Essays on career choice under risk and ambiguity

INAUGURAL-DISSERTATION

zur Erlangung des akademischen Grades eines Doktors der Wirtschafts- und Sozialwissenschaft (Dr. rer. pol.) der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Potsdam

vorgelegt von

Alexander Konon, M.Sc.

Erstgutachter: Prof. Dr. Alexander Kritikos
Zweitgutachter: Prof. Dr. Marco Caliendo
Drittgutachter: Prof. Dr. Rainald Borck

Januar 2018

Published online at the
Institutional Repository of the University of Potsdam:
URN urn:nbn:de:kobv:517-opus4-416462
http://nbn-resolving.de/urn:nbn:de:kobv:517-opus4-416462

Essays on career choice under risk and ambiguity

Dissertation

Alexander Konon

Acknowledgments

Many people contributed to this dissertation and my entire experience of writing it. I would especially like to thank my supervisor Alexander Kritikos, who provided guidance, support, and valuable comments on my projects. Also, I would like to thank Marco Caliendo and Rainald Borck for accepting to serve as second and third supervisors.

I wrote the thesis while being a graduate student at the DIW Berlin Graduate Center. The DIW provided an inspiring environment, in which I could discuss new ideas, and funding. I would like to thank Helmut Lütkepohl, the Graduate Center team, and members of the Graduate Center cohort of 2012. I would also like to thank Adam Lederer for proofreading my papers.

This dissertation would be impossible without constant support from my parents Elisabeth and Alexander Konon. Ivana Sidzimosvka provided me invaluable support in the last phase of the dissertation.

Abstract

This dissertation consists of five self-contained essays, in Chapters 2–6, addressing different aspects of career choices, especially the choice of entrepreneurship, under risk and ambiguity. The essays are preceded by a brief introduction in Chapter 1.

In Chapter 2, the first essay develops an occupational choice model with boundedly rational agents, who lack information, receive noisy feedback, and are restricted in their decisions by their personality, to analyze and explain puzzling empirical evidence on entrepreneurial decision processes. The model is capable of simultaneously explaining the empirical evidence on entrepreneurial entry and exit. In the model, individuals expecting their income to be highest when managed by themselves become entrepreneurs. Given incomplete information about their entrepreneurial abilities, the entrepreneurial entry decision is influenced by the personality metatrait Plasticity. Incomplete information and self-selection, resulting in adverse selection, explain why entrants have substantial overconfidence in their own abilities. After entry, entrepreneurs receive noisy feedback from the market. Depending on the metatrait Stability, entrepreneurs decide to either remain or leave the market. Our model generates behavioral and personality effects consistent with empirical findings. For instance, survival rates decrease at decreasing rates and it is possible to reproduce the earnings puzzle, i.e., a substantial share of entrepreneurs do not necessarily earn more than wage workers.

In the second essay, in Chapter 3, I contribute to the literature on entrepreneurial choice by providing a new perspective. Contemporary theoretical literature consists mostly of models that treat choice outcomes as either deterministic or risky. The essay proposes taking a different (slightly more realistic) perspective by constructing a general career choice model on the basis of the assumption that outcomes are partially ambiguous such that some reward distributions are unknown. The change in perspective yields some major advantages. Learning and career trajectories, which in general cannot be generated by models with deterministic or risky rewards, become a natural feature of the dynamic solution of sequential career choice problems. Consequently, the model accounts for the important phenomenon of mixed careers. Furthermore, earnings-puzzle-like observations can be explained by sufficiently high ambiguity aversion, as ambiguity aversion has a significant impact on learning.

The third essay, in Chapter 4, theoretically and empirically analyzes the impact of media on career choices, where information on entrepreneurship provided by the media is treated as an informational shock affecting prior beliefs. To theoretically examine media

effects, we construct a dynamic Bayesian occupational choice model with sequential decisions under ambiguity due to imperfect information. We show that sufficiently intensive positive media articles and reports about entrepreneurship increase the probability of self-employment and decrease the probability of wage work. Such reports counteract a potential bias against self-employment—caused by asymmetries in ambiguity preferences across occupations in the model. To test model predictions, we use an instrumental variable approach to identify causal media effects using US micro data and a self-constructed country-level macro panel with two different media variables. The empirical analysis shows that an increase in positive media articles and reports about entrepreneurs generates effects on choice probabilities that are consistent with the theoretical model: an increase in the probability of self-employment and a decrease in the probability of wage work.

The fourth essay, presented in Chapter 5, contains an empirical analysis of the effects of cyclical macro variables (GDP and unemployment) on innovative start-ups in Germany. We analyze whether start-up rates in different industries systematically change with business cycle variables using a fixed effects panel approach. We also account for potential endogeneity, effect dynamics, and spatial dependencies. We find mostly counter-cyclical correlations for non-innovative and innovative industries. Entries into large-scale industries, including the innovative part of the manufacturing sector, are more strongly influenced by changes in unemployment levels, while entries into small-scale industries, such as knowledge-intensive services, are mostly influenced by changes in GDP levels. Thus, business formation might have a stabilizing effect on the economy.

In the fifth, and last, essay in Chapter 6, we examine whether information on personality is useful for advice, using the example of career advice. Psychological research often develops profiles based on the average scores of groups such as successful entrepreneurs, where advice results from comparing the profile with the traits of potential entrepreneurs. Given a simple model with a choice problem, where model parameters are estimated with data from the German Socio-economic Panel, and using Monte Carlo methods, we show that under plausible conditions the recommendation performance of the common averages-scores approach can be inferior to the tossing of a coin. Other approaches (such as the direct estimation of outcome probabilities) deliver better performance and are more robust than advice using average scores.

The first, third, and fifth essay result from collaboration with Alexander Kritikos and the fourth essay is the result of collaboration with Michael Fritsch and Alexander Kritikos. The second essay is a single-authored paper.

Zusammenfassung

Diese Dissertation besteht aus fünf eigenständigen Aufsätzen (in Kapitel 2 bis 6), die sich mit verschiedenen Aspekten von Berufswahl, insbesondere der Wahl von Entrepreneurship, unter Risiko und Unsicherheit befassen. Den Aufsätzen vorangestellt ist eine kurze Einleitung (in Kapitel 1).

In Kapitel 2 entwickelt der erste Aufsatz ein Berufswahlmodell mit begrenztrationalen Agenten, die eingeschränkte Information haben, Feedback mit Rauschkomponenten erhalten und in ihren Entscheidungen durch ihre Persönlichkeit eingeschränkt sind, um einige bisher unerklärte Charakteristika von unternehmerischen Entscheidungsprozessen zu analysieren und zu erklären. Das Modell ist simultan in der Lage empirische Regelmäßigkeiten vom unternehmerischen Ein- und Ausstieg zu erklären. Modell-Individuen erwarten, dass ihr Einkommen am höchsten ist, wenn sie von sich selbst gemanagt werden. Angesichts unvollständiger Informationen über ihre unternehmerischen Fähigkeiten wird die unternehmerische Eintrittsentscheidung (Eintritt in die Selbständigkeit) von dem Persönlichkeitsmetamerkmal *Plasticity* beeinflusst. Unvollständige Information und Selbst-Selektion, die zu einer adversen Selektion führen, erklären warum neue Entrepreneure ihre Fähigkeiten erheblich überschätzen. Nach dem Eintritt in den Markt erhalten Unternehmer stochastisches Feedback vom Markt. In Abhängigkeit vom Metamerkmal *Stability* entscheiden sich Unternehmer entweder weiter ihr Unternehmen zu führen oder den Markt zu verlassen. Unser Modell erzeugt Verhaltens- und Persönlichkeitseffekte, die im Einklang mit empirischen Befunden stehen. Zum Beispiel sinken Überlebensraten mit abnehmenden Raten, die Persönlichkeit eines Unternehmers hat einen nicht-linearen Effekt auf den Entscheidungsverlauf und es ist möglich das Einkommenspuzzle zu reproduzieren: Ein wesentlicher Teil der Unternehmer verdient nicht unbedingt mehr als vergleichbare abhängig Beschäftigte.

Im zweiten Aufsatz in Kapitel 3 wird die Literatur zur Karrierewahl durch eine neue Perspektive ergänzt. Die theoretische Literatur besteht meist aus Modellen, die Entscheidungen als deterministisch oder als mit Risiko verbunden behandeln. Der Aufsatz schlägt vor eine andere (etwas realistischere) Perspektive einzunehmen und konstruiert ein allgemeines Modell auf der Grundlage der Annahme, dass Entscheidungsergebnisse (teilweise) unsicher sind, sodass einige Auszahlungsverteilungen unbekannt sind. Der Perspektivenwechsel bringt einige Vorteile mit sich. Zum Beispiel werden Lern- und Karriereverläufe, die im Allgemeinen nicht mit Modellen mit deterministischen oder riskanten Auszahlungen generiert werden können, zu einem intrinsischen Merkmal dynamischer Lösung von

sequentiellen Entscheidungsproblemen. Infolgedessen produziert das Modell gemischte Karrieren (Individuen wechseln mehrmals zwischen Entrepreneur und abhängiger Beschäftigung), die ein wichtiges empirisches Phänomen sind. Darüber hinaus können Beobachtungen, die mit dem Einkommenspuzzle konsistent sind, durch eine hohe Unsicherheitsaversion gepaart mit Überschätzung erklärt werden, da Unsicherheitsaversion einen starken Einfluss auf Lernen unter Unsicherheit hat.

Der dritte Aufsatz in Kapitel 4 analysiert theoretisch und empirisch die Auswirkung von Medien auf Berufsentscheidungen. Informationen über Entrepreneurship, die von den Medien zur Verfügung gestellt werden, werden als Informationsschocks behandelt, die vorherige Überzeugungen beeinflussen. Um Medieneffekte theoretisch zu untersuchen, konstruieren wir ein dynamisches, Bayesianisches Berufswahlmodell mit sequentiellen Entscheidungen unter Unsicherheit verursacht durch unvollständige Informationen. Wir zeigen, dass ausreichend intensive positive Medienartikel und Berichte über Unternehmer die Wahrscheinlichkeit von Selbständigkeit erhöhen und die Wahrscheinlichkeit von abhängiger Beschäftigung verringern. Solche Berichte reduzieren einen potenziellen „Bias“ gegen Selbständigkeit, der durch asymmetrische Unsicherheitsaversion – Aversion assoziiert mit Selbstständigkeit und abhängiger Beschäftigung unterscheiden sich – im Modell verursacht wird. Um Modellvorhersagen zu testen verwenden wir einen Ansatz mit Instrumentenvariablen, um kausale Medieneffekte mithilfe von US-Mikrodaten und einem selbst konstruierten Makropanel auf Länderebene mit zwei verschiedenen Medienvariablen zu identifizieren. Die empirische Analyse zeigt, dass eine Zunahme positiver Medienartikel und Berichte über Unternehmer Effekte auf Wahlwahrscheinlichkeiten generiert, die mit dem theoretischen Modell übereinstimmen. Die Wahrscheinlichkeit von Selbständigkeit erhöht sich und die Wahrscheinlichkeit von abhängiger Beschäftigung wird verringert.

Der vierte Aufsatz, der in Kapitel 5 vorgestellt wird, enthält eine empirische Analyse der Auswirkungen von zyklischen, regionalen Makrovariablen (Bruttoinlandsprodukt und Arbeitslosigkeit) auf innovative Start-Ups in Deutschland. Wir analysieren, ob der Konjunkturzyklus unterschiedliche Auswirkungen auf Start-Ups in innovativen und nicht-innovativen Industrien hat, und in Groß- und Kleinindustrien mit ökometrischen Modellen, die regionale Besonderheiten und die zeitliche Dimension berücksichtigen, aber auch für unbeobachtete räumliche Abhängigkeiten kontrollieren. Wir finden vorwiegend kontra-zyklische Effekte für nicht-innovative und innovative Industrien, wobei Eintritte in den innovativen Teil des verarbeitenden Sektors stärker durch Veränderungen der Arbeitslosigkeit beeinflusst werden, während Eintritte in wissensintensive Dienstleistungen vor allem durch Veränderungen des BIP-Niveaus beeinflusst werden. Zusätzlich gibt es einen klaren Größenaspekt. Eintritte in Kleinindustrien reagieren meistens kontrazyklisch auf Veränderungen des BIP, während Eintritte in Großindustrien meist kontrazyklisch auf Veränderungen der Arbeitslosigkeit reagieren. Somit könnte Unternehmensgründung eine stabilisierende Wirkung auf die Gesamtwirtschaft haben.

Im fünften und letzten Aufsatz in Kapitel 6 untersuchen wir am Beispiel der Berufsberatung, ob Informationen über die Persönlichkeit für Vorhersagen und Empfehlungen nützlich sind. In der psychologischen Forschung werden häufig Profile entwickelt, die auf den durchschnittlichen Persönlichkeitsmerkmalen von Gruppen wie erfolgreichen Unternehmern basieren. Empfehlungen resultieren dann aus dem Vergleich des Profils mit den Merkmalen potentieller Unternehmer. Unter Zuhilfenahme eines einfachen Modells, das mit Daten des Sozio-Ökonomischen Panels geschätzt und mit Monte-Carlo-Methoden untersucht wird, zeigen wir, dass unter plausiblen Bedingungen der auf Profilen basierende Ansatz durch den Wurf einer Münze geschlagen werden kann – die Münze hat eine höhere Trefferquote. Andere Ansätze (z.B. eine direkte Schätzung der Erfolgswahrscheinlichkeit) sind um einiges besser und vor allem robuster als Empfehlungen die Profile einsetzen.

Der erste, dritte und fünfte Aufsatz resultierten aus einer Kooperation mit Alexander Kritikos und der vierte Aufsatz ist das Ergebnis einer Zusammenarbeit mit Michael Fritsch und Alexander Kritikos. Der zweite Aufsatz ist in alleiniger Autorenschaft verfasst.

Contents

1 INTRODUCTION

Il n'est pas certain que tout soit incertain.

(Blaise Pascal, Pensées)

We experience the effects of imperfect information on a daily basis. Many decisions are based on estimates and not on full knowledge—an example is the simple decision of whether or not to take an umbrella when going out (if it is not already raining). Furthermore, even if we are able to obtain additional information, the informational signals we receive are likely to be noisy. Black (1986) argues that without noise financial markets would be impossible but, at the same time, noise makes them imperfect. Imperfect information can manifest itself in several ways. However, there are two interrelated manifestations of imperfect information that are of particular interest: risk and ambiguity. In this contribution, the difference between risk and ambiguity is defined as the difference between the distribution of outcomes and the (meta) distribution of outcome probabilities (see Fellner 1961; Frisch & Baron 1988; Camerer & Weber 1992). Consider a stochastic outcome variable (for instance, the monetary reward generated by some occupation) X, which is important for some decision problem. Assume that X can take only the two values "high" and "low." The distribution of X can be fully described by the probability that X takes the value "high," which we denote by ϕ. If ϕ is known, we are confronted by a decision problem with risk. If ϕ is unknown—even when there is some subjective distribution describing ϕ—the corresponding problem is affected by ambiguity.

Information on outcome probabilities is unlikely to be perfect for a number of relevant decision problems. One important problem setting, relevant for almost everyone, where the assumption of known outcome probabilities is unlikely to hold, is career choice. For instance, entrepreneurship is an ambiguous option *par excellence.* Being successful as entrepreneur is a process of trial and error (Manso 2016) with unknown outcome distributions. However, most approaches in the literature (e.g., Lucas 1978; Kihlstrom & Laffont 1979; Holmes & Schmitz 1990; Lazear 2005; Roessler & Koellinger 2012) abstract from career choice dynamics involving unknown outcome probabilities and only assume deterministic or risky outcomes.

This study, consisting of five essays, contributes to the existing literature on career (especially, entrepreneurial) choice by examining different aspects of occupational choices under ambiguity. As it is reasonable to treat ambiguity as a consequence of a lack of information (Camerer & Weber 1992), the study focuses on five aspects of information, where each aspect is treated in a separate essay. The following aspects are discussed:

1. choices on the basis of noisy entrepreneurial income feedback that is processed by heterogeneous entrepreneurs differing with respect to their personality;

2. the ability of ambiguity-averse individuals to correct false prior beliefs and to learn by transitioning between occupations (i.e., trial and error);

3. the effects of informational shocks in favor of an occupation (with a focus on self-employment);

4. the reaction of individuals to information on unanticipated shocks to macroeconomic variables such as GDP and employment; and

5. the usefulness of information on the entrepreneurial personality for career advice.

The first two aspects are mostly examined from a theoretical perspective, whereas the last three aspects are treated from an empirical or applied point of view.

Essay 1: Entrepreneurial decision making under incomplete information
The first essay, in Chapter 2, titled 'Entrepreneurial decision making under incomplete information,' presents a theoretical model with bounded rationality capable of replicating and explaining puzzling evidence on entrepreneurial decision processes. Rationality is bounded under three restrictions. First, it is assumed that individuals have incomplete information about entrepreneurial abilities and decisions are based on ability estimates. Second, individuals learn about the characteristics of their abilities once they ventured their businesses but the (market) feedback they receive, in form of entrepreneurial income, is noisy. Third, the decision process of an individual is affected by her personality, meaning that not all decision paths are feasible with the same probability. Personality is modeled, in line with psychological research, using the two metatraits of Stability, the stability of one's motivation, and Plasticity, a factor combining extraversion and openness to new experiences (Digman 1997; DeYoung, Peterson & Higgins 2002; DeYoung 2006).

Striking empirical features of entrepreneurial decisions are that individuals tend to systematically overestimate their own entrepreneurial abilities (Busenitz & Barney 1997; Camerer & Lovallo 1999; Koellinger, Minniti & Schade 2007), such that a substantial share of entrepreneurs fail shortly after start-up (Helmers & Rogers 2010; Quatraro & Vivarelli 2015). Moreover, some individuals remain in entrepreneurship although they could earn more in wage employment: The incomes of entrepreneurs are only sometimes higher than those of their counterparts in wage work (e.g., Hamilton 2000; Astebro & Chen 2014; Levine & Rubinstein 2016; Manso 2016). Research points to the entrepreneurial personality as one potential explanation (Holland 1997; Baum & Locke 2004; Zhao & Seibert

2006; Rauch & Frese 2007), but it also shows that although personality is important, its influence on decision making is non-linear—different traits affect entrepreneurial entry and survival (Caliendo, Fossen & Kritikos 2014).

To construct a choice model with learning dynamics, the static occupational choice model of Roessler & Koellinger (2012) is combined with the empirically tested learning algorithm of Camerer & Ho (1999). The model generates a number of simple mechanisms explaining empirical evidence. A simple selection mechanism to separate entrepreneurs from workers involves individuals becoming entrepreneurs because they assume that their entrepreneurial productivity, generating entrepreneurial income, is large enough to offset entrepreneurial costs and the alternative wage income. Hence, the distribution of estimated entrepreneurial productivities of entrants into entrepreneurship is truncated from below and the average of estimated abilities of entrants is pushed above the population average, where the population combines entrants and non-entrants. If actual entrepreneurial abilities are unknown, this results in a substantial overestimation of entrepreneurial abilities, found in empirical data, because the upward movement in average estimated ability levels does not reflect actual entrepreneurial abilities. Entrepreneurial overconfidence is generated even if estimates are right on average.

Entrepreneurial survival curves generated by the model replicate empirical observations: Survival probabilities decreasing at a decreasing rate because the market sorts out new entrepreneurs who do not possess abilities necessary for survival. As most sorting takes place in initial periods, when entrepreneurs receive feedback for the first time, the probability to fail is largest in this turbulent phase. In later periods, the probability to survive stabilizes. After entry, the Plasticity trait becomes less important than the Stability trait because survivors are mostly defined by the Stability trait, as its impact increases with estimation errors, whereas the effect of Plasticity is not influenced by errors, replicating the observation that different personality traits are correlated to entry and survival.

Some entrepreneurs in the model could earn more in wage employment but do not change the occupation. The reason is that entrepreneurial income is an imperfect signal and, thus, the best option is not identified with a probability of 100%. However, entrepreneurs exhibit a tendency to approach the optimal choice. The mechanisms in the model are sufficiently robust to relaxing assumptions, which is demonstrated in a simulation exercise.

Essay 2: Career choice under ambiguity

The second essay, in Chapter 3, titled 'Career choice under ambiguity,' uses the multi-armed Bayesian bandit problem to analyze career choices in an environment with unknown outcome probabilities (ambiguity). In such a setting, arms of a bandit machine can be used to represent choice options (entrepreneurship, wage work, and unemployment) with unknown payoff distributions, with individuals obtaining information about reward distributions by selecting arms and observing the outcome. The problem is to decide which arm to select in every period and, especially, when to switch between options conditional on new information (Gittins, Glazebrook & Weber 2011).

The bandit approach has a number of advantages. For instance, it allows discussing ambiguity, ambiguity preferences, and choices under ambiguity in a fully rational systematic Bayesian setting. Thus, the main difference between the model in Chapter 2 and 3 is that, in contrast to the model in Chapter 2, rationality is less bounded Chapter 3. More importantly: As ambiguity induces learning and learning involves switching between occupational options, the bandit model produces mixed careers (a mixture of spells in entrepreneurship, wage work, and unemployment), which are an important empirical phenomenon (Arthur 1994; Hall & Moss 1998; Valcour & Tolbert 2003; Arum & Müller 2004; Granrose & Baccili 2006; Sullivan & Baruch 2009; Biemann, Zacher & Feldman 2012) but cannot be captured by most existing occupational choice models. Furthermore, the model captures the trade-off between exploiting the best option known so far and exploring potentially better options.

Given that individuals are most likely ambiguity-averse (Ellsberg 1961), I use a decision rule with ambiguity preferences based on the mean-variance model of Maccheroni, Marinacci & Ruffino (2013) to compare optimal decisions (see Gittins & Jones 1974; Weber 1992; Gittins et al. 2011) and more realistic decisions affected by ambiguity aversion. The two main finding are, first, that ambiguity-averse individuals tend to not transition enough between occupations and to transition rather inefficiently compared to optimal behavior. Second, given entrepreneurial overconfidence, formalized as misleading prior knowledge, individuals with ambiguity aversion show a tendency of low learning efficiency. Due to this "trapped-by-overconfidence" effect, the correction of false prior beliefs can take exceptionally long. Inefficient learning, caused by ambiguity aversion, is a simple explanation for the earnings puzzle—the fact that a large share of entrepreneurs permanently remains in entrepreneurship although they could earn more in wage work. Thus, even with less strict boundaries on rationality than in Chapter 2, learning is not perfect and the best option might not be identified. Also, ambiguity-averse individuals oversample from less ambiguous options. Observable decision patterns are consistent with model predictions.

Essay 3: Media and occupational choice

The third essay, in Chapter 4, titled 'Media and occupational choice,' discusses the effects of informational shocks in a setting with ambiguity. In particular, it analyzes whether positive media articles and reports about entrepreneurs affect occupational choices. The focus is restricted to the choice between self-employment and wage work—voluntary unemployment is not discussed because it is hard to identify in data.

Using the hypothesis that media articles and reports are most likely to affect the meta distribution of outcome probabilities (reading an article about, for instance, a famous entrepreneur is unlikely to alter the riskiness of projects), the model in Chapter 3 is modified to generate predictions on media effects. More specifically, by exploiting Lazear's (2005) finding that jacks-of-all-trades are more likely to be entrepreneurs than specialists (also, see Wagner 2006), it is possible to alter the model in Chapter 3 such that choice probabilities are straightforward to derive.

Introducing informational shocks favoring entrepreneurship results in an increase in the probability to select self-employment and a decrease in the probability of wage work. This shift in probabilities improves decision performance if ambiguity aversion is asymmetric across options, such that there is higher aversion associated with self-employment than with wage work. The condition of asymmetric preferences is likely to hold as the self-employed can endure actual losses (their income can become negative due to debt), while the losses of wage workers are bounded by zero from below (the most disadvantageous outcome is job loss resulting in a zero wage).

Model predictions are tested using instrumental variable approaches (heteroskedastic IV probit and two-stage least squares with a transformed dependent variable) and two data sets: a micro panel based on the US National Health Interview Survey and a self-constructed macro panel on the basis of data from 38 countries provided by, *inter alia*, the World Bank, Transparency International, and the Global Entrepreneurship Monitor. To induce exogenous variations in media articles and reports favoring entrepreneurship, we use the occurrence of natural disasters in *other* regions and countries. This instrument is unlikely to directly influence occupational choices—for instance, a natural disaster in New Zealand does not have direct effects on entrepreneurial incentives in the Midwestern United States—but disasters generate top-priority news that crowd out other news, such as news about successful entrepreneurs. Data supports predictions derived from the theoretical model and effects are in line with previous findings. For example, in the macro panel, a one percentage point increase in the share of individuals noticing frequent reports about successful entrepreneurs increases the probability of self-employment by 0.5 percentage points and reduces the probability of wage work by 0.4 percentage points.

Essay 4: Business cycles and start-ups across industries

The fourth essay, in Chapter 5, titled 'Business cycles and start-ups across industries: An empirical analysis for Germany,' examines the impact of business cycles on new business formation in innovative and non-innovative industries, and in large- and small-scale industries, using the ZEW Founder Panel in combination with German administrative data. Theoretical arguments leave room for different effect directions, though previous results indicate that the business cycle indeed impacts start-ups (Parker 2012a; Koellinger & Thurik 2012). Furthermore, it is unclear whether effects differ between innovative and non-innovative industries, and industries with different minimal efficient sizes.

Two variables related to the general business cycle are of interest: GDP and unemployment. An effect is said to be pro-cyclical if entries are high when GDP is high or unemployment is low; counter-cyclical if entries are high when GDP is low or unemployment is high; and a-cyclical if entries do not react to a change in GDP or unemployment. As mentioned, theoretical predictions are rather ambiguous. Considering, for example, GDP effects, on the one hand, Rampini (2004) suggests that entries into entrepreneurship should be higher during boom periods, as expectations about the future are relatively optimistic during upswings, whereas, on the other hand, Francois & Lloyd-Ellis (2003) argue that entries are higher in recessions because production costs are relatively low (wages are depressed, interest rates are lower due to counter-cyclical monetary policy, and so on). To systematize the analysis, we develop three research questions, which can be also derived from a matching model on the basis of Mortensen & Pissarides (1994).

The research questions and the model account for the innovation potential of a firm and minimally required firm sizes, by specifying the amount of non-entrepreneurial labor necessary for production. Questions are as follows:

Q1 Are there differences in the effects of business-cycle variables with respect to entrepreneurial entry if we differentiate between innovative and non-innovative industries, and large- and small-scale sectors?

Q2 Do cyclical variations in GDP levels exhibit a pro-, counter-, or a-cyclical correlation with business entries conditional on innovation potential and minimal efficient size?

Q3 Which types of correlations (viz., pro-, counter-, or a-cyclical), conditional on size and innovation potential, can be expected for cyclical variations in unemployment?

The rationale behind the second question is that if optimism about the future is the dominant driving force, we should expect pro-cyclical effects of GDP on entry. However, if opportunity and resource costs (for instance, costs of hiring labor) are important, which is most likely to hold for large-scale industries, there should be a counter-cyclical relation between GDP and entrepreneurial entry. A similar line of reasoning is behind the third question, as a recession might create incentives to start a business because the probability of unemployment increases and labor becomes cheaper.

Business cycle effects are estimated with a fixed-effects panel model, where regional variations in Germany (NUTS 2 level) in gross entry into entrepreneurship are combined with variations over time (1996–2008) and cycles are isolated with a HP filter (Hodrick & Prescott 1997). All dependent variables are included with a time lag of one period, such that endogeneity problems are reduced. We also test a specification where the error term includes spatial lags and account for the most likely source of endogeneity of errors: an unobserved variable driving business entry and the cycle at the same time. We find a clear size effect. Entries into small-scale industries mostly react counter-cyclically to GDP, whereas entries into large-scale industries are mostly counter-cyclically affected by unemployment. We find nearly no support for pro-cyclical GDP effects, while counter-cyclical unemployment effects are common for all types of entrepreneurs. The exception are entries by entrepreneurs with a high innovation potential not employing a large amount of labor who do not react to the unemployment cycle. In general, innovative activities are more often than not started when economic conditions are in downturn.

ESSAY 5: ARE PERSONALITY PROFILES USEFUL FOR PREDICTION AND ADVICE
The fifth and last essay, in Chapter 6, titled 'Are personality profiles useful for prediction and advice? A Monte Carlo study,' assesses whether statistically significant differences in personality between groups can be used to produce valuable advice for individuals. As personality predicts occupational choices (Zhao & Seibert 2006; Rauch & Frese 2007; Zhao, Seibert & Lumpkin 2010), the essay concentrates on career advice. To examine recommendation performance, we construct a simple model on the basis of the bivariate normal distribution. There are two stochastic variables: personality and entrepreneurial fitness. The potential predictive power of personality stems from a correlation between the personality variable and fitness, where fitness determines whether an individual would be successful in entrepreneurship. The problem to solve is to determine whether a client (of a consultant) has sufficient entrepreneurial fitness by only using information on personality traits, when historical information on the joint distribution of personality and fitness is available.

The essay discusses the most common approach involving the construction of a representative personality profile of an entrepreneur and a non-entrepreneur (see, for instance, Obschonka, Schmitt-Rodermund, Silbereisen, Goslin & Potter 2013). This approach can be derived from a holistic concept of personality effects (Magnusson & Törestad 1993). Formally, the researcher (or consultant) computes the average personality scores of entrepreneurs and non-entrepreneurs, using historical data, and predicts whether a particular individual is an entrepreneur or not by examining the "distance" (see Cronbach & Gleser 1953) between the individual's personality and representative scores.

To benchmark the profile approach, we conduct a two-step procedure. In the first step, we estimate our model with data from the German Socio-economic Panel, which provides information on risk attitudes (personality trait) and individual differences in received self-employment income and wages (entrepreneurial fitness), where self-employment income and wages are observed for the same individual. In the second step, we run several Monte Carlo experiments. The robustness of our results is checked by using additional parameter combinations. We show that the common average-scores approach (representative personalities for entrepreneurs and non-entrepreneurs) can provide even worse recommendations than the toss of an unbiased coin. Thus, comparing personality scores with average scores of groups, which is done by internet-based questionnaires, although banks and investors are also periodically tempted to use personality inventories as part of their decision process, has a high risk of producing wrong predictions and recommendations.

In summary, this contribution establishes the importance of different aspects of information (noise in feedback, false prior beliefs, information favoring an occupation, information on shocks to the macroeconomic environment, and information on personality differences between groups of individuals) in a career-choice setting with risk and ambiguity. Moreover, it shows that different types of individuals (individuals with high or low motivational stability, individual with high or low ambiguity aversion, etc.) and firms (innovators and non-innovators, and firms with different minimal-size requirements) react to new information in a type-specific way. For instance, given the same feedback, an individual with a low intrinsic motivation might abandon her business, whereas an individual with a high motivation does not.

All of the essays are supplemented by appendices with proofs, additional information, and further results. Appendices, which are numbered in line with chapter numbers (for example, the appendix of Chapter 4 is Appendix 4) are collected in Chapter 7. Notations are consistent within an essay (chapter) but not necessarily consistent between essays.

Table 1.1 and 1.2 summarize research questions and results. The tables also show my own contribution, my contribution as a share of the overall workload, and publication status. Four of the five essays have co-authors, while the second essay is based on a single-authored paper.

TABLE 1.1. Overview: Chapter 2 and 3

Chapter	Title	Co-authors	Research question	Main results	Own contribution	Contribution share	Publication status
2	ENTREPRENEURIAL DECISION MAKING UNDER INCOMPLETE INFORMATION	Alexander Kritikos	Is it possible to construct a formal economic model replicating and explaining a sufficiently large number of puzzling observations on entrepreneurial choice dynamics?	Model with bounded rationality can explain and replicate five of the most striking features of entrepreneurial decision processes	Model, analytical results, and simulations	65%	Under review in European Economic Review (first round)
3	CAREER CHOICE UNDER AMBIGUITY		How are career choices affected by ambiguity, ambiguity preferences, and potentially misleading prior information (overconfidence)?	In a realistic setting with high ambiguity aversion, individuals inefficiently correct false beliefs and learn inefficiently in general, which is consistent with empirical findings (especially, the earnings puzzle)	Model, analytical results, simulations, and empirical tests	100%	Under review in Economic Theory (first round)

Table 1.2. Overview: Chapter 4, 5, and 6

Chapter	Title	Co-authors	Research question	Main results	Own contribution	Contribution share	Publication status
4	MEDIA AND OCCUPATIONAL CHOICE	Alexander Kritikos	Does media have an effect on occupational choices?	Media has theoretically and empirically a causal impact on occupational choices	Model, predictions, and estimations	75%	Ready for publication
5	BUSINESS CYCLES AND START-UPS ACROSS INDUSTRIES: AN EMPIRICAL ANALYSIS FOR GERMANY	Michael Fritsch and Alexander Kritikos	How does the business cycle influence new business formation across industries? Are there differences between innovative and non-innovative industries, and large and small scale industries?	Business cycle effects differ across industries; effects are mostly counter-cyclical such that innovative firms are more often started during economic downturns	Matching model, predictions, estimations, robustness checks	40%	Under revision in Journal of Business Venturing (second round)
6	ARE PERSONALITY PROFILES USEFUL FOR PREDICTIONS AND ADVICE? A MONTE CARLO STUDY	Alexander Kritikos	Can information on personality differences between groups be exploited to generate predictions and advice, given the example of career advice?	The common approach of comparing group averages results in low recommendation performance that is sometimes overperformed by a coin	Model, calibration, and simulations	60%	Ready for publication

2 ENTREPRENEURIAL DECISION MAKING UNDER INCOMPLETE INFORMATION

This chapter presents a dynamic occupational choice model with bounded rationality, accounting for individual personality and entrepreneurial learning, that is capable of simultaneously explaining the empirical evidence on entrepreneurial entry and exit. In our model, individuals expecting their income to be highest when managed by themselves become entrepreneurs. Incomplete information about entrepreneurial abilities and self-selection explain why entrants have overconfidence in their own abilities, influencing their decision to become entrepreneurs. After entry, entrepreneurs receive noisy feedback from the market. Depending on their personality entrepreneurs decide to either remain or leave the market. Our model generates behavioral effects consistent with empirical findings. For instance, survival rates decrease at decreasing rates, certain types of individuals persistently remain in entrepreneurship, and it is possible to reproduce the entrepreneurial earnings puzzle, i.e., a certain share of entrepreneurs earn less than similar wage workers in the short run.

2.1 Introduction

Many economic models of entrepreneurship assume that, based on complete information about their levels of entrepreneurial abilities, individuals make a 'once and for all' decision and either become entrepreneurs or employees for the rest of their working life (see, e.g., Lucas 1978; Kihlstrom & Laffont 1979; Holmes & Schmitz 1990; Lazear 2005; Roessler & Koellinger 2012). This kind of "die-hard" entrepreneur (Burke, FitzRoy & Nolan 2008) may exist, but this condition does not hold for the majority of individuals making the decision to become entrepreneurs. There is first evidence indicating that individuals might experiment with occupational choices and become entrepreneurs for a short period of time before deciding to return to a salaried job (see, *inter alia*, Manso 2016).

There are more stylized facts on entrepreneurial decision processes creating several puzzles. Individuals appear to systematically overestimate their own entrepreneurial abilities before establishing a business (Busenitz & Barney 1997; Camerer & Lovallo 1999; Koellinger et al. 2007). Therefore, a substantial number end their entrepreneurial adventure before it really started. Overall, around 50 percent of all young entrepreneurs leave this occupational form within the first five years after start-up; about half of these even after one year (Hyytinen & Rouvinen 2008; Helmers & Rogers 2010; Quatraro & Vivarelli

2015). Some other individuals persistently remain in entrepreneurship even though they could have earned more by returning to wage employment (e.g., Hamilton 2000; Astebro & Chen 2014; Levine & Rubinstein 2016). Confronted with this evidence, research points to the entrepreneurial personality as one potential explanation of these phenomena (Holland 1997; Baum & Locke 2004). Indeed, empirical research shows that personality seems to play an important role in entrepreneurship (Zhao & Seibert 2006; Rauch & Frese 2007). However, its influence on entrepreneurial decision making is non-linear and differs between entry, income from entrepreneurship (Hamilton, Papageorge & Pande 2017), and survival in entrepreneurship (Caliendo et al. 2014), opening up a new avenue to explain the puzzling empirical evidence.

From a bird's-eye view, one may easily get the impression that entrepreneurial decision processes are highly inefficient. Looking more deeply into the informational status of individuals making such decisions shows that they face many uncertainties. In particular, when they make this occupational choice for the first time, they will feel great uncertainty about their own entrepreneurial abilities—and research shows that there is a huge population variance in entrepreneurial abilities (Astebro & Chen 2014). Even more so, individuals may also face great uncertainty about the value of their idea (Kerr, Nanda & Rhodes-Kropf 2014), resulting in uncertainty about whether their newly established enterprise will survive in the market. As a consequence, there is *ex ante* genuine uncertainty about future entrepreneurial earnings. To a lesser extent, this kind of uncertainty also holds for serial entrepreneurs.

Uncertainty continues to exist once young entrepreneurs receive market feedback in terms of profits and losses for the first time. Theoretically, feedback allows learning about entrepreneurial abilities (Manso 2016). But, how should levels of abilities be inferred when, in the first year of the entrepreneurial career, the generated income is not negative but below the earnings in the last salaried position—as it often happens during the first year of an entrepreneurial adventure (Dillon & Stanton 2016)? Such feedback does not allow to perfectly determine whether the individual is an able entrepreneur or not. So, some individuals leave entrepreneurship and might make a type-1 error, they should have remained, while others make type-2 errors and remain although they should have left.

Thus, entering entrepreneurship is a decision under an uniquely high level of uncertainty. Individuals will come up against boundaries they cannot overcome, when they aim to reduce the level of uncertainty by acquiring additional information. Therefore, we argue that individuals make entrepreneurial decisions under bounded rationality (Simon 1959).

In this chapter, we ask whether it is possible to construct a model of occupational choice that captures such a decision process. By proposing a formal occupational choice model where individuals make decisions in an environment with bounded rationality, we show that we are capable of reproducing all of the puzzling observations on entrepreneurial entry, income, and survival. Backed by recent evidence, we construct a bounded-rationality model by introducing the following three features. First, we assume that individuals have *incomplete information* about their own entrepreneurial abilities. As a consequence, decisions are based on ability estimates. Second, individuals can learn about the characteristics of their abilities once they venture their businesses but the market feedback they receive is *noisy*. Third, the decision process of an individual is affected by her *personality* such that not all decision paths are feasible with the same probability.[1]

The model we construct extends the static approach of Roessler & Koellinger (2012) and combines it with the learning algorithm of Camerer & Ho (1999). This particular algorithm, tested with experimental data, is altered so that personality traits are allowed to interact with the learning process. A typical decision sequence generated by our model is as follows. First, traits influence the initial estimate of entrepreneurial abilities and, thus, of the initial decision for or against entrepreneurship. The initial decision is defined as an occupational choice under a restricted set of information—when individuals, at the beginning of their professional career, have no personal experience with dependent employment or entrepreneurship. Second, after entry entrepreneurs receive noisy (positive or negative) feedback from the market. Given feedback, respectively personal experience, with a certain occupational status, traits, different from those influencing the entry decision, determine information processing and, consequently, affect occupational choices based on new information. The decision of each individual can either be to keep their current status—remaining an entrepreneur or worker—or to change their status—to start or abandon an own business.

We show that incomplete information about own entrepreneurial abilities can result in a substantial overestimation of entrepreneurial productivity, even if individuals are not overconfident on average. Thus, overconfidence is not necessarily generated by a cognitive bias but can be produced by a simple selection mechanism. A further important insight of our model is that the heterogeneity of personality leads to different decision probabilities given the same specification of market feedback. This explains why entrepreneurs with a certain personality remain longer in the market than others. More specifically, it seems that motivational stability or perseverance, the reluctance to fundamentally revise the initial decision introduced by Duckworth, Peterson, Matthews & Kelly (2007), is important for entrepreneurial survival if individuals make estimation errors. Entrepreneurial entry is mostly affected by the breadth of thinking.

