_{i=1}^n \Psi(F_{Y_i}(y_i))\right),$$

where  $\lim_{z \rightarrow 0} \Psi(z) \rightarrow \infty$ ,  $\Psi(1) = 1$ ,  $\Psi'(z) < 0$ , and  $\Psi''(z) > 0$ . As Archimedean copulas often only possess one or two dependence parameters they are mostly used to model bivariate dependence structures. Alternatively, hierarchical models (see Jouini & Clemen 1996, p. 448)

$$AC_n(y_1, \dots, y_n | \vartheta_i, i = 1, \dots, n) = AC(AC_{n-1}(y_1, \dots, y_{n-1} | \vartheta_i, i = 1, \dots, n-1), y_n | \vartheta_n)$$

can be constructed for higher dimensions. Another feature is that some Archimedean copulas can reproduce positive dependence, only. This can be bypassed by  $C(F_{Y_1}(y_1), 1 - F_{Y_2}(y_2) | \vartheta)$ . The most dominant Archimedean copulas are summarized in table 48 (see Schmidt 2007, p. 23, or Jouini & Clemen 1996, p. 450).

$n = 2$	$\Psi(u_i)$	$C(u_1, u_2   \vartheta)$		
Clayton	$\frac{u_i^{-\vartheta} - 1}{\vartheta}$	$[\sum_{i=1}^n u_i^{-\vartheta} + n - 1]^{-1/\vartheta}$	$\vartheta \in (0, \infty)$	
			$\vartheta \rightarrow 0$	independence
			$\vartheta \rightarrow \infty$	comonotony
Gumble	$[-\ln u_i]^\vartheta$	$\exp\left(-\left[\sum_{i=1}^n [-\ln u_i]^\vartheta\right]^{-1/\vartheta}\right)$	$\vartheta \in [1, \infty)$	
			$\vartheta = 1$	independence
			$\vartheta \rightarrow \infty$	comonotony
Frank $_{n=2}$	$\ln\left(\frac{e^{-\vartheta u_i} - 1}{e^{-\vartheta} - 1}\right)$	$-\frac{1}{\vartheta} \ln\left(1 + \frac{[e^{-\vartheta u_1} - 1][e^{-\vartheta u_2} - 1]}{e^{-\vartheta} - 1}\right)$	$\vartheta \in (-\infty, \infty) \setminus 0$	
			$\vartheta \rightarrow 0$	independence
			$\vartheta \rightarrow \pm\infty$	comonotony
Frank $_{n>2}$		$\log_\vartheta\left(1 + \frac{[\vartheta^{u_1} - 1] \dots [\vartheta^{u_n} - 1]}{[\vartheta - 1]^{n-1}}\right)$	$\vartheta \in (0, 1)$	

Table 48: Archimedean copulas

The copula approach allows for a rather flexible generation of correlation structures between risk factors. These are not restricted to linearity. Even non-linear correlation is allowed. This flexibility requires the concept of concordance:

- *concordance*: high (low) values  $Y_1$  tend to realize with high (low) values of  $Y_2$ 
  - *comonotony*: perfect positive (non-)linear dependency between  $Y_1$  and  $Y_2$
- *discordance*: high (low) values  $Y_1$  tend to realize with low (high) values of  $Y_2$ 
  - *counter-comonotony*: perfect negative (non-)linear dependency between  $Y_1$  and  $Y_2$

The *Fréchet-Hoeffding boundaries* define the maximum range of correlation levels that can be described by a copula

$$W(u_1, \dots, u_n) \leq C(u_1, \dots, u_n) \leq \underbrace{M(u_1, \dots, u_n)}_{\text{comonotony}},$$

where  $W(u_1, \dots, u_n) = \max(\sum_{i=1}^n u_i - n + 1, 0)$  and  $M(u_1, \dots, u_n) = \min_i u_i$  are the lower and upper bounds. The upper bound corresponds to *comonotony* while *counter-comonotony* can only be reached by a copula in the bivariate case. The independence copula is  $\Pi(u_1, \dots, u_n) = \prod_{i=1}^n u_i$ .

Generally, the widespread bivariate *Pearson's correlation coefficient*  $\rho \in [\rho_{\min}, \rho_{\max}]$  is not capable to describe the full range of correlation between the Fréchet-Hoeffding boundaries as it only measures linear correlation. Perfect/ deterministic non-linear positive or negative correlation can not be identified because  $\rho_{\min} > -1$  and  $\rho_{\max} < 1$ . Moreover, Pearson's correlation is no suitable concordance measure as it additionally depends on the marginal distributions. In contrast, *rank correlations* only require measurability on an ordinal scale, are more robust to outliers, and solely depend on the copula function. In the copula context, Spearman's rank correlation coefficient  $\rho^S$  and Kendall's tau  $\tau$  are dominant.

- *Spearman's rank correlation* (see Schmidt 2007, p. 16)

$$\begin{aligned} \rho^S &= \text{Corr}_{\text{Person}}(F_{Y_1}(Y_1), F_{Y_2}(Y_2)) = 4 \int_0^1 \int_0^1 [C(u_1, u_2) - u_1 \cdot u_2] du_1 du_2 \\ &= 1 - \frac{6}{n[n^2 - 1]} \sum_{k=1}^K [\text{Rank}(y_{1k}) - \text{Rank}(y_{2k})]^2 \in [-1, 1]. \end{aligned}$$

calculates the Pearson's correlation between the quantiles of the risk factors, where  $y_{ik}$  is the  $k$ th observation of factor  $i = 1, 2$ . Technically, it measures the differences between the ranks. Hence, it requires the assumption of identical distances between the ranks.

- *Kendall's tau* (see Genest & Favre 2007, p. 351)

$$\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1 = \frac{4}{n[n - 1]} \lambda - 1 \in [-1, 1],$$

where  $\lambda = \sum_{k=1}^{K-1} \sum_{l=k+1}^K 1_{(\tilde{y}_{2k} < \tilde{y}_{2l})}$  and  $\tilde{y}_{2k}$  is the  $k$ th observation of risk factor  $Y_2$  when ordered according to  $Y_1$ . It can also be described by the difference between the probability for concordance and discordance and only measures whether the rank of  $Y_1$  is lower/ equal/ higher compared to the rank of  $Y_2$ . It at least needs the assumption that the tuples  $(Y_{1k}, Y_{2k})'$  and  $(Y_{1l}, Y_{2l})'$ , for  $k \neq l$ , are independent.

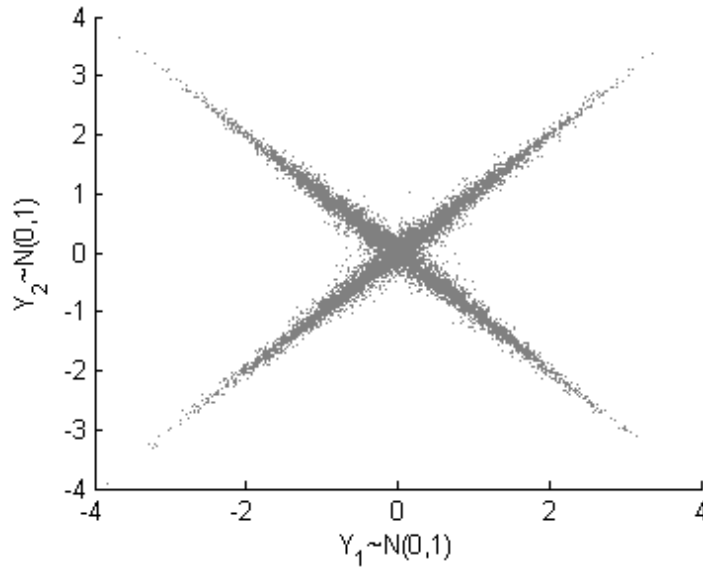


Figure 103: Student’s t copula for  $\nu = 0.1$  and zero (rank) correlation, i.e.  $\rho = \rho^{Spearman} = \tau = 0$

Unfortunately, even rank correlation measures cannot identify all possible correlation structures. Although all correlation produced by the multivariate Gaussian distribution is fully described by (rank) correlation measures, this is not true for the Student’s t copula which is also defined by the degree of freedom  $\nu$ . In figure 103 you can find a Student’s t copula ( $\nu = 0.1$ ) which displays strong dependency although Pearson’s correlation as well as Spearman’s and Kendall’s rank correlations are zero. Nevertheless, rank correlations allow for a stunning flexibility of correlations and can relatively easily converted in the more familiar Person’s rank correlation (or vice versa) (see table 49).

	Gaussian	Student’s t	Clayton	Gumble	Frank
$\rho^S$	$\frac{6}{\pi} \arcsin(\rho/2)$				$1 - \frac{12}{\vartheta} [D_1(\vartheta) - D_2(\vartheta)]$
$\tau$	$\frac{2}{\pi} \arcsin(\rho)$	$\frac{2}{\pi} \arcsin(\rho)$	$\frac{\vartheta}{\vartheta+2}$	$1 - \frac{1}{\vartheta}$	$1 - \frac{4}{\vartheta} [1 - D_1(\vartheta)]$
$D_k(\theta) = \int_0^\theta t^k / [\exp(t) - 1] dt$	... Debye function, $\rho$ ... Person’s correlation coefficient				

Table 49: rank correlations

Another important feature of copulas is their lower and upper tail dependence

$$\lambda_L = \lim_{\alpha \rightarrow 0^+} P(F_{Y_1}(Y_1) \leq \alpha | F_{Y_2}(Y_2) \leq \alpha) = \lim_{\alpha \rightarrow 0^+} \frac{C(\alpha, \alpha)}{\alpha}$$

$$\lambda_U = \lim_{\alpha \rightarrow 1^-} P(F_{Y_1}(Y_1) > \alpha | F_{Y_2}(Y_2) > \alpha) = \lim_{\alpha \rightarrow 1^-} \frac{1 - 2\alpha + C(\alpha, \alpha)}{1 - \alpha}$$

which describes the dependence structure of the copula for extreme joint lower and upper realizations (measured in quantiles) of the risk factors (see table 50).

	lower tail dependence $\lambda_L$	upper tail dependence $\lambda_U$
Gauss ( $\rho < 1$ )	0	0
Gauss ( $\rho = 1$ )	1	1
Student's t	$2 \cdot t_{v+1} \left( -\sqrt{v+1} \sqrt{\frac{1-\rho}{1+\rho}} \right)$	$2 \cdot t_{v+1} \left( -\sqrt{v+1} \sqrt{\frac{1-\rho}{1+\rho}} \right)$
Clayton	$2^{-1/\theta}$	0
Gumble	0	$2 - 2^{1/\theta}$
Frank	0	0

Table 50: lower & upper tail dependence

For a better understanding of the concepts of concordance and tail dependence we compare scatter plots of our basic copulas for a constant Kendall's  $\tau = 0.6$  (see figure 104). We can simply distinguish the different copulas by their ability to reproduce tail dependence. The

Gaussian and Frank copulas have independent tails while the Student's t copula can produce correlation in both tails. In contrast, the Clayton copula can generate lower and the Gumble copula upper tail dependence, only.

### A.8.2 Simulation of Copulas

The simulation of  $W$  correlated risk factors  $y^{(w)} = (y_1^{(w)}, \dots, y_n^{(w)})'$ , for  $w = 1, \dots, W$ , is easily done by Monte Carlo techniques. In a first step, you need to chose a suitable copula and a correlation level which should be measured by Kendall's rank correlation.<sup>366</sup> Then, one of two Monte Carlo algorithm needs to be programmed. For elliptical Gaussian or Student's t copulas the code is simple:

1. Set the desired Kendall's rank correlations for the risk factors,  $\{\tau(Y_i, Y_j)\}_{i,j=1,\dots,n}$  and transform it to Pearson's correlations  $\{\rho(Y_i, Y_j)\}_{i,j=1,\dots,n}$  by  $\rho = \sin(\tau \cdot \pi/2)$ . In the case of a Student's t copula, you need additionally to define a suitable degree of freedom  $v > 0$ .<sup>367</sup>
2. Sample  $W$  pairs of  $\{z_1^{(w)}, \dots, z_n^{(w)}\}$  from a multivariate Gaussian,  $N_n(\mathbf{0}, \Gamma)$ , or multivariate non-central Student's t distribution,  $t_v(\mathbf{0}, \Gamma)$ , with the respective vector of zeros  $\mathbf{0}$  and correlation (= covariance) matrix  $\Gamma$ .

<sup>366</sup>Kendall's  $\tau$  is more flexible than Pearson's  $\rho$  (see section A.8.1).

<sup>367</sup>The lower  $v$  the more probable become joint extreme events.

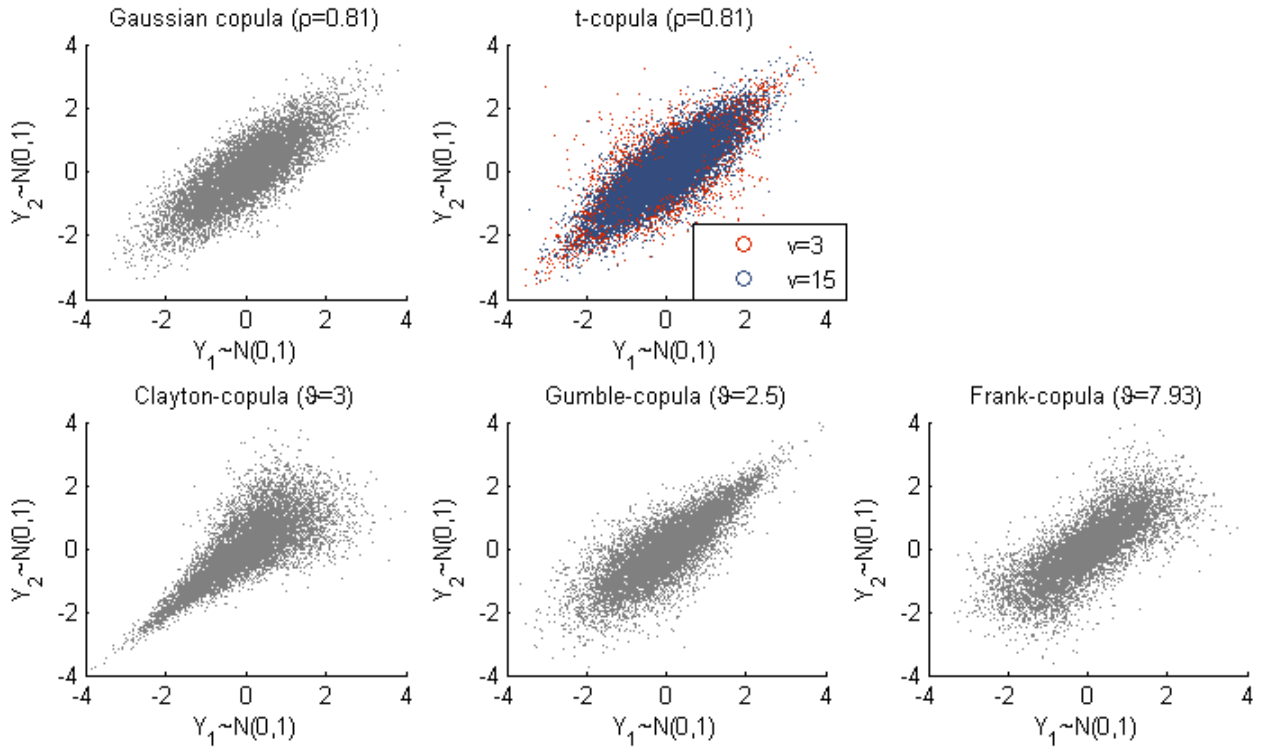


Figure 104: copulas and tail dependence

3. Generate  $W$  pairs of  $\{u_1^{(w)}, \dots, u_n^{(w)}\}$  by the transformations  $u_i^{(w)} = \Phi(z_i^{(w)} | 0, 1)$  or  $u_i^{(w)} = T_v(z_i^{(w)})$ , for  $i = 1, \dots, n$ .<sup>368</sup>
4. Generate  $W$  pairs of the risk factors  $\{y_1^{(w)}, \dots, y_n^{(w)}\}$  by the transformations  $y_i^{(w)} = F_{Y_i}^{-1}(u_i^{(w)})$ , where  $i = 1, \dots, n$  and  $F_{Y_i}^{-1}(\cdot)$  is the marginal inverse distribution of risk factor  $Y_i$ .

The generation of Archimedean copulas (Clayton, Gumble, and Frank) is similar but slightly more complex. Hence, we only present the simulation of bivariate Archimedean copulas for two correlated risk factors  $Y = (Y_1, Y_2)'$ . Hierarchical modeling can be used to generate higher dimensional copulas.

1. Set the desired rank correlation  $\tau(Y_1, Y_2)$ . The Clayton and Gumble copula can only model negative correlations by  $|\tau(Y_1, 1 - Y_2)|$ .