Learning heuristics are such that, similar to empirical evidence, survival probabilities

[1]Personality traits are recent introductions to economic theory. Non-standard preferences and bounded rationality are two ways of integrating personality traits (Almlund, Duckworth, Heckman & Kautz 2011). Moreover, Borghans, Duckworth, Heckman & ter Weel (2008) depict an array of ways of how to integrate cognitive and non-cognitive skills into economic theory.

of entrepreneurs decrease at a decreasing rate. Finally, we also show that entrepreneurs can earn less than similar wage workers because the noisy feedback signal might not convince some entrepreneurs, who would be better off as workers, to abandon entrepreneurship. In general, unknown factors and noise can produce deviations from the full-information equilibrium, the objectively best choice, although choices exhibit a tendency to approach the full-information solution.

There are few other contributions formalizing transitions between entrepreneurship and wage work. For instance, Manso (2016) develops a model where individuals are assumed to experiment with entrepreneurship, which is a central assumption in recent models (see Daly 2015; Dillon & Stanton 2016), discussed in Section 2.7 in more detail. We acknowledge that experimentation may play an important role. However, models using the notion of experimentation usually assume that decisions are made under full rationality and learning is optimal. This assumption, as we show later, comes at a certain cost. Entry as well as exit patterns—and learning and behavioral patterns in general—differ substantially from the ones actually observed. A fully rational experiment cannot account for the fact that personality influences how individuals learn (Gill & Prowse 2016) and, consequently, the learning outcome.[2] Furthermore, a fully rational experiment reduces the notion of uncertainty and behavior under uncertainty to a problem of how to optimally deal with lack of information and new information obtained in the course of learning. However, training programs for potential entrepreneurs that focus on providing information, do not appear to have an effect on entry decisions (Karlan & Valdivia 2011; Hirshleifer, McKenzie, Almeida & Ridao-Cano 2015; Fairlie, Karlan & Zinman 2015).[3] Concentrating exclusively on information and ignoring personality might, therefore, produce non-optimal policy suggestions—for instance, if support programs for business start-ups concentrate too much on providing new information and neglect other important aspects, such as selecting individuals with an entrepreneurship-prone personality profile.

The remainder of the chapter is organized as follows. Section 2.2 reviews previous research. Section 2.3 presents a basic outline of our model. In Section 2.4, we present the model's core—its static components. Section 2.5 introduces dynamics. In Section 2.6, we derive the main results. Section 2.7 discusses how our model relates to recent approaches, especially approaches using the notion of an experiment, and points to limitations. Section 2.8 summarizes. Appendix 2 provides proofs not included in the main text.

[2] Chapter 3 analyzes the impact of personality-related ambiguity preferences in an experimental setting.

[3] That is not to say that information has no effect at all. In Chapter 4, it is demonstrated that media reports affect occupational choice probabilities.

2.2 Previous research

Entrepreneurial decision making attracts considerable attention in entrepreneurship research. There are several threads of research on occupational choice, entrepreneurial abilities, personality and entrepreneurship, entrepreneurial learning, and personality and learning.

2.2.1 Occupational choice models and entrepreneurial abilities

An entry into entrepreneurship is, first of all, an occupational choice, which is the first research line we address here. Earlier models analyze variables that drive this choice and focus on entrepreneurial abilities from different angles. It is often assumed that people select their occupation according to its expected utility and start to pursue an own venture when this appears to be more rewarding than being a paid employee (Knight 1921; Evans & Jovanovic 1989; Taylor 1996).

In this context, Lucas (1978) emphasizes the entrepreneurial talent of individuals. Kihlstrom & Laffont (1979) highlight that entrepreneurs need to be more risk tolerant than workers. The learning model of MacDonald (1988) can be interpreted as a choice model with two kinds of entrepreneurs, with either low or high entrepreneurial abilities. According to his model, individuals with high abilities remain in entrepreneurship, while low-ability individuals return to wage employment after some time. Holmes & Schmitz (1990) point to the opportunity-seeking character of entrepreneurship. They assume that transforming entrepreneurial abilities into running a business is a random process inducing individuals with higher entrepreneurial abilities to have a higher probability of successfully running a firm. Lazear (2005) argues that as a consequence of personal preferences and talent, some people choose to be jacks-of-all-trades, where others specialize, while jacks-of-all-trades have more entrepreneurship-prone abilities. Roessler & Koellinger (2012) provide an explanation for the emergence of entrepreneurs by simultaneously determining occupational choice and job matching. They argue that those who are relatively unmanageable, while possibly being excellent managers of themselves and others, become entrepreneurs. Dillon & Stanton (2016) assume, in a dynamic model, that entrepreneurs always have the option to return to wage work. They show that the option to transition to wage work constitutes a substantial non-pecuniary benefit for entrepreneurs, which can explain why some entrepreneurs appear to accept incomes lower than wage-work incomes in cross-sectional data.

2.2.2 Imperfect information about entrepreneurial abilities

Entrepreneurial abilities are, thus, seen as key determinant of entrepreneurial choice and success. However, at the same time, research emphasizes that individuals have incomplete information about their entrepreneurial abilities before starting their own firms, which is the central assumption of our model and the second line of research we address here.[4] In recent research, starting a firm is also viewed as an experiment with an unknown outcome (e.g., Jovanovic 1982; MacDonald 1988; Hintermaier & Steinberger 2002; Vereshchagina & Hopenhayn 2009; Campanale 2010; Poschke 2013; Daly 2015; Manso 2016).

Also, there is a growing literature reporting that those who enter into entrepreneurship make self-assessment errors of a certain type, which can be described as overconfidence (Busenitz & Barney 1997; Camerer & Lovallo 1999; Bernardo & Welch 2001; Koellinger et al. 2007). The observation of overconfidence implies that information on entrepreneurial abilities must be imperfect, as in a setting with perfect information judgment mistakes are ruled out. Koellinger et al. (2007) show that individuals tend to rely on subjective perceptions rather than on objective success probabilities and that there is a positive bias: Nascent entrepreneurs tend to overestimate their own success probabilities (also, see De Meza & Southey 1996). Camerer & Lovallo (1999) establish a similar result in an experimental setting (a game of entry into a competitive market). They reveal that if subjects know that outcomes will depend on their own abilities, excess entry is more pronounced than in other settings, such that overconfidence is mostly a property of self-selected entry.[5] In a survey-based approach, Busenitz & Barney (1997) show that entrepreneurs tend to have a stronger confidence bias than managers in large organizations, thereby providing additional support for the phenomenon of generally overconfident entrepreneurs.

2.2.3 Personality and entrepreneurship

The review of findings on uncertainty about entrepreneurial abilities and overconfidence of entrepreneurs points to the fact that the personality of individuals may influence occupational choices. This leads to the third line of research we focus on.

Psychological research claims that personality is an essential determinant of occupational choices (Holland 1997). Moreover, personality theory asserts that the influence of personality variables on entrepreneurial decisions is mediated by the strategies and goals of the decision maker (Baum & Locke 2004). There is also an ongoing discussion with respect to the influence of personality traits on performance in various areas and on different levels. Cunha, Heckman, Lochner & Masterov (2005), for instance, develop a model depicting the development of abilities, including self-productivity, over the life cycle.

[4]Imperfect information in our model is close to Knightian uncertainty (Knight 1921), in the sense that uncertainty about estimates cannot be measured but new information, potentially correcting previous estimates, can be obtained by trial and error.

[5]Overconfidence has positive and negative aspects. A positive aspect is that overconfident individuals generate information that would not be available without them (Bernardo & Welch 2001).

Empirical evidence reveals that personality traits are important for entrepreneurial choice and success. The personality structure of entrepreneurs is distinct from that of managers and workers, either when measured by the psychological toolbox of the Big Five personality construct (Zhao & Seibert 2006) or when measured by a specific set of personality characteristics (Rauch & Frese 2007).

Further parallel research beyond the Big Five construct focuses on specific personality characteristics influencing individual decisions to stay in entrepreneurship. In particular, grit or perserverance is a personality characteristic defined by Duckworth et al. (2007, p. 1087–1088) as the "passion for long-term goals," entailing "working strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress." Thus, perserverance, or grit, unfolds effects that are related to motivational stability. The long-term goal of an individual could be entrepreneurship. A gritty individual will stick to this objective as long as possible. Empirical research shows that one characteristic setting apart successful from non-successful entrepreneurs, and non-entrepreneurs, is indeed perseverance (Gimeno, Folta, Cooper & Woo 1997; Markman & Baron 2003; Markman, Baron & Balkin 2005; Lowe & Ziedonis 2006; Wolfe & Patel 2016). Entrepreneurs select a path and continue despite setbacks (Holland & Shepherd 2013; Muehlfeld, Urbig & Weitzel 2017).

Recent analysis makes, however, clear that observing differences in personality characteristics between entrepreneurs and other population groups does not imply that these traits influence entrepreneurial entry and success in the same manner. It is necessary to separate the influence of personality traits on different entrepreneurial decisions. For example, Caliendo et al. (2014) show that different traits affect entrepreneurial entry and survival. Entry is positively influenced by traits like Openness to Experience and Extraversion, while survival is affected by traits like Agreeableness or Conscientiousness (also, see Ciavarella, Buchholtz, Riordan, Gatewood & Stokes 2004).

To make the diverging influence of personality on entrepreneurial decision making tractable, we reduce the number of factors as much as possible in our model, while keeping them meaningful. We follow Digman (1997) who proposed a higher order of traits referring to α and β factors. The α factor combines Agreeableness, Emotional Stability, and Conscientiousness, the traits mostly influencing entrepreneurial survival. The β factor is a combination of Extraversion and Openness, which influence the entry decision into entrepreneurship. DeYoung et al. (2002) call α factors 'Stability' traits and β factors 'Plasticity' traits. Stability and Plasticity can be interpreted as *meta*traits (Hirsh, DeYoung & Peterson 2009), such that we can refer to *the* Stability and Plasticity trait, respectively to the α and β factor. The diverging effects of the Big Five traits on entrepreneurial entry and survival are assumed to correspond to this higher order of traits.

The potential influence of further personality characteristics—such as grit on survival—will be captured in our model by the Stability factor. In such a way, we will explicitly account for the findings on the connection between personality (i.e., the Big Five and further characteristics) and entrepreneurship by allowing metatraits (Plasticity and Stability) to influence decisions.[6]

2.2.4 Entrepreneurial learning

Imperfect knowledge about the own entrepreneurial abilities leads to the question of whether individuals learn about them over time, and is the fourth line of research addressed here. MacDonald (1988) models choices with the help of an information accumulation process. Astebro, Chen & Thompson (2011) explain entrepreneurship by the existence of frictions in the assignment of individuals to tasks in firms. They argue that inefficient assignments lead to entrepreneurship, while in case of frictionless assignments there is no need for entrepreneurs.

Another theoretical approach related to our model is Braguinsky, Klepper & Ohyama (2012) who assume a three-period learning framework. In the first period, individuals draw an entrepreneurial idea of a certain quality (from a distribution) but do not gain any additional information. In the second period, individuals receive a noisy signal and update the idea's quality prior according to Bayesian rules. In the third period, individuals know the exact quality of the entrepreneurial idea. Assessing the realism of their learning algorithm, Braguinsky et al. (2012) state:

> In reality, feedback about the value of entrepreneurial ideas typically comes in the form of noisy realizations, such as earned profits. Consequently, entrepreneurs never truly learn the value of their ideas but become better informed about them over time as feedback accumulates. (Braguinsky et al. 2012, p. 873)

The main difference between our model—as we show in the next sections—and the model of Braguinsky et al. (2012) is that we do not abstract from these two real-world conditions. The assumption that feedback comes in the form of noisy earnings is fundamental to our model. As a consequence, it is not guaranteed that entrepreneurs learn the true value of their entrepreneurial idea but they can accumulate feedback.

[6]In the literature on personality, the two metatraits have been linked to the two neurotransmitters serotonin and dopamine (DeYoung et al. 2002; DeYoung 2006; Hirsh et al. 2009). Serotonin has been directly linked to the Stability trait (Manuck, Flory, McCaffery, Matthews, Mann & Muldoon 1998; Jang, Livesley, Angleitner, Riemann, Ando, Ono, Vernon & Hamer 2001; Hirsh et al. 2009). Serotonin limits aggression and maintains motivational stability (Hirsh et al. 2009), which is important for entrepreneurial survival. Dopamine has been related to the Plasticity trait (Depue & Collins 1999; DeYoung, Peterson & Higgins 2005; Harris, Wright, Hayward, Starr, Whalley & Deary 2005; Hirsh et al. 2009). Dopamine affects the breadth of thinking (Berridge & Robinson 1998; Panksepp 1998; Hirsh et al. 2009). An open mind is a necessary condition to generate entrepreneurial ideas and to decide to enter entrepreneurship.

2.2.5 *Personality and learning*

The impact of personality on learning, the fifth and last research line we examine, is not a well-researched area of economics but there are some results. For instance, Gill & Prowse (2016) find that personality affects learning in p-beauty contest experiments. More specifically, Gill & Prowse (2016) establish that more agreeable and emotionally stable individuals (both traits belonging to Stability, the α factor) learn faster.

The relation between personality and learning is better studied in educational and personality research. A general finding is that personality is related to learning styles (Furnham 1992; Geisler-Brenstein, Schmeck & Hetherington 1996; Gadzella, Ginther, Master & Guthrie 1997; Zhang 2003; Komarraju, Karau, Schmeck & Avdic 2011). In particular, Komarraju et al. (2011) show that Consciousness, Agreeableness, and Emotional Stability (all belonging to the metatrait Stability) are related to synthesis analysis, methodical study, fact retention, and elaborative processing. We incorporate the findings on personality and learning into our model by introducing a dependency between learning heuristics and personality traits.

2.2.6 *Survival dynamics and the earnings puzzle*

Last but not least, there are two further empirical findings completing the puzzles about entrepreneurial entry, exit, and success. First, focusing on exits from entrepreneurship, there is empirical evidence that entrepreneurial survival rates are only about 50% after five years (Helmers & Rogers 2010; Quatraro & Vivarelli 2015). The highest exit rates are in the first year after entry into self-employment showing a relatively high rate of "revolving-door entrepreneurs," while exit rates in the following years become lower and stabilize (Evans & Leighton 1989; Knaup 2005; Knaup & Piazza 2007; Georgarakos & Tatsiramos 2009). In other words, entry and exit are correlated generating "turbulence" (Quatraro & Vivarelli 2015, p. 279) such that firm survival is mostly threatened immediately after entry.

Second, parallel research analyzes the incomes of entrepreneurs. Several empirical studies find that median entrepreneurs do not earn more than the median wage-employed (Hamilton 2000; Williams 2000; Kawaguchi 2003; Hyytinen & Rouvinen 2008; Hartog, Van Praag & Van der Sluis 2010; Astebro, Braunerhjelm & Broström 2013; Hyytinen, Ilmakunnas & Toivanen 2013; Astebro & Chen 2014), leading to the question: Why do some individuals remain for too long in entrepreneurship if they could earn more in dependent employment?

2.2.7 *Summary and benchmarks*

The studies briefly reviewed in this section discuss several variables and constructs that may explain individuals' decisions to enter and to exit entrepreneurship. Among them, personality traits play a crucial role, with a differing influence on entrepreneurial entry and exit. To capture these observations, we construct a model where personality influences occupational choices, by influencing estimates and learning heuristics. Our model is a synthesis of the research lines discussed above, overcoming difficulties associated with each line. We combine the occupational choice model of Roessler & Koellinger (2012) with the experience-weighted attraction (EWA) learning algorithm of Camerer & Ho (1999) (also, Ho, Camerer & Chong 2007), who model learning as the development of the relative attraction of a particular option over time. A slight modification of the EWA algorithm allows us to introduce personality traits into the learning process.

To examine the explanatory performance of our model, we test whether it can replicate the following five central observations on the decision processes of entrepreneurs:

Entrepreneurial overconfidence
Entrants into entrepreneurship exhibit overconfidence with respect to their entrepreneurial abilities.

Shape of survival curves
Exit rates are high in initial periods after entry, but then decrease and stabilize the longer individuals remain in entrepreneurship.

Impact of personality on entry and survival
Different types of personality traits influence entry and exit decisions.

Grit and survival
More motivationally stable (grittier) individuals survive longer in entrepreneurship.

Earnings puzzle
For a substantial share of entrepreneurs, positive returns to entrepreneurial activities are absent in the short run, while those who remain entrepreneurs for a longer time period are able to produce positive returns.

2.3 Model outline

This section presents a brief outline of the whole model. Our main assumption is that individuals cannot fully anticipate to what extent they will succeed as entrepreneurs because they have incomplete information about their entrepreneurial abilities, but they can make an informed guess and learn by trial and error.

Formally, there are $1, \ldots, I$ individuals who can decide between entrepreneurship and employment, where employment entails working under someone else's supervision. In general, individuals are indexed by i. The set of all individuals is $N = \{1, 2, \ldots, I\}$. Every individual decides whether she starts or abandons entrepreneurship in every period $t > 0$, where $t = 0, \ldots, T$. An action decided upon in one period is executed in the next period.

Individuals are assumed to differ with respect to their personality. Following the discussion in Section 2.2.3, and to keep the model tractable, we assume that the complex measurement of personality (for instance, the Big Five approach) can be captured by two metatraits. The first metatrait Plasticity is denoted by $\beta_i \in \mathbb{T}$, where $\mathbb{T}$ is some subset of $\mathbb{R}$. Second, the Stability trait is given by $\alpha_i \in \mathbb{T}$. Traits are assumed to be time-invariant, such that $\alpha_{i,t} = \alpha_i$ and $\beta_{i,t} = \beta_i$, which is consistent with empirical findings. For example, Cobb-Clark & Schurer (2012) show that the Big Five traits of working-age adults are stable enough to include personality as a time-invariant input into economic models.

Our model consists of two parts: a static occupational choice model, the model's core, and a learning algorithm generating choice dynamics. The occupational choice model is an adaptation of Roessler & Koellinger (2012). This model has many interesting features. For instance, it can generate all common types of firms and entrepreneurs. Firms created by entrepreneurs can consist of solo entrepreneurs or of employers with a large number of additional employees; entrepreneurs can start a business out of necessity or can be opportunity seekers. We modify the Roessler & Koellinger (2012) model by assuming that, instead of known deterministic entrepreneurial abilities, abilities are unknown, an assumption that is consistent with empirical findings (Section 2.2.2). Given known deterministic abilities, the solution of the original model is a Nash equilibrium determining the occupational choice—wage worker or entrepreneur—of every individual $i \in N$ for all t. We refer to this particular solution of the static model as the full-information equilibrium.

However, individuals do not necessarily reach the full-information equilibrium because of imperfect information about entrepreneurial abilities and noisy signals. To model decision dynamics, we rely on the approach of Camerer, Ho & Chong (2008) to learning in games: the experienced-weighted attraction learning or EWA algorithm. EWA assumes that every option has an attraction, a numerical measure, and the option with the highest level of attraction is selected in every period (Camerer & Ho 1999). The attraction of the entrepreneur option is denoted by $\mathscr{E}_{i,t} \in \mathbb{R}$ and the attraction of the best non-entrepreneurial option by $\mathscr{W}_{i,t} \in \mathbb{R}$ for all i and t.

Attraction begins with prior values and is then updated given experience (Camerer & Ho 1999; Camerer et al. 2008) such that $\mathscr{E}_{i,0}$ depends on initial estimates of entrepreneurial abilities denoted by v_{ii}.[7] In line with empirical research on the impact of metatraits on decisions, discussed in Section 2.2.3, we assume that the initial estimate of entrepreneurial abilities is related to the Plasticity trait such that $v_{ii} = h_\beta(\beta_i)$, where $h_\beta : \mathbb{T} \to \mathbb{R}$ and

[7]To keep the model simple, the initial attraction of wage work, $\mathscr{W}_{i,0}$, does not depend on estimates, as we assume that there is no uncertainty associated with wages.

$h'_\beta > 0$. Put differently, the *breadth* of thinking influences how productive individuals believe to be in entrepreneurship.

The attraction of an option is updated according to the algorithm presented in Camerer & Ho (1999). The attraction in the current period depends on the former attraction of the option and the recently observed reward from selecting the option. Both values are weighted by experience weights (Camerer et al. 2008). We simplify the algorithm of Camerer & Ho (1999) such that it has only one parameter: the weight applied to previous attraction. This weight represents motivational stability, as motivation becomes unstable the more individuals rely on recent outcomes, and ignore previous data and their initial estimate of entrepreneurial abilities. As motivational stability is linked to the Stability trait (Section 2.2.3), we assume that the learning heuristic individuals apply depends on a function of Stability $h_\alpha(\alpha_i)$, where $h_\alpha : \mathbb{T} \to [0, 1)$ and $h'_\alpha > 0$. This assumption is also consistent with the empirical finding that personality influences learning styles (Section 2.2.5).

The functions h_α and h_β are rationalizing functions assigning each trait a certain space of action. Assuming that both functions are strictly increasing simplifies interpretations. Although we do not define explicit rationalizing functions, it should be possible to either find functions with required properties or to find proxies for the variables to conduct an econometric analysis.

The reward of employment is a known wage sequence $w_{i,0}, w_{i,1}, w_{i,2}, \ldots w_{i,T}$. Without doubt, it would be more realistic to assume that the wage is also not perfectly known. However, the model is able to generate realistic features of entrepreneurial decision processes even with this assumption only approximating real-world conditions. Entrepreneurship generates observable rewards $\phi_{i,1}, \phi_{i,2}, \ldots \phi_{i,T}$, which are not known beforehand. The relative attraction of wage work $\mathscr{W}_{i,t} - \mathscr{E}_{i,t}$ is, in stylized form, given by

$$\mathscr{W}_{i,t} - \mathscr{E}_{i,t} = d_{\mathscr{W}-\mathscr{E}}(h_\alpha(\alpha_i), h_\beta(\beta_i), \boldsymbol{\phi}_{i,t}, \mathbf{x}_{i,t})$$

$d_{\mathscr{W}-\mathscr{E}}$ is a function defined in Section 2.5. $\boldsymbol{\phi}_{i,t}$ is a vector of observed entrepreneurial incomes and $\mathbf{x}_{i,t}$ is a vector of known variables such as entrepreneurial costs and variables influenced by labor market interactions. Depending on the relative attraction of wage work, we obtain the following decision:

$$\mathbb{d}_{i,t} = \begin{cases} 1 & \text{if } d_{\mathscr{W}-\mathscr{E}}(h_\alpha(\alpha_i), h_\beta(\beta_i), \boldsymbol{\phi}_{i,t-1}, \mathbf{x}_{i,t-1}) \leqslant 0 \\ 0 & \text{if } d_{\mathscr{W}-\mathscr{E}}(h_\alpha(\alpha_i), h_\beta(\beta_i), \boldsymbol{\phi}_{i,t-1}, \mathbf{x}_{i,t-1}) > 0 \end{cases}$$

where $\mathbb{d}_{i,t}$ indicates whether entrepreneurship ($\mathbb{d}_{i,t} = 1$) or wage work ($\mathbb{d}_{i,t} = 0$) is selected in period t by individual $i \in N$. The set of all entrepreneurs in period t is $E_t = \{i \in N : \mathbb{d}_{i,t} = 1\}$ and the set of all workers is $W_t = \{i \in N : \mathbb{d}_{i,t} = 0\}$.

To map attraction into choice probabilities, Camerer & Ho (1999) and Camerer et al. (2008) use an exponential logit rule. We use an alternative and more intuitive way to introduce choice probabilities. As entrepreneurial abilities are unknown, individuals have to become entrepreneurs to learn whether they can be successful in entrepreneurship. Information on entrepreneurial abilities is provided by the market. The market evaluates entrepreneurial proficiency and reports it to the individual with some noise. We refer to the noisy market signal generated by the interaction between an entrepreneur and the market as market feedback.

As suggested by Braguinsky et al. (2012), feedback comes in form of noisy entrepreneurial earnings. Formally, instead of one $\boldsymbol{\phi}_{i,t}$, we have a number of realizations of $\boldsymbol{\phi}_{i,t}$ generated according to some distribution. If feedback was noise-free, entrepreneurs entering in the first period of the game would always find the full-information equilibrium (demonstrated in Appendix 2.B). Thus, given unknown entrepreneurial abilities but fully informative feedback individuals always identify the best option. With noise introduced by the distribution of $\phi_{i,t}$, feedback is not fully informative anymore and introduces random judgment mistakes—the use of an exponential logit rule, as preferred by Camerer et al. (2008), also induces judgment errors. Hence, the probability to select entrepreneurship or wage work is given by

$$\mathbb{P}(\mathbb{d}_{i,t} = 1) = \mathbb{P}(\mathbb{1}\{d_{\mathscr{W}-\mathscr{E}}(h_\alpha(\alpha_i), h_\beta(\beta_i), \boldsymbol{\phi}_{i,t-1}, \mathbf{x}_{i,t-1}) \leqslant 0\})$$

$$\mathbb{P}(\mathbb{d}_{i,t} = 0) = 1 - \mathbb{P}(\mathbb{d}_{i,t} = 1)$$

where randomness is introduced by $\boldsymbol{\phi}$ and the full-information equilibrium is not necessarily identified with a probability of 1.

To analyze our model, we use two approaches. First, by restricting labor market interactions, resulting in a simplified model version, we analytically replicate the five findings on entrepreneurial decision processes, the benchmarks in Section 2.2.7. Especially, the simplified model version allows deriving a straightforward expression for the conditional choice probability $\mathbb{P}(\mathbb{d}_{i,t} = 1|\mathbb{d}_{i,1} = 1)$.

However, as the labor market has a significant impact on decisions and to check whether some assumptions we use in the simplified model are critical, we also provide numerical results for a version of the model without any restrictions on interactions (and more plausible assumptions). The numerical simulation approach also allows to examine properties of the decision behavior of heterogeneous individuals, for instance, by following a cohort of entrepreneurs. We demonstrate that the simplified and the full model provide similar results, such that restrictions on interactions seem not to be crucial. While presenting our model, we provide all elements of the full model. Restrictions are explicitly mentioned.

2.4 The model's core

In this section, we depict the model's static core, adapted from Roessler & Koellinger (2012), explaining how firms are organized. We also derive earning functions. Additionally, the section analyzes the preferences of entrepreneurs with regard to the construction of their firms.

The basic components of the model are conditional productivities. We understand a conditional productivity as the productivity of an individual, say j, if she is supervised by another individual, say i. Let $v_{ij} \in \mathbb{R}^+$ where $i \neq j$ collect all productivities based on a collaboration of two individuals. Non-collaborative productivities, referred to as self-management productivities, are accounted for by $v_{ii} \in \mathbb{R}$ for all i.

In the model, the main difference between entrepreneurs and employees is that the latter have a manager, while the former operate on their own. Self-management productivities, v_{ii}, represent entrepreneurial abilities since they constitute the share of a firm's productivity level that can be exclusively traced back to the entrepreneur and cannot be replaced by the work of an employee.

We assume that two-person collaborations are the maximal degree of immediate mutual work interactions.[8] Consequently, the whole productivity system is described by v_{ij} and v_{ii} for all i. We can write the system of inter-dependent and self-dependent productivities as a matrix:

$$\mathfrak{P} = \begin{bmatrix} v_{11} & v_{12} & v_{13} & \cdots & v_{1I} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2I} \\ v_{31} & v_{32} & v_{33} & \cdots & v_{3I} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{I1} & v_{I2} & v_{I3} & \cdots & v_{II} \end{bmatrix} \tag{2.1}$$

$\mathfrak{P}$ is the true productivity matrix.

Following empirical findings on the imperfect character of information on entrepreneurial abilities, $\mathfrak{P}$ cannot be fully known (Section 2.2.2). In a setting with perfect information on $\mathfrak{P}$, there, for instance, would be no entrepreneurial overconfidence. Consequently, while assuming that productivities based on a collaboration between two different individuals are known because, for instance, they can be inferred from education levels[9], it is assumed that individuals lack knowledge on entrepreneurial abilities, represented by self-management productivities. Put differently, diag($\mathfrak{P}$) is unknown.

Actual self-management productivities depend on the Plasticity trait, β, and a number of unknown factors, denoted by $\theta \in \mathbb{R}$. The factor θ introduces uncertainty. Thus,

[8] Individual i can collaborate with j and k, but then j and k do not directly collaborate.

[9] One can argue that this particular assumption is too simplifying. However, as non-self-management productivities are used to determine wages and wages are usually known beforehand, it can be argued that the assumption is approximately true.

entrepreneurial abilities are given by $\upsilon_{ii} = h_\upsilon(\beta_i, \theta_i)$ where $h_\upsilon : \mathbb{T} \times \mathbb{R} \to \mathbb{R}$.[10] When θ is unknown, the only information on entrepreneurial abilities is given by Plasticity. This holds as long as individuals do not receive any feedback from the market because they were never entrepreneurs. We, thus, assume that individuals construct an estimate of entrepreneurial abilities on the basis of β. The estimate of entrepreneurial abilities is given by $v_{ii} = h_\beta(\beta_i)$ for all i.

On the basis of estimates, individuals construct

$$\mathfrak{A} = \begin{bmatrix} v_{11} & \upsilon_{12} & \upsilon_{13} & \cdots & \upsilon_{1I} \\ \upsilon_{21} & v_{22} & \upsilon_{23} & \cdots & \upsilon_{2I} \\ \upsilon_{31} & \upsilon_{32} & v_{33} & \cdots & \upsilon_{3I} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \upsilon_{I1} & \upsilon_{I2} & \upsilon_{I3} & \cdots & v_{II} \end{bmatrix} \tag{2.2}$$

$\mathfrak{A}$ is the estimated productivity matrix and common knowledge. A comparison between the main diagonals of $\mathfrak{P}$ and $\mathfrak{A}$ yields the estimation error due to lack of information. For a specific individual i, the estimation error is $\varepsilon_i = v_{ii} - \upsilon_{ii}$. Actual and estimated entrepreneurial abilities are both influenced by the Plasticity trait such that there will be dependencies between them. One way to capture these dependencies is to examine the correlation between estimates and actual values. Consider the following example.

Example 2.1. The correlation between actual and estimated entrepreneurial abilities determines the importance of the unknown factor θ relative to the Plasticity trait, β. A weak correlation implies that actual entrepreneurial performance is mostly determined by unknown factors, whereas a strong correlation implies that it is mostly driven by the Plasticity trait. Consider a simple linear relation where the actual value is generated by $\upsilon = h_\upsilon(\beta, \theta) = a\beta + b\theta$ and the estimate by $v = h_\beta(\beta) = a\beta$, where $a > 0$ and $b \geqslant 0$. β and θ are independent and $\mathbb{E}[\theta] = 0$ such that $\mathbb{E}[v] = \mathbb{E}[\upsilon] = a\mathbb{E}[\beta]$, i.e., estimates are correct on average. Then, $\lim_{b\to\infty} \mathrm{Cor}(v, \upsilon) = a\mathbb{E}[\beta]\mathbb{E}[\theta](a\mathbb{V}[\beta]^{\frac{1}{2}}\mathbb{V}[\theta]^{\frac{1}{2}})^{-1} = 0$ and there will be almost no correlation between the true and the estimated value if b, which determines the importance of the unknown factor, is sufficiently large. If $b = 0$, such that true values are exclusively determined by β, we get $\mathrm{Cor}(v, \upsilon) = 1$.

Rewrite $\mathfrak{A}$ as

$$\mathfrak{A} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_{I-1} & \mathbf{p}_I \end{bmatrix}$$

where $\mathbf{p}_i$ is a productivity column vector of individual i under different managers. Assume that each vector $\mathbf{p}_i$ allows us to create a distinct ranking $\upsilon_{\{1\}i} > \upsilon_{\{2\}i} > \cdots > \upsilon_{\{I\}i}$ where

[10] In reality, υ_{ii} depends on a multitude of known factors—all unknown factors are captured by θ—and not just two. However, to keep notation as simple as possible, we abstract from these, as known factors do not explain the difference between estimates and actual abilities, being generated by a *lack* of information.

$\{1\}i$ is the best manager for i, $\{2\}i$ is the second-best manager for i, and so on. The ranking refers to $\mathfrak{A}$. It is necessary to impose the following restriction on rankings (Roessler & Koellinger 2012): Given $\mathfrak{A}$, suppose that for a subset of N we can assign ranks from 1 to n such that $v_{12} > v_{22}, v_{23} > v_{33}, v_{34} > v_{44}, \ldots, v_{[n-1]n} > v_{nn}$. Given the previous ranking, $v_{n1} < v_{11}$ must hold. The assumption makes management transitive.[11]

We construct equilibrium firm organization in an intuitive way.[12] Assume that managerial assignments are hierarchical such that each individual will be managed by one another individual unless she is an entrepreneur; an entrepreneur is always managed by herself. Firms consist of units of optimally assigned individuals. A unit

$$\mathcal{U}_i^{(0)} = \{j \in N : v_{ij} > v_{kj} \text{ for all } j \text{ and } j, k \neq i\}$$

collects all individuals whose best manager is individual i, with the exception of i if i is an entrepreneur. Note that if individual i is not the best manager of anyone, we have $\mathcal{U}_i^{(0)} = \emptyset$. Units are unique, $\mathcal{U}_i^{(0)} \cap \mathcal{U}_j^{(0)} = \emptyset$ for $i \neq j$, since there is only one best manager for each individual. A potential entrepreneur is not part of any firm unit and believes to be her own best manager. Following Roessler & Koellinger (2012), we assume that at least one entrepreneur must exist, such that $E_1 \neq \emptyset$.[13]

Entrepreneurs are indifferent between receiving their income in wages and profit.[14] Entrepreneurial income is equated with the sum of profits and wages entrepreneurs pay to themselves. Let $\mathcal{F}_i^{(0)}$ denote a set containing all individuals in the firm of individual i, where firms are named after entrepreneurs. Ignoring dynamics, the estimated entrepreneurial income of individual i is given by

$$\hat{\pi}_{i,0} = v_{\{1\}i} + \sum_{j \in \mathcal{F}_i^{(0)} \setminus i} v_{\{1\}j} - \sum_{j \in \mathcal{F}_i^{(0)} \setminus i} w_j \tag{2.3}$$

where $v_{\{1\}i} = v_{ii}$ and w_j is a worker's wage. Equation (2.3) is formalized as an estimated income, as we only have estimates for v_{ii}. If individual i acts according to her estimated

[11] For instance, if 1 is the best manager of 2, 2 is the best manager of 3, and 3 is the best manager of 1, we will not be able to construct a proper firm. To do this, we, for example, can rule out the last relation "3 is the best manager of 1," which is done by the assumption for arbitrary long chains.

[12] A formal proof is provided by Roessler & Koellinger (2012). The game is a normal-form game, actions are simultaneous, with public information on (estimated) productivities and ultimatum wage offers where the solution is a Nash-equilibrium satisfying hierarchical assignments.

[13] An equivalent and more intuitive assumption is that there exists at least one individual k with an estimated self-productivity level above the critical value ξ_k, where ξ_k is defined in Proposition 2.1 (Section 2.5.4).

[14] Here, we follow Taylor (1996) who argues that higher expected earnings in self-employment than in paid employment are a key factor determining the utility of self-employment. In other words, we abstract from non-monetary incomes, also associated with entrepreneurial choices (Blanchflower 2000; Hundley 2001; Benz & Frey 2008a,b).

income, the deterministic part of her actual income can be obtained by replacing estimated productivity v_{ii} with actual productivity υ_{ii} in (2.3).

The output that can be unequivocally attributed to entrepreneur i is i's self-management productivity and i's managerial advantage over the second-best managers of individuals in her unit. Roessler & Koellinger (2012) demonstrate that this is the unique Nash-equilibrium outcome given hierarchical assignments. It follows that (2.3) can be written as

$$\hat{\pi}_{i,0} = v_{ii} + \sum_{j \in \mathcal{U}_i^{(0)}} \left(\upsilon_{\{1\}j} - \upsilon_{\{2\}j}\right) \tag{2.4}$$

Again, v_{ii} can be replaced by υ_{ii} to get true earnings.

There is no unique outcome for wages since entrepreneurs are indifferent between different schemes that do not change profits (Roessler & Koellinger 2012). The indirect contribution of a worker i to a firm j consists of her managerial advantage over the second-best manager of individuals in her unit. However, units can be moved and it is not enough to reward i for her managerial advantage only. If individual i is in an optimally constructed firm, her direct contribution to output is $\upsilon_{\{1\}i}$. For a firm without i's first-best manager, the direct contribution is $\upsilon_{\{n \neq 1\}i} < \upsilon_{\{1\}i}$, while the managerial advantage remains the same if we move i with her unit $\mathcal{U}_i^{(0)}$. Consider the following wage offer:

$$w_{i,0} = \upsilon_{\{2\}i} + \sum_{j \in \mathcal{U}_i^{(0)}} \left(\upsilon_{\{1\}j} - \upsilon_{\{2\}j}\right) \tag{2.5}$$

The intuition behind (2.5) is that i's net-contribution, the sum of direct and indirect contributions minus i's wage given by (2.5), to a firm with her first-best manager is $\upsilon_{\{1\}i} - \upsilon_{\{2\}i} > 0$. Her net-contribution to a firm without her first-best manager is $\upsilon_{\{n \neq 1\}i} - \upsilon_{\{2\}i} \leqslant 0$. Only the firm with i's second-best manager will be willing to match the offer in (2.5) since $\upsilon_{\{n\}i} - \upsilon_{\{2\}i} < 0$ for $n \neq 1, 2$ and profit from employing i is negative. All firms j where $\{1\}i, \{2\}i \notin \mathcal{F}_j^{(0)}$ will offer less than $w_{i,0}$. Hence, i gets the same highest offer twice—one from the firm with her first-best and one from the firm with her second-best manager—and chooses the firm with her first-best manager by assumption. Roessler & Koellinger (2012) demonstrate that (2.5) constitutes one, but not a unique, Nash-equilibrium, which will be used in further calculations.

It can be shown that entrepreneurs are indifferent toward the employment status of lower hierarchy employees:

Lemma 2.1. *An entrepreneur is indifferent between different firm structures as long as her own unit is left untouched, as only the unit of entrepreneurs has an effect on entrepreneurial incomes.*

Proof. See Appendix 2.A. ∎

In the dynamic setting with unrestricted interactions, we allow for firm restructurings resulting from the fact that some workers can become entrepreneurs or some entrepreneurs become workers. Restructurings can affect units low in firm hierarchy. According to Lemma 2.1, entrepreneurs do not necessarily have an incentive to fully compensate the manager of a unit low in firm hierarchy for a change in unit structure. This results in a wage regime where wages of managers low in firm hierarchy are not increased although there is an increase in management tasks since more individuals have to be managed.

2.5 Entry and exit decision dynamics

This section introduces dynamics by describing the decision process of workers and entrepreneurs in time. The decision process, we implement in the model, accounts for the fact that personality traits affect entrepreneurial decision making. We begin the section by introducing assumptions on dynamics. Second, we model firm restructuring capturing entrepreneurial entry and exit. After explaining firm restructuring, we discuss decision dynamics.

2.5.1 Wage structure and dynamic stability

Before analyzing entry and exit decisions, we make three assumptions. Wage renegotiations are costly and entrepreneurs may choose to keep allocation rules constant if this does not hurt the entrepreneur's income. Hence, in the first assumption, we fix the wage structure. More specifically, we assume that wages can be renegotiated in every period if the entrepreneur chooses to do so. However, if wages in $t = 0$ are structured according to a certain rule determining how to allocate firm income between workers and the entrepreneur, this structure must not be altered in $t > 0$. The assumption allows us to use wage equations already defined without the need to determine wage equilibria in every period.

Next, we add a dynamic extension to the static non-circularity assumption making management transitive. An alternative to entrepreneurship always exists such that *if* an entrepreneur j decides to abandon entrepreneurship, her second-best manager is not part of j's company, i.e., $\{2\}j \notin \mathcal{F}_j^{(t)}$ for all t. Note that the assumption is relatively weak because it only requires that certain conditions are met when an entrepreneur wants to exit but these conditions do not need to be fulfilled if an entrepreneur does not wish to abandon her business. Also, it can be easily relaxed by assuming that if the first-best manager is in the own company, one can go to the second-best manager and so on.

Last, we extend the static assumption of Roessler & Koellinger (2012) on the existence of entrepreneurs and assume that at least one entrepreneur always exists or $E_t \neq \emptyset$ for all $t > 1$.[15] This assumption, together with the previous assumption, also ensures, that an entrepreneur abandoning entrepreneurship always finds employment opportunities, as there is no social security system providing unemployment benefits, etc. in the model.