<sup>368</sup> $\Phi(\cdot | 0, 1)$  and  $u_i^{(w)} = T_v(\cdot)$  are the cdfs of the standard Gaussian and univariate Student's t distribution.



2. Find the respective dependence parameter  $\vartheta$  by<sup>369</sup>

$$\arg \min_{\vartheta} |\tau - \vartheta / [\vartheta + 2]| \quad (\text{Clayton})$$

$$\arg \min_{\vartheta} |\tau - [1 - 1/\vartheta]| \quad (\text{Gumble})$$

$$\arg \min_{\vartheta} |\tau - [1 - 4[1 - D_1(\vartheta)]/\vartheta]| \quad (\text{Frank})$$

3. Simulate  $W$  dependent quantiles  $\{u_1^{(w)}, u_2^{(w)}\}$  of the risk factors  $Y_1$  and  $Y_2$

- *Clayton*: generate  $W$  samples  $U_1^{(w)}, \tilde{U}_2^{(w)} \sim \text{Uni}(0, 1)$ , where  $U_1^{(w)}$  and  $\tilde{U}_2^{(w)}$  are independent. Subsequently, calculate

$$u_2^{(w)} = \left[ \left[ u_1^{(w)} \right]^{-\vartheta} \cdot \left[ \left[ \tilde{u}_2^{(w)} \right]^{-\vartheta/[\vartheta+1]} - 1 \right] + 1 \right]^{-1/\vartheta}.$$

- *Gumble*: generate  $W$  samples  $V^{(w)}, S^{(w)} \sim \text{Uni}(0, 1)$ , where  $V^{(w)}$  and  $S^{(w)}$  are independent. Subsequently, minimize

$$\tilde{u}^{(w)} = \arg \min_{\kappa} |s^{(w)} - \kappa [1 - \ln(\kappa) / \vartheta]|.$$

and calculate

$$u_1^{(w)} = \exp \left( [v^{(w)}]^{1/\vartheta} \ln \tilde{u}^{(w)} \right) \ \& \ u_2^{(w)} = \exp \left( [1 - v^{(w)}]^{1/\vartheta} \ln \tilde{u}^{(w)} \right).$$

- *Frank*: generate  $W$  samples  $U_1^{(w)}, S^{(w)} \sim \text{Uni}(0, 1)$ , where  $U_1^{(w)}$  and  $S^{(w)}$  are independent. Subsequently, calculate

$$u_2^{(w)} = -\ln \left( \frac{1 + s^{(w)} [1 - \exp(-\vartheta)]}{s^{(w)} \left[ \exp(-\vartheta u_1^{(w)}) - 1 \right] - \exp(-\vartheta u_1^{(w)})} \right) / \vartheta.$$

4. Generate  $W$  pairs  $\{y_1^{(w)}, y_2^{(w)}\}$  of the risk factors  $Y_1$  and  $Y_2$  by the transformations  $y_i^{(w)} = F_{Y_i}^{-1} \left( u_i^{(w)} \right)$ , where  $i = 1, 2$  and  $F_{Y_i}^{-1}(\cdot)$  is the marginal inverse distribution of risk factor  $Y_i$ .

<sup>369</sup>In Matlab  $1 - 4[1 - D_1(\vartheta)]/\vartheta$  is automatically calculated by `copulastat('Frank',theta)`.

## References

- [1] J. Aldrich. Fisher's "Inverse Probability" of 1930. *International Statistical Review/Revue Internationale de Statistique*, 68(2):155–172, August 2000.
- [2] F. Allen and D. Gale. Stock-Price Manipulation. *Review of Financial Studies*, 5(3):503–529, 1992.
- [3] T. G. Andersen, H.-J. Chung, and B. E. Sørensen. Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics*, 91(1):61–87, July 1999.
- [4] E. W. Anderson, L. P. Hansen, and T. J. Sargent. Robustness, Detection and the Price of Risk. *Working Paper*, March 2000.
- [5] J. E. Angus. The Probability Integral Transform and Related Results. *SIAM Review*, 36(4):652–654, December 1994.
- [6] J. S. Armstrong. Combining Forecasts. In J. S. Armstrong, editor, *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Springer, 2001.
- [7] S. Athanasoulis, R. Shiller, and E. van Wincoop. Macro Markets and Financial Security. *Economic Policy Review*, 5(1):21–39, April 1999.
- [8] R. J. Aumann. Agreeing to Disagree. *The Annals of Statistics*, 4(6):1236–1239, November 1976.
- [9] F. Autin, E. LePennec, and K. Tribouley. Thresholding methods to estimate the copula density. *Journal of Multivariate Analysis*, 101(1):200–222, January 2010. forthcoming.
- [10] T. Aven. *Risk Analysis: Assessing Uncertainties Beyond Expected Values and Probabilities*. Wiley, 2003.
- [11] BaFin. Mindestanforderungen an das Risikomanagement - MaRisk. *Rundschreiben 15/2009(BA) vom 14.08.2009*, August 2009.
- [12] M. Baker and J. Stein. Market liquidity as a sentiment indicator. *Journal of Financial Markets*, 7(3):271–299, June 2004.
- [13] G. Bamberg and F. Baur. *Statistik*. Oldenbourg Verlag, 11th edition, 2001.

- [14] L. Bauwens and M. Lubrano. Bayesian inference on GARCH models using the Gibbs sampler. *Econometrics Journal*, 1(1):23–46, June 1998.
- [15] L. Bauwens and M. Lubrano. Bayesian option pricing using asymmetric GARCH models. *Journal of Empirical Finance*, 9(3):321 – 342, August 2002.
- [16] L. Bauwens and J. Rombouts. Bayesian Inference for the Mixed Conditional Heteroskedasticity Model. *Econometrics Journal*, 10(2):408 – 425, June 2007.
- [17] M. Bemann. Verbesserung der Vergleichbarkeit von Schätzgüteregebnissen von Insolvenzprognosestudien. *Dresden Discussion Paper in Economics*, (08/05), August 2005.
- [18] E. Benz and S. Trück. Modeling the price dynamics of CO2 emission allowances. *Energy Economics*, 31(1):4 – 15, January 2009.
- [19] J. Berg, R. Forsythe, and T. A. Rietz. What Makes Markets Predict Well? In W. Albers, W. Güth, P. Hammerstein, B. Moldovanu, and E. v. Damme, editors, *Understanding Strategic Interaction - Essays in Honor of Reinhard Selten*. Springer Verlag, 1997.
- [20] J. Berg, F. Nelson, and T. Rietz. Accuracy and Forecast Standard Error of Prediction Markets. *Working Paper*, July 2003.
- [21] J. E. Berg and T. A. Rietz. Prediction Markets as Decision Support Systems. *Information Systems Frontiers*, 5(1):79–93, January 2003.
- [22] M. Berlemann. Politische Aktienmärkte als Alternative zur traditionellen Meinungsforschung: Die Wahlbörse zur Landtagswahl 1999 in Sachsen. *Working Paper*, 1999.
- [23] M. Berlemann. Forecasting Inflation. *Postdoctoral Lecture Qualification (TU Dresden)*, 2003. unpublished.
- [24] M. Berlemann. Forecasting the ECB’s Main Refinancing Rate - A Field Experiment. *Economics Letters*, 99(2):379–383, May 2008.
- [25] M. Berlemann, K. Dimitrova, and N. Nenovsky. Assessing Market Expectations on Exchange Rates and Inflation: A Pilot Forecasting System for Bulgaria. *William Davidson Institute Working Paper*, (759), March 2005.
- [26] M. Berlemann and F. Nelson. Forecasting Inflation via Electronic Markets. *Working Paper*, October 2002.

- [27] M. Berlemann and F. Nelson. Forecasting Inflation via Experimental Stock Markets - Some Results from Pilot Markets. *Ifo Working Papers*, (10), June 2005.
- [28] M. Berlemann and C. Schmidt. Predictive Accuracy of Political Stock Markets - Empirical Evidence from a European Perspective. *Dresden Discussion Paper Series in Economics*, (5/01), August 2001.
- [29] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.
- [30] N. H. Bingham and R. Kiesel. *Risk-Neutral Valuation - Pricing and Hedging of Financial Derivatives*. Springer-Verlag, 2nd edition, 2004.
- [31] BIS. Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements. *Bank for International Settlements*, January 1996.
- [32] BIS. Proposed revisions to the Basel II market risk framework. *Bank for International Settlements*, July 2008.
- [33] BIS. 79th Annual Report (1 April 2008 - 31 March 2009). *Bank for International Settlements*, June 2009.
- [34] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, 2nd edition, 2004.
- [35] F. Black. Bank funds management in an efficient market. *Journal of Financial Economics*, 2(4):323–339, December 1975.
- [36] R. J. Bloomfield, M. O’Hara, and G. Saar. How Noise Trading Affects Markets: An Experimental Analysis. *Review of Financial Studies*, 22(6):2275–2302, June 2009.
- [37] T. Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327, April 1986.
- [38] M. Bonamente, M. K. Joy, J. E. Carlstrom, E. D. Reese, and S. J. LaRoque. Markov Chain Monte Carlo Joint Analysis of Chandra X-Ray Imaging Spectroscopy and Sunyaev-Zel’dovich Effect Data. *The Astrophysical Journal*, 614(1):56–63, October 2004.
- [39] S. Borak, W. Haerdle, S. Trueck, and R. Weron. Convenience Yields for CO2 Emission Allowance Futures Contracts. *SFB 649 Discussion Paper*, (2006-076), November 2006.