2.5.2 *Firm restructuring*

A restructuring takes place if some individual leaves a firm to start her own company or some individual exits entrepreneurship and integrates into the firm of another individual. Call an individual who leaves a 'nascent entrepreneur' and an individual who integrates into another individual's firm an 'applicant.'[16]

Leaving from and applying to a firm are possible in periods $t > 1$. Before departure, a nascent entrepreneur j is part of a firm and correspondingly part of the unit of some $i \neq j$. Thus, we can define a set of nascent entrepreneurs depending on the unit they are leaving from and period of departure as $\mathcal{D}_i(t)$. If j leaves in period t from the unit of i, we denote this by $j \in \mathcal{D}_i(t)$. An equivalent holds for applicants. An applicant applies to a specific unit in the firm. We assume that she is immediately integrated into the unit if she chooses to apply. The set of applicants for the unit of i in period t is given by $\mathcal{A}_i(t)$.

The structure in period $t = 1$ is set in period $t = 0$ according to variables known in period zero, such as matrix $\mathfrak{A}$. Thus, we have a starting value $\mathcal{U}_i^{(0)}$ for all i. For the next periods, we have $\mathcal{U}_i^{(1)} = \mathcal{U}_i^{(0)}$ for all i because decisions in t determine actions in $t + 1$. Furthermore, we have

$$\mathcal{U}_i^{(t)} = \{\mathcal{U}_i^{(t-1)} \cup \mathcal{A}_i(t)\} \setminus \mathcal{D}_i(t) \tag{2.6}$$

for all i and all $t > 1$. Firms in period t, $\mathcal{F}_i^{(t)}$, can be constructed using restructured units.

An example is provided in Figure 2.1 where we see two consecutive periods T and $T + 1$. In period T, only the firm of individual 1 exists. In period $T + 1$, individual 6 becomes entrepreneur. Because 6 is the first-best manager of 7, she will take 7 with her. The unit of 3 is affected by individual 6 who is excluded from $\mathcal{U}_3^{(T)}$. The case of applicants works *vice versa* (by interpreting Figure 2.1 from right to left).

[15] A more intuitive and equivalent assumption is that actual entrepreneurial abilities of at least one individual k are such that $\psi_k(v_{ii}) \approx 1$, where ψ_k is given in Proposition 2.3 (Section 2.6.1).

[16] Note that for an applicant i with $i \in E_1$ failure in entrepreneurship implies a shift in productivity ranks. If individual i entered entrepreneurship in $t = 1$, she considered herself as her first-best manager. If individual i abandons entrepreneurship in $t > 1$, she cannot consider herself as her first-best manager anymore.

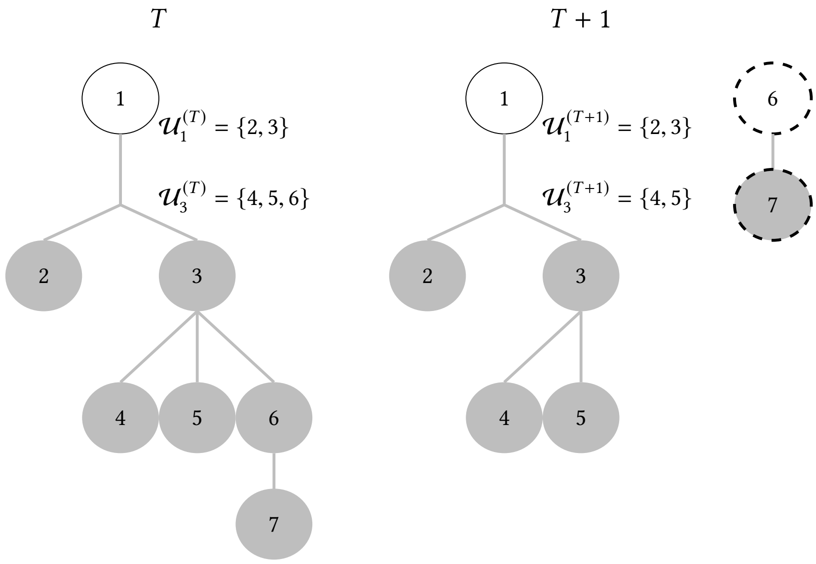


FIGURE 2.1. Restructuring example

2.5.3 *Wage regime*

When describing the model's core, we established that entrepreneurs are indifferent between different firm structures as long as their own units are left untouched (see Lemma 2.1). This result is now used to define a wage regime where some employees are denied opportunities and might decide to become entrepreneurs as a consequence.

Note that wages can be written as $w_{i,t} = w_{i,0} + \tilde{w}_{i,t}$ where $\tilde{w}_{i,t} = \sum_{k=1}^{t} \Delta w_{i,k}$ and $\Delta w_{i,k} = w_{i,k} - w_{i,k-1}$. $\Delta w_{i,k}$ is the wage difference between two consecutive periods, determined by the behavior of i's unit. Estimated entrepreneurial incomes can be written in a similar way such that $\hat{\pi}_{i,t} = \hat{\pi}_{i,0} + \tilde{\pi}_{i,t}$ where $\tilde{\pi}_{i,t} = \sum_{k=1}^{t} \Delta\pi_{i,k}$ and $\Delta\pi_{i,k} = \pi_{i,k} - \pi_{i,k-1}$. $\Delta\pi_{i,k}$ is the difference in entrepreneurial income between two consecutive periods.

Assume that a former entrepreneur is integrated into the firm of individual i's entrepreneur such that individual i is the new manager of the former entrepreneur. Consequently, individual i has to perform additional work. If individual i's entrepreneur compensates i for additional work, i is not denied any opportunities. Moreover, individual i would, obviously, always compensate herself for additional work if she was an entrepreneur. Hence: If wages are always renegotiated, we have $\Delta w_{i,k} = \Delta\pi_{i,k}$ and i's entrepreneur is willing to pay i the income change that i would pay to herself if she would be an entrepreneur.

However, individual i's entrepreneur has an incentive to not to fully renegotiate wages with every employee for the following reason. An entrepreneur always wants to keep the employees in her own unit since entrepreneurial income would be reduced if someone leaves this unit. Thus, entrepreneurs always fully adjust the wages of their units such that workers in these units will have no reason to start an own business. Application always increases output. In particular, application to a non-entrepreneur unit increases firm output. Yet, entrepreneurial income does not increase if the entrepreneur fully rewards additional managerial effort by the individual who manages the applicant. This hints at a strong incentive to not raise wages contingent upon application to a non-entrepreneurial unit.

Thus, we assume that entrepreneurs fully renegotiate wages with their own units such that wages are cut if there is less managerial work and increased if there is more. Entrepreneurs cut wages in all other units in their companies if necessary. However, entrepreneurs do not raise wages in units not managed by themselves.

As a consequence, we might have $\Delta w_{i,k} < \Delta \pi_{i,k}$ for workers low in firm hierarchy. In the empirical literature on entrepreneurship (see, for instance, Maritz 2004; Harding, Brooksbank, Hart, Jones-Evans, Levie, O'Reilly & et al. 2006; Caliendo & Kritikos 2009) there is usually a differentiation between entrepreneurs who enter because they see an opportunity or who enter because of necessity. If $\Delta w_{i,k} < \Delta \pi_{i,k}$, there might be an opportunity opening up to a worker that she can only exploit as an entrepreneur. Necessity entrepreneurs are a subgroup of the model's set of entrepreneurs who enter because they perceive their self-management productivity as the productivity generating the best outcome.

2.5.4 *Initial period decision*

In the initial period, individuals suffer from a lack of information. The only two sources of information are a wage offer and an estimated entrepreneurial income based on an estimate of entrepreneurial abilities.

Individuals discount future incomes by factor $\delta \in (0,1)$. Entrepreneurship generates a per-period cost of $c \in (0,1)$, such that only a share, $1-c$, of entrepreneurial income is available in every period. Wage work does not generate any costs. To decide which occupation to select, individuals are assumed to project income for each occupation. Projected incomes determine the attraction of an option in period zero. The projected income stream of wage work is

$$\mathscr{W}_{i,0} = \sum_{t=1}^{T} \delta^t w_{i,0} = \bar{\delta} w_{i,0} \tag{2.7}$$

where $\bar{\delta} \equiv \delta(1-\delta^T)(1-\delta)^{-1}$, $\bar{\delta} \approx \delta(1-\delta)^{-1}$ for a large T. Similarly, the projected income stream of the entrepreneur status is

$$\mathscr{E}_{i,0} = \sum_{t=1}^{T} \delta^t (1-c)\hat{\pi}_{i,0} = \bar{\delta}(1-c)\hat{\pi}_{i,0} \tag{2.8}$$

Proposition 2.1. *An individual i will select entrepreneurship in period 1 such that* $\mathbb{d}_{i,1} = 1$ *if* $\nu_{ii} \geqslant \xi_i$ *where*

$$\xi_i = \frac{1}{1-c} v_{\{2\}i} + \frac{c}{1-c} \sum_{j \in \mathcal{U}_i^{(0)}} \left(v_{\{1\}j} - v_{\{2\}j} \right) \tag{2.9}$$

Proof. An individual i will become entrepreneur in period 1 if $\mathscr{E}_{i,0} \geqslant \mathscr{W}_{i,0}$ and the proposition immediately follows from the condition. ∎

As estimated self-management productivities depend on Plasticity, Proposition 2.1 implies the following:

Corollary 2.1. *Only the Plasticity trait,* β, *is responsible for entrepreneurial entry in the first period.*

Proof. Follows from Proposition 2.1. ∎

Critical self-management productivities, ξ_i, in (2.9) are positive in costs, c, the productivity under the second-best manager, $v_{\{2\}i}$, and cumulated managerial advantage of i over the second-best managers of her unit, $\sum_{j \in \mathcal{U}_i^{(0)}} (v_{\{1\}j} - v_{\{2\}j})$. Thus, a necessary condition for entrepreneurship is that the estimated self-management productivity has to be the largest conditional productivity. If the estimated self-management productivity is large enough to offset costs, the sufficient condition for starting up a firm is fulfilled.

By replacing the estimate, ν_{ii}, by the actual value, v_{ii}, in Proposition 2.1, we obtain the full-information equilibrium defined as follows.

Lemma 2.2. *If individual i had perfect information in period 0 and feedback would be fully informative, she would select the following full-information equilibrium:*

$$\mathbb{d}_i^* = \begin{cases} 1 & \text{if } v_{ii} \geqslant \xi_i \\ 0 & \text{else} \end{cases} \qquad \text{for all } t$$

where $v_{ii} \geqslant \xi_i$ *is equivalent to* $(1-c)\pi_{i,0} \geqslant w_{i,0}$.

Proof. Use the condition that entrepreneurship is selected if $\mathscr{E}_{i,0} \geqslant \mathscr{W}_{i,0}$ where ν_{ii} is replaced by v_{ii} to obtain $\mathbb{d}_{i,1}^*$. Then, note that a reevaluation of a choice in period t in $t+1$ would always come to the same conclusion, as no *new* information can be obtained and entrepreneurial units are left intact by the wage regime if everything is known and deterministic. Thus, we must have $\mathbb{d}_{i,t}^* = \mathbb{d}_{i,1}^* = \mathbb{d}_i^*$. ∎

Consequently, in an environment with perfect information entrepreneurship is selected ($\mathbb{d}_i^* = 1$) if it generates a (weakly) higher net income than wage work. The full-information equilibrium can be used to assess learning efficiency: the ability of individuals to identify the best option.

2.5.5 *Decisions influenced by experience*

So far, decisions were based on common knowledge and estimated entrepreneurial abilities, while no learning took place. When introducing learning, we argue that information is provided by the market in form of stochastic feedback. Decision dynamics are driven by the EWA algorithm (Camerer & Ho 1999) based on option attractions, which are self-reinforcing such that the previous attraction of an option influences today's attraction.

2.5.5.1 *Market feedback*

If an individual selects entrepreneurship, she can observe the feedback of the market providing information on her entrepreneurial abilities. We assume that the market is correct in expectations but reports with noise. Denote market feedback in period t of individual i by $\phi_{i,t}$. Under our assumption, $\mathbb{E}[\phi_{i,t}] = \pi_{i,t}$. One way to satisfy this assumption is $\phi_{i,t} = \pi_{i,t} + u_{i,t}$ where for the noise variable we have $\mathbb{E}[u_{i,t}] = 0$ and $\mathbb{V}[u_{i,t}] > 0$.[17]

2.5.5.2 *Attraction formation*

The structure of attraction formation follows the one suggested by Camerer & Ho (1999). Each individual starts with the same initial horizon of experience $\mathcal{H}_0 = 1$. The horizon of experience is updated according to

$$\mathcal{H}_t = \mathcal{H}_{t-1} + 1 \tag{2.10}$$

such that experience increases by one "experience unit" in every period.

The attraction of being a worker is given by

$$\mathcal{W}_{i,t} = \frac{\mathcal{H}_{t-1}}{\mathcal{H}_t}\alpha_i \mathcal{W}_{i,t-1} + \frac{1}{\mathcal{H}_t} w_{i,t} \tag{2.11}$$

The parameter $\alpha_i \in [0, 1)$ describes how much information individual i extracts from her own working history. To formalize the attraction of entrepreneurship, which follows the same structure as attraction of being a worker, we consider two cases. First, if individual i is an entrepreneur in period t, we have

$$\mathcal{E}_{i,t|i\in E_t} = \frac{\mathcal{H}_{t-1}}{\mathcal{H}_t}\alpha_i \mathcal{E}_{i,t-1} + \frac{1-c}{\mathcal{H}_t}\phi_{i,t} \tag{2.12}$$

[17] By including the stochastic noise variable u, we simultaneously introduce failure by chance. Even an entrepreneur with a high true performance could decide to abandon entrepreneurship because feedback is below a critical threshold by chance.

Second, if individual i is a worker in period t, we have

$$\mathscr{E}_{i,t|i\in W_t} = \frac{\mathscr{H}_{t-1}}{\mathscr{H}_t}\alpha_i\mathscr{E}_{i,t-1} + \frac{1-c}{\mathscr{H}_t}\left(\phi_{i,t_-} + \tilde{\pi}_{i,t} - \tilde{\pi}_{i,t_-}\right)\mathbb{1}_t^E + \frac{1-c}{\mathscr{H}_t}\hat{\pi}_{i,t}\left(1-\mathbb{1}_t^E\right) \tag{2.13}$$

$$\mathbb{1}_t^E = \begin{cases} 1 & \text{if } \{\breve{t}\in(0,t) : i\in E_{\breve{t}}\}\neq\emptyset \\ 0 & \text{else}\end{cases}, \qquad t_- = \max\{\breve{t}\in(0,t) : i\in E_{\breve{t}}\}$$

where $\max\emptyset = 0$. If individual i was an entrepreneur before, we get $\mathbb{1}_t^E = 1$. Then, t_- is the last period of i's entrepreneurship such that i will take the feedback of this period ϕ_{i,t_-} and adjust it to the development of her unit between period t_- and the actual period t. If individual i was never an entrepreneur before, we get $\mathbb{1}_t^E = 0$, and the only information about entrepreneurship she has is her estimate of entrepreneurial income.

The parameter α_i controls how much information is extracted from previous estimates and observations. We interpret the parameter as an indicator of motivational stability or grit and link it to the Stability trait or, formally, $\alpha_i = h_\mathrm{\alpha}(\mathrm{\alpha}_i)$. We discuss this interpretation in the following example.

Example 2.2. By iterating attraction, it is straightforward to show that

$$\mathscr{W}_{i,t} - \mathscr{E}_{i,t} = \frac{\alpha_i^t}{t+1}\bar{\delta}[w_{i,0} - (1-c)\hat{\pi}_{i,0}] + G_t$$

where G_t is a function of potentially all model variables besides estimated entrepreneurial income if the individual received market feedback in period $1,\ldots,t$. Consider an individual whose estimate of entrepreneurial abilities is such that $(1-c)\hat{\pi}_{i,0} > w_{i,0}$. The individual will decide for entrepreneurship in period 1. Hence, the motivation of the individual is to be an entrepreneur. The impact of this motivation is influenced by the Stability trait parameter α_i. If, for instance, $\alpha_i = 0$, the initial motivation to be an entrepreneur is already forgotten in the first period after entry. However, if $\alpha_i < 1$ is large, the initial motivation to be an entrepreneur plays a permanent role, even in the last periods of the individual's career. Consequently, an increase in α_i stabilizes motivation. An alternative interpretation is that an increase in α increases perseverance, which is important for entrepreneurial survival and success (Gimeno et al. 1997; Markman & Baron 2003; Markman et al. 2005; Lowe & Ziedonis 2006).

The decision for entrepreneurship ($\mathbb{d}_{i,t} = 1$) or wage work ($\mathbb{d}_{i,t} = 0$) is given by

$$\mathbb{d}_{i,t} = \begin{cases} 1 & \text{if } \mathscr{E}_{i,t-1} \geqslant \mathscr{W}_{i,t-1} \\ 0 & \text{if } \mathscr{W}_{i,t-1} > \mathscr{E}_{i,t-1}\end{cases} \tag{2.14}$$

Decisions can deviate from the full-information equilibrium. Furthermore, given different realizations of feedback, there exist different decision paths.

In this section, we introduce all dynamic components of our model. Dynamics start with an initial period where individuals without any work experience decide if they want to be workers or entrepreneurs. After the initial period, those who chose to become entrepreneurs are confronted with feedback from the market. Feedback is then used to construct dynamic attractions, where the Stability trait determines the weight assigned to previous experience. Entrepreneurship is selected if it is at least as attractive as wage work.

2.6 Analysis of decision processes

This section presents our main results. We examine our model's selection mechanism from two different perspectives. The first perspective results from the imposition of the following restriction on labor market interactions:

Assumption 2.1.1. In the restricted model, interactions from an individual *i*'s point of view are such that $\Delta w_{i,t} = \Delta\pi_{i,t} = 0$ for all t.

Note that Assumption 2.1.1 does not rule out unit restructurings in general. It only requires that the unit managed by the individual we examine does not change. The second perspective does not impose any restrictions on labor market interactions:

Assumption 2.1.2. In the unrestricted model, interactions from an individual *i*'s point of view are such that $\Delta w_{i,t} \lesseqgtr 0$, $\Delta\pi_{i,t} \lesseqgtr 0$, and $\Delta w_{i,t} \leqslant \Delta\pi_{i,t}$.

The main difference between the two model versions is as follows:

Lemma 2.3. *In the restricted model version (under Assumption 2.1.1), individuals who start as workers never become entrepreneurs after their initial choice,* $\mathbb{P}(\mathbb{d}_{i,t} = 1|\mathbb{d}_{i,1} = 0) = 0$ *for* $t > 1$. *Furthermore, once entrepreneurship is abandoned, individuals do not return to entrepreneurship. In the unrestricted version (under Assumption 2.1.2), all transitions are possible.*

Proof. See Appendix 2.B. ∎

The model with restricted interactions is analytically tractable, while the unrestricted version is not. Hence, first, analytical results for the restricted version are provided. We prove that the model replicates the five benchmark observations mentioned in Section 2.2.7.

We, then, simulate the unrestricted version. Simulations allow us to analyze the decisions of heterogeneous individuals with mixed career histories (individuals switching from entrepreneurship to wage work and the other way around). Thus, we can determine if labor market interactions substantially influence results. We also exploit the flexibility of the simulation approach to introduce liability constraints, rationalizing entrepreneurial costs, and to replace some assumptions by more realistic ones. We demonstrate that the basic, analytically derived, selection mechanism of our model is fairly general and robust to the inclusion of additional, plausible, influencing factors; and robust to relaxing simplifying assumptions.

2.6.1 *Analysis of model with restricted labor market interactions: Reproducing benchmarks*

Entrepreneurs tend to overestimate their own entrepreneurial abilities (Busenitz & Barney 1997; Camerer & Lovallo 1999; Koellinger et al. 2007). In our model, overconfidence is generated under fairly general conditions. Assume that estimated and true self-management productivities are generated by a bivariate normal distribution with correlation parameter $\gamma \in [0, 1)$. This is a simple way to formalize the relation between estimates and actual values. The correlation determines the accuracy of self-assessments: $\gamma = 0$ is zero accuracy and $\gamma \approx 1$ is very precise self-assessment (see the example in Section 2.4). Let υ denote actual and ν estimated self-management productivities of the population, where both are stochastic variables. We assume that estimates are correct on average such that $\mathbb{E}[\upsilon] = \mathbb{E}[\nu]$. Let υ_E and ν_E denote actual and estimated self-management productivities of entrants into entrepreneurship in the initial period. Then, the following result holds:

Proposition 2.2. *Assume that non-self-management productivities are such that for the critical values, defined in Proposition 2.1, we have $\xi_1 = \xi_2 = \cdots = \xi_I = \xi$. Entrepreneurs will, on average, exhibit overconfidence in entrepreneurial abilities, resulting in $\mathbb{E}[\nu_E] > \mathbb{E}[\upsilon_E]$, if the variance of estimated self-management productivities is sufficiently large such that $\mathbb{V}[\nu]^{\frac{1}{2}} > \gamma$. Given that $\mathbb{V}[\nu]^{\frac{1}{2}} > \gamma$ holds, overconfidence increases in costs, the average productivity under the second-best manager given fixed managerial advantages, and the average managerial advantage given fixed productivities under the second-best manager. Furthermore, given $\mathbb{V}[\nu]^{\frac{1}{2}} > \gamma$, overconfidence decreases in the correlation γ.*

Proof. See Appendix 2.C.1. ■

Given that, according to Proposition 2.1, estimated self-management productivities of entrants into entrepreneurship must be large enough to offset costs, the productivity under the second-best manager, and managerial advantages, entrants will come from the right tail of the distribution of estimated self-management productivities, which is mostly responsible for the result in Proposition 2.2. Thus, an overestimation of entrepreneurial abilities, found in field and experimental data, is a straightforward implication of the selection mechanism of our model.

An interesting feature of Proposition 2.2 is that it does not rely on different treatments of different types of information. A common approach to explain overconfidence is that individuals tend to take too much credit for successes, while under-weighting failures (see, for instance, Gervais & Odean 2001). Also, the explanation of overconfidence does not need the assumption of post-decisional bolstering (Cooper & Woo 1988), the exaggeration of the attractiveness of an option after it has been selected. In our model, overconfidence is simply generated by the fact that there is an alternative to entrepreneurship generating a certain income. Given that only individuals who assume that entrepreneurship is better than the alternative enter entrepreneurship, which is a very simple selection mechanism, the distribution of estimated abilities of entrants can have a mean that is significantly above the mean of actual abilities.

A further empirical regularity found in data is that survival rates decrease at decreasing rates such that entrepreneurial survival is mostly threatened immediately after entry (Evans & Leighton 1989; Knaup 2005; Knaup & Piazza 2007; Georgarakos & Tatsiramos 2009; Helmers & Rogers 2010; Quatraro & Vivarelli 2015). Our model produces survival curves with the shape found in empirical data. To derive the corresponding result, we make the following assumption with respect to the noise variable:

Assumption 2.2.1. Let $U_{i,t} = \sum_{k=1}^{t} \alpha_i^{t-k} u_{i,k}$. Feedback takes a path between a “bad” path

$$U_{i,t}^- = \alpha_i^{t-1} u_i^- + \alpha_i^{t-2} u_i^- + \cdots + \alpha_i^1 u_i^- + \alpha_i^0 u_i^- = \sum_{k=1}^{t} \alpha_i^{t-k} u_i^- = \frac{1-\alpha_i^t}{1-\alpha_i} u_i^-$$

and a "good" path

$$U_{i,t}^+ = \alpha_i^{t-1} u_i^+ + \alpha_i^{t-2} u_i^+ + \cdots + \alpha_i^1 u_i^+ + \alpha_i^0 u_i^+ = \sum_{k=1}^{t} \alpha_i^{t-k} u_i^+ = \frac{1-\alpha_i^t}{1-\alpha_i} u_i^+$$

where $u_i^+ > 0$ and $u_i^+ + u_i^- = 0$. All paths are taken with an equal probability.

Note that $\mathbb{E}[U_{i,t}] = 0$, i.e., noise does not introduce any bias. Using the dot notation for derivatives with respect to time, we establish the following result:

Proposition 2.3. *Assume that individual i strictly preferred entrepreneurship in period 1 and that feedback noise is generated in line with Assumption 2.2.1. Furthermore, assume that individual i is not completely unfit for entrepreneurship such that* $\pi_{i,0} > 0$. *Then, for the probabilities that individual i is an entrepreneur in period* $t+1 > 1$, *we have*

$$\dot{\mathbb{P}}(\mathbb{d}_{i,t+1} = 1 | \mathbb{d}_{i,1} = 1) < 0, \quad \ddot{\mathbb{P}}(\mathbb{d}_{i,t+1} = 1 | \mathbb{d}_{i,1} = 1) > 0$$

$$\mathbb{P}(\mathbb{d}_{i,t+1} = 1 | \mathbb{d}_{i,1} = 1) \underset{t\to\infty}{\longrightarrow} \psi_i = \frac{w_{i,0} - (1-c)u_i^+ - (1-c)\pi_{i,0}}{2(c-1)u_i^+}$$

Proof. See Appendix 2.C.2. ■

Thus, the probability of an entrepreneur to survive mostly decreases right after entry, in the 'turbulent' phase, because the entrepreneur receives feedback for the first time in this period, while the initial attraction of entrepreneurship is partially forgotten (see the example in Section 2.5.5.2). After the first periods, the probability to survive stabilizes, at probability ψ_i, in the 'consolidation' phase.

Our model produces personality trait effects following in the empirically established direction (Markman & Baron 2003; Markman et al. 2005; Lowe & Ziedonis 2006) in a non-linear way (Ciavarella et al. 2004; Caliendo et al. 2014). Let

$$\eta_x = \frac{\partial \mathbb{P}(\mathbb{d}_{i,t+1} = 1 | \mathbb{d}_{i,1} = 1)}{\partial x}$$

denote the effect of an arbitrary variable on the probability to be an entrepreneur in period $t+1$ given that the individual strictly prefers to be entrepreneur in period 1. We obtain the following result:

Proposition 2.4. *The probability to remain entrepreneur is positive in trait-related variables such that $\eta_{\hat{\pi}_{i,0}}, \eta_{\alpha_i} > 0$. However, the effect of Stability (α) is amplified by estimation errors, while the effect of estimated entrepreneurial income (β) is not:*

$$\frac{\partial \eta_{\hat{\pi}_{i,0}}}{\partial \varepsilon_i} = 0, \quad \frac{\partial \eta_{\alpha_i}}{\partial \varepsilon_i} > 0$$

Proof. See Appendix 2.C.3. ■

Corollary 2.2. *The impact of traits can be different for entry and survival. Entry is only driven by the Plasticity trait, β. Given a large enough error in the estimation of entrepreneurial abilities, survival will be mostly driven by differences in Stability, α, as its effect is amplified by errors, while the effect of the Plasticity trait is not.*

Proof. Follows from Corollary 2.1 and Proposition 2.4. ■

To see why Corollary 2.2 holds, consider two examples. First, assume that an entrepreneur correctly estimated her entrepreneurial abilities. Then, initial motivation and feedback work in the same direction by confirming the individual's estimate and motivational stability is not an important factor for entrepreneurial survival. Second, assume that an entrepreneur made an over-estimation error, which is likely given Proposition 2.2.

Then, feedback will contradict the initial estimate. However, if the individual assigns a sufficiently large weight to her initial estimate, if she is sufficiently motivationally stable, she might decide to continue being an entrepreneur. As all entrepreneurs have Plasticity suggesting entrepreneurship but only some entrepreneurs are sufficiently motivationally stable, the group of survivors will be mostly determined by Stability and not Plasticity. This mechanism applies to entrepreneurs overestimating their abilities but with entrepreneurship as the full-information equilibrium (entrepreneurship is their objectively best option) but it also applies to entrepreneurs whose full-information equilibrium is wage work, leading to the last important observation on entrepreneurial decision processes.

The following observations attracted considerable attention in the empirical literature: In the short run, many entrepreneurs do not exit despite the option of earning more as employees. The median entrepreneur in cross-sectional analysis earns less than the median salaried employee (e.g., Hamilton 2000; Astebro & Chen 2014). In the long run, entrepreneurs can realize an income premium when their incomes are compared to similar employees (Daly 2015; Manso 2016). Our model provides an intuitive explanation for this observation.

To simplify notations, assume that entrepreneurial costs are included in incomes. Assume that, besides some residual firm that can employ failed entrepreneurs, there are two entrepreneurs in period 1: individual k and l, who both strictly prefer entrepreneurship in the first period. Individual k estimated her entrepreneurial performance correctly such that $\hat{\pi}_{k,0} = \pi_{k,0}$ and $\pi_{k,0} > w_{k,0}$. Individual l substantially overestimated her entrepreneurial performance such that $\hat{\pi}_{l,0} \gg \pi_{l,0}$ and $\pi_{l,0} \ll w_{l,0}$. Furthermore, assume, without loss of generality, that $\alpha_k = 0$. Let Δ_t denote the earning differential between entrepreneurial and contrafactual wage income, as perceived by an outside observer with perfect information. Define the differential as follows:

$$\Delta_t \equiv \begin{cases} \Delta_t^- = \pi_{k,0} + \pi_{l,0} - (w_{k_0} + w_{l,0}) < 0 & \text{if } k \text{ and } l \text{ remain entrepreneurs} \\ \Delta_t^{--} < \Delta_t^- & \text{if } k \text{ and } l \text{ exit} \\ \Delta_t^\dagger = \pi_{l,0} - w_{l,0} < 0 & \text{if } k \text{ exits and } l \text{ remains entrepreneur} \\ \Delta_t^+ = \pi_{k,0} - w_{k_0} > 0 & \text{if } k \text{ remains entrepreneur and } l \text{ exits} \end{cases}$$

The earning differential is positive, which is the most desirable outcome, if individual k remains entrepreneur and l exits; and it is negative in every other scenario. The most desirable scenario corresponds to the full-information equilibrium.

For the probability of observing a non-negative earning differential, such that individuals earn more as entrepreneurs than as wage workers, the following result holds:

Proposition 2.5. *The probability of a positive earning differential is strictly increasing in* *time. However, the probability is bounded such that*

$$q_k \underline{q} \leqslant \mathbb{P}(\Delta_t > 0) < q_k \overline{q} \quad \text{for } t > 1$$

where $\overline{q} > \underline{q}$ *and* $\overline{q}$ *is decreasing in* $\pi_{l,0}$*. Thus, if* $\pi_{l,0}$ *is sufficiently large,* $q_k \overline{q}$ *might be substantially below 1.*

Proof. See Appendix 2.C.4. ∎

The proposition has two important implications. First, learning sorts out individuals unfit for entrepreneurship, as the probability of a positive earning differential increases in time. This also implies that learning shows a tendency to approach the full-information equilibrium even with noisy feedback. Second, the probability of a positive earning differential is, however, bounded and the boundary can be below 1, depending on model parameters. According to Proposition 2.5, average feedback, revealing actual entrepreneurial abilities, might not be sufficiently low to fully convince an entrepreneur who should select wage work, because it is her full-information equilibrium, to abandon entrepreneurship.

The probability of a negative earning differential, which is consistent with the earnings puzzle, is highest in the turbulent phase after entry. In reality, there will be smooth transitions between the turbulent phase and the consolidation phase due to overlapping cohorts of entrants. The number of entrepreneurs in the turbulent phase is always larger than in the consolidation phase—a general prediction of our model. Hence, lower earnings of entrepreneurs compared to workers may be observed because the characteristics of entrepreneurs are mostly driven by turbulent-phase entrepreneurs. Accordingly, in cross-sectional data there should be a high share of turbulent-phase entrepreneurs, which would, for instance, explain the findings of Hamilton (2000), while the result from the consolidation phase confirms findings of Manso (2016).

To summarize, the basic mechanisms of our model produce outcomes that are consistent with empirical evidence on entrepreneurial decision processes.

2.6.2 Analysis of model with unrestricted labor market interactions: A Monte Carlo study

The propositions derived using the simplified version of our model ignore labor market interactions and use a number of simplifying assumptions. It can be demonstrated that analytically derived mechanisms are robust to the inclusion of labor market interactions and using more realistic assumptions for the source of entrepreneurial costs and the feedback process. We discuss the robustness of our analytical findings in a simulation exercise.

2.6.2.1 Simulation setup

Introducing liability constraints

Up to now, model entrepreneurs could accumulate losses without any restrictions. To impose a restriction on losses, we assume that individuals face liability (budget) constraints. If an individual is willing to start a firm, a financial institution provides capital. An entrepreneur cannot operate without capital, as without capital she cannot pay wages in a period with sufficiently negative feedback. The financial institution computes the capital requirement[18] according to

$$K_{i,\tau_i} = \frac{\delta(\delta^{\tau_i} - \delta^T)c}{(1-\delta)(1+\mu)}\hat{\pi}_{i,\tau_i} \tag{2.15}$$

where $\tau_i \in [0, T)$ is the period in which i applies for capital; she becomes entrepreneur in $\tau_i + 1$. $\mu > 0$ is the financial institution's mark-up to diversify risk, in particular, to account for potential estimation errors of individual i, and establish profit, computed according to the institution's criteria. Individuals have no incentive to misreport $\hat{\pi}_{i,\tau_i}$ since the financial institution can verify it by combining information on matrix $\mathfrak{A}$ and data on past restructurings (both common knowledge).

To exclude different fraud schemes, the following rules are perfectly enforced: (#1) Entrepreneurs are committed to transfer a constant income share of c to the financial institution in every period of entrepreneurship. (#2) Capital must not be consumed. (#3) If an entrepreneur exits, capital is returned to the financial institution. (#4) If losses occur, a share c is covered by the financial institution but the financial institution immediately subtracts it from K. The rest, $1 - c$, has to be covered by the entrepreneur. (#5) If capital becomes non-positive, the entrepreneur is forced to exit (a restriction on losses), as any additional sufficiently negative feedback period would result in wage default.

Distributional assumptions

We use univariate normal, bivariate normal, and truncated normal distributions to generate our model's parameters. Estimated and true self-management productivities are drawn from a bivariate normal distribution with a non-negative correlation. Non-self-management productivities are generated by a truncated normal distribution with zero as lower boundary. We abstract from the question of how interactions between two individuals are related to self-management performance by assuming that non-self-management productivities are independent from true and estimated self-management productivities.

[18]Given that entrepreneurs transfer income share c to the institution and the current estimate of entrepreneurial income $\hat{\pi}_{i,\tau_i}$, the financial institution expects to receive $K_{i,\tau_i} = \sum_{t=\tau_i+1}^{T} \delta^t c\hat{\pi}_{i,\tau_i} = \delta(\delta^{\tau_i} - \delta^T)(1-\delta)^{-1}c\hat{\pi}_{i,\tau_i}$. To account for estimation errors, etc., and to generate profit, the institution requires to receive $K_{i,\tau_i}(1+\mu)$, where $\mu > 0$ is a mark-up, while it provides capital K_{i,τ_i}.

Furthermore, individuals who are good at managing a particular individual are not necessarily doing well as universal managers. Thus, non-self-management productivities are assumed to be independent from each other.

The average individual is better in managing others than herself such that the mean of non-self-management productivities is larger than the population mean of self-management productivities. Entrepreneurial abilities are assumed to be more heterogeneous than management abilities such that the variance of true and estimated self-management productivities is larger than the variance of non-self-management productivities. Estimates are assumed to be more spread out than true values such that the variance of estimated self-management productivities is larger than the variance of true productivities. The Stability parameter, α, is independently generated by a truncated normal distribution—with zero as lower and 1 as upper boundary.

We also replace Assumption 2.2.1 with a slightly more realistic noise process:

Assumption 2.2.2. Noise is i.i.d. normal with $\mathbb{E}[u_{i,t}] = 0$ and $\mathbb{V}[u_{i,t}] = \sigma^2_{u,i}$ for all i and t. For the variance of noise (and feedback in general), we have $\sigma_{u,i} = \tilde{u}|v_{ii}|$, such that noise varies by some fraction $\tilde{u} > 0$ of individual i's true productivity.

Numerical assumptions

Table 2.D.1 (in Appendix 2.D) collects all numerical assumptions. The assumption on the correlation between estimated and actual self-management productivities results in two different scenarios. In the *first scenario*, estimated and actual self-management productivities are not correlated, corresponding to a low accuracy of self-assessment. In the *second scenario*, estimated and actual self-management productivities are strongly, but not perfectly, correlated, representing high accuracy of self-assessment.

Simulation procedure

Simulations are based on (2.4), (2.5), (2.6), (2.7), (2.8), (2.11), (2.12), (2.13), (2.14), and (2.15); Assumption 2.1.2 and 2.2.2; the rules ensuring that liability constraints hold; and Table 2.D.1. (Code is provided upon request.) We compute 1,000 simulations for each scenario. The simulation procedure is briefly described in Appendix 2.D.

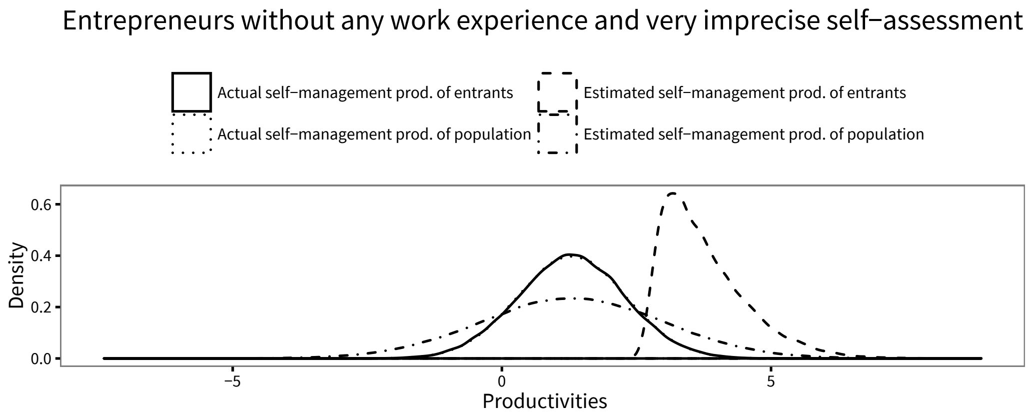
2.6.2.2 Simulated self-selection into and out of entrepreneurship and earnings

We focus on the two scenarios, differing in the degree of self-assessment errors, defined in the previous subsection. It is common to examine self-selection into and out of entrepreneurship by inspecting differences between groups—such as employees, entrepreneurs without work experience, nascent entrepreneurs with employee experience, surviving entrepreneurs, and so on—to determine which variables are important. We do the same but, in addition, we also condition on the degree of self-assessment errors, as, according to our model, this variable is also highly important.

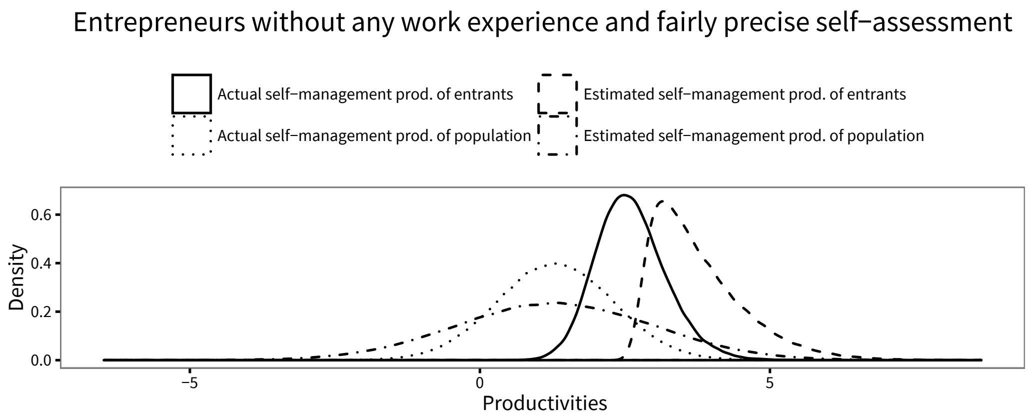
First, we discuss self-selection into entrepreneurship of individuals without any experience and with employee experience. Second, we examine the survival curve of a cohort. Third, we analyze the impact of traits on entrepreneurial survival. Then, we scrutinize earning differentials, where we are especially interested in the dynamics of earnings distributions of entrepreneurs and their contrafactual earnings distributions in wage work. All consecutive figures are generated with simulated data.

Entry and overconfidence

Figure 2.2 presents results on entry into entrepreneurship without any work experience—



(a) First scenario with low accuracy of self-assessment



(b) Second scenario with high accuracy of self-assessment

Figure 2.2. Population and entrepreneurs without any work-experience

entrepreneurs entering in the first period—by depicting distributions of estimated and true self-management productivities of entrants without experience. These productivities can be compared to distributions of the overall population, which are also depicted in Figure 2.2. Figure 2.2a shows the first scenario and Figure 2.2b the second scenario.