- [40] P. Bossaerts, L. Fine, and J. Ledyard. Inducing liquidity in thin financial markets through combined-value trading mechanisms. *European Economic Review*, 46(9):1671–1695, October 2002.
- [41] P. Bossaerts and C. Plott. The CAPM in thin experimental financial markets. *Journal of Economic Dynamics and Control*, 26(7-8):1093–1112, July 2002.
- [42] P. Bossaerts and C. Plott. Basic Principles of Asset Pricing Theory: Evidence from Large-Scale Experimental Financial Markets. *Review of Finance*, 8(2):135–169, June 2004.
- [43] J. Boudoukh, M. Richardson, Y. Shen, and R. F. Whitelaw. Do Asset Prices Reflect Fundamentals? Freshly Squeezed Evidence from the OJ Market. *Journal of Financial Economics*, 83(2):397–412, February 2007.
- [44] E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli. Copulas for Finance - A Reading Guide and Some Applications. *Working Paper*, July 2000.
- [45] G. v. Bruggen, M. Spann, G. L. Lilien, and B. Skiera. Institutional Forecasting: The Performance of Thin Virtual Stock Markets. *SSRN eLibrary*, May 2006.
- [46] M. K. Brunnermeier and S. Nagel. Hedge Funds and the Technology Bubble. *Journal of Finance*, 59(5):2013–2040, October 2004.
- [47] D. V. Budescu and T. S. Wallsten. Consistency in Interpretation of Probabilistic Phrases. *Organizational Behavior and Human Decision Processes*, 36(3):391–405, December 1985.
- [48] D. V. Budescu, S. Weinberg, and T. S. Wallsten. Decisions based on numerically and verbally expressed uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, 14(2):281–294, May 1988.
- [49] F. O. Bunnin, Y. Guo, and Y. Ren. Option pricing under model and parameter uncertainty using predictive densities. *Statistics and Computing*, 12(1):37–44, January 2002.
- [50] K. Burdzy. *The Search for Certainty: On the Clash of Science and Philosophy of Probability*. World Scientific Publishing Company, 2009.
- [51] K. P. Burnham and D. R. Anderson. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2):261–304, November 2004.

- [52] T. Cahlik, A. Gersl, M. Hlavacek, and M. Berlemann. Market Prices as Indicators of Political Events Evidence from the Experimental Market on the Czech Republic Parliamentary Election in 2002. *IES Working Paper*, (77), 2005.
- [53] J. Cai. A Markov Model of Switching-Regime ARCH. *Journal of Business & Economic Statistics*, 12(3):309–316, July 1994.
- [54] C. Camerer and K. Weigelt. Information Mirages in Experimental Asset Markets. *Journal of Business*, 64(4):463–493, October 1991.
- [55] C. F. Camerer. Can Asset Markets Be Manipulated? A Field Experiment with Race-track Betting. *The Journal of Political Economy*, 106(3):457–482, June 1998.
- [56] B. P. Carlin and S. Chib. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(3):473–484, 1995.
- [57] C. Cech. Copula-based top-down approaches in financial risk aggregation. *Working Paper (University of Applied Sciences of bfi Vienna)*, (32), December 2006.
- [58] N. H. Chan and H. Y. Wong. *Simulation Techniques in Financial Risk Management*. Wiley-Interscience, 2006.
- [59] N. T. Chan, E. Dahan, A. W. Lo, and T. Poggio. Experimental Markets for Product Concepts. *MIT - artificial intelligence laboratory*, (149), July 2001.
- [60] C. W. Chen, R. Gerlach, and M. K. So. Bayesian model selection for heteroskedastic models. In G. K. Siddhartha Chib, William Griffiths and D. Terrell, editors, *Bayesian Econometrics*, volume 23 of *Advances in Econometrics*, pages 567 – 594. 2008.
- [61] K.-Y. Chen, L. R. Fine, and B. A. Huberman. Forecasting Uncertain Events with Small Groups. *Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 58–64, 2001.
- [62] R. T. Clemen. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559 – 583, October/ December 1989.
- [63] R. T. Clemen and T. Reilly. Correlations and Copulas for Decision and Risk Analysis. *Management Science*, 45(2):208–224, February 1999.
- [64] R. T. Clemen and R. L. Winkler. Limits for the Precision and Value of Information from Dependent Sources. *Operations Research*, 33(2):427–442, March/ April 1985.

- [65] R. T. Clemen and R. L. Winkler. Combining Economic Forecasts. *Journal of Business & Economic Statistics*, 4(1):39–46, January 1986.
- [66] R. T. Clemen and R. L. Winkler. Combining Probability Distributions From Experts in Risk Analysis. *Risk Analysis*, 19(2), April 1999.
- [67] J. H. Cochrane. *Asset Pricing*. Princeton University Press, 2001.
- [68] D. Colander, M. Goldberg, A. Haas, K. Juselius, A. Kirman, T. Lux, and B. Sloth. The financial crisis and the systemic failure of academic economics. *Critical Review*, 21(2):249 – 267, June 2009.
- [69] P. Congdon. Bayesian model choice based on Monte Carlo estimates of posterior model probabilities. *Computational Statistics & Data Analysis*, 50(2):346 – 357, January 2006.
- [70] P. Congdon. Model weights for model choice and averaging. *Statistical Methodology*, 4(2):143–157, April 2007.
- [71] B. Cowgill, J. Wolfers, and E. Zitzewitz. Using Prediction Markets to Track Information Flows: Evidence from Google. *Working Paper*, January 2008.
- [72] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2):385–408, March 1985.
- [73] E. Dahan, A. W. Lo, T. Poggio, N. Chan, and A. Kim. Securities Trading of Concepts (STOC). *Working Paper*, May 2007.
- [74] T. Darsinos and S. E. Satchell. Bayesian Analysis of the Black-Scholes Option Price. *Cambridge Working Papers in Economics*, January 2001.
- [75] G. Daskalakis, D. Psychoyios, and R. N. Markellos. Modeling CO2 Emission Allowance Prices and Derivatives: Evidence from the European Trading Scheme. *Journal of Banking & Finance*, 33(7):1230–1241, July 2009.
- [76] B. de Finetti. *Theory of probability: A critical introductory treatment*, volume 1. John Wiley & Sons, 1974.
- [77] B. de Finetti. *Theory of probability: A critical introductory treatment*, volume 2. John Wiley & Sons, 1975.
- [78] B. de Finetti. Probability: Beware of Falsifications. In H. Kyburg, editor, *Studies in Subjective Probability*, volume 2, pages 192–224. Krieger Publ. Co., 1976.

- [79] B. de Finetti. Foresight: Its Logical Laws, Its Subjective Sources. In S. Kotz and N. L. Johnson, editors, *Breakthroughs in Statistics*, volume 1, pages 134–174. Springer, 1997.
- [80] A. Delbecq and A. Van de Ven. A group process model for problem identification and program planning. *The Journal of Applied Behavioral Science*, 7(4):466–492, Summer 1971.
- [81] P. Dellaportas, J. J. Forster, and I. Ntzoufras. On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12(1):27–36, January 2002.
- [82] P. Diaconis, S. Holmes, and R. Montgomery. Dynamical Bias in the Coin Toss. *SIAM review*, 49(2):211–235, May 2007.
- [83] E. Dimson and M. Mussavian. A brief history of market efficiency. *European Financial Management*, 4(1):91–103, March 1998.
- [84] D. Draper. Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):45–97, 1995.
- [85] J. Duan. The GARCH option pricing model. *Mathematical Finance*, 5(1):13–32, January 1995.
- [86] R. F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007, July 1982.
- [87] R. F. Engle and V. K. Ng. Measuring and Testing the Impact of News on Volatility. *The Journal of Finance*, 48(5):1749–1778, December 1993.
- [88] B. Eraker. MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business & Economic Statistics*, 19(2):177–191, April 2001.
- [89] B. Eraker. Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance*, 59(3):1367–1403, June 2004.
- [90] B. Eraker, M. Johannes, and N. Polson. The Impact of Jumps in Volatility and Returns. *The Journal of Finance*, 58(3):1269–1300, June 2003.
- [91] EU. The EU Emissions Trading Scheme: EU action against climate change. *European Commission*, 2009.
- [92] E. F. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383–417, May 1970.