We can observe that self-selection in the initial period, the period without any first-hand knowledge, into entrepreneurship is based solely on estimated self-management productivities. If estimated self-management productivities are not correlated with true productivities (Figure 2.2a), entrepreneurs entering in the first period have a higher average estimated self-management productivity than the overall population, while their distribution of true self-management productivities is indistinguishable from the distribution of the population. If estimated and true productivities are positively correlated (Figure 2.2b), the group of entrepreneurs entering in the initial period will have a higher average estimated *and* true productivity. However, estimated values are still substantially larger than the true values. Hence, in both scenarios entry is influenced by the Plasticity trait, which drives estimates. Furthermore, overconfidence in the own abilities is present even when there is a strong correlation between estimated and true values.

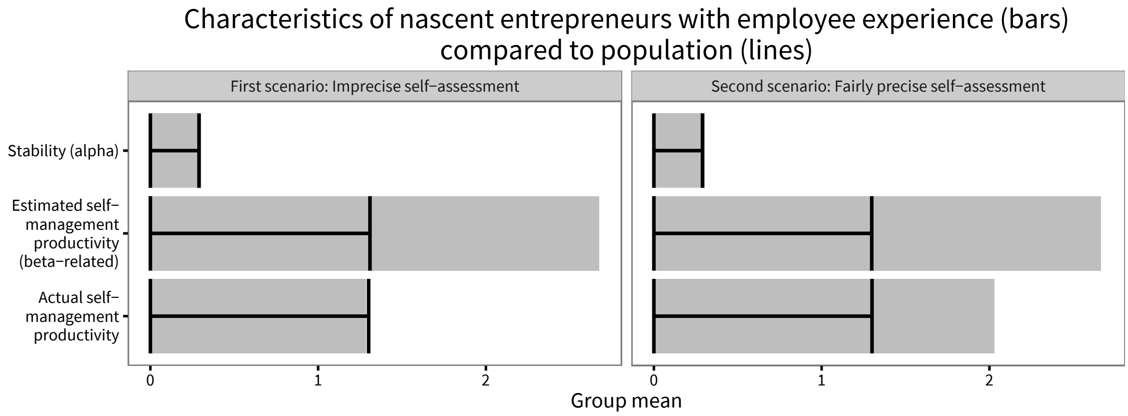


FIGURE 2.3. Nascent entrepreneurs without entrepreneurial experience but with employee experience

Figure 2.3 present results on nascent entrepreneurs without entrepreneurial experience but with experience as employees. This type cannot be discussed using the simplified version of our model. The figure compares the group means of nascent entrepreneurs (gray bars) to population means (black lines) with respect to Stability (α), and estimated and actual self-management productivities, which are related to Plasticity (β). We see that the statements for entrepreneurs without any work experience also hold for nascent entrepreneurs with employee experience. Although opportunity entrepreneurs have an average level of stability near population levels, they differ from the overall population in one regard: Their estimated self-management productivities are approximately twice as large as the overall population analogue. Also, entrepreneurial abilities are overestimated in both scenarios, as average estimated abilities are larger than average true abilities. Hence, we conclude that the full model is consistent with Proposition 2.1 and 2.2, including Corollary 2.1.

Survival dynamics

Switching from the individual perspective to a group-based view, Figure 2.4 presents sur-

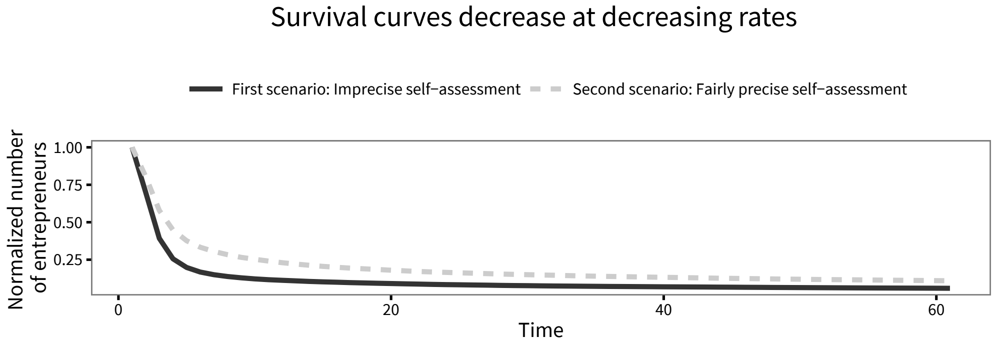


Figure 2.4. Survival dynamics

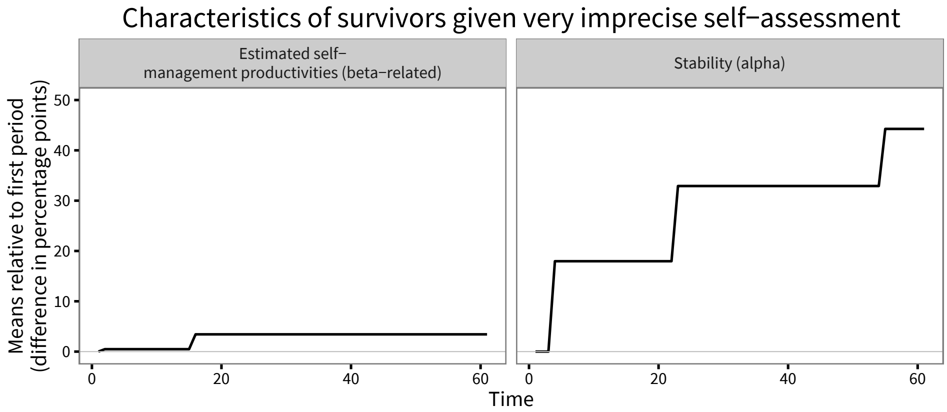
vival dynamics by following a *cohort* of entrepreneurs over $T + 1 = 61$ periods. In the figure, the number of entrepreneurs in each period is normalized by the number of entrants in the first period such that we can examine the share of the original cohort still in business in each period $t = 2, \ldots, T + 1$.

As one can see in Figure 2.4, the probability of survival strongly decreases in the first periods after entry, in the turbulent phase, and stabilizes in later periods, in the consolidation phase. This corresponds to survival-curve shapes in reality. The shape of the curves in Figure 2.4 is also consistent with Proposition 2.3, based on individual survival probabilities. Intuitively, reducing self-assessment errors increases the probability to survive for the whole cohort, which corresponds to the difference between the first and the second scenario in Figure 2.4.

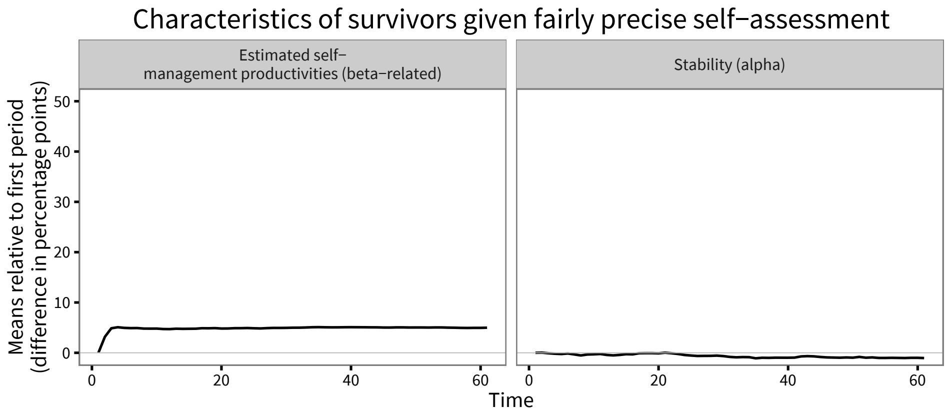
Non-linear impact of traits

The impact of personality on survival can be analyzed with normalized means. We compute the means of survivors' trait-related variables given that entrepreneurs survived over t periods and normalize them by the mean in the first period. The interpretation is straightforward. If, for example, the mean of the Stability parameter in period $t > 1$ is substantially above zero, this implies that survivors have a Stability trait level above the level of entrants.

Figure 2.5 shows normalized means of each trait-related variable of the survivor group as a function of time. In line with Proposition 2.4 and Corollary 2.2, survival is mostly driven by the Stability trait if estimation errors are sufficiently large (Figure 2.5a), while Stability of the survivors' group is fairly stable over time if estimation errors are small (Figure 2.5b).



(a) First scenario with low accuracy of self-assessment



(b) Second scenario with high accuracy of self-assessment

FIGURE 2.5. Change in trait-related variables of survivors over time

EARNING DIFFERENTIALS

The observation that many entrepreneurs earn less than similarly qualified wage workers is often based on a cross-sectional comparison of entrepreneurial earnings with the contrafactual wages. The simulation data generated by our model (we use data from the first scenario) allows for a direct comparison of these two groups without the need to adjust for selection bias.

Figure 2.6 addresses earning differentials by depicting distributions of contrafactual wage income and net entrepreneurial earnings in three different periods. Distributions for $t = 2$ are most similar to empirical distributions presented by Hamilton (2000). The periods $t = 1$ and $t = 2$ belong to the turbulent phase. For $t \in \{1, 2\}$, we see that the individual receiving the median entrepreneurial income would be better off as an employee. In $t = 50$, we are in the consolidation phase. In this phase, most, though not all, entrepreneurs who could get a higher wage income made an exit from entrepreneurship as a consequence of learning. In $t = 50$, the median income of an entrepreneur is strictly

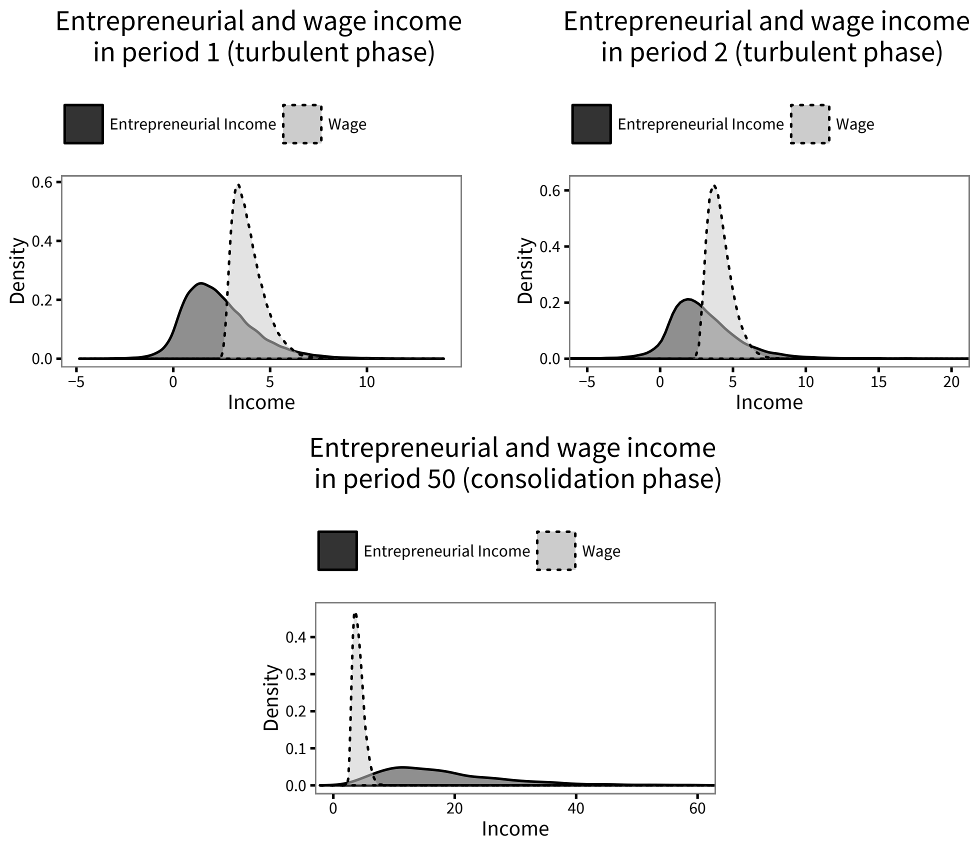


FIGURE 2.6. Earnings distributions over time

better than the median wage income. Thus, if cohorts overlap and observations are mostly driven by turbulent-phase entrepreneurs, as this is the largest group of entrepreneurs, the model produces earning differentials that are consistent with empirical observations. Furthermore, the dynamics of earnings distributions are consistent with Proposition 2.5—and the more recent empirical observations of, for instance, Manso (2016).

Concluding, it can be said that the mechanisms described in Proposition 2.1–2.5 are general enough such that they do not critically depend on simplifying assumptions made in the process of obtaining them.

2.7 Related approaches, limitations, and future research

2.7.1 Related approaches

In the literature, there are several approaches that, although partly focusing on different questions, aim to explain transitions between entrepreneurship and salaried jobs. These approaches fall into two broad categories: models that emphasize experimentation and models that concentrate on a different aspect.

Experimentation-based approaches

Vereshchagina & Hopenhayn (2009) develop a real option value argument that might explain the absence of a risk premium for entrepreneurs. By assuming that entrepreneurs are self-financed, they show that there is an incentive for poor entrepreneurs to invest in risky projects. In contrast to the risky environment of Vereshchagina & Hopenhayn (2009), we assume uncertainty. Moreover, we account for personality and use empirically tested learning heuristics. Furthermore, the assumption that entrepreneurs are self-financed restricts the applicability of the model, as a substantial share of entrepreneurs use outside capital (Banerjee & Duflo 2014). Borrowing constraints are endogenous in the unrestricted version of our model, where potential entrepreneurs can borrow capital from a financial institution.

Manso (2016) construct a simple model where individuals are not sure about their entrepreneurial abilities and can experiment to obtain additional information. However, as in case of Vereshchagina & Hopenhayn (2009), Manso's model is constructed under the assumption that the probabilities of an outcome are known—it assumes risk and not uncertainty. It also does not account for risk aversion, which is a personality-related variable, working against the incentive to experiment, as entrepreneurship is more risky than wage work.

The approach of Dillon & Stanton (2016), although it has a somewhat different aim, is similar to our model in the sense that it develops a dynamic model of entrepreneurship that incorporates learning, heterogeneous preferences of individuals with respect to entrepreneurship, and initial uncertainty about entrepreneurial incomes. A related approach is Miller (1984) who uses the multi-armed bandit problem to model entrepreneurship-related experimentation. In both approaches, individuals enter entrepreneurship because the higher uncertainty of entrepreneurial rewards, compared to wage work, creates an incentive for optimal learners to experiment with the option of entrepreneurship. Both contributions use the concept of optimal learning without controlling for learning efficiency. This has three major limitations, which are avoided by our model.

First, these models are silent on the topic of personality. However, personality plays a major role for occupational choices. If personality is accounted for, long-term entrepreneurial earnings could be above or below the long-term earnings of workers, strongly but not completely, depending on the share of individuals with a certain personality in a certain market, region, or industry sector. Empirical evidence shows that there is behaviorally induced heterogeneity in perseverance in entrepreneurship (see, e.g., Muehlfeld et al. 2017). In contrast to experimentation-based approaches, our model captures this evidence fairly well.

Second, individuals do not necessarily learn in an optimal way. Even in an experimental setup with a simple multi-armed bandit game, capturing the trade-off between experimentation and exploitation in occupational choice, players tend to be myopic, dislike uncertainty, and play non-optimal options for too long (Meyer & Shi 1995; Anderson 2001; Gans, Knox & Croson 2007; Steyvers, Lee & Wagenmakers 2009). In our model, we are able to match empirical and theoretical learning dynamics and personality effects with a single algorithm—a result that cannot be obtained if optimal learning is applied.

Third, experimentation approaches tend to substantially overestimate the probability of entering entrepreneurship. For instance, assuming that individuals are forward-looking income maximizers who experiment with entrepreneurship, the model of Dillon & Stanton (2016) predicts that "55% of paid workers will choose to become entrepreneurs next year when only 2% do" (Dillon & Stanton 2016, p. 24). To reduce the probability of entrepreneurship to a reasonable level, Dillon & Stanton (2016) introduce three additional variables: entry costs, unobserved heterogeneity in preferences, and taste shocks. The latter two, especially unobserved preferences, are rather intransparent variables. In our model, the determinants of entry are transparent. As laid out in Proposition 2.1, given fixed (entry) costs, the overall probability of the decision to become entrepreneur depends on the distribution of known management and estimated self-management abilities, which in turn are influenced by the distribution of Plasticity in the population. Thus, in contrast to experimentation, our model can generate realistic entry probabilities without using unobserved preferences and taste shocks.[19]

Non-experimentation-based approaches

Astebro et al. (2011) use the assumption that entrepreneurship results from non-optimal assignments of workers to firms, which generates effects similar to the earnings puzzle. As a consequence, there would be no entrepreneurs if assignments were frictionless. Our model does not require frictions to produce entrepreneurs. Furthermore, Astebro et al. (2011) is restricted to explaining the occurrence of necessity entrepreneurs, as only the inability of an individual to find a good match induces her to become an entrepreneur. In our model, necessity and opportunity entrepreneurs are both present and earnings-puzzle-like effects do not rely on frictions.

2.7.2 Limitations

Despite the fact that our model is able to relax crucial assumptions, comparing it to earlier and parallel research, there are two obvious limitations. First, the model still needs some assumptions to make it tractable. The most critical assumption is that labor market interactions can be ignored to a certain extent. However, we are able to show that the mechanisms we proposed are sufficiently robust by relaxing critical assumptions in a

[19] In Chapter 3, I introduce a flexible experimentation-based model that does not require taste shocks but only ambiguity preferences.

simulation exercise. Secondly, the way we model feedback might be too simplistic. In particular, our assumptions on the distribution of noise might impose a very simple structure on feedback dynamics. For instance, it might be more realistic to assume that noise shows clusters of volatility, or that negative feedback tends to be followed by negative feedback.

2.7.3 Future research

The model provides a dynamic learning algorithm predominantly operating with observable variables and trait-rationalizing variables. As traits can be measured, an obvious first extension for future research would be to directly estimate the model parameters with representative data, for instance with US data like the Panel Study of Income Dynamics (PSID), or data provided by the German Socio-economic Panel (GSOEP). Both data sets contain rich data including patterns of employment and income over time as well as determinants of personality characteristics. For instance, the GSOEP delivers information on the Big Five, which could be used to estimate the higher order personality inventory.

Secondly, beyond direct empirical tests, our model could be used to discuss serial entrepreneurship, which would involve relaxing the assumption of restricted labor market interactions in the simplified model version. Failure and success by chance is another topic for further research. It might also be interesting to use ability and feedback distributions with heavier tails, ensuring the appearance of Bill-Gates-type entrepreneurs, to discuss inequality, as entrepreneurship and income distributions are linked (Lloyd-Ellis & Bernhardt 2000).

Thirdly, the model also allows investigating how certain policy measures, such as programs supporting business creation, may influence entrepreneurial decision making. One unique feature of our model is that it would reveal how different types of potential entrepreneurs react to policy measures, given that our model covers behaviorally induced heterogeneity in the entry decision and in the pace of learning.

Last, but not least, the most radical model innovation to be discussed in future research is to further relax the common knowledge assumptions in the productivity matrix (Equation 2.1) and not only assume incomplete information in the diagonal but for the complete matrix.

2.8 Summary

In this chapter, we develop a bounded-rationality occupational choice model, where individuals are not fully aware of their entrepreneurial abilities but learn from noisy market feedback, and are restricted in their reactions to feedback by their personality. The model allows us to discuss self-selection into and out of entrepreneurship and is capable to explain empirical evidence on the two processes. With respect to personality traits, we follow earlier research to make the model tractable and use higher order personality factors to sort traits into two metatraits. The Plasticity trait (β) influences the self-perception and self-selection of individuals into occupations. The Stability trait (α) determines how quickly an individual reacts if faced with a new stimulus, respectively it influences the stability of an individual's initial motivation.

The model generates all types of entrepreneurs: from solo entrepreneurs to entrepreneurs running large companies with many employees. Individuals become entrepreneurs because they assume that their entrepreneurial productivity, generating entrepreneurial income, is large enough to offset entrepreneurial costs and the alternative wage income. As a consequence, the distribution of estimated entrepreneurial productivities of entrants into entrepreneurship is truncated from below pushing the average of estimated abilities of entrants above the population average. If actual entrepreneurial abilities are unknown, this results in a substantial overestimation of entrepreneurial abilities, found in empirical data, because the upward movement in average estimated ability levels does not reflect actual entrepreneurial abilities. The mechanism generates entrepreneurial overconfidence even if estimates are right on average, such that individuals do not have to be generally overconfident but overconfident individuals have a higher probability to become entrepreneurs. Thus, a simple self-selection mechanism can lead to overconfident nascent entrepreneurs.

Entrepreneurial survival curves produced by our model are plausibly shaped with survival probabilities decreasing at a decreasing rate. As individuals self select into entrepreneurship, the market sorts out fresh entrepreneurs who do not possess the "right" abilities to survive. The sorting mechanism does most of its work in initial periods—in these periods entrepreneurs receive market feedback for the first time—following entry such that the probability to fail is largest in this turbulent phase. In later periods, the probability to survive stabilizes.

After entry, the Plasticity trait becomes less important than the Stability trait. If self-assessment of entrepreneurial performance is sufficiently bad, the group of survivors is mostly defined by the Stability trait, as its impact increases with estimation errors, whereas the effect of Plasticity is not influenced by errors. Thus, traits affecting entry and survival can be different because of unknown entrepreneurial abilities and noisy feedback. Furthermore, if estimation errors are present, being more motivationally stable (or having higher perseverance) makes entrepreneurs more resilient such that survival probabilities increase.

Some entrepreneurs could earn more in wage employment but may not change the occupation to realize it. However, the number of individuals who stay in entrepreneurship, even if their counterfactual wage income is larger, decreases with time. The reason is that entrepreneurial income is an imperfect signal that comes in form of noisy market feedback and, thus, the best option is not identified with a probability of 100%. As this holds for each new cohort of entrepreneurs, there will always be a large number of entrepreneurs who should switch to wage work but stay in entrepreneurship if cohorts overlap. This explanation of the earnings puzzle does not require non-standard assumptions on preferences such as an additional non-pecuniary benefit of entrepreneurship (e.g., Hundley 2001).

Thus, all results of our model are consistent with the puzzling observations on entrepreneurial decision processes. A main insight of our analysis is that the driving force behind most of the results are effects of self-assessment errors due to a difference between perceived and true entrepreneurial abilities. Moreover, the relation of traits to entrepreneurial survival is rather complex since no trait matters unconditionally. The prototype successful entrepreneur—if it is aimed to build such a construct—cannot be described by a static definition, but needs the “right” traits in the right situation.

3 CAREER CHOICE UNDER AMBIGUITY

Contemporary theoretical literature on entrepreneurial choice consists mostly of models that treat choice outcomes as either deterministic or risky. This chapter proposes taking a new perspective by constructing a general career choice model on the basis of the assumption that outcomes are partially ambiguous such that some reward distributions are unknown. The change in perspective yields some major advantages. Learning and career trajectories, which in general cannot be generated by models with deterministic or risky rewards, become a natural feature of the dynamic solution of sequential career choice problems. As a consequence, the model accounts for the important phenomenon of mixed careers. Furthermore, earnings-puzzle-like observations can be explained by sufficiently high ambiguity aversion, as ambiguity aversion has a significant impact on learning.

3.1 Introduction

Modeling that explains career choices is a central focus of labor economics, entrepreneurship research, and management science. In the literature, there are various formal models (e.g., Lucas 1978; Kihlstrom & Laffont 1979; Holmes & Schmitz 1990; Lazear 2005; Roessler & Koellinger 2012) explaining why some individuals become wage workers and why some decide to start an own business. However, most models assume that once a decision for entrepreneurship or wage work is made, individuals stay in their selected occupation until the end of their careers. A consequence of the assumption that decisions are final is that the theoretical literature neglects mixed career trajectories, where careers can include a mixture of spells in entrepreneurship, wage work, and unemployment. Mixed careers are, however, common in empirical data (Ferber & Waldfogel 1998; Williams 2000; Arum & Müller 2004). As career patterns shift from long-term employment in one firm to greater mobility between firms and occupations (Arthur 1994; Hall & Moss 1998; Valcour & Tolbert 2003; Granrose & Baccili 2006; Sullivan & Baruch 2009; Biemann et al. 2012), mixed careers can be assumed to become even more common and important in the future. An additional empirical regularity that is also hard to explain with most contemporary career choice models is the earnings puzzle: That is, the observation that a large share of entrepreneurs could earn more in dependent employment but they do not choose to do so (e.g., Hamilton 2000; Astebro & Chen 2014).

In this contribution, it is argued that altering one modeling assumption can significantly reduce the gap between theory and empirical evidence. Instead of the common assumption that career choices are made in a deterministic or risky choice environment,[20] this contribution assumes that the choice environment has elements of ambiguity. More formally, let $\mathbb{O}$ denote the finite set of alternative career choices, such that there are $|\mathbb{O}|$ options, where each choice is assigned a reward distribution with density f_i for $i \in \mathbb{O}$. In a deterministic or risky environment, distributions $f_1, f_2, \ldots, f_{|\mathbb{O}|}$ are either degenerate or known. In an environment with ambiguity, the assumption used in this chapter's model, there exists at least one f_i, where $i \in \mathbb{O}$, that is *unknown*.

A good example is entrepreneurship. The assumption that income distributions in entrepreneurship are perfectly known is very restrictive since the distribution of rewards from entrepreneurship depends on many factors that individuals cannot fully anticipate: Innovative entrepreneurship is essentially all about trying something never done before. A similar argument applies to the distribution of rewards from wage work. Wage workers do not have full control over their careers. The probability of being promoted or of losing one's job can depend on factors that are not fully transparent to workers. Job matching has a random component. Furthermore, many skills wage workers need are acquired through formal training in schools or universities in controlled settings. Formal training and real-world situations can be quite different. For instance, the stress level in real-world situations can be much higher than in controlled settings. Consequently, some important skills affecting rewards in wage work can only be revealed by actually performing some tasks on the job.

In this vein, Antonovics & Golan (2012) show that empirical occupational-choice patterns and wage growth are consistent with the assumption that jobs only gradually reveal information about unknown workers' skills. However, unknown distributions do not imply that there is no information at all. Usually, there is some prior information about options and this information is updated as new information is obtained.

The model constructed in this chapter is an adaption of the multi-armed Bayesian bandit problem to general career choice. A bandit model is a stylized representation of a sequential decision problem where arms of a bandit machine represent options with potentially unknown payoff distributions. Individuals obtain information about the reward distribution of a particular option by selecting it and observing the outcome. An individual has to decide which option to select in every period and, especially, when to switch between options conditional on new information (Gittins, Glazebrook & Weber 2011).

A career choice model based on the bandit approach has a number of interesting features. First, with the help of the model ambiguity can be discussed in a systematic Bayesian framework. Second, assuming that some choice outcomes are ambiguous automatically

[20] The few models not using one of the two assumptions are discussed in Section 3.2.

induces learning. As learning processes are paralleled by transitions between occupations to obtain information, the model generates career trajectories including mixed careers. Third, introducing ambiguity aversion into the model is a simple approach to discuss learning efficiency—the ability to identify the objectively best career option—which is important for the explanation of the earnings puzzle. For example, it is intuitively plausible that highly ambiguity-averse individuals avoid experiments. Furthermore, it turns out that the earnings puzzle can be produced by a combination of misleading prior knowledge, characterized by overconfidence with respect to entrepreneurship, and ambiguity aversion reducing learning efficiency. Effectively, ambiguity-averse entrepreneurs can be trapped by entrepreneurial overconfidence. Lastly, the bandit is the simplest theoretical device to depict the trade-off between exploitation and exploration inherent in career choice, as the problem of finding the best occupational option—to explore—and generating a high-enough income—to exploit—must be solved simultaneously.

The remainder of the chapter is organized as follows. Section 3.2 reviews previous research relevant for the model. Section 3.3 depicts the model's setup. Section 3.4 discusses decision rules. Section 3.5 gives an in-depth examination of choice behavior in a partially ambiguous decision environment. Section 3.6 illustrates the model's mechanisms and discusses empirical evidence on model predictions. Section 3.7 concludes. Appendix 3 provides proofs not included in the main text, additional results, and empirical evidence on the two most important model predictions.

3.2 Related research

The model in this chapter combines, and extends, three strands of research. More specifically, I build upon research on career choice, research on multi-armed bandits, and research on ambiguity aversion.

3.2.1 Career choice: Theory and empirics

This section discusses three aspects of research on career choice: choice theory, the earnings puzzle, as well as empirical evidence on career trajectories.

3.2.1.1 Career choice theory

Choice models provide different sophisticated explanations for why individuals become entrepreneurs or wage workers. Lucas (1978) puts emphasis on entrepreneurial talent. Kihlstrom & Laffont (1979) show that entrepreneurs need to be more risk tolerant than wage workers. Holmes & Schmitz (1990) point out that entrepreneurs are opportunity seekers. Lazear (2005) argues that entrepreneurs are jacks-of-all-trades, while workers are specialists.[21] Roessler & Koellinger (2012) explain the emergence of entrepreneurs in a job matching model, where individuals who are best managed by themselves and not by others become entrepreneurs. All of the aforementioned models have one central feature. As they operate with deterministic or risky rewards, they cannot explain career trajectories by construction. By assuming that reward distributions of some career options are unknown and that individuals are Bayesian learners reducing ambiguity by selecting unknown options, my contribution to the theory of career choice consists of the introduction of a new model that generates realistic career trajectories.

There are few career choice models with learning components in the literature. For instance, in the model of MacDonald (1988) predictions on the distribution of rewards are derived from an information accumulation process. However, existing learning models do not answer an important question: What determines learning efficiency? In the model presented in this chapter, learning efficiency is a function of the psychological variable ambiguity aversion. In this way, my model not only explains how individuals learn but also why different individuals can arrive at different conclusions given the same learning time and information—an important element in the explanation of an observation like the earnings puzzle.

3.2.1.2 Earnings puzzle

The earnings puzzle is based on the counterintuitive observation that a large share of entrepreneurs could earn more in wage work (Hamilton 2000; Williams 2000; Kawaguchi 2003; Hyytinen & Rouvinen 2008; Hartog, Van Praag & Van der Sluis 2010; Astebro, Braunerhjelm & Broström 2013; Hyytinen, Ilmakunnas & Toivanen 2013; Astebro & Chen 2014), raising the question of why entrepreneurship is selected when wage work generates higher payoffs. Moreover, entrepreneur cohorts tend to permanently earn less than employees (Astebro & Chen 2014), such that income differences are not reduced over time. To my best knowledge, the only model with learning components able to replicate the earnings puzzle is Astebro et al. (2011). However, their model only addresses necessity entrepreneurship, as, in their model, the only reason to start an own business is the inability of the market to frictionlessly assign workers to firms or workers to tasks.

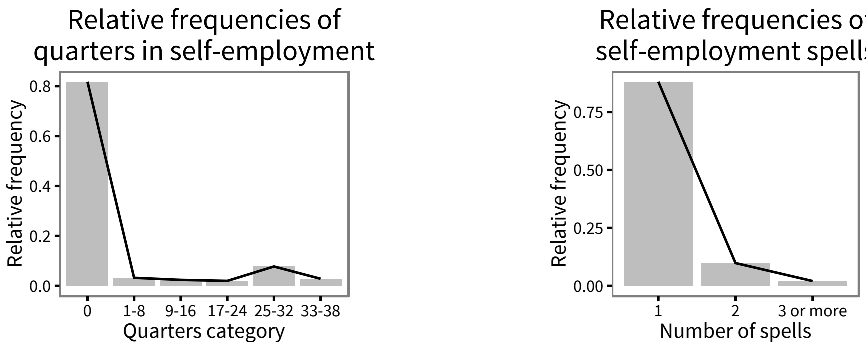
In this chapter, it is argued that a simple explanation for the earnings puzzle, very much overlooked in the literature, is that entrepreneurs are overconfident (Busenitz & Barney 1997; Camerer & Lovallo 1999; Bernardo & Welch 2001; Koellinger et al. 2007) and ambiguity-averse, with the synergy between entrepreneurial overconfidence and ambiguity aversion playing the central role in the explanation of the puzzle.

[21] See Wagner (2006) and Stuetzer, Goethner & Cantner (2012) for empirical evidence on Lazear's predictions.

3.2.1.3 *Mixed careers*

Career choices are not final decisions. Burke et al. (2008) show that pure entrepreneurial careers—an individual is an entrepreneurs over her whole career—are outnumber by mixed careers—individuals combining entrepreneurship and wage work spells. Similarly, Williams (2000) finds that a substantial share of US employees has previous self-employment experience (also, see Ferber & Waldfogel 1998; Arum & Müller 2004).

Figure 3.1a shows the relative frequency of quarters (three months) in self-employment on the basis of calculations provided by Burke et al. (2008), drawing data from the UK National Child Development Study for the years 1991–2000. Individuals



(a) Relative frequencies of time spent in self-employment

(b) Relative frequencies of self-employment spells conditional on at least one spell

Source: See Burke et al. (2008), especially Table 1 and 2 based on UK National Child Development Study (data for the years 1991–1999/2000).

FIGURE 3.1. Observations on self-employment spells

with zero quarters in self-employment are "die-hard wage workers" who did not have an own business in the given time frame. The category of 33–38 quarters in entrepreneurship represents "die-hard entrepreneurs" who were mostly self-employed in the given time frame. As we can see in Figure 3.1a, there is a significant share—approximately 10%—of individuals who are not in any "die-hard category;" they are neither pure wage workers nor pure entrepreneurs.

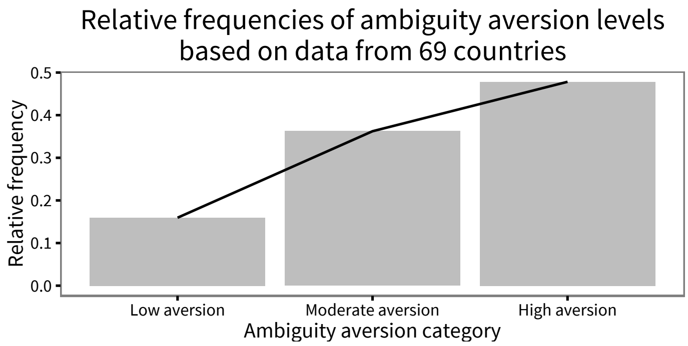
Figure 3.1b depicts the relative frequency of self-employment spells conditional on at least one spell (the data source is the same as in Figure 3.1a). The number of self-employment spells is a proxy for transitions between occupations. According to Figure 3.1b, the overwhelming majority of entrepreneurs generates only one spell. Yet, approximately 12% generate more than one spell. Figure 3.1 leads to the following conclusion: Although mixed careers are relatively common and should be accounted for, the observable number of transitions between occupations is relatively low.

3.2.2 Multi-armed bandits

The multi-armed bandit problem provides a particularly suitable method to discuss ambiguity in career choice. Economic theory is aware of the bandit problem and there are some adaptations. For instance, Rothschild (1974) uses a two-armed bandit to demonstrate that price distributions can exist even when costs and demand are identical across firms and markets. Jovanovic (1979) shows that employee turnover can be an unavoidable consequence of learning processes. Bergemann & Hege (2005) employ bandits to examine best financing rules for research projects with unknown success outcomes and unknown length. My contribution to research on multi-armed bandits is to apply the bandit approach to general career choice. This approach allows for the derivation of results on transitions and learning efficiency not obtainable in career choice models without ambiguous components.

3.2.3 Ambiguity aversion and its impact

The most prolific study evaluating cross-cultural differences in ambiguity avoiding behavior was conducted by Hofstede, Hofstede & Minkov (2010), who, based on questions answered by IBM employees,[22] construct country-specific indices capturing attitudes toward ambiguity.[23] Figure 3.2 shows relative frequencies of the categories "low," "moderate," and "high" ambiguity aversion, on the basis of ambiguity aversion indices from 69 different countries.



Note: Figure compiled on the basis of k-means clustering (Hartigan & Wong 1979) presented in Appendix 3.F (Table 3.F.2). Data is taken from Hofstede (2015).

FIGURE 3.2. Relative frequencies of low, moderate, and high ambiguity aversion

As can be clearly seen in Figure 3.2, the high-aversion group dominates, while the

[22]Respondent groups not consisting of IBM employees generate results strongly correlated to results obtained by questioning IBM employees (Huang 2008).
[23]The cross-cultural dimension is better researched than intra-cultural differences in ambiguity preferences.

low-aversion group has the lowest relative frequency. Thus, accounting for *high* ambiguity aversion is important.

Research shows that ambiguity aversion has an impact on a number of variables. Some examples include Huang (2007), showing that countries with higher ambiguity aversion export less to distant countries. Huang (2008) provides evidence that countries with higher ambiguity aversion grow slower in industrial sectors where information is scarce. Ramírez & Tadesse (2009) demonstrate that firms in countries with higher ambiguity aversion hold more cash. Inklaar & Yang (2012) show that financial crises have a stronger negative effect on investment in countries with higher ambiguity aversion.

My contribution to the literature on the impact of ambiguity aversion is a new effect: Ambiguity aversion can influence career decisions and, in particular, ambiguity aversion might explain a phenomenon like the earnings puzzle.

3.2.4 Summary and contribution

Summarizing strands of research relevant for the chapter at hand, it can be shown that most models in the literature on career choice operate with deterministic or risky rewards and do not aim to explain multiple transitions between occupations. By combining the strand of research on career choice and the strand of research on multi-armed bandits, I introduce a model with unknown stochastic rewards that is able to generate multiple transitions, producing realistic career trajectories under realistic assumptions. Furthermore, contemporary career choice models do not provide a satisfactory explanation for the earnings puzzle. To obtain results consistent with the earnings puzzle, this chapter draws on research on ambiguity-aversion effects and contributes to the literature by showing that sufficiently high ambiguity aversion results in earnings-puzzle-like observations.

3.3 The model

This section builds a career choice model with a representative individual and partially unknown reward distributions. As this type of models can be most precisely be expressed in the language of multi-armed bandit problems, terminology alternates between career choice and gambling throughout the chapter. First, I introduce the building blocks of the model: rewards. Then, the model's setup is introduced.

3.3.1 Rewards as building blocks

Consider an arbitrary occupational option producing a per-period reward. Rewards are stochastic but the distribution of rewards is unknown. It is obvious that unknown stochastic rewards must be treated differently than deterministic rewards. However, unknown stochastic rewards also have substantially different implications than known risky rewards.

First, given known risky rewards, learning about reward distributions is not necessary. If distributions are unknown, if there is ambiguity, learning is essential and improves decisions. Furthermore, learning can be influenced by preferences toward ambiguity.[24] Second, given a criterion like expected rewards, in an environment with risk, the best option is known from the beginning—for instance, the best option is the option with the highest expected reward. In an environment with ambiguity, the best option is unknown from the beginning and must be identified in the course of learning. Third, in an environment with known risky rewards, beliefs do not play a role. However, if reward distributions are unknown, prior beliefs can significantly influence choice outcomes.

Instead of monetary incomes, the model uses success outcomes (whether the individual was successful in a certain period or not), generated on the basis of success probabilities. This outcome formulation has several advantages. First, the *structure* of the problem is not affected by the alternative formulation, as the gambler has to learn and to decide which occupation to select given that incomes are drawn from unknown distributions. A problem setting with monetary outcomes would have the same structure, and—as we are interested in a solution for the "structural" problem—the solution would also be similar. Second, monetary rewards can be transformed into success probabilities, such that decisions based on expected values and on the basis of success probabilities do not contradict each other in general.[25] Third, in contrast to a formulation with monetary rewards, the formulation using success outcomes is broader. For instance, it can also account for non-monetary aspects of rewards, which is not possible if expected values are the only criteria.[26] Last, but not least, the formulation greatly simplifies the analysis, as we only have

[24] Therefore, concepts such as subjective expected utility (Savage 1954), which do not differentiate between risk and ambiguity, are not appropriate to model decisions under ambiguity.

[25] There are several types of transformations, but a simple one is as follows. Let Ψ denote monetary rewards. Furthermore, let $\bar{\Pi} \in \mathbb{R}^+$ denote an income benchmark. The individual considers a period as a success if income is above the benchmark and as a nonsuccess if not. Thus, the success probability is given by

$$\phi = \mathbb{P}(\Psi > \bar{\Pi})$$

As the distribution of Ψ is unknown, the success probability, ϕ, is also unknown. In Appendix 2.A.1, I derive two representation results showing that decisions based on expected values and on the basis of success probabilities conform, such that they rank two payoff distributions the same, given a log-normal payoff (see Lopez & Servén 2006) and a certain, strictly positive, income benchmark. In particular, it can be shown that ranking distributions according to their means can be interpreted as ranking them according to their success probabilities given a certain income benchmark (but not the other way around).

[26] Blanchflower (2000), Hundley (2001), and Benz & Frey (2008a,b) demonstrate that entrepreneurs experience greater job satisfaction than wage workers. To account for this empirical phenomenon, we could, for instance, assume that the benchmark (as defined in Footnote 25) for entrepreneurship is smaller than the benchmark for wage work.

to deal with one parameter: the success probability. In case of, for example, normally distributed rewards, we would have to account for updates of the mean and the variance. This would only increase the complexity of the problem without necessarily providing new results.

3.3.2 Setup

Think of a casino with a bandit. The bandit has three arms. Arm $\mathfrak{w}$ represents wage work; arm $\mathfrak{e}$ is entrepreneurship; and arm $\mathfrak{u}$ is unemployment. Thus, there is a set of occupational options $\mathbb{O} \equiv \{\mathfrak{e}, \mathfrak{w}, \mathfrak{u}\}$. Sometimes, it is more convenient to use a smaller set of options $\mathbb{O}_{-\mathfrak{u}} \equiv \mathbb{O} \setminus \mathfrak{u}$.

A gambler can play the bandit in every time period from 1 until infinity, where time periods are indexed by $t \in \{0, 1, 2, \ldots\}$. The three arms are independent. Only one arm can be selected in a particular period (an individual cannot have more than one occupation at the same time) and one arm must be pulled (an individual must have some occupation including unemployment). If the gambler selects an option $i \in \mathbb{O}$ in some period, I also say that the gambler samples from $i \in \mathbb{O}$. Switching from one arm to another does not result in any *direct* costs. If one arm is pulled, the remaining arms rest and cannot be observed (an individual can only gain information about an occupation by selecting it). Pulling an arm yields a reward: either a success, represented by $\pi = 1$, or a nonsuccess, represented by $\pi = 0$. Wage work generates an i.i.d. reward sequence $\left\{\Pi_{\mathfrak{w},t}\right\}_{t=1}^{\infty}$ (π is a realization of Π) drawn from a Bernoulli distribution with probability mass function

$$f(\pi_{\mathfrak{w}}; \phi_{\mathfrak{w}}) = \phi_{\mathfrak{w}}^{\pi_{\mathfrak{w}}}(1 - \phi_{\mathfrak{w}})^{1-\pi_{\mathfrak{w}}} \qquad \text{for } \pi_{\mathfrak{w}} \in \{0, 1\}$$

where $\phi_{\mathfrak{w}} \in (0, 1)$ is the success probability of wage work. Similarly, entrepreneurship generates an i.i.d. reward sequence $\left\{\Pi_{\mathfrak{e},t}\right\}_{t=1}^{\infty}$ drawn from a Bernoulli distribution with probability mass function $f(\pi_{\mathfrak{e}}; \phi_{\mathfrak{e}})$ where $\phi_{\mathfrak{e}} \in (0, 1)$ is the success probability of entrepreneurship. Unemployment generates a nonsuccess with probability 1 in all periods.

$\Pi_{\mathfrak{w}}$, $\Pi_{\mathfrak{e}}$, and the unemployment reward $\Pi_{\mathfrak{u},t} = 0$ cannot be observed in the same period. What can be observed is the result of using one particular arm (selecting a particular occupational option) in every period represented by the sequence $\{\Pi_t\}_{t=1}^{\infty}$. The gambler geometrically discounts rewards with a factor $\delta \in [0, 1)$, where $1 - \delta$ is the probability that at any given time the casino will close—equivalent to death, retirement, or inability to work.

Parameters $\phi_{\mathfrak{w}}$ and $\phi_{\mathfrak{e}}$ are unknown. It is assumed that there is some prior knowledge such that a prior on $\phi_{\mathfrak{w}}$, respectively a prior on $\phi_{\mathfrak{e}}$, can be specified. In fact, not knowing at least one parameter makes the decision problem more interesting since it does not allow for too simple decision strategies. In order to explore, the gambler might be compelled to give up some allegedly secure income. Let $\mathbb{x}_{i,0}$ denote initial, imperfect, prior information about option $i \in \mathbb{O}_{-\mathfrak{u}}$. Unemployment is a deterministic option such that information about it is always perfect. Imperfect information is updated after every decision round. Consequently, there is information $\mathbb{x}_{i,t}$ about every option $i \in \mathbb{O}_{-\mathfrak{u}}$ in every period $t \in \{0, 1, \ldots\}$. Let $\mathbb{X}_t = (\mathbb{x}_{\mathfrak{e},t}, \mathbb{x}_{\mathfrak{w},t})$ denote all information in a certain period.

The formation process of beliefs is as follows. Let $\phi \in (0, 1)$ denote the unknown success probability of an arbitrary occupational option, excluding unemployment. Assume that the prior on ϕ is beta—the beta distribution has the advantage that it is a proper conjugate prior in this setting. Hence, we have

$$\phi|\alpha, \beta \sim \mathfrak{B}(\alpha, \beta)$$

where $\mathfrak{B}$ denotes the beta distribution, and $\alpha > 0$ and $\beta > 0$ are two hyperparameters. The beta distribution has density $\mathrm{B}(\alpha, \beta)^{-1}\phi^{\alpha-1}(1 - \phi)^{\beta-1}$, where $\mathrm{B}(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1 - x)^{\beta-1}\mathrm{d}x$ is the beta function.

How will prior beliefs change if the gambler observes a one-period success or nonsuccess? According to Bayes' rule, beliefs about success probability ϕ are updated according to

$$\mathfrak{B}(\alpha_t, \beta_t) \longrightarrow \begin{cases} \mathfrak{B}(\alpha_{t+1} = \alpha_t + \pi, \beta_{t+1} = \beta_t + 1 - \pi) & \text{if } \pi \text{ observed} \\ \mathfrak{B}(\alpha_{t+1} = \alpha_t, \beta_{t+1} = \beta_t) & \text{if } \pi \text{ not observed} \end{cases} \tag{3.1}$$

As the hyperparameters α and β fully describe the distribution of the success probability, we get $\mathbb{x}_{i,t} = (\alpha_{i,t}, \beta_{i,t})$ for all $i \in \mathbb{O}_{-\mathfrak{u}}$.

The hyperparameters and their updates are straightforward to interpret. Think of α_t and β_t as of two checklists evolving over time. Parameter α represents the number of successes and β the number of nonsuccesses. Given a success, α is increased by 1 and β stays constant. Given a nonsuccess, α stays constant and β is increased by 1. Initial parameters $\mathbb{x}_0 = (\alpha_0, \beta_0)$ are determined by pseudo observations (e.g., some historical data). A pseudo observation is, for instance, the success history of a peer who is considered to be similar with respect to some aspects (e.g., skills) important for career choice. However, pseudo observations are not necessarily generated by the same reward distribution as actual observations.

The checklist character of the hyperparameters allows the construction of simple measures. Let $\gamma(\mathbb{x}) = \alpha + \beta$. Since every observed period either increases α or β, $\gamma(\mathbb{x})$ is a deterministic measure of the amount of information. The expected success probability is the number of successes divided by the number of observations, $\mu(\mathbb{x}) = \int_0^1 \phi f(\phi; \mathbb{x})\mathrm{d}\phi = \alpha\gamma(\mathbb{x})^{-1}$. The variance of the information about the probability to succeed can be measured by

$$\mathsf{v}(\mathbb{x}) = \int_0^1 \{\phi - \mu(\mathbb{x})\}^2 f(\phi;\mathbb{x})\mathrm{d}\phi = \frac{1-\mu(\mathbb{x})}{1+\gamma(\mathbb{x})}\mu(\mathbb{x})$$

Note that if the number of successes is much smaller than the number of non-successes, the gambler will expect a nonsuccess and a success will *increase* the variance, v. A similar statement holds for the effect of a nonsuccess.

Individuals are required to select a deterministic decision rule $\{d_t\}_{t=1}^{\infty}$ that specifies how to choose in every period. $d_t \in \mathbb{O}$ indicates the choice in period t. Rules continuously mixing are not allowed but mixing in a specific situation (for instance, a tie) is permitted. d_t can depend on all previous decisions $d_{t-1}, d_{t-2}, \ldots, d_1$ and all past observations $\pi_{t-1}, \pi_{t-2}, \ldots, \pi_1$ that are both captured by information $\mathbb{X}_t$, where the initialization $\mathbb{X}_0$ does not necessarily depend on actual decisions; and some additional parameter Λ reflecting preferences, etc. It is assumed that the gambler follows an index rule. In every period $t \geqslant 0$, the gambler constructs an index $\mathbf{r}_i(\mathbb{X}_t;\Lambda) \in \mathbb{R}$ for every option $i \in \mathbb{O}$ and selects the option with the largest index in the next period. Let

$$\mathbb{O}_t^* = \arg\max_{i\in\mathbb{O}} \{\mathbf{r}_\mathfrak{e}(\mathbb{X}_t;\Lambda), \mathbf{r}_\mathfrak{w}(\mathbb{X}_t;\Lambda), \mathbf{r}_\mathfrak{u}(\mathbb{X}_t;\Lambda)\}$$

and

$$d_t^{[\mathbf{r}]} = d_t(\mathbf{r}_\mathfrak{e}(\mathbb{X}_{t-1};\Lambda), \mathbf{r}_\mathfrak{w}(\mathbb{X}_{t-1};\Lambda), \mathbf{r}_\mathfrak{u}(\mathbb{X}_{t-1};\Lambda))$$

Thus, we have

$$d_t^{[\mathbf{r}]} = \begin{cases} \mathbb{O}_{t-1}^* & \text{if } |\mathbb{O}_{t-1}^*| = 1 \\ \text{each } i \in \mathbb{O}_{t-1}^* \text{ with probability } \frac{1}{|\mathbb{O}_{t-1}^*|} & \text{if } |\mathbb{O}_{t-1}^*| > 1 \end{cases} \tag{3.2}$$

Equation (3.2) can be further simplified by assuming that if $|\mathbb{O}_{t-1}^*| > 1$ and $\mathfrak{u} \in \mathbb{O}_{t-1}^*$, we should select $d_t^{[\mathbf{r}]} = \mathfrak{u}$—as unemployment is a no-effort option, while all other option require a certain effort.

Every decision $d_1^{[\mathbf{r}]}, d_2^{[\mathbf{r}]}, d_3^{[\mathbf{r}]}, \ldots$ will be associated with a reward $\Pi(d_t^{[\mathbf{r}]}) \in \{0,1\}$. The success history of an individual's career is given by $\sum_{t=1}^{\infty} \delta^{t-1}\Pi(d_t^{[\mathbf{r}]})$. In this chapter, the setup above is used to answer the following five questions: (a) What is the optimal strategy in a decision environment with elements of ambiguity? (b) How do individuals constrained by ambiguity aversion decide and what properties does their behavior exhibit, especially in comparison to the optimal rule? (c) How does ambiguity aversion affect the propensity to select relatively less ambiguous options? The deterministic option $\mathfrak{u}$ is of particular interest. (d) How does ambiguity aversion affect the correction of overconfidence with respect to entrepreneurship (learning efficiency)? (e) How are transitions from one occupation to another affected by ambiguity aversion?

The answers to the five questions provide a benchmark in the form of optimal decisions; they provide intuition into negative effects of ambiguity aversion; they provide explanations for observations like the earnings puzzle; and they also generate a number of testable predictions.

3.4 Decision rules solving career choice problems

How will individuals solve their career choice problems given that outcomes are partially ambiguous? The first part of this section examines optimal decisions maximizing the expected sum of rewards given prior information. The second part of the section uses an extended version of the mean-variance model (Maccheroni et al. 2013) to construct a decision rule accounting for ambiguity preferences.

3.4.1 A rule that maximizes expected rewards

The most simple approach to evaluate the impact of preferences toward ambiguity is to compare ambiguity-preferences-affected decisions to optimal decisions. A decision rule is said to be optimal if it induces decisions $d_1, d_2, \ldots$ maximizing the expected sum of discounted successes

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1}\Pi(d_t)|\mathbb{X}_0\right] \tag{3.3}$$

given prior information $\mathbb{X}_0 = ((\alpha_{\mathfrak{e},0}, \beta_{\mathfrak{e},0}), (\alpha_{\mathfrak{w},0}, \beta_{\mathfrak{w},0}))$. The best explanation of how to derive an optimal rule maximizing (3.3), which is simultaneously a proof of optimality, is presented by Weber (1992) and reproduced in Gittins et al. (2011).

Lemma 3.1. *The optimal decision rule is an index rule selecting the option $i \in \mathbb{O}$ with the largest index $\mathbf{g}_i(\mathbb{X}_t)$ in every period $t+1$.*[27] *Indices can be computed separately for each option such that $\mathbf{g}_i(\mathbb{X}_t) = \mathbf{g}(\mathbb{x}_{i,t})$. The index is the solution of the following optimal stopping problem:*

$$\mathbf{g}(\mathbb{x}_{i,t}) = \sup_{t_s > t}\left\{\frac{\mathbb{E}\left[\sum_{k=t+1}^{t_s} \delta^{k-1}\Pi_{i,k}|\mathbb{x}_{i,t}\right]}{\mathbb{E}\left[\sum_{k=t+1}^{t_s} \delta^{k-1}|\mathbb{x}_{i,t}\right]}\right\} \tag{3.4}$$

Proof. See Gittins et al. (2011) or Weber (1992). ■

Gittins & Jones (1974) were the first to solve the Bayesian multi-armed bandit problem with Bernoulli outcomes. Hence, we can refer to $\mathbf{g}(\mathbb{x})$ as the Gittins index. Given an optimal stopping time t_s, the Gittins index in (3.4) is the ratio between the expected sum of discounted rewards and the expected sum of "discounted time." In general, the problem

[27] Note that rule $\mathbf{g}$ is not simply the optimal index rule but also the best deterministic strategy.

in (3.4) does not have a known closed-form solution such that we have to rely on numerical approximations—Gittins et al. (2011) present different ways to approximate the index.

The Gittins index has three important properties:[28]

(G1) The Gittins index always stays on a winner such that $\mathbf{g}(\alpha + 1, \beta) > \mathbf{g}(\alpha, \beta)$.

(G2) The Gittins index is bounded such that $\mu(\mathbb{x}) \leqslant \mathbf{g}(\mathbb{x}) \leqslant \mu(\mathbb{x}) + \delta(1 - \delta)^{-1}\sqrt{\mathsf{v}(\mathbb{x})}$.

(G3) The optimal rule does not result in perfect learning. There is a nonzero probability that the best option is not selected for an infinite number of periods.

The fact that optimal learning and perfect information are not equivalent (property G3), has one important implication. Counterfactuals constructed under the assumption of perfect information, which is usually done when researchers examine phenomena like the earnings puzzle, are not an appropriate benchmark. In a realistic setting, decision makers first need to learn what the counterfactual is. Since optimal learning does not result in the perfect identification of the best option, even optimally behaving individuals might not decide on the basis of the best counterfactual, if success probabilities are unknown.

3.4.2 A rule that accounts for ambiguity aversion: A portfolio analysis approach

The optimal rule does not account for ambiguity-aversion—most individuals tend to be averse to ambiguity (Ellsberg 1961)—or for preferences toward ambiguity in general. Without doubt, there are numerous ways to formalize ambiguity and ambiguity preferences (see, e.g., Schmeidler 1989; Gilboa & Schmeidler 1989; Bewley 2002; Ghirardato, Maccheroni & Marinacci 2004; Klibanoff, Marinacci & Mukerji 2005; Maccheroni, Marinacci & Rustichini 2006; Siniscalchi 2009). Additionally, ambiguity-preferences-affected behavior can manifest in different facets. The choice of definitions and concepts below is guided by considerations of simplicity and analytical tractability.

3.4.2.1 Ambiguity and mean-variance portfolio choice

As a first step in the construction of a rule accounting for ambiguity preferences, consider the following definition of ambiguity adapted from Maccheroni et al. (2013) and conceptually similar to traditional definitions of risk (Tobin 1958):

Definition 3.1. An option $i \in \mathbb{O}$ is said to be ambiguous in period t if and only if $\mathsf{v}(\mathbb{x}_{i,t}) > 0$. If an option $j \in \mathbb{O}$ is not ambiguous, we have $\mathsf{v}(\mathbb{x}_{j,t}) = 0$ for all t. If $\mathsf{v}(\mathbb{x}_{i,t}) > \mathsf{v}(\mathbb{x}_{j,t})$,

[28] For property G1, see Bellman (1956). For property G2 and G3, see Brezzi & Lai (2000)

where $i \neq j$, we say that, in period $t \in \{0, 1, 2, \ldots\}$, option i is more ambiguous than option j and *vice versa*.

Note that ambiguity changes given information updates: The posterior distribution from the previous period is the prior in the current period—hence, the time dependency of (relative) ambiguity. Using the variance v to measure ambiguity is relatively straightforward since estimation errors are bounded by it.[29]

Assume that ambiguity preferences can be represented by a parameter $\lambda \in \mathbb{R}$. λ is a psychological primitive. It can reflect some measure of ambiguity tolerance (Sherman 1974) or a measure of ambiguity avoiding behavior like the index developed by Hofstede et al. (2010). λ might also be associated with the Openness factor in the Five Factor Model of personality (see McCrae & John 1992; John & Srivastava 2001). Without loss of generality, assume that ambiguity neutrality is represented by $\lambda = 0$. Ambiguity affinity is represented by $\lambda \in \mathbb{R}^-$. Finally, ambiguity aversion is represented by $\lambda \in \mathbb{R}^+$. Ambiguity neutrality can be used as a reference point to construct absolute ambiguity aversion: An individual 1 is assumed to be more ambiguity averse than an individual 2 if $\lambda_1 > \lambda_2$ where $\lambda_1, \lambda_2 \in \mathbb{R}^+$. For the remainder of the chapter, in line with findings presented in Section 3.2, it is assumed that ambiguity aversion is the most realistic preference toward ambiguity. However, ambiguity affinity and neutrality can be used to test the empirical validity of a decision rule, which is done in the next subsection.

To model decisions affected by ambiguity preferences, I rely on the cornerstone of portfolio choice: the mean-variance criterion, originally developed by Markowitz (1952). Maccheroni et al. (2013) extend the Arrow-Pratt approach, such that it can be applied in an environment with ambiguity, and obtain the following mean-variance-type certainty equivalent of an ambiguous option:

$$\boldsymbol{\sigma}(\mathbb{x}_{i,t}; \lambda) = \mu(\mathbb{x}_{i,t}) - \frac{\lambda}{2}\mathsf{v}(\mathbb{x}_{i,t}) \tag{3.5}$$

where $i \in \mathbb{O}$, λ represents ambiguity preferences, and individuals are assumed to be risk-neutral. Hence, gamblers are assumed to always select the option with the highest certainty equivalent such that

$$\mathbb{O}_t^{[\sigma]} = \arg\max_{i \in \mathbb{O}} \left\{\boldsymbol{\sigma}(\mathbb{x}_{\mathfrak{e},t}; \lambda), \boldsymbol{\sigma}(\mathbb{x}_{\mathfrak{w},t}; \lambda), \boldsymbol{\sigma}(\mathbb{x}_{\mathfrak{u},t}; \lambda) = 0\right\}$$

[29] According to Chebyshev's inequality, the probability that an estimation error, $|\phi_i - \mu(\mathbb{x}_{i,t})|$, is larger than or equal to some $\Delta > 0$, is bounded by $\mathbb{P}(|\phi_i - \mu(\mathbb{x}_{i,t})| \geqslant \Delta) \leqslant \mathsf{v}(\mathbb{x}_{i,t})/\Delta$ for $\Delta > \mathsf{v}(\mathbb{x}_{i,t})$. The boundary, $\mathsf{v}(\mathbb{x}_{i,t})\Delta^{-1}$, decreases if ambiguity decreases.

$$d_t^{[\sigma]} = \begin{cases} \mathbb{O}_{t-1}^{[\sigma]} & \text{if } |\mathbb{O}_{t-1}^{[\sigma]}| = 1 \\ \text{each } i \in \mathbb{O}_{t-1}^{[\sigma]} \text{ with probability } \frac{1}{|\mathbb{O}_{t-1}^{[\sigma]}|} & \text{if } |\mathbb{O}_{t-1}^{[\sigma]}| > 1 \end{cases} \tag{3.6}$$

Using mean-variance preferences in Equation (3.5) and decision rule (3.6) has the advantage of analytical tractability, but there are also at least three disadvantages. First, the mean-variance rule $\boldsymbol{\sigma}$ is myopic, which greatly simplifies computations. A myopic gambler (investor) is a standard assumption in multi-period portfolio choice problems (see, e.g., Ait-Sahalia & Brandt 2001; Jagannathan & Ma 2003; Acharya & Pedersen 2005; Hong, Scheinkman & Xiong 2006). Yet, it might be the case that results are mostly driven by myopia and not by ambiguity aversion. However, it is relatively easy to isolate the effect of myopia. In the Appendix, it is demonstrated that myopia is not the driving force behind the results (Appendix 3.C.5). Furthermore, it is shown (in Appendix 3.C.5.3) that a myopic and a global mean-variance criterion do not necessarily contradict if ambiguity aversion is sufficiently high and that, therefore, time inconsistencies can be avoided even when using a myopic decision rule.

Second, the mean-variance rule $\boldsymbol{\sigma}$ ignores risk aversion, as gamblers are assumed to be risk-neutral. Consequently, the effects presented in this chapter might not be generated by ambiguity aversion but by risk preferences. This particular concern is addressed in Appendix 3.D, where, by compounding the Bernoulli distribution with an expected success probability, the parameter ϕ, distributed according to a beta distribution to generate a measure of risk, it is demonstrated that risk preferences are unlikely to produce choice patterns resembling the earnings puzzle.

Third, mean-variance preferences tend to violate monotonicity (Maccheroni, Marinacci, Rustichini & Taboga 2009). But, as the mean-variance rule is applied to a Bernoulli distribution, a violation of monotonicity is not a substantial problem. Monotonicity cannot be violated if options have unknown success probabilities. A violation can occur if gamblers gamble for the sake of gambling such that they prefer an unknown option to a sure success. Gambling for the sake of gambling is ruled out by ambiguity aversion as

$$\max\{\boldsymbol{\sigma}(\mathbb{x}_{\mathfrak{e},t};\lambda), \boldsymbol{\sigma}(\mathbb{x}_{\mathfrak{w},t};\lambda)\} < 1 \quad \text{for all } t$$

if $\lambda > 0$. There is one relevant case where monotonicity is violated. Strictly speaking, it is irrational to select unemployment, as it delivers a nonsuccess in every state, while for a non-unemployment option there exists one state where it generates a success. However, this problem can be easily circumvented by assuming that unemployment generates a success with a known non-zero probability $\phi_{\mathfrak{u}} > 0$ since then monotonicity is not violated.[30] In such a case, unemployment is still absorbing (see S1) and gamblers with sufficiently high

[30] Given a known success probability, the index of unemployment is $\boldsymbol{\sigma}(\mathbb{x}_{\mathfrak{u},t};\lambda) = \phi_{\mathfrak{u}}$.

ambiguity aversion will oversample from unemployment (see Proposition 3.1 and 3.4). I use $\phi_{\mathfrak{u}} = 0$ because it simplifies notations.

The mean-variance rule has three important properties (demonstrated in Appendix 3.A.2):

(S1) If at some point in time it is weakly preferred to select unemployment, unemployment will be selected in all consecutive periods or $\mathbb{P}(d_{t+1}^{[\sigma]} = \mathfrak{u} | d_t^{[\sigma]} = \mathfrak{u}) = 1$, respectively $\mathbb{P}(d_{t+1}^{[\sigma]} \in \mathbb{O}_{-\mathfrak{u}} | d_t^{[\sigma]} = \mathfrak{u}) = 0$.

(S2) If a success was not expected due to a large number of previous nonsuccesses, the gambler might not stay on a winner.

(S3) Even when all prior observations $\mathbb{x}_{i,0}$ for all $i \in \mathbb{O}_{-\mathfrak{u}}$ are sampled from a Bernoulli distribution with a correct success probability ϕ_i, the mean-variance rule will with a nonzero probability never identify the best option.

Properties of the mean-variance and the optimal rule are quite similar—for example, both rules do not lead to perfect learning—but the properties of the mean-variance rule might be considered as slightly more realistic. For instance, for gamblers it might not make sense to stay on a winner—the optimal rule always stays on a winner—if a win is confusing because it was unexpected (S2 *versus* G1).

Careers are path-dependent (Bernhardt, Morris, Handcock & Scott 2001) and the mean-variance rule also shows many features of path dependency: There is an absorbing state (unemployment[31]); decisions depend on previous decisions and are self-reinforcing; due to chance, the best option might never be identified.

3.4.2.2 *Empirical validity of the mean-variance rule: A Monte Carlo study*

Experimental evidence on Bernoulli bandits is rather scarce. Most of the studies find deviations from optimal behavior (e.g., Meyer & Shi 1995; Anderson 2001). There are two particularly interesting findings provided by Steyvers, Lee & Wagenmakers (2009) and Gans, Knox & Croson (2007).

Steyvers et al. (2009) calculate different behavioral characteristics of individuals playing with Bernoulli bandits. A characteristic is, for instance, the number of times individuals selected arms with fewer success and nonsuccesses than alternative arms, which can be interpreted as exploratory behavior. Steyvers et al. (2009) could not establish significant correlations between characteristics and personality. Appendix 3.B.1 successfully replicates this finding in a simulated experiment using the setting of Steyvers et al. (2009)

[31] Verbruggen, van Emmerik, Van Gils, Meng & de Grip (2015) find that underemployment has negative long-term effects on career success. A similar effect is present in my model, as an individual selecting unemployment effectively limits her future set of options.

and the mean-variance rule. Interestingly enough, the result mostly holds if the share of ambiguity-averse individuals is sufficiently large in the group under examination.

Gans et al. (2007) apply different decision rules to data generated by the decisions of individuals who played with Bernoulli bandits. They establish a ranking of the explanatory performance of different rules. The ranking (of three rules differing in their basic assumptions) is successfully replicated in Appendix 3.B.2 using decision data generated by the mean-variance rule. It is only possible to replicate the ranking if ambiguity aversion is sufficiently pronounced. Thus, the mean-variance rule with ambiguity aversion is supported by available experimental evidence on Bernoulli bandits.

In summary, two decision rules are considered. The first rule is individually optimal, whereas the second rule allows for the influence of preferences toward ambiguity. The latter rule is more realistic than the optimal rule. However, the optimal rule is a proper benchmark for decision making and, especially, the efficiency of learning.

3.5 Analysis of career trajectories

Having introduced the model's setup, an individually optimal decision rule, and a decision rule accounting for ambiguity preferences, this section examines different aspects of career choices, in this way, presenting the main findings of the chapter. There are four interesting decision aspects deserving a closer examination. First, as unemployment has special features, compared to wage work and entrepreneurship that are neither unambiguous nor absorbing, the propensity to select unemployment is of interest. Second, as entrepreneurs are usually overconfident, it is interesting how ambiguity aversion impacts the ability to correct false beliefs. Third, the main motivation of the chapter is that career choices are not final decisions. Thus, an interesting aspect is the link between transitions and ambiguity preferences. Fourth, an aspect of interest are sampling frequencies. As ambiguity aversion may result in over- or undersampling from some options and sampling frequencies are observable, the analysis of sampling produces testable predictions.

This section consists of two parts. Propensities to select unemployment and overconfidence are discussed in the first part of the section. Transitions and sampling frequencies are hard to discuss in the full model because the potential number of transitions is infinite. In such a case, a simple approach is to modify the full model to make it suitable to discuss a certain aspect. Thus, in the second part of the section, I add the assumption of *fast learning*, restricting the maximal number of transitions, to discuss transitions and sampling.

A counter argument to such an approach is that a modification might *over*emphasize some aspects that would not play a large role without the modification. For instance, ambiguity aversion may have a large impact on transitions in the modified model but a negligible impact in the full model. For this reason, I simulate a version of the full model in the next section.

3.5.1 *Propensity to select unemployment and the correction of overconfidence*

With respect to the propensity to select unemployment, the following proposition can be obtained:

Proposition 3.1. *Assume that $\mathbb{x}_{\mathfrak{e},0}$ and $\mathbb{x}_{\mathfrak{w},0}$ are generated by a fixed number of draws $\gamma_{\mathfrak{e},0} > 1$ and $\gamma_{\mathfrak{w},0} > 1$ from two independent Bernoulli distributions. Let $\mathcal{S}_t^{[\mathbf{r}]} = \mathbb{P}(d_t^{[\mathbf{r}]} = \mathfrak{u})$ denote the propensity to select the sure option unemployment in some period $t \in \{1, 2, \ldots\}$. Sufficiently high ambiguity aversion leads to a higher than optimal propensity to select unemployment in any period $t \in \{1, 2, \ldots\}$:*

$$\mathcal{S}_t^{[\sigma]}(\lambda) > \mathcal{S}_t^{[\mathbf{g}]}$$

Furthermore, if ambiguity aversion is sufficiently high, no option besides unemployment is ever selected such that

$$\mathcal{S}_t^{[\sigma]}(\lambda) = 1$$

Proof. $d_t^{[\sigma]}$ has a complex distribution depending on prior information $\mathbb{X}_0$ and actual decisions generated by the mean-variance rule such that $\mathcal{S}_t^{[\sigma]}$ cannot be directly computed. However, using S1, it is obvious that the share of choices for unemployment in any period t is bounded from below by the share of choices for unemployment in period 1: If $p_{\mathfrak{u},1} \in [0, 1]$ is the probability to select unemployment in period 1, we must have

$$p_{\mathfrak{u},1} \leqslant \mathbb{P}(d_t^{[\sigma]} = \mathfrak{u}) \leqslant 1$$

as unemployment is absorbing. Therefore, it is enough to show that $p_{\mathfrak{u},1} > \mathbb{P}(d_t^{[\mathbf{g}]} = \mathfrak{u}) = 0$ given sufficiently high ambiguity aversion and $p_{\mathfrak{u},1} = 1$, such that $\mathbb{P}(d_t^{[\sigma]} = \mathfrak{u}) = 1$, given a high level of aversion. For a more detailed argument, see Appendix 3.C.1. ∎

Summarizing the proposition, it can be said that if the level of ambiguity aversion is sufficiently high, ambiguity-averse individuals will select unemployment with a higher than optimal probability and that ambiguity aversion can completely paralyze the individual such that her choice set is restricted to unemployment.

We know that entrepreneurs differ from the general population (Zhao & Seibert 2006; Rauch & Frese 2007). A particularly interesting characteristic of entrepreneurs is overconfidence (Busenitz & Barney 1997; Camerer & Lovallo 1999; Bernardo & Welch 2001;

Koellinger et al. 2007). Griffin & Varey (1996) argue that overconfidence expresses itself in two different ways:

> [Type-1 overconfidence] One type, the most dramatic, is optimistic overconfidence, the tendency to overestimate the likelihood that one's favored outcome will occur. (Griffin & Varey 1996, p. 228)
>
> [Type-2 overconfidence] A second type of overconfidence is the overestimation of one's knowledge (more generally, the overestimation of the validity of one's judgment) … . (Griffin & Varey 1996, p. 228)

Consequently, overconfidence with respect to entrepreneurship can be defined as follows: An individual is said to be overconfident with respect to entrepreneurship if (type 1) she assigns a higher expected success probability to entrepreneurship than the actual probability and (type 2) is less ambiguous about success in entrepreneurship than success in all alternatives besides unemployment.[32]

Overconfidence as such does not automatically lead to an observation like the earnings puzzle. First, individuals must believe that entrepreneurship is the best option available at some point in time, such that they become entrepreneurs, which is apparently true for entrepreneurs earning less in entrepreneurship than in wage work. Second, overestimating the success probability of entrepreneurship does not rule out that entrepreneurship is the best option. Therefore, assume that wage work has a higher success probability than entrepreneurship. This condition obviously holds for entrepreneurs who could earn more as wage workers. Given this constellation, beliefs about the superiority of entrepreneurship are objectively false. Hence, overconfidence with respect to entrepreneurship relevant for the earnings puzzle can be summarized as follows:

Definition 3.2. Overconfidence relevant for the earnings puzzle fulfills the following four conditions. (O1) Individuals overestimate the success probability of entrepreneurship, corresponding to type-1 overconfidence. (O2) Individuals select entrepreneurship over unemployment and wage work with a high probability. (O3) Individuals underestimate the ambiguity associated with successes in entrepreneurship relative to successes in wage work, corresponding to type-2 overconfidence. (O4) For individuals affected by the earnings puzzle, wage work would be the (objectively) better option.

Note that the definition of overconfidence is fully rational. Individuals draws rational, Bayesian-updating based, conclusions from pseudo observations. However, pseudo observations can be misleading, which is what generates overconfidence in the model. Formalizing the conditions is straightforward. O1 can be expressed as $\mu(\mathbb{x}_{e,0}) > \phi_e$, O3 as

[32]The condition of low perceived ambiguity does not have to hold for the general population but only for overconfident individuals.

$\mathsf{v}(\mathbb{x}_{\mathrm{e},0}) < \mathsf{v}(\mathbb{x}_{\mathfrak{w},0})$, and O4 as $\phi_{\mathfrak{w}} > \phi_{\mathrm{e}}$. To induce a high probability to select entrepreneurship, it is enough to assume that the success probability of entrepreneurship is estimated sufficiently high such that O2 can be formalized as follows: For $\mu(\mathbb{x}_{\mathrm{e},0}) - \mu(\mathbb{x}_{\mathfrak{w},0}) = \Delta\mu$, $\Delta\mu > 0$ and $\Delta\mu$ is sufficiently large.

Assume that, driven by overconfidence, as given in Definition 3.2, the individual decides for entrepreneurship in the first period. After the individual selects entrepreneurship, she observes a signal in the form of success observations $\pi_{\mathrm{e},1}, \pi_{\mathrm{e},2}, \ldots$. To make the signal as clear as possible, assume that it consists only of nonsuccesses—such a signal is consistent with O4 in Definition 3.2. The correction of high overconfidence requires a transition to wage work.[33] Proposition 3.2 describes how transitions leading to a potential correction of overconfidence and ambiguity aversion are related:

Proposition 3.2. *Assume that an individual who is overconfident with respect to entrepreneurship (in line with Definition 3.2) receives a clear signal,* $\mathbf{s}_{\mathrm{e},t} = (\pi_{\mathrm{e},1} = 0, \pi_{\mathrm{e},2} = 0, \ldots, \pi_{\mathrm{e},t} = 0)$, *such that she overestimates the success probability of entrepreneurship and that wage work might be superior to entrepreneurship. Furthermore, assume that the following three conditions hold:*

(C1) $\mu(\mathbb{x}_{\mathrm{e},0}) > \mu(\mathbb{x}_{\mathfrak{w},0}) + \delta(1-\delta)^{-1}\sqrt{\mathsf{v}(\mathbb{x}_{\mathfrak{w},0})}$ *(consistent with O2, respectively type-1 overconfidence);*

(C2) $\vartheta_{\mathrm{e}} < \mathsf{v}(\mathbb{x}_{\mathfrak{w},t})$ *for all* t, $\sqrt{\vartheta_{\mathrm{e}}} < (1-\delta)\delta^{-1}\mu(\mathbb{x}_{\mathfrak{w},0})$, *and* $\mu(\mathbb{x}_{\mathfrak{w},0})[\mathsf{v}(\mathbb{x}_{\mathfrak{w},0}) - \vartheta_{\mathrm{e}}](\mathsf{v}(\mathbb{x}_{\mathfrak{w},0})\sqrt{\vartheta_{\mathrm{e}}} + \mu(\mathbb{x}_{\mathfrak{w},0})[\mathsf{v}(\mathbb{x}_{\mathfrak{w},0}) - \vartheta_{\mathrm{e}}])^{-1} > \delta$ *(requires a sufficiently large amount of pseudo observations with respect to entrepreneurship and consistent with O3, respectively type-2 overconfidence), where*

$$\vartheta_{\mathrm{e}} \equiv \max_{k \geqslant 0} \mathsf{v}(\alpha_{\mathrm{e},0}, \beta_{\mathrm{e},0} + k)$$

(C3) $\lambda < 2\mu(\mathbb{x}_{\mathfrak{w},0})\mathsf{v}(\mathbb{x}_{\mathfrak{w},0})^{-1}$ *(such that wage work is superior to unemployment).*

Let $\mathcal{T}_{\mathbf{g}}$ *denote the period where an individual following a Gittins-index strategy experiences the incentive to abandon entrepreneurship and transition to wage work for the first time. Let* $\mathcal{T}_{\sigma}$ *denote the equivalent period for an individual maximizing utility given ambiguity aversion. If ambiguity aversion is sufficiently high such that*

$$\lambda > 2\frac{\delta\sqrt{\max_{k\geqslant 0}\mathsf{v}(\alpha_{\mathrm{e},0}, \beta_{\mathrm{e},0} + k)}}{(1-\delta)[\mathsf{v}(\mathbb{x}_{\mathfrak{w},0}) - \max_{k\geqslant 0}\mathsf{v}(\alpha_{\mathrm{e},0}, \beta_{\mathrm{e},0} + k)]}$$

we have

$$\mathcal{T}_{\sigma} > \mathcal{T}_{\mathbf{g}}$$

[33] Though, one transition might not be enough because the individual might return to the inferior option entrepreneurship. However, false beliefs cannot be corrected without a transition at all.

and $\mathbf{s}_{\mathfrak{e},t}$ *signaling the inferiority of entrepreneurship, compared to wage work, needs longer than optimal to induce a transition to wage work. Furthermore, given two levels of ambiguity aversion* $\lambda', \lambda > 0$*, an individual with higher ambiguity aversion will need longer to transition to entrepreneurship such that*

$$\mathcal{T}_{\sigma}(\lambda') > \mathcal{T}_{\sigma}(\lambda)$$

if $\lambda' > \lambda$*.*

Proof. Note that both rules will transition to wage work, such that $\mathcal{T}_{\mathbf{g}}$ and $\mathcal{T}_{\sigma}$ both exist, as, given signal $\mathbf{s}_{\mathfrak{e},t}$, $\mathbf{g}_{\mathfrak{e},t}$, $\boldsymbol{\sigma}_{\mathfrak{e},t} \to 0$ when $t \to \infty$ (see Appendix 3.A.2) given that $\mathbf{g}_{\mathfrak{w},0}, \boldsymbol{\sigma}_{\mathfrak{w},0} > 0$. To establish $\mathcal{T}_{\sigma} > \mathcal{T}_{\mathbf{g}}$, show that $\mathcal{T}_{\mathbf{g}} \in (0, \mathcal{T}_{\mathbf{g}}^{+}]$, $\mathcal{T}_{\sigma} \in [\mathcal{T}_{\sigma}^{-}, \infty)$, and $\mathcal{T}_{\sigma}^{-} > \mathcal{T}_{\mathbf{g}}^{+}$ given some $\lambda > 0$ consistent with C3. The second part of the proposition is, then, straightforward to demonstrate. For detailed descriptions of proofs for both parts, see Appendix 3.C.2. ■

The conditions C1, C2, and C3 are relatively weak as they only require that overconfidence (of both types) is sufficiently pronounced and that wage work is considered to be superior to unemployment. In the proposition, there are two important insights. First, individuals with sufficiently high ambiguity aversion will correct their false beliefs less efficiently than optimally-behaving individuals. Second, higher ambiguity aversion reduces the efficiency of the correction of false beliefs.

The general logic behind the latter result relies on type-2 overconfidence and is as follows. Assume that we have two individuals, $\mathfrak{A}$ and $\mathfrak{B}$, with exactly the same initial prior observations but one individual, say $\mathfrak{B}$, is more ambiguity-averse than the other. If both individuals observe signal $\mathbf{s}_{\mathfrak{e},t}$, their expected success probabilities will decrease by exactly the same amount. So, type-1 overconfidence is reduced by the same degree. However, due to type-2 overconfidence, the more ambiguity-averse individual $\mathfrak{B}$ will transition to wage work later than $\mathfrak{A}$. The reason behind is that, given type-2 overconfidence leading to lower ambiguity in entrepreneurship than in wage work, the more ambiguity-averse individual $\mathfrak{B}$ dislikes the fact that wage-work success is relatively more ambiguous than entrepreneurial success more than $\mathfrak{A}$. Consequently, a larger number of nonsuccesses in entrepreneurship is required to convince the more ambiguity-averse individual $\mathfrak{B}$ to transition to wage work than individual $\mathfrak{A}$.