- [93] E. F. Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3):283–306, September 1998.
- [94] Fed. The Supervisory Capital Assessment Program: Design and Implementation. *Board of Governors of the Federal Reserve System*, April 2009.
- [95] Fed. The Supervisory Capital Assessment Program: Overview of Results. *Board of Governors of the Federal Reserve System*, May 2009.
- [96] J.-D. Fermanian and O. Scaillet. Some Statistical Pitfalls in Copula Modeling for Financial Applications. *FAME Working Paper*, (108), March 2004.
- [97] S. Fillenbaum, T. S. Wallsten, B. L. Cohen, and J. A. Cox. Some Effects of Vocabulary and Communication Task on the Understanding and Use of Vague Probability Expressions. *The American Journal of Psychology*, 104(1):35–60, Spring 1991.
- [98] D. Fink. A Compendium of Conjugate Priors. *Working Paper*, May 1997.
- [99] I. Fischer and N. Harvey. Combining forecasts: What information do judges need to outperform the simple average? *International Journal of Forecasting*, 15(3):227–246, July 1999.
- [100] R. Forsythe, F. Nelson, G. Neumann, and J. Wright. Anatomy of an experimental stock market. *American Economic Review*, 82(5):1142–1161, December 1992.
- [101] R. Forsythe, T. A. Rietz, and T. W. Ross. Wishes, expectations and actions: A survey on price formation in election stock markets. *Journal of Economic Behavior & Organization*, 39(1):83–110, May 1999.
- [102] J. Fox. *The Myth of the Rational Market: History of Risk, Reward and Delusion on Wall Street*. Harper Collins, 2009.
- [103] A. Frieze, R. Kannan, and N. Polson. Sampling from Log-Concave Distributions. *Annals of Applied Probability*, 4(3):812–837, August 1994.
- [104] A. Garg, T. S. Jayram, S. Vaithyanathan, and H. Zhu. Generalized Opinion Pooling. *IBM Almaden Research Center*, 2004.
- [105] P. H. Garthwaite, J. B. Kadane, and A. O’Hagan. Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association*, 100(470): 680–701, June 2005.

- [106] H. Garz, A. Haas, R. Ötsch, P. Wirtz, and S. Zank. German power utilities - caught in the CO2 trap? *WestLB Study*, September 2009.
- [107] J. Geanakoplos and H. M. Polemarchakis. We Can't Disagree Forever. *Journal of Economic Theory*, 28(1):192–200, 1982.
- [108] A. E. Gelfand and A. F. M. Smith. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409, June 1990.
- [109] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Crc Pr Inc, 1 edition, 1995.
- [110] A. Gelman, G. O. Roberts, and W. R. Gilks. Efficient Metropolis Jumping Rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 5, pages 599–607. Oxford Science Publications, 1996.
- [111] A. Gelman and D. B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457–472, November 1992.
- [112] C. Genest and A.-C. Favre. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, 12(4):347–368, July/ August 2007.
- [113] J. Geweke. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In J. Geweke, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 4, pages 164–193. Oxford University Press, 1992.
- [114] J. F. Geweke. Bayesian comparison of econometric models. *Working Paper*, September 1995.
- [115] C. J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):473–483, November 1992.
- [116] G. Gigerenzer. *Bauchentscheidungen*. Goldmann Verlag, 2008.
- [117] S. Gjerstad. Market Dynamics in Edgeworth Exchange. *Economic Science Laboratory (University of Arizona)*, January 2005.
- [118] M. Goerz. Bayesianismus und Falsifizierbarkeit. *Working Paper*, October 2006.
- [119] I. J. Good. Rational Decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14(1):107–114, 1952.

- [120] J. R. Graham. Is a Group of Economists Better Than One? Than None? *The Journal of Business*, 69(2):193–232, April 1996.
- [121] P. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82(4):711–732, December 1995.
- [122] J. Grossklags and C. Schmidt. Artificial Software Agents on Thin Double Auction Markets - A Human Trader Experiment. *International Conference on Intelligent Agent Technology*, October 2003.
- [123] S. J. Grossman. An Introduction to the Theory of Rational Expectations Under Asymmetric Information. *The Review of Economic Studies*, 48(4):541–559, October 1981.
- [124] S. J. Grossman and J. E. Stiglitz. On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70(3):393–408, June 1980.
- [125] H. Gzyl, E. ter Horst, and S. Malone. Towards a Bayesian framework for option pricing (Part II). *Estudio IESA*, (31), October 2006.
- [126] R. S. Gürkaynak and J. Wolfers. Macroeconomic Derivatives: An Initial Analysis of Market-Based Macro Forecasts, Uncertainty, and Risk. *Federal Reserve Bank of San Francisco Working Paper Series*, (26), September 2005.
- [127] M. Haas, S. Mittnik, and M. S. Paoletta. A New Approach to Markov-Switching GARCH Models. *Journal of Financial Econometrics*, 2(4):493–530, 2004.
- [128] R. W. Hahn and P. C. Tetlock. A New Approach for Regulating Information Markets. *Journal of Regulatory Economics*, 29(3):265–281, May 2006.
- [129] J. D. Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357–384, March 1989.
- [130] J. D. Hamilton and R. Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1/2):307 – 333, September/ October 1994.
- [131] J. Hannig. On Fiducial Inference: the good, the bad and the ugly. *Working Paper*, February 2006.
- [132] J. Hansen, C. Schmidt, and M. Strobel. Manipulation in political stock markets - preconditions and evidence. *Applied Economics Letters*, 11(7):459 – 463, June 2004.

- [133] L. P. Hansen. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, 50(4):1029–1054, July 1982.
- [134] R. Hanson. Idea Futures: Encouraging an Honest Consensus. *Extropy*, 3(2):7–17, Winter 1992.
- [135] R. Hanson. Disagreement is unpredictable. *Economics Letters*, 77(3):365–369, November 2002.
- [136] R. Hanson. Eliciting Objective Probabilities via Lottery Insurance Games. Technical report, George Mason University Economics, 2002.
- [137] R. Hanson. Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation. Technical report, George Mason University Economics, January 2002.
- [138] R. Hanson. Combinatorial Information Market Design. *Information Systems Frontiers*, 5(1):107–119, January 2003.
- [139] R. Hanson. Designing Real Terrorism Futures. *Public Choice*, 128(1/2):257–274, July 2006.
- [140] R. Hanson. Foul Play in Information Markets. In R. W. Hahn and P. C. Tetlock, editors, *Information Markets: A New Way of Making Decisions*, pages 126–141. AEI Press, 2006.
- [141] R. Hanson. Shall We Vote on Values, But Bet on Beliefs? *Working Paper*, October 2007.
- [142] R. Hanson. The Policy Analysis - A Thwarted Experiment in the Use of Prediction Markets for Public Policy. *Innovations: Technology, Governance, Globalization*, 2(3): 73–88, Summer 2007.
- [143] R. Hanson. Idea futures: Encouraging an honest consensus. *Extropy*, 8, 2008. updated version of 1992 paper.
- [144] R. Hanson and R. Oprea. Manipulators Increase Information Market Accuracy. *Working Paper*, July 2004.
- [145] R. Hanson, R. Oprea, and D. Porter. Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior & Organization*, 60(4):449–459, June 2006.

- [146] G. W. Harrison. Expected Utility Theory and the Experimentalists. *Empirical Economics*, 19(2):223–253, March 1994.
- [147] J. Hartung. *Statistik: Lehr- und Handbuch der angewandten Statistik*. Oldenbrough Verlag, 12th edition, 1999.
- [148] A. Harvey, E. Ruiz, and N. Shephard. Multivariate Stochastic Variance Models. *The Review of Economic Studies*, 61(2):247–264, April 1994.
- [149] A. C. Harvey and N. Shephard. Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business & Economic Statistics*, 14(4):429–434, October 1996.
- [150] F. A. v. Hayek. The Use of Knowledge in Society. *The American Economic Review*, 35(4):519–530, September 1945.
- [151] P. Heidelberger and P. D. Welch. Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, 31(6):1109–1144, November/ December 1983.
- [152] J. S. Henneke, S. T. Rachev, and F. J. Fabozzi. MCMC based estimation of Markov switching ARMA-GARCH models. *Working Paper*, February 2006.
- [153] D. Hirshleifer. Investor Psychology and Asset Pricing. *The Journal of Finance*, 56(4): 1533–1597, August 2001.
- [154] J. W. Hopman. Using Forecasting Markets to Manage Demand Risk. *Intel Technology Journal*, 11(2):127–135, May 2007.
- [155] S. Hora and M. Jensen. Expert judgement elicitation. *SSI report*, (19), September 2002.
- [156] A. I. Huffcutt and D. J. Woehr. Further analysis of employment interview validity: a quantitative evaluation of interviewer-related structuring methods. *Journal of Organizational Behavior*, 20(4):549–560, July 1999.
- [157] J. Hull and A. White. The Pricing of Options on Assets with Stochastic Volatilities. *The Journal of Finance*, 42(2):281–300, June 1987.
- [158] J. C. Hull. *Options, Futures, and Other Derivatives*. Prentice Hall, 6th edition, 2006. ISBN 0136015867.
- [159] IPCC. Emission Scenarios. *IPCC Special Report*, 2000.

- [160] IPCC. Climate change and water: IPCC Technical Paper VI. *Intergovernmental Panel on Climate Change*, June 2008.
- [161] B. Jacobsen, J. Potters, A. Schram, F. van Winden, and J. Wit. (In)accuracy of a European political stock market: The influence of common value structures. *European Economic Review*, 44(2):205–230, February 2000.
- [162] E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics*, 12(4):371–389, October 1994.
- [163] E. Jacquier, N. G. Polson, and P. E. Rossi. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics*, 122(1):185 – 212, September 2004.
- [164] C. Jaeger, J. Krause, A. Haas, R. Klein, and K. Hasselmann. A Method for Computing the Fraction of Attributable Risk Related to Climate Damages. *An International Journal*, 28(4):815–823, July 2008.
- [165] I. L. Janis. *Victims of Groupthink: A psychological study of foreign-policy decisions and fiascoes*. Houghton Mifflin Company, 1972.
- [166] S. F. Jarner and G. O. Roberts. Polynomial Convergence Rates of Markov Chains. *The Annals of Applied Probability*, 12(1):224–247, February 2002.
- [167] R. A. Jarrow. Market Manipulation, Bubbles, Corners, and Short Squeezes. *The Journal of Financial and Quantitative Analysis*, 27(3):311–336, September 1992.
- [168] M. Johannes and N. Polson. MCMC Methods for Continuous-Time Financial Econometrics. *Working Paper*, December 2003.
- [169] K. John and R. Narayanan. Market Manipulation and the Role of Insider Trading Regulations. *The Journal of Business*, 70(2):217–247, April 1997.
- [170] G. L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1: 299–320, 2004.
- [171] M. N. Jouini and R. T. Clemen. Copula Models for Aggregating Expert Opinions. *Operations Research*, 44(3):444–457, May/ June 1996.
- [172] J. B. Kadane and R. L. Winkler. De Finetti’s Methods of Elicitation. In R. Viertl, editor, *Probability and Bayesian Statistics*, volume 1. Springer, 1987.

- [173] J. B. Kadane and R. L. Winkler. Separating Probability Elicitation From Utilities. *Journal of the American Statistical Association*, 83(402):357–363, June 1988.
- [174] D. Kahneman and A. Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–292, March 1979.
- [175] D. Kahneman and A. Tversky. On the study of statistical intuitions. *Cognition*, 11(2):123–141, March 1982.
- [176] K. Kalovcova. Why Betting and Prediction Markets Work (Not) Well: An Inventory Of Open Questions. *CERGE-EI Discussion Paper Series, Charles University, Prague*, (178), March 2007.
- [177] S. Kaplan. ‘Expert information’ versus ‘expert opinions’. Another approach to the problem of eliciting/ combining/using expert knowledge in PRA. *Reliability Engineering & System Safety*, 35(1):61 – 72, January 1992.
- [178] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović. Power law and exponential decay of inter contact times between mobile devices. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 183–194, 2007. ISBN 978-1-59593-681-3.
- [179] R. E. Kass and A. E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795, June 1995.
- [180] J. M. Keynes. *The General Theory of Employment, Interest and Money*. Palgrave Macmillan (New Ed edition), 2007.
- [181] S. Kim, N. Shephard, and S. Chib. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3):361–393, July 1998.
- [182] J. Kitzinger. Qualitative research: introducing focus groups. *British Medical Journal*, 311(7000):299–302, July 1995.
- [183] klimazwei. Climate Protection and Adaptation: Results of the klimazwei Research Programme. 2009.
- [184] G. Klir and M. Wierman. *Uncertainty-based information: Elements of generalized information theory*. Springer Verlag, 1999.
- [185] F. Knighth. *Risk, Uncertainty and Profit*. Boston and New York, 1921.

- [186] K. R. Koch. Gibbs sampler by sampling-importance-resampling. *Journal of Geodesy*, 81(9):581–591, September 2007.
- [187] S. G. Kou and M. E. Sobel. Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls. *Political Analysis*, 12(3):277–295, Summer 2004.
- [188] O. Krahl and J. Wagner. Szenarioanalysen im Kreditrisikomanagement. *Zeitschrift für das gesamte Kreditwesen*, 62(3):128–131, February 2009.
- [189] J. Krause. CO2-Emission Reduction Options for the German Automotive Industry: Insights from Expert Interviews. *PhD Thesis (forthcoming)*, May, 31 2010. unpublished.
- [190] C.-M. Kuan. Econometric Theory II. *National Taiwan University, Department of Economics and Department of Finance (lecture notes)*, chapter 9, 2007. <http://idv.sinica.edu.tw/ckuan/pdf/et01/>.
- [191] A. S. Kyle. Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335, November 1985.
- [192] A. S. Kyle. Informed Speculation with Imperfect Competition. *The Review of Economic Studies*, 56(3):317–355, July 1989.
- [193] M. Kynn. The 'heuristics and biases' bias in expert elicitation. *Journal of the Royal Statistical Society: Series A*, 171(1):239–264, January 2008.
- [194] J. O. Ledyard. Designing Information Markets for Policy Analysis. In R. W. Hahn and P. C. Tetlock, editors, *Information Markets: A New Way of Making Decisions*, pages 13–32. AEI Press, 2006.
- [195] B.-S. Lee and B. F. Ingram. Simulation estimation of time-series models. *Journal of Econometrics*, 47(2/3):197 – 205, February 1991.
- [196] A. Leigh, J. Wolfers, and E. W. Zitzewitz. What Do Financial Markets Think of War in Iraq? *SIEPR Discussion Paper*, (02-28), March 2003.
- [197] I. Levi. Imprecise and indeterminate probabilities. *Risk, Decision and Policy*, 5(2): 111–122, June 2000.
- [198] H. Li, M. T. Wells, and C. L. Yu. A Bayesian Analysis of Return Dynamics with Lévy Jumps. *Review of Financial Studies*, 21(5):2345–2378, September 2008.



- [199] D. Lindley. Reconciliation of Probability Distributions. *Operations Research*, 31(5): 866–880, September/ October 1983.
- [200] D. V. Lindley. *Making Decisions*. Wiley, 2nd edition, 1985.
- [201] J. Linnainmaa. The Limit Order Effect. *Working Paper*, October 2007.
- [202] X. Liu. Euler Discretization in Estimating Continuous-Time Financial Models via MCMC Methods. *Working Paper*, January 2006.
- [203] C. G. Lord, L. Ross, and L. L. M. R. Lepper, Mark R. Charles G.; Ross. Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2109, November 1979.
- [204] R. E. Lucas. Econometric Policy Evaluation: A Critique. *Carnegie-Rochester Conference Series on Public Policy*, pages 19–46, 1976.
- [205] S. Luckner. Prediction Markets: How Do Incentive Schemes Affect Prediction Accuracy? *Dagstuhl Seminar Proceedings (Negotiation and Market Engineering)*, (06461), 2006.
- [206] S. Makridakis. The Art and Science of Forecasting: An Assessment and Future Directions. *International Journal of Forecasting*, 2(1):15–39, January/ March 1986.
- [207] B. G. Malkiel. The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*, 17(1):59–82, Winter 2003.
- [208] S. Malone and E. ter Horst. Towards a Bayesian framework for option pricing. *Estudio IESA*, (27), February 2006.
- [209] N. G. Mankiw, R. Reis, and J. Wolfers. Disagreement about Inflation Expectations. *BER Macroeconomics Annual*, 18:209–248, 2003.
- [210] C. F. Manski. Interpreting the predictions of prediction markets. *Economics Letters*, 91(3):425–429, June 2006.
- [211] D. W. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, June 1963.