Also, Proposition 3.2 delivers a simple explanation for the earnings puzzle, the permanent selection of entrepreneurship although wage work is the better option. As high ambiguity aversion is empirically likely (Figure 3.2), individuals with a sufficiently high level of ambiguity aversion may need an exceptionally long period of time to correct high overconfidence with respect to entrepreneurship. Given this trapped-by-overconfidence effect, observing some finite time interval of an individual's career can lead to the conclusion that she permanently selects an non-optimal option. Theoretically, even if ambiguity aversion is high, a correction may occur at some point. However, the time period needed for a correction can be so long that the observation of a reasonably long interval might not be enough to record the correction event.

3.5.2 Transitions and sampling

The analysis of transitions and sampling requires the following assumption of fast learning:[34]

Definition 3.3. In an environment with fast learning, an individual learns if she is successful in an occupation one period after selecting the occupation for the first time.

Using Definition 3.3, there is still ambiguity about success probabilities, $\mathsf{v}_i > 0$ for all $i \in \mathbb{O}_{-\mathfrak{u}}$, until an option is selected once. However, an option selected for the second time is no longer ambiguous. The individual is successful in occupation $i \in \mathbb{O}_{-\mathfrak{u}}$, and receives $\pi_i = 1$ in all periods the occupation is selected, with probability $\phi_i \in (0,1)$, and the individual is not successful, and receives $\pi_i = 0$, with probability $1 - \phi_i$. The expected success probability is $\mu_i \in (0,1)$ for all $i \in \mathbb{O}_{-\mathfrak{u}}$, where it is assumed that $\mu_{\mathfrak{e}} \neq \mu_{\mathfrak{w}}$. The number of prior observations is γ_i for all $i \in \mathbb{O}_{-\mathfrak{u}}$. μ_i, γ_i, and v_i are computed on the basis of prior observations $\mathbb{x}_i$, where $\mathsf{v}_{\mathfrak{w}} \neq \mathsf{v}_{\mathfrak{e}}$ and $\mu_{\mathfrak{w}}/\mathsf{v}_{\mathfrak{w}} \neq \mu_{\mathfrak{e}}/\mathsf{v}_{\mathfrak{e}}$. Furthermore, I assume that it is commonly known that an option selected twice must be selected in all consecutive periods, i.e., an option selected twice is absorbing.

Given this particular setting, the number of decision sequences is restricted, thus it is possible to determine when and why ambiguity-averse gamblers do not exhibit optimal transition and sampling behavior. First, notice the following:

Lemma 3.2. *Let $\mathbf{g}_{i,h_i}$ denote the Gittins index of option $i \in \mathbb{O}$ selected h_i times. Given Definition 3.3, we obtain the following Gittins indices:*

$$\mathbf{g}_{\mathfrak{u},h_{\mathfrak{u}}} = 0 \qquad \textit{for all } h_{\mathfrak{u}} \tag{3.7}$$

$$\mathbf{g}_{i,0} = \frac{\mu_i}{\delta(\mu_i - 1) + 1} \qquad \textit{for all } i \in \mathbb{O}_{-\mathfrak{u}} \tag{3.8}$$

For the second time an option $i \in \mathbb{O}_{-\mathfrak{u}}$ is selected, we obtain

$$\mathbf{g}_{i,h_i} = \pi_i \in \{0,1\} \qquad \textit{for } h_i > 0 \tag{3.9}$$

Proof. Equation (3.7) follows because unemployment is not ambiguous and the unemployment reward is always zero. Equation (3.9) is a consequence of the assumption of

[34] The approach is inspired by Berninghaus & Seifert-Vogt (1987), which uses a similar method in a migration model.

fast learning, as an option selected once is no longer ambiguous and the gambler receives $\pi_i = 1$ or $\pi_i = 0$ for all remaining periods.

To derive (3.8), use the optimal-stopping formulation of the Gittins index (Equation [3.4]). First, note that it does not make sense to stop before observing the reward π_i at least once. After observing the reward, we either continue to play option i if $\pi_i = 1$ and obtain $\sum_{t=1}^{\infty} \delta^{t-1}\pi_i = 1/(1-\delta)$, which, from the point of view of the gambler, occurs with probability μ_i, or we stop after observing the reward once if $\pi_i = 0$ and obtain zero, which, from the perspective of the gambler, occurs with probability $1 - \mu_i$. By dividing by the expected sum of discounted time, we get

$$\mathbf{g}_{i,0} = \frac{\mu_i(1-\delta)^{-1}}{\mu_i(1-\delta)^{-1} + 1 - \mu_i} \qquad \text{for all } i \in \mathbb{O}_{-\mathfrak{u}}$$

which corresponds to (3.8).

■

In contrast to the model without fast learning, the Gittins index of an initially ambiguous option can become zero, as an option selected twice is no longer ambiguous. Consequently, a gambler following the Gittins rule might at some point become indifferent between unemployment and all available non-unemployment options. Let us assume that in such a case the gambler will select unemployment, as unemployment is a minimal-effort option. Furthermore, we have:

Lemma 3.3. *A Gittins-index strategy is optimal in the setting given by Definition 3.3, conditional on prior information $\mathbb{X}_0$.*

Proof. Let OSFL denote the optimal strategy given the assumption of fast learning. OSFL prescribes the following choices: Let $i^{(1)} = \arg\max_{i \in \mathbb{O}_{-\mathfrak{u}}} \mu_i$. In the first period, select the option with the largest expected success probability $\mu_{i^{(1)}} = \max\{\mu_{\mathfrak{e}}, \mu_{\mathfrak{w}}\}$. Learn everything about the reward of $i^{(1)}$. If $\pi_{i^{(1)}} = 1$, continue to select option $i^{(1)}$ forever and obtain $\delta/(1-\delta)$, which is the best possible outcome. If $\pi_{i^{(1)}} = 0$, switch to the option with the second-largest expected success probability $i^{(2)} = \mathbb{O}_{-\mathfrak{u}} \setminus i^{(1)}$ and observe the reward of $i^{(2)}$. If $\pi_{i^{(2)}} = 1$, continue to play $i^{(2)}$ and obtain $\delta^2/(1-\delta)$, which is the second-best possible outcome. If $\pi_{i^{(2)}} = 0$, all options have the same reward of zero such that it does not matter which option is selected. OSFL yields

$$\mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1}\Pi_t | \mathbb{x}_{\mathfrak{e}}, \mathbb{x}_{\mathfrak{w}}\right] = \frac{\delta\mu_{i^{(1)}}}{1-\delta} + \frac{\delta^2(1-\mu_{i^{(1)}})\mu_{i^{(2)}}}{1-\delta}$$

which cannot be improved by any other deterministic strategy.[35]

[35]To see, note, first, that it does not make sense to start with option $i^{(2)}$ or $\mathfrak{u}$, as both yield less than $\delta\mu_{i^{(1)}}/(1-$

The Gittins index prescribes a strategy equivalent to OSFL: In the first period, we have $\mathbf{g}_{i,0} > \mathbf{g}_{u,h_u} = 0$ for all $i \in \mathbb{O}_{-u}$ and all h_u; and $\mathbf{g}_{i,0} > \mathbf{g}_{j,0}$ if $\mu_i > \mu_j$ or $\mathbf{g}_{j,0} > \mathbf{g}_{i,0}$ if $\mu_j > \mu_i$ for $i, j \in \mathbb{O}_{-u}$ where $i \neq j$. Hence, $i^{(1)}$ is selected. In the second period, we have $\mathbf{g}_{i,1} > \mathbf{g}_{j,0}$ for $i, j \in \mathbb{O}_{-u}$ if $\pi_i = 1$ and $\mathbf{g}_{i,1} < \mathbf{g}_{j,0}$ if $\pi_i = 0$ for $i, j \in \mathbb{O}_{-u}$ where $i \neq j$. Hence, option $i^{(1)}$ is selected if $\pi_{i^{(1)}} = 1$ and $i^{(2)}$ is selected if $\pi_{i^{(1)}} = 0$, and so forth. ■

The mean-variance rule sets the following indices:

$$\boldsymbol{\sigma}_{i,0} = \mu_i - \frac{\lambda}{2}\mathrm{v}_i \qquad \text{for all } i \in \mathbb{O}_{-u} \tag{3.10}$$

$$\boldsymbol{\sigma}_{i,h_i} = \pi_i \in \{0, 1\} \qquad \text{for all } h_i > 0 \tag{3.11}$$

for a non-unemployment option $i \in \mathbb{O}_{-u}$ and

$$\boldsymbol{\sigma}_{u,h_u} = 0 \qquad \text{for all } h_u \tag{3.12}$$

for unemployment.

Using Lemma 3.2 and mean-variance-rule indices (3.10), (3.11), and (3.12), we can derive all plausible decision sequences. Table 3.1 compiles all decision sequences generated by the optimal or the mean-variance rule—or both. For instance, sequence DS4 should be read as follows. In the first period, the gambler selects wage work. In the second period, the gambler transitions to entrepreneurship. In the third period, the gambler again selects entrepreneurship. As an option selected twice must be selected in all consecutive periods, entrepreneurship is selected in all periods $t > 3$. There are only 9 plausible sequences. Some of the sequences, like DS1 and DS6, are produced by the mean-variance rule only.

Fast learning allows for the derivation of two central results on the relation between ambiguity aversion and transition intensity, and on the relation between ambiguity aversion and sampling frequencies. The first result is that ambiguity aversion leads to suboptimal transition intensity:

Proposition 3.3. *Let $E_t^{[\mathbf{r}]} = \{d_t^{[\mathbf{r}]} \neq d_{t-1}^{[\mathbf{r}]}\}$ for $t > 1$ denote a transition event generated by rule $\mathbf{r}$. The overall number of actual transitions given rule $\mathbf{r}$ takes a value in the finite set of possible transition numbers $\mathcal{M}^{[\mathbf{r}]}$ such that*

$$\sum_{t=2}^{\infty} \mathbb{1}\{E_t^{[\mathbf{r}]}\} \in \mathcal{M}^{[\mathbf{r}]} \subset \{0, 1, 2\}$$

δ). If $\pi_{i^{(1)}} = 0$, it does not make sense to switch to option u or to continue with $i^{(1)}$, as both actions yield zero. Consequently, the gambler should switch to $i^{(2)}$, as $\mu_{i^{(2)}}\delta^2/(1-\delta) > 0$. If $\pi_{i^{(1)}} = 1$, it does not make sense to switch from $i^{(1)}$, as $i^{(1)}$ yields $\delta/(1-\delta)$, while $i^{(2)}$ and u both generate a smaller expected reward—$i^{(2)}$ generates $\mu_{i^{(2)}}\delta/(1-\delta) < \delta/(1-\delta)$ (as $\mu_{i^{(2)}} < 1$) and u generates zero.

TABLE 3.1. All decision sequences (DS) that are plausible given fast learning

	Decision sequence	Number of transitions	Generated by which rule
(DS1)	$d_1 = \mathfrak{u} \to d_2 = \mathfrak{u} \to \dots$	0	$\boldsymbol{\sigma}$ but not $\mathbf{g}$
(DS2)	$d_1 = \mathfrak{w} \to d_2 = \mathfrak{w} \to \dots$	0	$\boldsymbol{\sigma}$ and $\mathbf{g}$
(DS3)	$d_1 = \mathfrak{e} \to d_2 = \mathfrak{e} \to \dots$	0	$\boldsymbol{\sigma}$ and $\mathbf{g}$
(DS4)	$d_1 = \mathfrak{w} \to d_2 = \mathfrak{e} \to d_3 = \mathfrak{e} \to \dots$	1	$\boldsymbol{\sigma}$ and $\mathbf{g}$
(DS5)	$d_1 = \mathfrak{e} \to d_2 = \mathfrak{w} \to d_3 = \mathfrak{w} \to \dots$	1	$\boldsymbol{\sigma}$ and $\mathbf{g}$
(DS6)	$d_1 = \mathfrak{w} \to d_2 = \mathfrak{u} \to d_3 = \mathfrak{u} \to \dots$	1	$\boldsymbol{\sigma}$ but not $\mathbf{g}$
(DS7)	$d_1 = \mathfrak{e} \to d_2 = \mathfrak{u} \to d_3 = \mathfrak{u} \to \dots$	1	$\boldsymbol{\sigma}$ but not $\mathbf{g}$
(DS8)	$d_1 = \mathfrak{w} \to d_2 = \mathfrak{e} \to d_3 = \mathfrak{u} \to d_4 = \mathfrak{u} \to \dots$	2	$\boldsymbol{\sigma}$ and $\mathbf{g}$
(DS9)	$d_1 = \mathfrak{e} \to d_2 = \mathfrak{w} \to d_3 = \mathfrak{u} \to d_4 = \mathfrak{u} \to \dots$	2	$\boldsymbol{\sigma}$ and $\mathbf{g}$

Note: The interpretation of, for instance, DS9 is that entrepreneurship is selected in the first period, wage work is selected in the second period, unemployment is selected in the third period, and unemployment is selected in the fourth period. If an option is selected for the second time (e.g., unemployment in the forth period in DS9), it must be selected forever.

where $\sum_{t=2}^{\infty} \mathbb{1}\{E_t^{[\mathbf{r}]}\} < \infty$ *as a consequence of fast learning. Let* $\mathcal{M}^{[\mathbf{g}]}$ *denote the set of transition numbers generated by the Gittins index and* $\mathcal{M}_\lambda^{[\boldsymbol{\sigma}]}$ *the set generated by an individual maximizing ambiguity-aversion-affected utility with aversion level* λ*. There exists a sufficiently high level of ambiguity aversion such that the ambiguity-averse individual transitions less than optimal or*

$$\max \mathcal{M}_\lambda^{[\boldsymbol{\sigma}]} < \max \mathcal{M}^{[\mathbf{g}]}$$

Furthermore, an increase in ambiguity aversion can further restrict transitions such that

$$\max \mathcal{M}_{\lambda'}^{[\boldsymbol{\sigma}]} < \max \mathcal{M}_\lambda^{[\boldsymbol{\sigma}]}$$

where λ *is such that* $\max \mathcal{M}_\lambda^{[\boldsymbol{\sigma}]} > 0$ *and* λ' *is sufficiently larger than* λ*.*

Proof. The proposition can be proven with the aid of Table 3.1 by considering which decision sequences will be generated given different level of ambiguity aversion. For the full proof, see Appendix 3.C.3. ∎

The established link between ambiguity aversion and transition intensity—higher ambiguity aversion results in less transitions—is plausible as empirically observable ambiguity aversion is rather high (Figure 3.2) and empirically observable transition intensity is low (Figure 3.1b).

Secondly, it can be demonstrated that ambiguity aversion leads to oversampling from less ambiguous options:

Proposition 3.4. *Let* $\mathcal{F}_i^{[\mathbf{r}]} = \sum_{t=1}^{\infty} \mathbb{1}\{d_t^{[\mathbf{r}]} = i\}$ *denote the number of times rule* $\mathbf{r}$ *samples from option* $i \in \mathbb{O}$*, the sampling frequency of the option. Given prior information* $\mathbb{x}_{\mathfrak{e}}$ *and* $\mathbb{x}_{\mathfrak{w}}$*, and without loss of generality, assume that entrepreneurship is the most ambiguous option, such that* $\mathsf{v}_{\mathfrak{e}} > \mathsf{v}_{\mathfrak{w}}$ *and* $\mu_{\mathfrak{e}}\mathsf{v}_{\mathfrak{e}}^{-1} < \mu_{\mathfrak{w}}\mathsf{v}_{\mathfrak{w}}^{-1}$*. Assume that either* $\mu_{\mathfrak{e}} > \mu_{\mathfrak{w}}$ *or* $\mu_{\mathfrak{w}} > \mu_{\mathfrak{e}}$*, where both events occur with a positive probability. Let* $\mathcal{F}_i^{[\sigma]}(\lambda)$ *denote the number of times an ambiguity-averse gambler samples from option* $i \in \mathbb{O}$ *given ambiguity-aversion level* λ*. Given a sufficiently high level of ambiguity aversion,* λ*, the ambiguity-averse gambler oversamples with a positive probability from options with less ambiguity than entrepreneurship such that*

$$\mathbb{P}(\mathcal{F}_{\mathfrak{w}}^{[\sigma]}(\lambda) > \mathcal{F}_{\mathfrak{w}}^{[\mathbf{g}]} \vee \mathcal{F}_{\mathfrak{u}}^{[\sigma]}(\lambda) > \mathcal{F}_{\mathfrak{u}}^{[\mathbf{g}]}) > 0$$

Proof. As we have the initial relation $\mu_{\mathfrak{e}} > \mu_{\mathfrak{w}}$ or $\mu_{\mathfrak{w}} > \mu_{\mathfrak{e}}$, it is straightforward to derive sampling distributions of the optimal rule (with the aid of Table 3.1) for the two different unique initial relations. To obtain the result, compare optimal sampling distributions to distributions generated by the mean-variance rule conditional on different levels of ambiguity aversion. For the full argument, see Appendix 3.C.4. ∎

The results provided in Proposition 3.1, 3.2, 3.3, and 3.4 do not rely on myopia, which is shown in Appendix 3.C.5. Moreover, myopic and global mean-variance preferences do not necessarily contradict in an environment with fast learning and relatively high ambiguity aversion, which is relevant for Proposition 3.3 and 3.4, and is demonstrated in Appendix 3.C.5.3.

3.6 Illustration of mechanisms and empirical evidence

This first part of this section consists of a simulation, which is, first of all, used to illustrate the model's mechanisms. However, it is also used to verify that results derived in the model with fast learning (see Definition 3.3) also apply to the full model. The second part of the section briefly discusses empirical evidence on model predictions.

The simulation requires the assumption of unexpected termination:

Definition 3.4. A career choice model is said to have the property of unexpected termination if it stops after a fixed number of $\mathsf{T} \in \mathbb{N}$ periods, while gamblers assume that they are playing a model with an infinite time horizon, such that they do not anticipate termination time T.

In effect, the model with unexpected termination time corresponds to observing a finite time interval of an infinitely played bandit game. First, I introduce the simulation setup, including two simulation scenarios. Then, choice patterns generated by the model are presented and analyzed.

3.6.1 Simulation setup

To simulate optimal decision behavior, I rely on an approximation of the Gittins index provided by Brezzi & Lai (2002). The simulation proceeds as follows.[36] In the first step, the algorithm generates rewards given some set of parameters $\phi_{\mathfrak{w}}$ and (or) $\phi_{\mathfrak{e}}$. In the second step, the algorithm applies the Gittins-index and the mean-variance rule to data simulated in the first step and produces decision sequences given some set of parameters like ambiguity aversion, λ, priors, $\mathbb{x}_{\mathfrak{e},0}$ and $\mathbb{x}_{\mathfrak{w},0}$, etc. For each combination of parameters, the two steps are sequentially repeated 10,000 times.

I employ two distinct scenarios to examine different aspects of career choices in an ambiguous decision environment. The *first scenario* examines the impact of ambiguity aversion on transition intensity (Proposition 3.3) and sampling frequencies (Proposition 3.4). The *second scenario* analyzes the effect of ambiguity aversion on the correction of overconfidence with respect to entrepreneurship (Proposition 3.2).

TABLE 3.2. Simulation scenarios

Parameter	Relation	Assumption
(a) First scenario: Transitions and sampling		
T	$=$	100
δ	$=$	0.97
λ	$\in$	$\{1, 8, 15, \ldots, 50, 57, 64\}$
$\phi_{\mathfrak{e}}, \phi_{\mathfrak{w}}$	$\in$	$\{0.1, 0.3, 0.5, 0.7, 0.9\}$
$\alpha_{\mathfrak{e},0}, \beta_{\mathfrak{e},0}, \alpha_{\mathfrak{w},0}, \beta_{\mathfrak{w},0}$	$\in$	$\{5, 6\}$
(b) Second scenario: Correction of overconfidence		
δ	$=$	0.97
λ	$\in$	$\{0.1, 1.1, 2.1, \ldots, 7.1, 8.1, 9.1\}$
$\phi_{\mathfrak{e}}$	$\in$	$\{0.05, 0.10, 0.15\}$
$(\alpha_{\mathfrak{e},0}, \beta_{\mathfrak{e},0})$	$=$	$(10, 5)$
$(\alpha_{\mathfrak{w},0}, \beta_{\mathfrak{w},0})$	$=$	$(3, 6)$

The first scenario is interesting because it illustrates the effects of ambiguity aversion on career choice probabilities and career dynamics (transitions). Table 3.2a summarizes all assumption on model parameters that are necessary to examine these effects. In the first scenario, after $\mathsf{T} = 100$ periods gamblers are forced to retire. Proposition 3.4 assumes that one option is more ambiguous than all alternative options. By symmetry of assumptions in Table 3.2a, we can assign the highest level of ambiguity to either entrepreneurship or wage work without changing results. When examining sampling frequencies, it is assumed that entrepreneurship is the most ambiguous option.

[36]The simulation code is available upon request from the author.

The second scenario illustrates the problem of the correction of high overconfidence with respect to entrepreneurship. Numerical assumptions are presented in Table 3.2b. Given assumptions in Table 3.2b, entrepreneurship is selected by the Gittins index and the mean-variance rule in the first period. As the only variable of interest is the earliest period where individuals abandon entrepreneurship, there is no fixed termination period. Parameter combinations are consistent with overconfidence with respect to entrepreneurship, as stated in Definition 3.2.

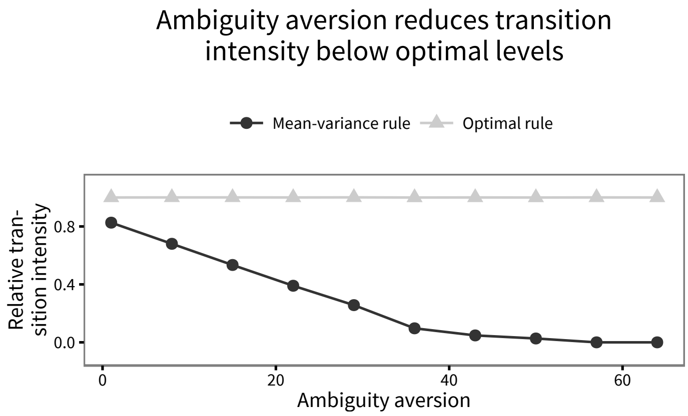
It is, in fact, rather difficult to provide a simulation scenario where the mean-variance rule abandons entrepreneurship. As the mean-variance index of entrepreneurship converges to the true probability to succeed in entrepreneurship (see Appendix 3.A.2), we have to impose $\phi_{\mathfrak{e}} < \boldsymbol{\sigma}(\mathbb{x}_{\mathfrak{w},0})$. Otherwise, the probability of a mean-variance-induced correction of high overconfidence given a reasonably long observation period is small.

Lastly, note that the signal used in the second scenario, where each signal component is $\pi_{\mathfrak{e},t} = 1$ with probability $\phi_{\mathfrak{e}} > 0$ or $\pi_{\mathfrak{e},t} = 0$ with probability $1-\phi_{\mathfrak{e}} > 0$, is more inconclusive, and, therefore, more realistic than the complete failure signal used in Proposition 3.2.

3.6.2 *Simulated decisions*

3.6.2.1 *First scenario: Transitions and sampling*

Figure 3.3 depicts simulation averages of transition intensity, i.e., the number of transitions divided by the maximal number, normalized by average optimal levels, conditional on different levels of ambiguity aversion. The figure reveals that ambiguity-averse indi-

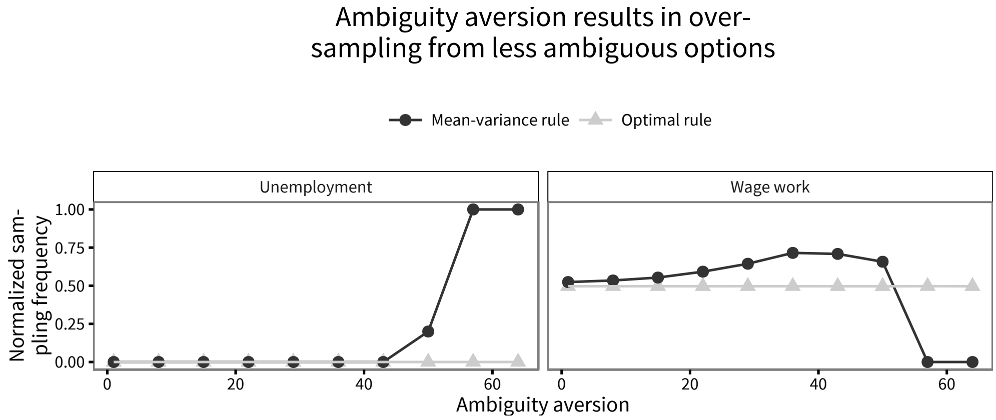


Note: The figure presents simulation *averages* of transition intensity normalized by the optimal average level of transition intensity.

FIGURE 3.3. Transition intensity (first simulated scenario)

viduals transition less than optimally and that the gap between optimal and ambiguity-aversion-affected number of transitions increases if ambiguity aversion increases. Very high ambiguity aversion results in no transitions at all. As the same result, Proposition

3.3, is obtained under fast learning, it can be concluded that ambiguity-aversion effects are consistent across all model versions discussed. The relations in Figure 3.3 are consis-



Note: The figure presents simulation *averages* of sampling frequencies. Frequencies are normalized by the overall number of observation, i.e., T = 100. It is assumed that entrepreneurship is the most ambiguous option. Consequently, the figure presents sampling frequencies for wage work and unemployment, as these options are less ambiguous than entrepreneurship.

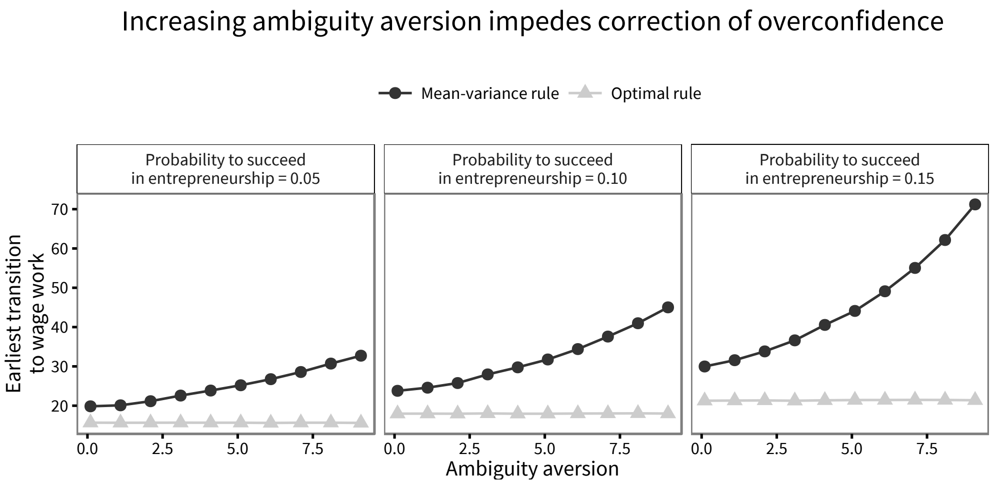
FIGURE 3.4. Oversampling from less ambiguous options (first simulated scenario)

tent with observations on actual transitions (Figure 3.1b) and actual levels of ambiguity aversion (Figure 3.2). In reality, high ambiguity aversion goes along with a small number of transitions. In the model, increasing ambiguity aversion reduces the number of transitions such that the aforementioned empirical observation is reproduced.

In Figure 3.4, it is assumed that entrepreneurship is the most ambiguous option. To examine sampling from less ambiguous options, the figure depicts simulation averages of sampling frequencies, normalized by the overall number of observations, of unemployment and wage work. The figure shows that high ambiguity aversion leads to oversampling from unemployment and wage work—sufficiently high aversion results in sampling from unemployment only. Lower ambiguity aversion does not result in oversampling from unemployment but oversampling from wage work is still present. However, oversampling almost disappears if the level of ambiguity aversion is very small, such that individuals are almost ambiguity-neutral. The results are in line with Proposition 3.4 derived under fast learning.

3.6.2.2 Second scenario: Correcting overconfidence

Figure 3.5 shows simulation averages of the earliest transition period (transitions from entrepreneurship to wage work), given that gamblers are overconfident with respect to entrepreneurship in early periods of their careers. In the figure, it is demonstrated that ambiguity-averse individuals transition later than optimal to wage work and that increasing ambiguity aversion delays the transition period, as exemplified in Proposition 3.2.



Note: The figure presents simulation averages of the earliest transition period (from entrepreneurship to wage work).

FIGURE 3.5. Correction of false beliefs (second simulated scenario)

An additional result is that ambiguity-aversion-affected transition time is increasing in the "vagueness" of the signal. To see, note that, given prior information, the gambler rather expects a success than a nonsuccess. If we consider a setting with a success probability of entrepreneurship of 0.05, the signal the gambler receives is rather clear as in most periods entrepreneurship generates a nonsuccess. In this setting, an ambiguity-averse individual maximally needs less than two times longer than optimal to correct overconfidence. If we make the signal slightly more inconclusive by increasing the success probability of entrepreneurship from 0.05 to 0.15, an ambiguity-averse individual maximally needs more than 3.3 times longer than optimal to correct overconfidence with respect to entrepreneurship.

Thus, a smaller overconfidence bias can be *harder* to correct than a larger one. In general, a non-optimal correction of overconfidence leading to earnings-puzzle-like observations is a likely scenario if entrepreneurs are overconfident and ambiguity-averse.

In Appendix 3.D, I show that risk aversion produces transition patterns differing from ambiguity aversion. Risk aversion is unlikely to produce the earnings puzzle because given entrepreneurial overconfidence it does not sufficiently impact the transition time to wage work.

To summarize, an increase in ambiguity aversion induces an increase in the propensity to select a relatively less ambiguous option and impedes the correction of false beliefs. Furthermore, the optimal rule and the mean-variance rule with a sufficiently low level of ambiguity aversion produce transitions between occupations across all model versions, reproducing the empirical observation that many individuals have mixed careers consisting of entrepreneurship and wage work spells (e.g., Burke et al. 2008). In the bandit model, high ambiguity aversion is associated with a small number of transitions between occupations, which is consistent with empirical observations. Finally, as the capacity to correct false beliefs is limited by ambiguity aversion and entrepreneurs tend to be overconfident (e.g., Busenitz & Barney 1997; Camerer & Lovallo 1999; Bernardo & Welch 2001; Koellinger et al. 2007), they have too positive beliefs with respect to entrepreneurial success, ambiguity aversion provides an explanation for the earnings puzzle.

3.6.3 Evidence on model predictions

The bandit model generates a number of testable predictions. Two especially interesting predictions include, first, that higher ambiguity aversion will result in a higher probability to select a "sure" option; and, secondly, that higher ambiguity aversion will reduce the learning efficiency of entrepreneurs. Note that the two predictions hold at different levels: The first prediction applies to general career choices, while the second prediction is specific for entrepreneurs.

As entrepreneurship is usually more ambiguous than wage work (see, e.g., Sullivan & Peterson 1991) and unemployment because information about it is scarce (Kelley, Brush, Greene & Litovsky 2013), a testable version of the first prediction is that, compared to a low level of ambiguity aversion, a higher aversion level is associated with a lower probability to select entrepreneurship because the propensity to select an option with low ambiguity increases. In Appendix 3.E.1, the prediction is tested using a clustering model (Banfield & Raftery 1993; Fraley & Raftery 2007), data from the Global Entrepreneurship Monitor (GEM 2015), and research of Hofstede et al. (2010) on cultural dimensions. I find support for the prediction in data (Appendix 3.E.1).

A testable version of the second prediction is that higher ambiguity aversion results in *higher* entrepreneurial survival rates, as increasing ambiguity aversion reduces learning efficiency and the probability to transition from one occupation to another. The second prediction is especially interesting because the underlying effect is also responsible for the earnings puzzle. The prediction is tested in Appendix 3.E.2. On the basis of data from the OECD (2015), Hofstede et al. (2010), and the World Bank (2015), I estimate effects of ambiguity aversion on entrepreneurial survival rates using a beta regression[37] (Ferrari & Cribari-Neto 2004; Kosmidis & Firth 2010) and two-stage-least-squares models. The beta regression is employed to account for the fact that the dependent variable is a share. In the two-stage-least-squares models, a logit transformation is applied to the dependent variable and ambiguity aversion is instrumented by (i) a variable indicating whether the religious majority in the country is Roman Catholic[38] or (ii) the exposure to natural disasters (Guha-Sapir, Below & Hoyois 2016). Both instruments have sufficient strength but

[37] For further studies employing the relatively new beta regression approach, see, e.g., De Paola, Scoppa & Lombardo (2010) and Buntaine (2011).
[38] Roman Catholic countries tend to be more ambiguity averse (Huang 2008).

there is no indication for an endogeneity problem. Support for the second prediction can be established in all regression models (Appendix 3.E.2).

However, there is a problem with data availability. The only established measure of ambiguity aversion that I am aware of is Hofstede's (2010) index, which is discussed in Section 3.2 (also, see Appendix 3.F). Hofstede's index is available for a number of countries but the index does not vary over time. Thus, relying on Hofstede's index has the downside that one can only exploit the cross-country dimension where data availability is significantly restricted.

3.7 Summary and conclusion

Ambiguity is an unavoidable feature of daily life. Career decisions are also confronted with ambiguous rewards. Yet, practically all approaches modeling career choices in the literature discuss choices under the assumption that choice outcomes are either deterministic or risky. The aim of this chapter is to take a renewed perspective on career choices by constructing a model where, in contrast to existing literature, it is assumed that some choice outcomes are ambiguous—the distributions of rewards in entrepreneurship and wage work are unknown.

This chapter provides a tractable and flexible formulation of a Bayesian multi-armed bandit model with three arms representing the options wage work, entrepreneurship, and unemployment. Decisions are derived from either an optimal rule or a decision rule accounting for ambiguity aversion; the rule accounting for ambiguity aversion is the well-established mean-variance preferences model and is consistent with available experimental results on the behavior of actual individuals playing bandits. The model generates career trajectories ranging from pure wage workers and entrepreneurs to individuals mixing spells in wage work, entrepreneurship, and unemployment. As mixed careers, which are not generated by most contemporary career choice models, are common and likely to become more important as traditional career paths (a worker is employed in one firm for a long period of time) gradually become less frequent, the bandit model is a particularly appropriate approach to analyze career choices.

With respect to effects of ambiguity preferences, I show that sufficiently high ambiguity aversion has a number of negative consequences such as insufficient transitions between occupational options and oversampling from less ambiguous options, where the benchmark is individually optimal behavior. Furthermore, sufficiently high ambiguity aversion introduces inefficiencies into the correction of false beliefs of individuals exhibiting strong overconfidence. In conjunction with the observation that entrepreneurs tend to be overconfident, inefficient learning explains an observation like the earnings puzzle.

Regarding observable decision patterns, the bandit model predicts that high-aversion societies have lower entrepreneurship levels than low-aversion societies and that, under plausible conditions, firm death rates should be higher in low-aversion than in high-aversion societies, as the former learn more efficiently than the latter. Although there are some restrictions on data availability, existing data supports both predictions.

The model presented in this chapter has some advantages over alternative career choice models. A main advantage of the model is that career trajectories are a natural feature of the dynamic solution of individual career choice problems and that learning efficiency is linked to a simple psychological variable in the form of ambiguity aversion. Still, there are also shortcomings requiring further research. For instance, in the model, I postulate occupation-specific probabilities to succeed but success probabilities should rather not be treated as primitives, making it necessary to introduce lower-level determinants of occupational successes.

4 MEDIA AND OCCUPATIONAL CHOICE

This chapter addresses the question of whether media influences occupational choices. To theoretically examine media effects, we construct a dynamic Bayesian occupational choice model with sequential decisions under ambiguity due to imperfect information. We show that sufficiently intensive positive media attention for entrepreneurship increases the probability of self-employment and decreases the probability of wage work. To test our model, we use an instrumental variable approach to identify causal media effects using US micro data and a country-level macro panel with two different media variables. We find that an increase in positive media articles and reports about entrepreneurs generates effects on choice probabilities that are consistent with our model.

4.1 Introduction

Recent research shows that media has a significant impact on a number of economic, social, and political outcomes. For instance, by blaming persons or institutions for violating certain rules or shaming them for being under-performers, the media is able to play a corporate governance role for publicly traded companies (Zingales 2000; Dyck, Volchkova & Zingales 2008). Media also influences voting behavior (Della Vigna & Kaplan 2007) and voter turnout (Gentzkow 2006). In general, media can shape the image of public figures, institutions, groups, and individuals.

One group of individuals attracting a significant amount of media attention in recent times are entrepreneurs; including reports and TV shows in many countries that celebrate entrepreneurs as "heroes."[39] One clear aim of media shows is formulated by the investor Frank Thelen, judge on the weekly German TV show *Die Höhle der Löwen*:

> This show is a crash course on start-ups for the living room. I would like to show how

[39]For example, in March 2009, the *Economist* presented a special issue on entrepreneurs under the title "Global heroes." Along the same lines, the German newspaper *Frankfurter Allgemeine Sonntagszeitung* published specials on the start-up scene in Berlin (in June 2014 and August 2015) and Tel Aviv (in October 2015). There are also entrepreneurship-related television series such as *How I Made My Millions* (CNBC channel), *CNBC Titans* (CNBC channel), *Dragons' Den* (BBC channel), as well as *Shark Tank* (on the ABC network in the USA) and its German adaptation *Die Höhle der Löwen* (VOX channel), which portray a variety of successful entrepreneurs in different situations. There are many other positive shows and reports about entrepreneurs all around the world.

> start-ups work, thus opening this path for more people. (Frankfurter Allgemeine Zeitung 2017)

Thelen reports of countless letters from viewers of his TV show claiming that the show created impulses to venture own businesses (Frankfurter Allgemeine Zeitung 2017). In this contribution, we ask, therefore, whether media articles and reports about entrepreneurs have an effect on the occupational choices of media consumers. Thus, we are not interested in the effect on those who were placed on the pedestal but in the effects channeled through media consumption. To be more specific, we analyze, from a theoretical and empirical point of view, whether media attention for successful entrepreneurs (e.g., articles, reports, and shows), which is unlikely to change actual probabilities to succeed in self-employment (the distribution of outcomes) but may change beliefs (the distribution of subjective outcome probabilities), influences occupational choice decisions.

To theoretically examine potential effects of positive media attention for entrepreneurs, we construct a dynamic occupational choice model with Bayesian learning. In our model, individuals select an occupation given that they are only imperfectly informed—the outcomes of their choices are subject to ambiguity. We compare optimal choices to choices of ambiguity-averse individuals, where ambiguity aversion may vary across occupational options. We show that, if ambiguity aversion associated with self-employment is higher than aversion linked to wage work, there is a bias against self-employment, in the sense that self-employment is selected with a lower than optimal probability.

By assuming that individuals making occupational choices also use information from the media, which is accessible at negligibly low cost, we integrate positive media reports about entrepreneurship as informational shocks. Based on our model, we derive two predictions. We establish that sufficiently intensive positive media reports about entrepreneurs increase the likelihood to select self-employment, while the probability to select wage work is reduced.

The two predictions are empirically tested with two different data sets. The first data set (we refer to it as the “micro panel”) is based on the US National Health Interview Survey, providing rich individual-level information on occupational status and income, as well as on various demographic and socio-economic characteristics. Media consumption of positive articles and reports about entrepreneurs is approximated by the regional frequency of the search item 'famous entrepreneurs' provided by the Google Trends tool. We use the number of natural disasters, from the International Disaster Database, in non-US regions to introduce an exogenous variation in the media variable. Natural disasters in non-US regions represent natural-experiment-type exogenous shocks that are not related to driving factors of occupational choice in the US but affect media reports, as disasters generate top-priority news. Taking into account potential heteroskedasticity, which, if unaccounted for, leads to inconsistent estimates, effect directions are identified with a heteroskedastic IV probit approach.