- [212] M. Mather and M. K. Johnson. Choice-Supportive Source Monitoring: Do Our Decisions Seem Better to Us as We Age? *Psychology and Aging*, 15(4):596–606, December 2000.
- [213] M. Mather, E. Shafir, and M. K. Johnson. Misremembrance of options past: Source monitoring and choice. *Psychological Science*, 11(2):132–138, March 2000.
- [214] D. McFadden. A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica*, 57(5):995–1026, September 1989.
- [215] D. H. Meadows, D. L. Meadows, J. Randers, and W. W. Behrens. *The limits to growth: a report for the Club of Rome's project on the predicament of mankind*. Universe Books New York, 1972.
- [216] A. Meirowitz and J. A. Tucker. Learning From Terrorism Markets. *Perspectives on Politics*, 2(2):331–336, June 2004.
- [217] K. L. Mengersen and R. L. Tweedie. Rates of Convergence of the Hastings and Metropolis Algorithms. *The Annals of Statistics*, 24(1):101–121, February 1996.
- [218] R. C. Merton. Option Pricing when Underlying Stock Returns Are Discontinuous. *Journal of Financial Economics*, 3(1/2):125–144, January/ March 1976.
- [219] R. C. Merton. On estimating the expected return on the market : An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361, December 1980.
- [220] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. H. Teller, and E. Teller. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6):1087–1091, June 1953.
- [221] R. Meyer and J. Yu. BUGS for a Bayesian analysis of stochastic volatility models. *The Econometrics Journal*, 3(2):198–215, December 2000.
- [222] P. Milgrom and N. Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, 26(1):17–27, February 1982.
- [223] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill Series in Probability and Statistics. McGraw-Hill Companies, 3rd edition, 1974.
- [224] D. L. Morgan. Focus groups. *Annual Review of Sociology*, 22(1):129–152, August 1996.

- [225] M. Morgan and D. W. Keith. Improving the way we think about projecting future energy use and emissions of carbon dioxide. *Climatic Change*, 90(3):189–215, October 2008.
- [226] S. Morris. The Common Prior Assumption in Economic Theory. *Economics and Philosophy*, 11(2):227–253, October 1995.
- [227] E. Moulines, R. Douc, G. Fort, G. Roberts, and J. . Rosenthal. Rate of convergence, polynomial ergodicity and computable bounds for Markov Chains. *CIRM Presentation*, September 2001. <http://www.ceremade.dauphine.fr/~xian/EMoulines.pdf>.
- [228] K. Murphy. Kalman filter toolbox for matlab. internet page for download, 1998/2004. <http://www.cs.ubc.ca/~murphyk/Software/Kalman/kalman.html>.
- [229] J. F. Muth. Rational Expectations and the Theory of Price Movements. *Econometrica*, 29(3):315–335, July 1961.
- [230] R. Nadeau, E. Cloutier, and J.-H. Guay. New Evidence about the Existence of a Bandwagon Effect in the Opinion Formation Process. *International Political Science Review / Revue internationale de science politique*, 14(2):203–213, April 1993.
- [231] R. F. Nau. De Finetti was right: Probability does not exist. *Theory and Decision*, 51(2/4):89–124, December 2001.
- [232] M. Nesvisky. Macroeconomic Derivatives. *The NBER Digest*, August 2006.
- [233] J. Noh, R. Engle, and A. Kane. Forecasting Volatility and Option Prices of the S&P 500 Index. *The Journal of Derivatives*, 2(1):17–30, Fall 1994.
- [234] M. Nöth, C. F. Camerer, C. R. Plott, and M. Weber. Information Aggregation in Experimental Asset Markets: Traps and Misaligned Beliefs. *Social Science Working Paper*, (1060), April 1999.
- [235] M. Nöth and M. Weber. Information Aggregation with Random Ordering: Cascades and Overconfidence. *The Economic Journal*, 113(484):166–189, January 2003.
- [236] A. O’Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain Judgements: Eliciting Experts’ Probabilities (Statistics in Practice)*. Wiley, 2006.
- [237] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6th edition, 2007.

- [238] K. Oliven and T. A. Rietz. Suckers Are Born but Markets Are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market. *Management Science*, 50(3):336–351, March 2004.
- [239] J. Ondrus, C. Gaspoz, and Y. Pigneur. Technology Foresight for IT Investment: Multi-Criteria Decision-Making versus Prediction Markets. *pre-Icis AIM*, 2007.
- [240] M. Onischka. Diversifikation von ökonomisch signifikanten Klimarisiken. *Climate2008 Conference*, 2008.
- [241] S. Ostrover. Employing Information Markets to Achieve Truly Collaborative Sales Forecasting. *The Journal of Business Forecasting*, 24(1):9–12, Spring 2005.
- [242] M. Ottaviani and P. N. Sørensen. Outcome Manipulation in Corporate Prediction Markets. *Journal of European Economic Association*, 5(2/3):554–563, April/May 2007.
- [243] M. S. Paoella and L. Taschini. An Econometric Analysis of Emission Trading Allowances. *Journal of Banking and Finance*, 32(10):2022–2032, October 2008.
- [244] D. L. Passmore and E. D. Cebeci. Potential of Idea Futures Markets in Educational Technology. *Working Paper*, 2003.
- [245] A. G. Patt and D. P. Schrag. Using Specific Language to Describe Risk and Probability. *Climatic Change*, 61(1/2):17–30, November 2003.
- [246] J. G. Pedersen. Fiducial Inference. *International Statistical Review / Revue Internationale de Statistique*, 46(2):147–170, August 1978.
- [247] D. M. Pennock. A dynamic pari-mutuel market for hedging, wagering, and information aggregation. *Proceedings of the 5th ACM conference on Electronic commerce*, pages 170–179, 2004.
- [248] D. M. Pennock. Implementing Hanson’s Market Maker. October 2006. <http://blog.oddhead.com/2006/10/30/implementing-hansons-market-maker/> (last revised April 2, 2008).
- [249] D. M. Pennock, C. L. Giles, and F. A. Nielsen. The real power of artificial markets. *Science*, 291(5506):987–988, February 2001.
- [250] M. Peterson. *An Introduction to Decision Theory*. Cambridge Introductions to Philosophy. Cambridge University Press, 2009.

- [251] M. Pitt and N. Shephard. Filtering Via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association*, 94(446):590–591, June 1999.
- [252] C. R. Plott. Markets as Information Gathering Tools. *Southern Economic Journal*, 67(1):2–15, July 2000.
- [253] C. R. Plott and S. Sunder. Efficiency of Experimental Security Markets with Insider Information: An Application of Rational-Expectations Models. *The Journal of Political Economy*, 90(4):663–698, August 1982.
- [254] C. R. Plott, J. Wit, and W. C. Yang. Parimutuel betting markets as information aggregation devices: experimental results. *Economic Theory*, 22(2):311–351, September 2003.
- [255] P. M. Polgreen, F. D. Nelson, and G. R. Neumann. Use of Prediction Markets to Forecast Infectious Disease Activity. *Clinical Infectious Diseases*, 44(2):272–279, January 2007.
- [256] M. C. Pollard. Bayesian Value-at-Risk and Capital Charge Puzzle. *Working Paper*, November 2007.
- [257] M. C. Pollard. Markov Chain Monte Carlo Analysis of Option Pricing Models. *Working Paper*, August 2007.
- [258] R. A. Posner. *A failure of capitalism: the crisis of '08 and the descent into depression*. Harvard University Press, 2009.
- [259] O. Pourret, P. Naim, and B. Marcot. *Bayesian networks: A practical guide to applications*. Wiley, 2008.
- [260] S. J. Press. *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*. Wiley, 2002.
- [261] S. T. Rachev, J. S. J. Hsu, B. S. Bagasheva, and F. J. Fabozzi. *Bayesian Methods in Finance*. Frank J. Fabozzi Series. Wiley, 2008.
- [262] A. E. Raftery. Approximate Bayes Factors and Accounting for Model Uncertainty in Generalised Linear Models. *Biometrika*, 83(2):251–266, June 1996.
- [263] R. Ranjan and T. Gneiting. Combining Probability Forecasts. *Technical Report: Department of Statistics, University of Washington*, (543), October 2008.

- [264] H. v. Restorff. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologie Forschung*, 18(1):299–234, December 1933.
- [265] P. W. Rhode and K. S. Strumpf. Historical Presidential Betting Markets. *Journal of Economic Perspectives*, 18(2):127–142, Spring 2004.
- [266] P. W. Rhode and K. S. Strumpf. Manipulating Political Stock Markets: A Field Experiment and a Century of Observational Data. *Working Paper*, January 2007.
- [267] T. A. Rietz. Behavioral Mis-pricing and Arbitrage in Experimental Asset Markets. *Working Paper*, June 2005.
- [268] C. Robert and J.-M. Marin. On some difficulties with a posterior probability approximation technique. *Bayesian Analysis*, 3(2):427–442, June 2008.
- [269] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7(1):110–120, February 1997.
- [270] G. O. Roberts and R. L. Tweedie. Geometric Convergence and Central Limit Theorems for Multidimensional Hastings and Metropolis Algorithms. *Biometrika*, 83(1):95–110, March 1996.
- [271] R. Roll. Orange juice and weather. *The American Economic Review*, 74(5):861–880, December 1984.
- [272] D. L. Ronis and J. F. Yates. Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes*, 40(2):193–218, October 1987.
- [273] M. Rosenblatt. Remarks on a Multivariate Transformation. *The Annals of Mathematical Statistics*, 23(3):470–472, September 1952.
- [274] E. S. Rosenbloom and W. Notz. Statistical Tests of Real-Money versus Play-Money Prediction Markets. *Electronic Markets*, 16(1):63 – 69, February 2006.
- [275] S. M. Ross. *Stochastic Processes*. Wiley, 2nd edition, 1995.
- [276] D. B. Rubin. Using the SIR algorithm to simulate posterior distributions. In M. H. Bernardo, K. M. Degroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, volume 3. Oxford University Press, 1988.

- [277] F. Sanders. On Subjective Probability Forecasting. *Journal of Applied Meteorology*, 2(2):191–201, April 1963.
- [278] L. J. Savage. *The Foundations of Statistics*. Courier Dover Publications, 2nd edition, 1972.
- [279] D. S. Scharfstein and J. C. Stein. Herd Behavior and Investment. *The American Economic Review*, 80(3):465–479, June 1990.
- [280] R. Schlittgen and B. H. J. Streitberg. *Zeitreihenanalyse*. Oldenbrough Verlag, 3rd edition, 1989.
- [281] T. Schmidt. Coping with Copulas. In J. Rank, editor, *Copulas - From Theory to Applications in Finance*, pages 3–34. Risk Books, 2007.
- [282] P. J. H. Schoemaker. The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. *Journal of Economic Literature*, 20(2):529–563, June 1982.
- [283] S. L. Scott. Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century. *Journal of the American Statistical Association*, 97(457):337–351, March 2002.
- [284] E. Servan-Schreiber, J. Wolfers, D. M. Pennock, and B. Galebach. Prediction Markets: Does Money Matter? *Electronic Markets*, 14(3):243 – 251, September 2004.
- [285] R. J. Shiller. *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks*. Oxford University Press, 1993.
- [286] R. J. Shiller. *The subprime solution: how today's global financial crisis happened, and what to do about it*. Princeton University Press, 2008.
- [287] A. Shleifer. *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford University Press, 2000.
- [288] B. Siliverstovs, R. Ötsch, C. Kemfert, C. C. Jaeger, A. Haas, and H. Kremers. Climate change and modelling of extreme temperatures in Switzerland. *Stochastic Environmental Research and Risk Assessment*, June 2009. published online.
- [289] D. Sornette. Dragon-Kings, Black Swans and the Prediction of Crises. *Swiss Finance Institute Research Paper Series*, 36(9), September 2009.
- [290] M. Spann and B. Skiera. Internet-Based Virtual Stock Markets for Business Forecasting. *Management Science*, 49(10):1310–1326, October 2003.

- [291] M. Spiegel and A. Subrahmanyam. Informed Speculation and Hedging in a Noncompetitive Securities Market. *Review of Financial Studies*, 5(2):307–329, February 1992.
- [292] S. D. Spiegelhalter, N. G. Best, B. P. Carlin, and A. V. D. Linde. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.
- [293] C. R. Sunstein. Deliberation and Information Markets. In R. W. Hahn and P. C. Tetlock, editors, *Information Markets: A New Way of Making Decisions*, pages 67–100. AEI Press, 2006.
- [294] J. Surowiecki. *The Wisdom of Crowds*. Doubleday, 2004. ISBN 0385721706.
- [295] N. N. Taleb. *The black swan*. Random House, 2007.
- [296] P. C. Tetlock. Liquidity and Prediction Market Efficiency. *Working Paper*, March 2008.
- [297] R. H. Thaler and W. T. Ziemba. Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries. *The Journal of Economic Perspectives*, 2(2):161–174, Spring 1988.
- [298] L. Tierney. Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, 22(4):1701–1728, December 1994.
- [299] L. Tierney and J. B. Kadane. Accurate Approximations for Posterior Moments and Marginal Densities. *Journal of the American Statistical Association*, 81(393):82–86, March 1986.
- [300] A. Tversky and D. Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, September 1974.
- [301] G. Tziralis and I. Tatsiopoulos. Prediction Markets: An Extended Literature Review. *The Journal of Prediction Markets*, 1(1):75–91, February 2007.
- [302] M. Uhrig-Homburg and M. Wagner. Derivative Instruments in the EU Emissions Trading Scheme - An Early Market Perspective. *Energy & Environment*, 19(5):635–655, September 2008.
- [303] M. Uhrig-Homburg, Marliese. und Wagner. Futures Price Dynamics of CO2 Emission Certificates: An Empirical Analysis. *Working Paper*, August 2007.
- [304] O. Vasicek. An Equilibrium Characterisation of the Term Structure. *Journal of Financial Economics*, 5(2):177–188, November 1977.



- [305] A. H. v. d. Ven and A. L. Delbecq. The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *The Academy of Management Journal*, 17(4):605–621, December 1974.
- [306] R. L. Vislocky and J. M. Fritsch. Improved Model Output Statistics Forecasts through Model Consensus. *Bulletin of the American Meteorological Society*, 76(7):1157–1164, July 1995.
- [307] T. Von Ghyczy, C. von Clausewitz, B. von Oetinger, and C. Bassford. *Clausewitz on strategy: inspiration and insight from a master strategist*. John Wiley & Sons, 2002.
- [308] R. Von Mises and H. Geiringer. *Probability, Statistics and Truth*. Dover Publications, 1981.
- [309] D. Vose. *Risk Analysis: A Quantitative Guide*. John Wiley & Sons, Ltd, 3rd edition, 2008.
- [310] M. W. Wagner. *CO<sub>2</sub>-Emissionszertifikate: Preismodellierung und Derivatebewertung*. PhD thesis, Faculty of Economics, Universität Karlsruhe (TH), 2007.
- [311] P. C. Wason. On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, 12(3):129–140, July 1960.
- [312] M. Weitzman. Subjective Expectations and Asset-Return Puzzles. *American Economic Review*, 97(4):1102–1130, September 2007.
- [313] G. Winkler. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Springer, 2nd edition, 2004.
- [314] R. L. Winkler. Combining Probability Distributions from Dependent Information Sources. *Management Science*, 27(4):479–488, April 1981.
- [315] R. L. Winkler and A. H. Murphy. 'Good' Probability Assessors. *Journal of Applied Meteorology*, 7(5):751–758, October 1968.
- [316] J. Wolfers. New Uses for New Macro Derivatives. *FRBSF Economic Letters*, (2006-21), August 2006.
- [317] J. Wolfers and E. Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, Spring 2004.

- [318] J. Wolfers and E. Zitzewitz. Five Open Questions about Prediction Markets. In R. W. Hahn and P. C. Tetlock, editors, *Information Markets: A New Way of Making Decisions*, pages 37–66. AEI Press, 2006.
- [319] J. Wolfers and E. Zitzewitz. Interpreting Prediction Market Prices as Probabilities. *Working Paper*, January 2007.
- [320] J. Yu. MCMC Methods for Estimating Stochastic Volatility Models with Leverage Effects: Comments on Jacquier, Polson and Rossi (2002). *Working Paper*, November 2002.
- [321] J. Yu. On leverage in a stochastic volatility model. *Journal of Econometrics*, 127(2): 165–178, August 2005.
- [322] R. Zeckenhauser. The Economics of Catastrophes. *Journal of Risk and Uncertainty*, 12(2/3):113–140, May 1996.
- [323] R. Ötsch. Bayesianisches Risikomanagement - eine Einführung. *Working Paper*, June 2008.