The second data set (the "macro panel") is a self-constructed country-level panel using, *inter alia*, data on choice frequencies from the World Bank and data on media reports from the Global Entrepreneurship Monitor (GEM). In the macro panel, media consumption of positive articles and reports about entrepreneurs is approximated by the working-age-population share of individuals noticing frequent reports about successful entrepreneurs, provided by the GEM. Natural disasters in other countries are used to construct a sufficiently strong instrument for the media variable in the second data set. We estimate instrumental variable regressions for the probability of self-employment and wage work. In both data sets, including two different media variables, we are, thus, capable to identify causal effects.

Based on two IV regressions, we find, in support of our hypotheses, that positive media reports about entrepreneurs significantly increase the probability of selecting self-employment and reduce the probability of wage work. Using linear probability models, our macro panel also allows us to approximate effect sizes. We find that a one percentage point increase in the media variable increases the probability of self-employment by 0.5 percentage points and decreases the probability of wage work by 0.4 percentage points. Effect sizes are consistent with previous findings on persuasion effects in the literature.

Overall, we contribute to the existing literature in two ways. First, we develop an occupational choice model operating under ambiguity that allows for a direct assessment of media effects on choice probabilities. Secondly, we provide first empirical evidence for media effects on occupational choices.

The remainder of the chapter is organized as follows. In Section 4.2, we review previous research related to our approach. Section 4.3 presents the theoretical model. In Section 4.4, we analyze the effects of media on occupational choices in the theoretical model. Section 4.5 provides empirical results. In Section 4.6, we summarize and conclude. The Appendix contains proofs, additional information, and supplementary results.

4.2 Previous research

This section provides an overview of two strands of research, which we aim to combine and extend. First, we provide a brief review of the existing empirical evidence on the impact of media on individual and institutional behavior. Second, as we build upon research on decisions under ambiguity, we discuss theoretical concepts related to the so-called multi-armed bandit problem, which is a simple way to model decisions with ambiguous outcomes. We, then, outline our research approach.

4.2.1 Research on media impact

Recent research shows that media significantly impacts the decisions of individuals and institutions. For instance, Dyck et al. (2008) reveal that media shaming of corporate governance violations increases the probability of their reversal in Russia. Della Vigna & Kaplan (2007) establish that the conservative Fox News Channel convinced a substantial share of its non-Republican viewers to vote Republican in US presidential elections between 1996 and 2000. Enikolopov, Petrova & Zhuravskaya (2011) find that media impacts voting behavior in Russia. Gentzkow (2006) establishes a negative effect of television on voter turnout in the United States. Dyck, Moss & Zingales (2008) demonstrate that media coverage of certain topics (such as poverty) can affect the voting behavior of US Congressional Representatives and Senators (also, see Besley & Burgess 2002; Strömberg 2004; Eisensee & Strömberg 2007).

Parallel research shows that media can also significantly influence the decisions of media consumers in many other fields, including areas like attitudes, social norms, consumption, crime, education, family issues, health, or migration. For instance, La Ferrara, Chong & Duryea (2012) find that women seeing more soap operas choose to have less children. Hennighausen (2015) reveals that East Germans who were not exposed to West German TV more often believed that success in life is a matter of luck and not of effort. Less clear effects are found with respect to the exposure of children to TV, in particular the effects on their later educational development.[40]

4.2.2 Modeling decisions under ambiguity with multi-armed bandits

Individuals making occupational choices are assumed to operate in an *ambiguous* environment, where we use the standard definition of ambiguity caused by imperfect information (Fellner 1961; Frisch & Baron 1988; Camerer & Weber 1992):

> Ambiguity is uncertainty about probability, created by missing information that is relevant and could be known. (Camerer & Weber 1992, p. 330)

As an alternative, the literature also uses the notion of 'uncertainty' (Camerer & Weber 1992). According to the definition above, uncertainty is synonymous with an ambiguous choice environment.[41]

Occupational choices are characterized by ambiguity of choice outcomes. For instance, the launch of a new product by an entrepreneur is usually associated with ambiguity regarding market reaction. Consequently, theoretical and empirical models in

[40] For an excellent overview of positive and negative effects of media on educational outcomes and further results in other areas, see Della Vigna & La Ferrara (2015).

[41] Note that this statement does not hold for Knightian uncertainty (Knight 1921), which is immeasurable in principle.

the literature, such as Jovanovic (1979), MacDonald (1988), Hintermaier & Steinberger (2002), Vereshchagina & Hopenhayn (2009), Campanale (2010), Poschke (2013), and Manso (2016), treat starting a business as an experiment with an unknown outcome. The dynamics of wage growth are also consistent with the assumption that information about unknown workers' skills is only gradually revealed to the employer (Antonovics & Golan 2012). Thus, career decisions are not final but rather a process of trial and error, with learning resulting in transitions between occupations.

By allowing to model different types of sequential decisions under ambiguity, the bandit problem, a fairly general framework, accounts for the most prominent features of occupational choices, beyond a deterministic or risky choice environment. The conventional description of the general $\mathfrak{O}$-armed bandit problem is as follows. Assume that, in a casino, there are $\mathfrak{O} \in \mathbb{N}$ one-armed bandits that can be played by a gambler. Pulling one arm results in a reward generated by some distribution that is usually unknown, making the decision environment ambiguous. However, pulling one arm and observing the outcome provides information about the underlying reward distribution, such that the gambler can learn. Reward distributions are usually assumed to be different across the $\mathfrak{O}$ arms but there may exist dependencies between them. Given some time horizon and an objective function (for instance, the expected sum of rewards), the gambler must decide which of the $\mathfrak{O}$ arms to play; how many times to play each arm; and in which order to play them. Gittins et al. (2011) provide an extensive overview on Bayesian multi-armed bandit problems and corresponding problem solutions.

Rothschild (1974) uses a two-armed bandit to analyze equilibrium price distributions given that firms have imperfect knowledge about demand functions and need to experiment to find the profit-maximizing price. Jovanovic (1979) examines employee turnover as a consequence of learning processes. Bergemann & Hege (2005) use bandits to examine best financing rules for research projects with unknown length and success. Antonovics & Golan (2012) investigate career patterns of workers under the condition that skills are unknown. Konon (2016) examines how ambiguity preferences influence occupational choices under ambiguity.

4.2.3 Our research approach

In this contribution, we combine the two strands of research above. To the best of our knowledge, we are the first to theoretically and empirically analyze the impact of media on occupational choices. In order to analyze how media may influence occupational choices, we develop an occupational choice model, where we use a two-armed bandit with ambiguous arm-specific reward distributions, Bayesian learning, and a joint prior distribution for reward probabilities. Media is embedded as informational shocks. Our formulation of the problem builds upon the work of Bradt, Johnson & Karlin (1956) and Konon (2016). We show that in such a setting with ambiguity, even rather unspecific information from the media is theoretically able to affect choices by influencing beliefs about how likely success in self-employment is in general. Put differently, media—for instance, an article about a famous entrepreneur—might not be able to affect the distribution of outcomes, to change risk, but it may change the meta distribution of outcome probabilities, constituting the beliefs of an individual.[42]

The distinctive feature of our model is that, given our assumptions on the prior distribution and outcome distributions, we are able to derive simple expressions for the distributions of choices. Consequently, we can directly examine properties of theoretical choice probabilities, whereas most of the aforementioned papers using bandits only derive general properties of strategies. This allows us to compare theoretical choice distributions to observable choice probabilities.

Based on a proposition resulting from our theoretical model, we derive two hypotheses allowing us to examine the central question of this chapter: the question of whether media influences occupational choices. The hypotheses are tested with two different data sets, a micro panel from the US that allows investigating individual behavior and a macro panel combining information from a larger number of countries. Using a micro and a macro panel allows us to provide results on media effects within a country and between-country effects.

4.3 A model of occupational choices under ambiguity

To theoretically analyze under what conditions articles and reports affect occupational choices, we construct an occupational choice model with ambiguity and learning. Choices are driven by decision rules. In the following, we differentiate between two potential decision rules. The first rule maximizes expected success outcomes and, thus, allows deriving individually optimal behavior. The second decision rule is based on the assumption that individuals dislike ambiguity (Ellsberg 1961) and, as they prefer to avoid ambiguous situations, they may decide for options that are not optimal in terms of income expectations but better in terms of ambiguity avoidance. The section first introduces the model's setup, followed by a discussion of the basic assumptions of the model, and an examination of decision rules and individual behavior.

[42] Consuming media articles and reports about successful or famous entrepreneurs cannot affect choices in a setting with deterministic outcomes and perfect information by construction. Risk taking is an important factor underlying entrepreneurial activities (e.g., Djankov, Qian, Roland & Zhuravskaya 2006; White, Thornhill & Hampson 2006; Vereshchagina & Hopenhayn 2009), but the reading of an article about a famous and successful entrepreneur is unlikely to reduce the entrepreneurial risk of an individual, as information provided by the article is not specific enough to be relevant for the individual's future business. Consequently, even in a setting with stochastic outcomes but known outcome distributions—a setting with risk—media is unlikely to have any effect.

4.3.1 Setup

There are two occupational options an individual can choose from: self-employment $\mathcal{S}$ and paid work $\mathcal{W}$. Let $\mathbb{O} \equiv \{\mathcal{S}, \mathcal{W}\}$ denote the set of all available options. In reality, there is the alternative of unemployment as well. However, an active choice of voluntary unemployment cannot be identified in the micro and macro data we use, and is hard to identify in data in general. Therefore, we restrict our attention to self-employment and wage work. Yet, the model can be easily extended to account for unemployment.

Every option in $\mathbb{O}$ is associated with an i.i.d. reward sequence $\{\Omega_{i,n}\}_{n=1}^{N}$, where after a fixed and known period $N > 1$ the individual retires. Each reward sequence is based on a reward distribution F, such that $\Omega_{i,n}$ is generated by F with an option-specific parameter ϕ_i. Rewards come in form of occupational successes and failures, where $\omega = 1$ represents a success and $\omega = 0$ a failure (henceforth Assumption 4.1, discussed further below). Thus, reward distributions are Bernoulli (F is Bernoulli) and ϕ_i is the probability of succeeding in occupation $i \in \mathbb{O}$. A success is generated with probability ϕ_i and a failure occurs with probability $1 - \phi_i$.

We impose the following restrictions. For the probabilities to succeed in wage work and self-employment, we have $\phi_{\mathcal{W}}, \phi_{\mathcal{S}} \in (0, 1)$. Furthermore, we assume that

$$\phi_{\mathcal{W}} + \phi_{\mathcal{S}} = 1 \tag{4.1}$$

such that individuals may either be successfully self-employed or successful wage workers, but success probabilities in wage work and self-employment are almost never the same. This assumption (henceforth Assumption 4.2) is based on Lazear (2005) and is discussed in the next subsection. Individuals know that their probability to succeed in self-employment is decreasing in the probability to succeed in wage work, i.e., (4.1) is common knowledge.

Information is assumed to be imperfect. Thus, the probabilities to succeed in wage work and self-employment are both unknown, implying ambiguity (henceforth Assumption 4.3, justified below). However, individuals have some prior knowledge. Furthermore, individuals obtain additional information about an option $i \in \mathbb{O}$ by selecting it and observing the outcome, reward drawn from $F(\phi_i)$.

Prior knowledge is given by successes in wage work $a_{\mathcal{W},0} \in \mathbb{N}^+$ and self-employment $a_{\mathcal{S},0} \in \mathbb{N}^+$ that individuals draw from a set of historical data. Historical data can be represented by reward observations of other individuals, such as parents and peers[43] (e.g., Minniti 2005; Bosma, Hessels, Schutjens, Van Praag & Verheul 2012[44]).

Prior distributions are Dirichlet. The Dirichlet distribution is a proper conjugate prior for probabilities, where the condition $\phi_{\mathcal{S}} = 1 - \phi_{\mathcal{W}}$ holds, and has density

$$\varphi(\phi_{\mathcal{S}}, \phi_{\mathcal{W}}; a_{\mathcal{S},n}, a_{\mathcal{W},n}) = \check{\Gamma}(a_{\mathcal{S},n}, a_{\mathcal{W},n})\phi_{\mathcal{S}}^{a_{\mathcal{S},n}-1}\phi_{\mathcal{W}}^{a_{\mathcal{W},n}-1}, \quad \check{\Gamma}(x_1, x_2) \equiv \frac{\Gamma(x_1 + x_2)}{\Gamma(x_1)\Gamma(x_2)} \tag{4.2}$$

$a_{\mathcal{S},n}$ and $a_{\mathcal{W},n}$ are parameters of the distribution, and Γ is the gamma function. Given no actual observations of rewards but some set of historical data, a success probability ϕ_i for $i \in \mathbb{O}$ obeys the following distribution:

$$\varphi(\phi_i; a_{\mathcal{S},0}, a_{\mathcal{W},0}) = \check{\Gamma}(a_{\mathcal{S},0}, a_{\mathcal{W},0})\phi_i^{a_{i,0}-1}(1 - \phi_i)^{a_0 - a_{i,0}-1}, \quad a_0 \equiv a_{\mathcal{S},0} + a_{\mathcal{W},0} \tag{4.3}$$

Actual observations (i.e., non-historical data) change prior distributions according to Bayes' law. Assume that in some period $n > 0$ wage work is selected and the reward $\omega_{\mathcal{W},n} \in \{0, 1\}$ is observed. Successes in wage work until period n are given by $a_{\mathcal{W},n-1}$. Successes in self-employment are given by $a_{\mathcal{S},n-1}$. Then, the posterior distribution (given a Dirichlet prior, the posterior is also Dirichlet) of the probability to succeed in wage work is

$$\varphi(\phi_{\mathcal{W}}; a_{\mathcal{S},n}, a_{\mathcal{W},n}) = \check{\Gamma}(a_{\mathcal{S},n}, a_{\mathcal{W},n})\phi_{\mathcal{W}}^{a_{\mathcal{W},n}-1}(1 - \phi_{\mathcal{W}})^{a_n - a_{\mathcal{W},n}-1} \tag{4.4}$$

where $a_{\mathcal{W},n} = a_{\mathcal{W},n-1} + \omega_{\mathcal{W},n}$, $a_{\mathcal{S},n} = a_{\mathcal{S},n-1} + 1 - \omega_{\mathcal{W},n}$, and $a_n \equiv a_{\mathcal{S},n} + a_{\mathcal{W},n}$. The posterior distribution of the probability to succeed in self-employment is obtained in a similar way.

The general setup of the model, established above, introduces a sequential decision problem: Individuals have to decide which occupation to select in every period $n = 1, \dots, N$. Let $\mathfrak{d}_n \in \mathbb{O}$ denote the decision in period n. We assume that individuals use a decision rule that determines the probabilities to select an option. In other words, a decision rule generates $\mathbb{P}(\mathfrak{d}_n = i)$ for all $i \in \mathbb{O}$ and all n.

The analysis of the model consists of two steps. First, we demonstrate that sufficiently high ambiguity aversion generates choice probabilities that differ from the optimal probability maximizing the expected sum of individual successes. Second, we show that media can change choice probabilities and, in particular, media is able to reduce or even eliminate the difference between optimal and ambiguity-aversion-affected probabilities. However, before analyzing decisions, we first discuss modeling assumptions.

[43] Information from parents, spouses, and peers can either encourage or discourage a certain occupational choice. For instance, there is anecdotal evidence that wives and parents tend to block the pursuit of entrepreneurship in Japan, while American parents tend to encourage entrepreneurial activities (Fifield 2016).

[44] Note, however, that Bosma et al. (2012) use the concept of role models, which is much richer than our concept of information because besides information role models also provide support and guidance.

4.3.2 Assumptions

Our model's setup rests on three basic modeling assumptions. The first assumption determines the type of rewards by restricting it to successes and failures. The assumption is helpful for two reasons. First, it simplifies modeling. Second, as it is relatively easy to find or construct a measure of the number of media reports on successful entrepreneurs, it makes it possible to conduct an empirical analysis.

Assumption 4.1. Occupational options produce rewards in form of periodical occupational successes or failures according to some distribution. Occupational options may differ with respect to their ability to deliver successes such that reward distributions can be different across options.

Occupational successes can be defined in various ways. A simple definition is that a success is achieved when an individual reaches a self-set monetary income benchmark. More formally, let Π denote the monetary reward generated by an arbitrary occupation. Let F_Π denote the corresponding continuous distribution function of monetary rewards. Furthermore, let B_Π denote a self-set income benchmark. A success occurs if the monetary income is above the benchmark. Consequently, the probability of a success is

$$\phi = \mathbb{P}(\Pi > B_\Pi) = 1 - \int_{-\infty}^{B_\Pi} f_\Pi(\pi)\mathrm{d}\pi$$

Our model is constructed under the assumption that success probabilities are unknown. This is fully consistent with the definition above if we assume that the distribution of monetary rewards, F_Π, is unknown, such that $\mathbb{P}(\Pi > B_\Pi)$ cannot be directly computed.

The type of rewards fixed by Assumption 4.1 does not contradict the standard way to assess rewards or incomes using the expected value—an option is "better" if it yields a higher expected income. The following example demonstrates this conjunction for the most common distribution of incomes: the log-normal (see Lopez & Servén 2006).

Example 4.1. Assume that Π has a log-normal distribution such that $\mathbb{E}[\log \Pi] = \mu_\Pi$ and $\mathbb{V}[\log \Pi] = \sigma_\Pi^2$. Assume that there are two options where $\mu_{\Pi,1} > \mu_{\Pi,2}$, while $\sigma_{\Pi,1} = \sigma_{\Pi,2} = \sigma_\Pi$. As $\mathbb{E}[\Pi] = \exp(\mu_\Pi + \sigma_\Pi^2/2)$, option 1 generates a higher expected income than 2. Let the benchmark be sufficiently large such that $B_\Pi > 1$ (for instance, larger than one unit of money). The success probability of an arbitrary option is $\phi = 1 - \mathbb{P}(\Pi \leqslant B_\Pi) = 1/2 - 1/2\,\mathrm{erf}([\log B_\Pi - \mu_\Pi]2^{-\frac{1}{2}}\sigma_\Pi^{-1})$ where erf is the Gauss error function. Hence, we get

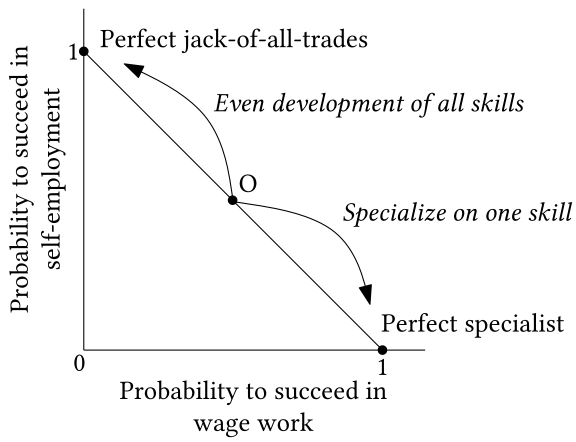
$$\phi_1 - \phi_2 = \frac{1}{2}\left[\mathrm{erf}\left(\frac{\log B_\Pi - \mu_{\Pi,2}}{\sqrt{2}\sigma_\Pi}\right) - \mathrm{erf}\left(\frac{\log B_\Pi - \mu_{\Pi,1}}{\sqrt{2}\sigma_\Pi}\right)\right]$$

Using the properties of the error function, it is easy to show that $\phi_1 - \phi_2 > 0$ for $B_\Pi > 1$. Consequently, $\mathbb{E}[\Pi_1] > \mathbb{E}[\Pi_2]$ transforms into $\phi_1 > \phi_2$.

The second assumption establishes how success probabilities are related.

Assumption 4.2. Individuals are either productive in self-employment or in paid employment but almost never both at exactly the same level.

Lazear (2005) theoretically and empirically shows that the self-employed are rather jacks-of-all-trades than specialists (also, see Wagner 2006; Stuetzer et al. 2012). Assumption 4.2 builds on this finding.[45] In our model, the probability of succeeding in wage work is implicitly assumed to increase in specialization. A specialist with much-needed skills will experience high rewards in wage work but low rewards in self-employment since highly developing a particular skill is not possible without neglecting all other skills. Figure 4.1 explains how jacks-of-all-trades and specialists are related to each other.



Consider point "O." If the individual decides to specialize on one skill, she will increase her probability to succeed in wage work but simultaneously decrease her probability to succeed in self-employment. An even development of all skills will decrease the probability to succeed in wage work but increase the probability to succeed in self-employment. However, it is not possible to increase both probabilities at the same time.

FIGURE 4.1. Jacks-of-all-trades and specialists

The third assumption introduces imperfect information.

Assumption 4.3. The probabilities to succeed in wage work and self-employment are unknown.

[45] Lazear's (2005) approach is empirically tested but it does not capture some types of individuals. For instance, individuals who are strongly restricted in their choice of occupation, for example, due to severe poverty or disabilities, must be excluded from the analysis. Furthermore, a certain level of *basic* education is necessary for wage work and self-employment, such that the assumption holds conditional on basic education levels.

There are several reasons for why reward distributions are unknown. The reward from entrepreneurship depends on many factors that individuals cannot control or fully anticipate. If an entrepreneur launches a new product, she may make a certain prediction about how the market will react to it, but there is still substantial ambiguity about market success—highly innovative products often tend to be rejected by the market.

The reward from wage work is unknown because wage workers do not have full control over their careers. The probability of losing a job or being promoted is usually not perfectly known. Furthermore, there is evidence that skills are also unknown. Antonovics & Golan (2012) show that patterns of occupational choices and wage growth are consistent with the assumption that jobs only gradually reveal information about unknown workers' skills.[46] Consequently, if some important skill influencing outcomes in wage work can only be revealed by actually doing some tasks, there will be ambiguity about outcomes and rewards.

4.3.3 Decisions: Rules and strategies

Given the setup depicted above, individuals are assumed to follow an occupational strategy based on a decision rule. As in standard economic theory, we assume that in every period individuals assign a measure of utility to every option in $\mathbb{O}$ and select the option with the highest utility. In the context of multi-armed bandits, researchers label such an approach as index strategy. An equivalent formulation is that individuals will decide based on relative utility. For instance, to decide between wage work and self-employment, individuals assign utility (index) $\mathsf{u}_{\mathcal{W}}$ to wage work and $\mathsf{u}_{\mathcal{S}}$ to self-employment, and decide for wage work if relative utility $\mathsf{u}_{\mathcal{W}} - \mathsf{u}_{\mathcal{S}}$ is weakly positive and for self-employment else.

We discuss two ways to formalize utility. The first approach is relative unbiased utility, which we use as a benchmark. The second approach is relative biased utility, which incorporates ambiguity preferences.

4.3.3.1 Unbiased utility

The idea behind the construct of relative unbiased utility is to select the most promising option, the option with the largest expected success probability, in every period by relying on priors and actual observations. A strategy exclusively concentrating on expected successes effectively ignores ambiguity because deviations from expected success proba-

[46] Rewards of wage workers depend on their skills and the ability of employers to correctly assess these skills and set a corresponding wage. There is evidence that the productivity of young workers is an unknown variable for employers such that employers need some time to learn about the skills of their workers (e.g., Mansour 2012). Hence, ambiguity in wage work might be two-sided: In addition to ambiguity on the workers' side, there is a strong indication for ambiguity on the employers' side.

bilities are not assigned any relevance. Utility is unbiased because the only motive behind decisions is to always select the best option in expectations.

Relative unbiased utility is equivalent to a simple maximization of per-period expected rewards. In every period n, distributions of probabilities to succeed in wage work and self-employment are given by densities $\varphi(\phi_{\mathcal{W}}; a_{\mathcal{W},n}, a_{\mathcal{S},n})$ and $\varphi(\phi_{\mathcal{S}}; a_{\mathcal{W},n}, a_{\mathcal{S},n})$. Hence, given information in period n, expected probabilities to succeed are

$$\mu_{i,n} = \int_0^1 \phi_i \varphi(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n}) \mathrm{d}\phi_i \quad \text{for } i \in \mathbb{O} \tag{4.5}$$

Equation (4.5) is updated in every period since either $a_{\mathcal{W},n-1} < a_{\mathcal{W},n}$ or $a_{\mathcal{S},n-1} < a_{\mathcal{S},n}$ (never both). The individual, then, selects the option that promises a success with the highest probability, while the option with the highest expected success probability may change as new information is obtained.

This decision rule corresponds to the following strategy: In period $n > 0$, use the expected success probability of wage work $\mu_{\mathcal{W},n-1}$ and the expected success probability of self-employment $\mu_{\mathcal{S},n-1}$ to construct $\mathfrak{u}_n = \mu_{\mathcal{S},n-1} - \mu_{\mathcal{W},n-1}$. Select self-employment if $\mathfrak{u}_n > 0$ and wage work if $\mathfrak{u}_n \leqslant 0$.

4.3.3.2 Biased utility

As demonstrated by Ellsberg (1961), individual decisions are not entirely based on expected outcomes but there is also a tendency to avoid ambiguity. For instance, assume that we have two options with exactly the same success probability, but the first option is more ambiguous than the second. It is reasonable to assume that an ambiguity-averse individual will exhibit a tendency to select the less ambiguous option over the more ambiguous one. Accordingly, decisions may not only be motivated by good performance in expectations but also by ambiguity avoidance. In such a case, relative utility can be biased because, besides good decision performance, preferences toward ambiguity also influence decisions.

Ambiguity can be defined as the variance of the distributions of success probabilities (Maccheroni et al. 2013):

$$\mathsf{v}_{i,n} = \int_0^1 \{\phi_i - \mu_{i,n}\}^2 \varphi(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n}) \mathrm{d}\phi_i \quad \text{for } i \in \mathbb{O} \tag{4.6}$$

An option $i \in \mathbb{O}$ is ambiguous in period n if $\mathsf{v}_{i,n} > 0$. Relative biased utility extends the idea of unbiased utility by simultaneously accounting for expected success probabilities and ambiguity.

Before introducing relative biased utility, it is necessary to discuss properties of preferences toward ambiguity. First, ambiguity aversion might be context-dependent. In our model, wage work and self-employment represent two different contexts. Wage workers are usually not directly responsible for covering damages. The self-employed do not have a buffer, in form of managers or employers, and must take the full responsibility for their actions (they have sufficient "skin in the game"). Put differently, self-employment can generate actual losses, while a wage is always non-negative, assuming that job loss corresponds to a zero wage. Hence, ambiguity in self-employment might be perceived as different (for instance, more problematic) than ambiguity in paid employment.

Second, Soto, John, Gosling & Potter (2011) show that individuals become more open to experience as they accumulate experience, a psychological effect consistent with the assumption that ambiguity aversion might be affected by (positive) experience. Third, and last, decisions tend to be self-reinforcing.[47] Positive psychological effects of successes constitute a simple approach to induce partially self-reinforcing decision patterns.

Hence, we postulate a last assumption establishing the existence of preferences toward ambiguity and their properties.[48]

Assumption 4.4. Individuals can be ambiguity-averse. Furthermore, ambiguity aversion can vary across occupations, and reactions to ambiguity may change over time depending on experience. However, the following conditions hold: (a) An individual who is ambiguity-averse never becomes ambiguity-affine or ambiguity-neutral. (b) In addition to informational effects, successes can have a psychological effect. A success can increase self-confidence. As a result, occupational-specific ambiguity aversion can effectively decrease.

The underlying idea of relative biased utility is as follows. Let $\mathrm{w}_{\phi-\mu}(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n}, \theta_i)$ denote a function weighting potential deviations of actual success probabilities from the expected probability. We assume that the weight $\mathrm{w}_{\phi-\mu}$ depends on information $(a_{\mathcal{W},n}, a_{\mathcal{S},n})$ and a parameter $\theta_i \in \mathbb{R}$ representing option-specific preferences toward ambiguity. The utility of an option is given by

$$\eta_{i,n} = \mu_{i,n} + \xi_{i,n} \quad \text{for } i \in \mathbb{O} \tag{4.7}$$

[47] For the high explanatory performance of algorithms with choice-reinforcement components, see, for instance, Camerer & Ho (1999).

[48] A decision maker confronted with a risky option knows the distribution of the outcome (see, e.g., Holm, Opper & Nee 2013) and the known variance of the outcome can be used to measure risk (Tobin 1958). A decision in an ambiguous choice environment has to cope with the fact that outcome probabilities are unknown (Ellsberg 1961). This conceptual difference has an important implication with respect to the difference between risk and ambiguity preferences. Ambiguity preferences evaluate the distribution of outcome probabilities, which might change as new information is obtained, whereas risk preferences evaluate the known distribution of outcomes.

$$\xi_{i,n} = \int_0^1 \{\phi_i - \mu_{i,n}\} \mathrm{w}_{\phi-\mu}(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n}, \theta_i)\varphi(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n})\mathrm{d}\phi_i$$

The rationale behind Equation (4.7), which is similar to the mean-variance rule in portfolio choice (see Maccheroni et al. 2013; Thimme & Völkert 2015), is that individuals will base their decisions on expected probabilities to succeed, represented by μ, but they will also anticipate potential mistakes, represented by $\phi - \mu$, which they might dislike, caused by the imperfect character of information.[49]

We use the following weighting function:

$$\mathrm{w}_{\phi-\mu}(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n}, \theta_i) \equiv 1 - \frac{\theta_i}{a_{i,n}} \left\{\phi_i - \mu_i(a_{\mathcal{W},n}, a_{\mathcal{S},n})\right\} \quad \text{for } i \in \mathbb{O} \tag{4.8}$$

To verify that $\mathrm{w}_{\phi-\mu}$ is an appropriate weight, consider three types of preferences toward ambiguity.

Ambiguity neutrality

Let $\theta_i = 0$ represent ambiguity neutrality. The weight is given by $\mathrm{w}_{\phi-\mu} = 1$. Hence:

$$\xi_{i,n} = \int_0^1 \phi_i\varphi(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n})\mathrm{d}\phi_i - \mu_{i,n}\int_0^1 \varphi(\phi_i; a_{\mathcal{W},n}, a_{\mathcal{S},n})\mathrm{d}\phi_i = 0$$

Consequently, $\eta_{i,n} = \mu_{i,n}$ such that relative unbiased and biased utility are equivalent, i.e., decisions exclusively concentrate on expected success probabilities.

Ambiguity affinity

Let $\theta_i \in \mathbb{R}^-$ represent ambiguity affinity. Conditional on ambiguity affinity, success probabilities above the expected probability μ are assigned a weight larger than 1, while success probabilities below the expected success probability μ are assigned a weight smaller than 1. Thus, $\mathrm{w}_{\phi-\mu}$ emphasizes the following aspect of ambiguity: The true probability to succeed might be higher than the expected probability. More ambiguity will increase the utility of an option as

$$\frac{\partial}{\partial \mathrm{v}_{i,n}}\xi_{i,n} = -\frac{\theta_i}{a_{i,n}} > 0 \quad \text{for } i \in \mathbb{O} \text{ and all } n$$

Ambiguity aversion

Let $\theta_i \in \mathbb{R}^+$ represent ambiguity aversion. In this case $\mathrm{w}_{\phi-\mu}$ will emphasize negative estimation errors, i.e., the fact that the true probability to succeed might be smaller than the expected probability μ, by assigning a weight smaller than 1 if $\phi - \mu > 0$ and a weight

[49] Kahn & Sarin (1988) construct a similar representation—with a different weighting function—as an extension of subjective expected utility.

larger than 1 if $\phi - \mu < 0$. More ambiguity also decreases the utility of an option as

$$\frac{\partial}{\partial \mathsf{v}_{i,n}} \xi_{i,n} = -\frac{\theta_i}{a_{i,n}} < 0 \quad \text{for } i \in \mathbb{O} \text{ and all } n \tag{4.9}$$

such that $\eta_{i,n} < \mu_{i,n}$.

Note that Assumption 4.4 holds since (4.9) never changes sign. An ambiguity-averse individual never becomes ambiguity-affine or ambiguity-neutral. Moreover, observed successes in an occupation have a self-confidence effect since $-\theta_i(a_{i,n} + 1)^{-1} > -\theta_i a_{i,n}^{-1}$, such that given more successes individuals will react less negatively to more ambiguity. Finally, note that in consistency with Assumption 4.4, ambiguity preferences can be different across occupations since $\theta_i \in \mathbb{R}^+$ does not rule out $\theta_{\mathcal{S}} > \theta_{\mathcal{W}}$ (or $\theta_{\mathcal{W}} > \theta_{\mathcal{S}}$). For the remainder of the chapter, we assume ambiguity aversion or $\theta_i \in \mathbb{R}^+$ for all $i \in \mathbb{O}$.

A strategy grounded in relative biased utility can be described as follows. In period $n > 0$, use the subjective utility of wage work $\eta_{\mathcal{W},n-1}$ and the subjective utility of self-employment $\eta_{\mathcal{S},n-1}$ to construct $\mathfrak{b}_n = \eta_{\mathcal{S},n-1} - \eta_{\mathcal{W},n-1}$. If $\mathfrak{b}_n > 0$, select self-employment. If $\mathfrak{b}_n \leqslant 0$, select wage work.

4.3.4 *Individual behavior*

A decision strategy induces a behavioral pattern. We assume that behavioral patterns are fully specified by the probabilities to select an option $i \in \mathbb{O}$ in some arbitrary period n. An important feature of our model, setting it apart from bandit models in the literature (e.g., Rothschild 1974; Jovanovic 1979; Bergemann & Hege 2005; Antonovics & Golan 2012; Konon 2016), is that it allows for the derivation of theoretical choice probabilities, which can, in principle, be directly compared to their empirical counterparts.

Behavior (probabilities to make a specific choice) induced by relative unbiased and biased utility is as follows.

Lemma 4.1. *Let $\mathfrak{d}^{\mathfrak{u}}$ denote a choice made by unbiased utility and $\mathfrak{d}^{\mathfrak{b}}$ a choice made by biased utility. Unbiased utility selects wage work with probability*

$$\mathbb{P}(\mathfrak{d}^{\mathfrak{u}}_n = \mathcal{W}) = \mathbb{P}(\mathfrak{u}_n \leqslant 0) = H(\tau^{\mathfrak{u}}_n; n, \phi_{\mathcal{S}}), \quad \tau^{\mathfrak{u}}_n = -\frac{n + a_{\mathcal{W},0} - a_{\mathcal{S},0}}{2}$$

and self-employment with probability $\mathbb{P}(\mathfrak{d}^{\mathfrak{u}}_n = \mathcal{S}) = 1 - H(\tau^{\mathfrak{u}}_n; n, \phi_{\mathcal{S}})$, where $H(x; n, \phi)$ is the cumulative distribution function of the binomial distribution given period n and success probability ϕ. Biased utility selects wage work with probability

$$\mathbb{P}(\mathfrak{d}^{\mathfrak{b}}_n = \mathcal{W}) = \mathbb{P}(\mathfrak{b}_n \leqslant 0) = H(\tau^{\mathfrak{b}}_n; n, \phi_{\mathcal{S}})$$

$$\tau_n^{\mathfrak{b}} = -\frac{a_{\mathcal{S},0}\theta_{\mathcal{W}} - (n + a_{\mathcal{W},0})\theta_{\mathcal{S}} - (a_n^2 + a_n)n - (a_n^2 + a_n)a_{\mathcal{W},0} + a_{\mathcal{S},0}a_n^2 + a_{\mathcal{S},0}a_n}{\theta_{\mathcal{W}} + \theta_{\mathcal{S}} + 2a_n(a_n + 1)}$$

whereas the probability to select self-employment is $\mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{S}) = 1 - H(\tau_n^{\mathfrak{b}}; n, \phi_{\mathcal{S}})$.

Proof. See Appendix 4.A.1. ■

Decisions, respectively strategies, are evaluated according to a simple criterion: the number of successes they produce. A straightforward evaluation criterion is the expected number of successes given by

$$C \equiv \mathbb{E}\left[\sum_{n=1}^{N} \Omega(\mathfrak{d}_n) | a_{\mathcal{S},0}, a_{\mathcal{W},0}\right] \tag{4.10}$$

where $\Omega(\mathfrak{d}_n) \in \{0, 1\}$ is the reward given choice $\mathfrak{d}_n$. C only evaluates individual decision performance, abstracting from welfare effects and other non-individual criteria.

We establish the following property for the behavioral patterns of unbiased utility:

Lemma 4.2. *Behaving according to relative unbiased utility maximizes* C *such that unbiased utility is an optimal strategy given* $(a_{\mathcal{S},0}, a_{\mathcal{W},0})$*. By implication, behavior induced by relative unbiased utility is optimal.*

Proof. See Appendix 4.A.2. ■

For biased utility, we obtain the following result:

Proposition 4.1. *In general, behaving according to relative biased utility does not maximize* C *such that behavior is not optimal. Wage work will be selected with a higher than optimal probability if the ambiguity aversion associated with self-employment is higher than the aversion associated with wage work or* $\theta_{\mathcal{S}} > \theta_{\mathcal{W}}$*, where* $\theta_{\mathcal{S}}$ *is sufficiently large. The same applies to self-employment that is selected with a higher than optimal probability if the ambiguity aversion associated with wage work is higher than the ambiguity aversion associated with self-employment or* $\theta_{\mathcal{W}} > \theta_{\mathcal{S}}$*, where* $\theta_{\mathcal{W}}$ *is sufficiently large.*

Proof. See Appendix 4.A.3. ■

In this section, we demonstrate that the individually optimal strategy, maximizing the expected sum of occupational successes, is to always select the option with the highest expected success probability. The optimal strategy prescribes to exclusively concentrate on expected successes and to ignore potential errors in form of deviations of the true success probability from the expected value.

However, individuals making occupational choices might not be able to fully ignore errors, where the possibility of errors represents ambiguity. Therefore, we introduce a second decision strategy that accounts for ambiguity and, more specifically, ambiguity aversion, while also allowing for ambiguity aversion to differ across occupational choices. The introduction of ambiguity aversion reveals that there might be a bias for or against a particular occupation if ambiguity aversion is asymmetric across occupational options. This particular bias—the difference in choice probabilities between the optimal strategy and a strategy accounting for ambiguity aversion—is necessarily produced by a sufficiently high level of asymmetric ambiguity aversion in our model but cannot be tested with the data available to use.

4.4 Impact of media on decision patterns

Does media change choice probabilities (behavior)? Moreover, when does the consumption of articles and reports favoring entrepreneurship by ambiguity-averse individuals decrease deviations from optimal behavior, thereby improving decisions? To answers both questions, we analyze the impact of media and illustrate the model's mechanism by depicting (potentially positive) effects of media on occupational choices. We also derive two predictions that can be empirically tested.

4.4.1 Media as an informational intervention

Media articles and reports are denoted by $\mathfrak{m} \in \mathbb{N}$. $\mathfrak{m}$ is an informational intervention that does not affect probabilities to succeed. One of the simplest ways to formalize such an informational shock in favor of self-employment is to assume that in period $n = 0$ individuals are shown $\mathfrak{m} > 0$ additional successes in self-employment. Given media, instead of prior information $a_{\mathcal{S},0}$, individuals base their decisions on $\mathring{a}_{\mathcal{S},0} = a_{\mathcal{S},0} + \mathfrak{m}$, where $\mathring{a}_{\mathcal{S},0} > a_{\mathcal{S},0}$, while prior information about wage work is not directly affected. Media intensity is measured by the size of $\mathfrak{m}$, i.e., an increase in $\mathfrak{m}$ is interpreted as an increase in intensity.

Using the definition of media introduced above, media effects with respect to behavior are as follows.

Proposition 4.2. *Let $\mathring{\mathfrak{d}}_n$ denote a choice affected by media, whereas the choice without media impact is $\mathfrak{d}_n$. Given sufficient media intensity $\mathfrak{m} > 0$ and the two decision rules established (viz., unbiased and biased utility), media increases the probability to select self-employment and decreases the probability of wage work such that $\mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{u}} = \mathcal{S}) > \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{S})$ and $\mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{u}} = \mathcal{W}) < \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{W})$, and $\mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}} = \mathcal{S}) > \mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{S})$ and $\mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}} = \mathcal{W}) < \mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{W})$ for all n, where the effect requires $\theta_{\mathcal{S}} \geqslant \theta_{\mathcal{W}}$ in case of biased utility.*

Proof. See Appendix 4.A.4. ∎

The number of settings where positive media articles and reports about entrepreneurship might, theoretically, have a positive effect on occupational choices is restricted. For instance, if there is already a bias for self-employment, as it happens when $\theta_W > \theta_S$ (see Proposition 4.1), increasing the number of self-employed is unnecessary. Yet, there is one setting where media does have normatively positive effects.

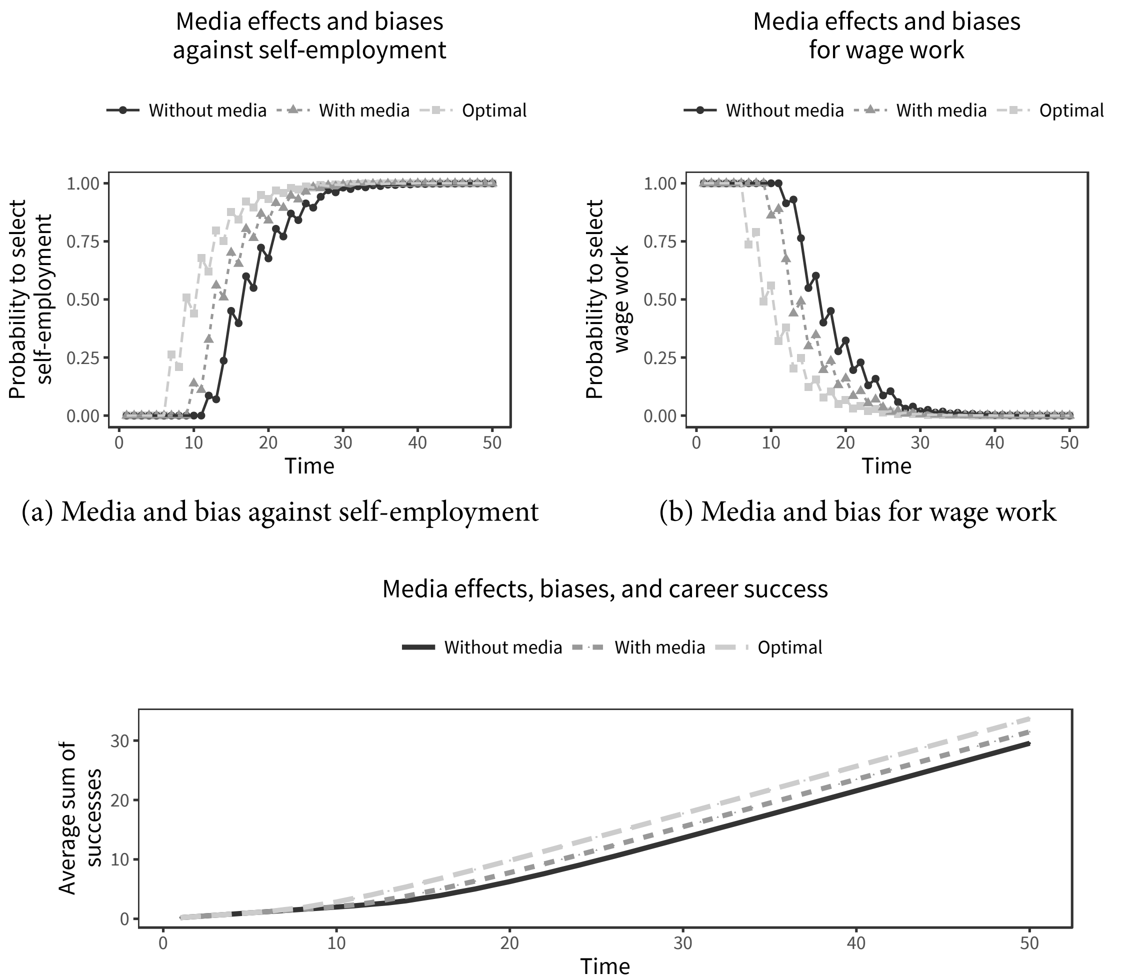
Proposition 4.3. *Assume that ambiguity aversion in self-employment is higher than ambiguity aversion in wage work, $\theta_S > \theta_W$, such that individual decisions are biased against self-employment. In such a setting, there always exists a level of media intensity such that the bias against self-employment is reduced. However, too intensive media effects might also create a bias for self-employment.*

Proof. See Appendix 4.A.5. ∎

To build intuition on the model's mechanism, consider a simple numerical example demonstrating how media influences (and improves) decisions.

Example 4.2. Assume that we could observe a sufficiently high number of alternative decision histories, allowing us to evaluate choice distributions, of an individual who retires after 50 periods. The individual's true probability to succeed in wage work is 20% and the probability to succeed in self-employment is 80%. Both probabilities are unknown to the individual making decisions. Furthermore, before her own career, the individual could observe the careers of two relatives. One relative was successful in self-employment over 5 periods, whereas the other relative was successful in wage work over 10 periods. Consequently, initial information suggests an expected success probability of 33% in self-employment and 67% in wage work.

Assume that the individual is not particularly ambiguity-averse but that ambiguity aversion with respect to self-employment is substantially higher than with respect to wage work such that $\theta_W = 100$ and $\theta_S = 5\theta_W$. The individual selects self-employment with a lower than optimal probability, which is depicted in Figure 4.2a, and wage work with a higher than optimal probability, which is depicted in Figure 4.2b, because her ambiguity preferences bias her toward wage work. Now, assume that the individual watched TV reports about successful entrepreneurs. She decided that two reports ($\mathfrak{m} = 2$) were trustworthy. Hence, media only slightly changes the expected probability to succeed in self-employment, which increases by about 8 percentage points, and the probability to succeed in wage work, which decreases by 8 percentage points. Yet, even the small change increases the probability to become self-employed, respectively reduces the probability to select wage work, as shown in Figures 4.2a and 4.2b. As a consequence of incorporating information from the media into her beliefs, the individual becomes more successful, as demonstrated in Figure 4.2c, where the sum of successes given media effects clearly dominates successes without media effects.



(a) Media and bias against self-employment

(b) Media and bias for wage work

(c) Media and career success

FIGURE 4.2. Media corrects a bias against self-employment

Thus, in our model, positive media articles and reports about entrepreneurship increase the probability to select self-employment, whereas the probability to select wage work decreases. Furthermore, if ambiguity aversion in self-employment is sufficiently higher than ambiguity aversion associated with wage work, there will be a bias against self-employment. Positive media articles and reports about entrepreneurship may reduce biases involving an informational "push" towards self-employment.

4.4.2 Predictions

The theoretical model allows us to formulate the following two predictions, on the basis of Proposition 4.2, with respect to marginal effects of media on occupational choices:

Hypothesis 4.1. In the wake of consuming media articles and (TV) reports with positive attitudes toward entrepreneurship, the probability of self-employment increases.

Hypothesis 4.2. The consumption of media articles and (TV) reports about successful entrepreneurs reduces the probability of selecting wage work.

4.5 Empirical evidence on media effects

4.5.1 A minimalistic model of media consumption

Before presenting data and regression approaches, we introduce a simple model of media consumption linking individual consumption of articles and reports about entrepreneurship to natural disasters, substantiating the first stage of our regressions.

Let vector $\mathbf{m} = [\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_k, \mathfrak{m}, \mathcal{U}]^\top$ denote media consumption consisting of reports not related to entrepreneurship, $\mathcal{R}_1, \ldots, \mathcal{R}_k$, positive reports about entrepreneurship, $\mathfrak{m}$, and urgent news, $\mathcal{U}$. Daily consumption time is restricted to 24 hours. Hence, we can safely assume that $\mathbf{1}^\top \mathbf{m} = \bar{m}$, where the consumption limit $\bar{m} > 0$ is fixed and $\mathbf{1}$ is a vector of ones with $k+2$ elements. Without loss of generality, assume that the only urgent news are news about disasters. Disasters induce a variation in $\mathcal{U}$ such that $\mathbb{V}[\mathcal{U}] > 0$. If $\mathbb{V}[\mathcal{U}] > 0$, we also have $\mathbb{V}[\bar{m} - \mathcal{U}] > 0$. Using $\mathbf{1}^\top \mathbf{m} = \bar{m}$, we obtain

$$\mathbb{V}[\bar{m} - \mathcal{U}] = \mathbb{V}\left[\sum_{l=1}^{k} \mathcal{R}_l\right] + \mathbb{V}[\mathfrak{m}] + 2\mathrm{Cov}\left[\sum_{l=1}^{k} \mathcal{R}_l, \mathfrak{m}\right] \tag{4.11}$$

Thus, if the variation induced by natural disasters is not completely absorbed by articles and reports not related to entrepreneurship and consumption is fixed at some level, two rather weak conditions, disasters will induce a variation in the consumption of positive articles and reports about entrepreneurship. This relation can be tested—by testing for instrument strength.

Whether disasters increase or decrease the consumption of articles and reports about entrepreneurship depends on the correlation between non-entrepreneurship-related news and stories about entrepreneurship. There might be a compensation effect, bad news (natural disasters) are compensated by reading success stories about entrepreneurs, or a crowding-out effect, individuals concentrate on bad news and reduce the consumption of stories about entrepreneurs. Our results (first-stage regressions) provide evidence for crowding-out effects.

4.5.2 Micro panel

4.5.2.1 Data description

Except for the media variable and its instrument, our micro panel is based on data from the Integrated Health Interview Series (IHIS; Minnesota Population Center and State Health Access Data Assistance Center 2016), which is in turn based on the National Health Interview Survey (NHIS). NHIS is an annual survey that has been conducted since 1957. NHIS mostly provides information on health but the survey also provides data on occupational choice and variables important for the choice such as previous income, work experience, education, access to finance, etc.[50] We only consider adults (18–65 years old) who are employed (either wage workers or self-employed) in the period 2004–2015.[51] Observations are either available at the individual level or at the level of US regions, as used by the United States Census Bureau, which are: Northeast, North Central (Midwest), South, and West. The panel is a repeated cross section. Given our restrictions on age and occupational status, 10,851 observations are available. Further information on the micro panel is provided in Appendix 4.B.1. The variables used in the empirical analysis are as follows.

Occupational status

The dependent variable is binary (1 if an individual has a certain occupational status and zero else). The most common occupational status is wage work, which is shown in Figure 4.B.2 (Appendix 4.B.1). There are no striking differences in occupational shares between regions but the North Central (Midwest) region tends to have a smaller self-employment share than other regions.

First media variable: Consumption of articles about famous entrepreneurs

The consumption of entrepreneurial success stories is approximated by the regional frequency of the search item ’famous entrepreneurs’ in Google. Data is provided by the Google Trends tool.[52] The tool provides results at the US state level, which are aggregated to obtain searches at the region level. Since results are always measured relative to the state with the most searches (which is normalized to 100), only effect directions can be identified.

The instrument: Number of natural disasters

The consumption of articles about famous entrepreneurs might be endogenous. Therefore, we instrument it by the number of natural disasters in non-US regions, as natural disasters are exogenous to occupational choice but are usually covered in media reports, thus, affecting the consumption of articles about entrepreneurs (see Section 4.5.1). Data

[50] For further information on the panel, see https://www.cdc.gov/nchs/nhis/.
[51] Google Trends data is available starting 2004.
[52] Available under www.google.com/trends/.

on natural disasters is collected by the Centre for Research on the Epidemiology of Disasters (Guha-Sapir et al. 2016). We only consider natural or "complex"[53] disasters, while specifically excluding technological disasters, as the latter type is caused by human action and is less likely to be exogenous. Regions with disasters are assigned based on geographical and cultural proximity but we avoid assigning a region that is too close to the US region. The Northeast region is assigned disasters in Mexico; the North Central region is assigned disasters in Australia and New Zealand; the South region is assigned disasters in South America; finally, the West region is assigned disasters in Western Europe.

We include several major determinants of occupational choice identified in the previous literature, including a number of demographic characteristics, capital income, educational levels, work experience, physical and mental health, as well as personality.

Demography

Demographic controls include age, gender, whether the individual was born in the United States, and ethnicity.

Income, education, and work

We also control for earnings during the previous year; whether the individual usually works full time; educational attainment, ranging from "never attended school" to "obtained a doctoral degree;" and years on main or longest, or last job. Furthermore, we control for whether the individual received public assistance or food stamps in the previous year, and if the individual has access to the financial market, approximated by whether the individual earned dividends from stocks or mutual funds in the previous year.

Physical health

Health is controlled by a general health variable (increase indicates decreasing health) and more specifically by whether the individual has any activity limitations.

Mental health and personality

We also control for mental health by including a set of variables capturing individuals' answers to the question of whether everything felt like an effort in the past 30 days, whether feelings interfered with life; and how often the individual felt hopeless, nervous, restless, sad, or worthless. Note that mental health also partially captures personality traits, as traits are linked to the probability of depression and anxiety (Klein, Kotov & Bufferd 2011).

4.5.2.2 *Identifying media effects in the micro panel*

In the micro panel based on US data, we have observations at the level of individuals indexed by k and at the level of US regions indexed by r. At the individual level, the

[53] A complex disaster includes famines for which drought was not the main cause.

"panel" is a repeated cross section such that individuals (and their number) change from period to period, where time is indexed by n. The dependent variable is dichotomous. An individual k from region r in period n can be self-employed, $d_{k,r,n,\mathcal{S}} = 1$, or not self-employed, $d_{k,r,n,\mathcal{S}} = 0$. Furthermore, an individual can be a wage worker, $d_{k,r,n,\mathcal{W}} = 1$, or not a wage worker, $d_{k,r,n,\mathcal{W}} = 0$. As we only consider individuals who are employed, we must have $d_{k,r,n,\mathcal{W}} = 1$ if $d_{k,r,n,\mathcal{S}} = 0$ and $d_{k,r,n,\mathcal{S}} = 1$ if $d_{k,r,n,\mathcal{W}} = 0$.

To analyze media effects, we use a probit model. Let $d_{k,r,n,i} = \mathbb{1}\{d^*_{k,r,n,i} > 0\}$ for $i \in \mathbb{O}$ where $d^*_{k,r,n,i}$ is an unobserved latent variable. The latent variable is modeled as

$$d^*_{k,r,n,i} = \mathsf{e}_{r,i} + \kappa_{1,i} M^{[1]}_{r,n} + \boldsymbol{\rho}^\top_{1,i} \mathbf{x}^{[1]}_{k,r,n} + \nu^{[1]}_{k,r,n,i} \tag{4.12}$$

where $\mathsf{e}_{r,i}$ is an option-specific fixed region effect and $\mathbf{x}^{[1]}_{k,r,n}$ are individual- and region-specific controls. $\kappa_{1,i}$ is the reaction of the latent variable and, thus, the individual choice variable, to the regional consumption of positive media articles about entrepreneurs $M^{[1]}_{r,n}$.[54] $M^{[1]}_{r,n}$ is constructed on the basis of Google Trends data revealing information on the dynamics of the search item 'famous entrepreneurs.' Unfortunately, $\kappa_{1,i}$ does not allow for the identification of effects sizes, due to the construction of the media variable, but effect directions can be easily identified.

The error term is likely heteroskedastic. For instance, there is a gender gap in entrepreneurship (Wagner 2007). If women react differently than men to incentives to become self-employed, the variance of the error cannot be equal across all individuals. However, even though the choice model is normalized, heteroskedasticity results in biased parameter estimates in a probit model (Yatchew & Griliches 1985), which is, for instance, not the case in a linear model, where coefficients are still unbiased under heteroskedasticity. A straightforward approach to account for heteroskedasticity-related issues is to explicitly model its determinants (Alvarez & Brehm 1995) by including a subset of covariates in the error variance specification. Therefore, we assume that

$$\nu^{[1]}_{k,r,n,i} \sim \text{Normal}(0, \exp\{\boldsymbol{\nu}^\top \mathbf{z}_{k,r,n}\}) \tag{4.13}$$

where determinants of heteroskedasticity in $\mathbf{z}_{k,r,n}$ and covariates in $\mathbf{x}^{[1]}_{k,r,n}$ can partially overlap.

To account for a potential endogeneity of media consumption, we instrument it by disasters in other (non-US) regions $D^{[1]}_{r,n}$ (the construction of the instrument is explained in Section 4.5.2.1), yielding the following first stage:

$$M^{[1]}_{r,n} = \breve{\mathsf{e}}_r + \varkappa_1 D^{[1]}_{r,n} + \breve{\boldsymbol{\rho}}^\top_1 \mathbf{x}^{[1]}_{k,r,n} + \breve{\nu}^{[1]}_{k,r,n} \tag{4.14}$$

[54]Note that by properties of the dependent variable, we must have $\kappa_{1,\mathcal{S}} = -\kappa_{1,\mathcal{W}}$ and $\boldsymbol{\rho}_{1,\mathcal{S}} = -\boldsymbol{\rho}_{1,\mathcal{W}}$.

where $\breve{e}_r$ is a region fixed effect.

4.5.2.3 *Estimation and results of micro panel models*

Micro panel models are estimated by maximum likelihood. As choice incentives might vary between genders (Wagner 2007), we include gender in the variance model, in Equation (4.13). In addition, we also use the following covariates to model the variance: region fixed effects, age, ethnicity, health, and education. The variance determinants were selected from a larger set on the basis of statistical significance and plausibility. For instance, it is plausible that health matters for the reaction to incentives to become self-employed or wage worker, as does the age of the individual (Caliendo et al. 2014).

Table 4.1 presents first-stage estimation results. Instrument strength does not pose a problem (the F-statistic is 41.382). Disasters are negatively correlated to the consumption of articles about famous entrepreneurs, corresponding to a crowding-out effect, which is in line with the minimalistic model of media consumption, constructed in Section 4.5.1. The crowding out is consistent with previous research. There is, for example, evidence that humans are predisposed to focus on negative information, because the costs of ignoring negative information outweigh the benefits of positive information (Soroka & McAdams 2015). Such a negativity bias is a reasonable heuristic if costs and benefits from different types of information are asymmetric.

Table 4.2 shows results generated by heteroskedastic IV probit.[55] We observe that women are less likely to become self-employed, which is consistent with previous results (Cowling & Taylor 2001; Wagner 2007; Caliendo et al. 2014). Being female also has a significant effect on the variance of choices. Furthermore, in line with previous findings, work experience (years on job) and receiving dividends (access to the financial market) both increase the probability of self-employment (Blanchflower & Oswald 1998; Gompers, Lerner & Scharfstein 2005; Elfenbein, Hamilton & Zenger 2010), while the effect of age follows an inverse u-shaped relationship (Caliendo et al. 2014). Thus, the coefficients of non-media variables confirm earlier findings.

Turning now to the influence of our first media variable, the consumption of articles about famous entrepreneurs, we can see that the consumption of articles about famous entrepreneurs significantly increases the probability of selecting self-employment and reduces the probability of selecting wage work. The effects in Table 4.2 support Hypothesis 4.1 and 4.2.

[55] Estimation results indicate an endogeneity issue: Results with IV and without IV, given in Table 4.C.1 (Appendix 4.C), substantially differ.

TABLE 4.1. First stage of micro panel model, where dependent variable is media attention

Variable	Coefficient	SE
Number of natural disasters	-0.008^{***}	(0.001)
Age	0.005	(0.003)
Age^2	−0.000	(0.000)
Female†	−0.006	(0.011)
Born in US†	0.049^{***}	(0.017)
Non-white†	-0.042^{***}	(0.013)
Full-time work†	0.038^{***}	(0.013)
Earnings	-0.006^{**}	(0.003)
Got dividends†	0.098^{***}	(0.019)
Got food stamps†	0.501^{***}	(0.030)
Got welfare†	-0.174^{***}	(0.049)
Education	−0.001	(0.002)
Years on job	0.001	(0.001)
Limitations†	−0.027	(0.016)
Health‡	−0.001	(0.006)
Effort	-0.013^{***}	(0.005)
Feelings interfered with life	−0.002	(0.007)
Hopeless	-0.016^{**}	(0.008)
Nervous	-0.014^{***}	(0.005)
Restless	−0.003	(0.005)
Sad	0.016^{**}	(0.006)
Worthless	0.008	(0.008)
North Central†	0.163^{***}	(0.031)
Northeast†	-0.709^{***}	(0.036)
West†	-1.271^{***}	(0.030)
Constant	1.581^{***}	(0.075)
	10,851 obs.; $R^2 = 0.550$	

Notes: F-statistic for instrument weakness with heteroskedasticity-robust errors: 41.382; †dummy variable; ‡increase indicates more health problems; ***significant at the 1%-level; **significant at the 5%-level; standard errors in parentheses are heteroskedasticity-consistent.

TABLE 4.2. IV probit estimates of marginal effects in micro panel model, where dependent variable is choice dummy

	Self-employment		Wage work	
Variable	Coefficient	SE	Coefficient	SE
Consumption of articles about famous entrepreneurs[§]	0.611***	(0.196)	−0.611***	(0.196)
Age	0.178***	(0.034)	−0.178***	(0.034)
Age^2	−0.002***	(0.000)	0.002***	(0.000)
Female[†]	−1.912***	(0.525)	1.912***	(0.525)
Born in US[†]	−0.191	(0.100)	0.191	(0.100)
Non-white[†]	−0.234	(0.158)	0.234	(0.158)
Full-time work[†]	−0.405***	(0.108)	0.405***	(0.108)
Earnings	0.004	(0.013)	−0.004	(0.013)
Got dividends[†]	0.237**	(0.119)	−0.237**	(0.119)
Got food stamps[†]	−0.369**	(0.177)	0.369**	(0.177)
Got welfare[†]	0.167	(0.251)	−0.167	(0.251)
Education	0.019	(0.017)	−0.019	(0.017)
Years on job	0.073***	(0.016)	−0.073***	(0.016)
Limitations[†]	0.060	(0.091)	−0.060	(0.091)
Health[‡]	0.012	(0.059)	−0.012	(0.059)
Effort	0.034	(0.028)	−0.034	(0.028)
Feelings interfered with life	−0.045	(0.041)	0.045	(0.041)
Hopeless	−0.071	(0.047)	0.071	(0.047)
Nervous	0.018	(0.031)	−0.018	(0.031)
Restless	0.071**	(0.031)	−0.071**	(0.031)
Sad	−0.035	(0.038)	0.035	(0.038)
Worthless	−0.070	(0.049)	0.070	(0.049)
North Central[†]	−0.456**	(0.197)	0.456**	(0.197)
Northeast[†]	0.302	(0.198)	−0.302	(0.198)
West[†]	0.610**	(0.260)	−0.610**	(0.260)
Constant	−5.656***	(0.763)	5.656***	(0.763)
Variable: Variance model		Coefficient	SE	
North Central[†]		−0.042	(0.070)	
Northeast[†]		−0.100	(0.083)	
West[†]		0.057	(0.073)	
Age		0.013***	(0.003)	
Female[†]		0.584***	(0.109)	
Non-white[†]		−0.083	(0.067)	

Health‡	−0.052**	(0.026)
Education	−0.003	(0.008)
	10,851 obs.	

Notes: §Normalized media consumption is instrumented by number of natural disasters in other regions and countries; †dummy variable; ‡increase indicates more health problems; ***significant at the 1%-level; **significant at the 5%-level; standard errors in parentheses are heteroskedasticity-consistent.

4.5.3 Macro panel

4.5.3.1 Data description

To check whether results are robust, we also use, in addition to the micro data model, an empirical model based on an unbalanced country-level macro panel. Effects on the probability of self-employment and wage work are estimated on the basis of 38 countries. In sum, there are 170 joint observations. The panel is fairly representative, as both developed and developing countries are included.[56] We use annual country-level data from four different sources: the Global Entrepreneurship Monitor, the World Bank, Transparency International, and the Centre for Research on the Epidemiology of Disasters. We only consider the 2003–2012 time period. The minimum of observed periods is 3 and the maximum is 10. Appendix 4.B.2 shows data characteristics. The following variables are used in our regressions:

Shares of self-employed and wage workers
The dependent variables of our regression models are relative choice frequencies, or empirical probabilities. We approximate relative choice frequencies by the share of wage workers and self-employed provided by the World Bank. In Figure 4.B.3 (Appendix 4.B.2) it is shown that there is a substantial variation in choice frequencies across countries. For instance, the maximum country average of the probability of self-employment is 52% (in Peru), while the minimum level is 7% (in the US).

Second media variable: Media attention for entrepreneurship
The media variable in the macro panel differs from the one in the micro model. Instead of deriving it from Google Trends data, we approximate media intensity by "media attention for entrepreneurship," surveyed by the Global Entrepreneurship Monitor (GEM). Media attention is measured by the percentage of the population aged 18–64 who report that in their country there are frequent media reports about successful new businesses. The

[56]The following countries are included: Argentina, Australia, Belgium, Brazil, Canada, Chile, Colombia, Croatia, Ecuador, Greece, Hong Kong, Hungary, Iran, Ireland, Israel, Italy, Jamaica, Japan, South Korea, Latvia, Malaysia, Mexico, the Netherlands, New Zealand, Norway, Peru, Poland, Romania, Russia, Serbia, Singapore, Slovenia, Sweden, Switzerland, the UK, the USA, Uruguay, and Venezuela.

advantage of the indicator provided by the GEM compared to other indicators is that it measures perceptions so that we can be sure that decision makers are aware of positive reports about entrepreneurship.

The instrument: Number of natural disasters

As in the micro model, we instrument the media variable by the number of natural disasters in other countries. In the macro panel, countries are paired randomly and the combination yielding the strongest instrument is selected, as described in Section 4.5.3.2. Note that we have more observations of natural disasters than of media attention for entrepreneurship such that the pool of countries with disasters is larger than 38.

Fear of entrepreneurial failure

Countries differ with respect to their attitudes towards entrepreneurial failure. As noted by the *Economist*:

> If you start a company in London or Paris and go bust, you have just ruined your future; do it in Silicon Valley and you have simply completed your entrepreneurial training. (The Economist 1997, p. 17)

Being afraid to fail, and the associated stigma, can prevent an individual from becoming self-employed. Hence, we control for country-specific attitudes towards failure by including the percentage of the population aged 18–64 perceiving good opportunities for business who indicate that fear of failure would prevent them from setting up a business. The fear of failure rate is provided by the GEM.

Ease of doing business

Annual indicators of ease of doing business measuring a country's regulatory environment[57] are provided by the World Bank Group (Doing Business project). The higher the indicator value is, the easier is doing business. In our sample, ease of doing business mostly reflects the difference between developed and developing countries—doing business tends to be easier in developed countries. The difference in economic development is important for occupational choice as, compared to developed countries, developing countries are exposed to higher unemployment levels, have lower levels of wage work, and higher levels of self-employment (see, e.g., Chen & Doane 2008; Gindling & Newhouse 2012). Unfortunately, using annual indicators would greatly reduce the number of available observations. Therefore, we, first, take country-specific averages and, then, construct two groups based on these country averages with k-means clustering: a group of countries where doing business is relatively easy and a group where it is relatively difficult. This classification is

[57] The regulatory environment includes components such as starting a business, dealing with construction permits, getting electricity, registering property, getting credit, protecting minority investors, paying taxes, trading across boarders, enforcing contracts, and resolving insolvency (see World Bank Group 2015).

assumed to hold for all periods 2003–2012, even if annual ease of doing business was not observed in some periods.[58]

Corruption

The Corruption Perceptions Index, taking values on the interval $[0, 10]$, is annually provided by Transparency International. The higher the index value is, the *less* corruption is perceived. We include a measure of corruption in our regressions because our data includes developing countries and previous research shows the relative importance of institutional constraints impeding development in developing economies (Goedhuys & Sleuwaegen 1999; Ardagna & Lusardi 2010; Quatraro & Vivarelli 2015). Furthermore, Anokhin & Schulze (2009) demonstrate that corruption hampers innovation and entrepreneurship.

Other controls

In addition to the aforementioned variables, we use the following controls that could also affect occupational choices: GDP (per capita), GDP growth, inflation, and the real interest rate. All four covariates are provided by the World Bank.

4.5.3.2 *Identifying media effects in the macro panel*

We have data for a set of countries indexed by j. Each country j is observed over some periods indexed by n. The number of observed periods is allowed to differ across countries.[59] With respect to the dependent variable, we observe two shares for each j and n. The first share, $p_{j,n,\mathcal{S}} \in (0, 1)$, is the share of self-employed individuals (the empirical probability of the choice 'self-employment') in the working-age population. The second share, $p_{j,n,\mathcal{W}} \in (0, 1)$, is the share of wage workers in the working-age population. We also refer to $p_{j,n,i}$ as the relative frequency of occupation $i \in \mathbb{O}$.

Let $\mathfrak{L}(p) \equiv \log(p[1-p]^{-1})$ denote the logit transformation function, where $p \in (0, 1)$ is a proportion. The transformation maps a share on the real line. To model a relative choice frequency, we use the following linear model:

$$\mathfrak{L}(p_{j,n,i}) = \mathsf{c}_i + \kappa_{2,i} M^{[2]}_{j,n} + \boldsymbol{\rho}^{\top}_{2,i} \mathbf{x}^{[2]}_{j,n} + \nu^{[2]}_{j,n,i} \tag{4.15}$$

$\mathbf{x}^{[2]}_{j,n}$ are time- and country-specific covariates. c_i is an option-specific constant. $\kappa_{2,i}$ is the option-specific effect of media, i.e., the effect of most interest. $\exp(\kappa_{2,i})$ corresponds to the relative change in odds given a one unit increase in media attention, when all the

[58] This assumption makes sense if relative ease of doing business is sufficiently stable over time. In Appendix 4.B.3, we examine stability with available data and find a strong tendency of countries to remain in one group.

[59] However, we require that $n \geqslant 3$ for all j such that the effects of time-variant variables can be distinguished from the impact of time-invariant covariates.

remaining variables are held constant. To approximately examine effect sizes, we also use a linear probability model, where the left hand side of (4.15) is $p_{j,n,i}$.

Equation (4.15) is the second stage of our regression. As the media variable in (4.15) might be endogenous, $M_{j,n}^{[2]}$ and the error term $v_{j,n,i}^{[2]}$ may be correlated, we instrument media attention by the number of natural disasters, denoted by $D_{j,n}^{[2]}$. To ensure that the exclusion restriction holds, we only use disaster data from other countries.[60] The first stage is as follows:[61]

$$M_{j,n}^{[2]} = \breve{c} + \varkappa_2 D_{j,n}^{[2]} + \breve{\boldsymbol{\rho}}_2^\top \mathbf{x}_{j,n}^{[2]} + \breve{v}_{j,n}^{[2]} \tag{4.16}$$

To generate an instrument with sufficient strength, we use the following three-step approach:

STEP 1. Each country i in the panel is randomly, without repetitions, assigned another country $\mathfrak{a}_i$ with disasters resulting in assignment matrix

$$\mathcal{A}_\mathsf{r} = \begin{bmatrix} \text{Country 1} & \mathfrak{a}_1 \\ \text{Country 2} & \mathfrak{a}_2 \\ \vdots & \vdots \\ \text{Country 38} & \mathfrak{a}_{38} \end{bmatrix}$$

The assignment procedure is repeated R times, such that we obtain the general assignment matrix

$$\mathbf{A} = \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 & \cdots & \mathcal{A}_{\mathsf{R}-1} & \mathcal{A}_\mathsf{R} \end{bmatrix}$$

STEP 2. For each assignment $\mathcal{A}_\mathsf{r}$ in $\mathbf{A}$, a first-stage F-statistic, to assess instrument strength, is computed (Staiger & Stock 1997). Statistics account for heteroskedasticity or clustering at the country level.

STEP 3. The combination with the best F-statistic result given potential heteroskedasticity, conditional on sufficient instrument strength in case of errors clustering at the country level, is selected.

Instrument strength is considered as sufficient if the first-stage partial F-statistic is substantially larger than 10 (Staiger & Stock 1997; Stock & Yogo 2005). Besides the best instrument, the three-step approach will generate a number of country pairings with sufficient strength (a large F-statistic). These combinations can be used to test whether results

[60] $D_{j,n}^{[2]}$ captures disasters in a country assigned to j but different from j.

[61] We do not transform $M^{[2]}$—given our data, $M^{[2]} \in (0, 1)$—as this would limit interpretations. However, our main results, the outcome of the test of the two central model predictions, does not depend on the transformation of $M^{[2]}$.

depend on a particular combination of countries or are robust to using different country pairs.

4.5.3.3 *Estimation and results of macro panel models*

To estimate our macro panel models, we use two-stage least squares (with heteroskedasticity-robust standard errors and errors clustered at the country level). The best pairing of countries on the basis of 10,000 random assignments is given in Table 4.C.3 (Appendix 4.C). Table 4.4 shows first-stage results.

TABLE 4.4. First stage of macro panel model, where dependent variable is media attention

Variable	Coefficient	SE
Number of natural disasters	$-1.442^{***[***]}$	(0.155)
Doing business is relatively easy†	5.819	(3.174)
Fear of entrepreneurial failure	-29.984^{**}	(12.160)
Inflation	0.346	(0.211)
GDP	0.000	(0.000)
GDP growth	$0.966^{***[***]}$	(0.223)
Real interest rate	$0.523^{***[***]}$	(0.076)
Lack of corruption‡	0.895	(0.725)
Constant	$64.653^{***[***]}$	(6.425)
	170 obs.; $R^2 = 0.418$	

Notes: F-statistic for instrument weakness with heteroskedasticity-robust errors: 86.929; F-statistic for instrument weakness with errors clustered at country level: 161.737; †dummy is 1 if yes and zero else; ‡increase indicates less corruption; ***significant at the 1%-level; **significant at the 5%-level; $^{[***]}$significant at the 1%-level with country-level clustering; $^{[**]}$significant at the 5%-level with country-level clustering; standard errors in parentheses are heteroskedasticity-consistent.

According to first-stage F-statistics, which are both larger than 100, the instrument is sufficiently strong independent of whether we use heteroskedasticity-robust errors or cluster errors at the country level. As in the micro panel model, natural disasters are negatively correlated to the media variable: There is a crowding-out effect. An additional natural disaster is associated with a reduction in media attention for entrepreneurship of 1.4 percentage points.

Table 4.5 presents results generated by IV regressions, with Table 4.4 as first stage.[62] As positive media reports about entrepreneurship increase the probability of self-employment, we find support for Hypothesis 4.1. Table 4.5 also provides support for Hypothesis 4.2: Media reports about entrepreneurial success reduce the probability of wage

[62]Two-stage least squares results significantly differ from OLS results, given in Table 4.C.2 (Appendix 4.C). OLS tends to underestimate effects.

Table 4.5. IV estimates of marginal effects in macro panel model, where dependent variable is transformed choice share

	Self-employment		Wage work	
Variable	Coefficient	SE	Coefficient	SE
Media attention for entrepreneurship§	0.028$^{***[***]}$	(0.004)	−0.021$^{***[***]}$	(0.003)
Doing business is relatively easy†	−0.678$^{***[**]}$	(0.158)	0.473***	(0.121)
Fear of entrepreneurial failure	−0.014	(0.667)	0.163	(0.532)
Inflation	−0.023	(0.015)	0.017	(0.012)
GDP	0.000***	(0.000)	0.000$^{***[**]}$	(0.000)
GDP growth	−0.008	(0.014)	0.016	(0.011)
Real interest rate	−0.005	(0.004)	0.001	(0.003)
Lack of corruption‡	−0.047	(0.033)	0.074***	(0.023)
Constant	−2.025$^{***[***]}$	(0.392)	0.958	(0.301)
		170 obs.		

Notes: §Media attention is instrumented by number of natural disasters in other countries; †dummy is 1 if yes and zero else; ‡increase indicates less corruption; ***significant at the 1%-level; **significant at the 5%-level; $^{[***]}$significant at the 1%-level with country-level clustering; $^{[**]}$significant at the 5%-level with country-level clustering; standard errors in parentheses are heteroskedasticity-consistent.

work. The effects are driven by differences *between* countries, as after the inclusion of country fixed effects (not presented here[63]) media effects become insignificant.

The exponential of the coefficient of media attention can be interpreted as an effect on odds[64] when all other variables are held constant. A one percentage point increase in positive media attention for entrepreneurship increases the odds of self-employment by 2.8% and decreases the odds of wage work by 2.1%.

4.5.3.4 *Robustness*

Assignment of countries

Using a different assignment of countries with natural disasters produces similar effects. In Figure 4.3, we show the estimated effects on self-employment and wage work (all significant at the 5%-level using errors clustered at the country level) of the 15 best unique assignments.

The minimum effect on self-employment is 0.01 (effect on odds: increase by 0.92%), the maximum is 0.04 (effect on odds: increase by 4.52%), whereas the average effect is 0.02 (effect on odds: increase by 2.51%). The minimum effect on wage work is −0.01 (effect on

[63] Results available from the authors on request.

[64] The odds of occupation $i \in \mathbb{O}$ are $p_i(1-p_i)^{-1}$.

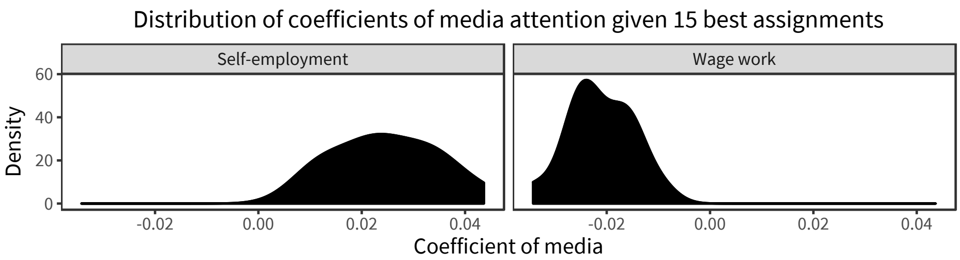


FIGURE 4.3. Using different country pairings, resulting in different instruments

odds: decrease by 0.89%), the maximum is −0.03 (effect on odds: decrease by 3.40%), while the average effect is −0.02, which is the effect established with the best assignment.

Beta model as alternative to logit transformation

There is an open concern that our models might be misspecified because the logit transformation does not fully remove skewness from our dependent variables. For instance, the distribution of transformed wage work shares in Figure 4.B.4 (Appendix 4.B.2) is clearly skewed. Thus, our results might be mostly driven by modeling assumptions.

To reduce the danger of model misspecification (especially, the danger that results are driven by skewness), we model the original, non-transformed, shares with beta regressions.[65] The beta regression, proposed by Ferrari & Cribari-Neto (2004), accommodates skewness and heteroskedasticity, as values near zero and 1 have typically a smaller variance than other values in the (0, 1) interval. We employ a two stage procedure. The first stage is (4.16), estimated in Table 4.4, while the second stage is the beta regression. We use estimated residuals from the first stage as an additional predictor in the second stage (Newey 1987; Terza, Basu & Rathouz 2008).[66] Second stage confidence intervals are bootstrapped with clustering at the country level in line with Efron (1987).

Beta regression results presented in Table 4.6 clearly support Hypothesis 4.1 and 4.2, as media attention for entrepreneurship significantly increases the average share of the self-employed and reduces the average share of wage workers. Thus, the micro and macro panel models, based on two different data sets, support Hypothesis 4.1 and 4.2, derived from our theoretical model of career choice under ambiguity.

[65] In line with suggestions of Ferrari & Cribari-Neto (2004), the beta distribution is parameterized in terms of its mean and precision (a large precision corresponds to a small variance). A linear combination of predictors is linked to the mean by a logit link. Consequently, a positive estimated coefficient of a predictor can be interpreted as a positive effect on the average share and *vice versa*.

[66] Using predicted values from the first stage yields numerically very similar media effects.

TABLE 4.6. IV beta estimates of marginal effects on original shares using macro panel, where dependent variable is original choice share

	Self-employment		Wage work	
Variable	Coefficient	95% CI	Coefficient	95% CI
Media attention for entrepreneurship	0.028**	0.02, 0.04	−0.021**	−0.03, −0.01
Residuals from first stage	−0.017**	−0.03, −0.01	0.013**	0.00, 0.02
Doing business is relatively easy†	−0.592**	−0.90, −0.28	0.433**	0.18, 0.68
Fear of entrepreneurial failure	−0.306	−1.62, 0.87	0.398	−0.54, 1.42
Inflation	−0.028**	−0.07, −0.00	0.019	−0.00, 0.05
GDP	0.000**	0.00, 0.00	0.000**	0.00, 0.00
GDP growth	−0.009	−0.05, 0.02	0.018	−0.00, 0.05
Real interest rate	−0.006	−0.01, 0.00	0.002	−0.01, 0.01
Lack of corruption‡	−0.057	−0.11, 0.00	0.077**	0.04, 0.12
Constant	−1.711**	−2.56, −0.98	0.801**	0.18, 1.41
Precision parameter	36.263**	27.89, 41.43	46.497**	34.59, 54.74
		170 obs.		

Notes: †Dummy is 1 if yes and zero else; ‡increase indicates less corruption; **significant at the 5%-level; confidence intervals (CI) are bootstrapped (2,000 replications) at the country level.

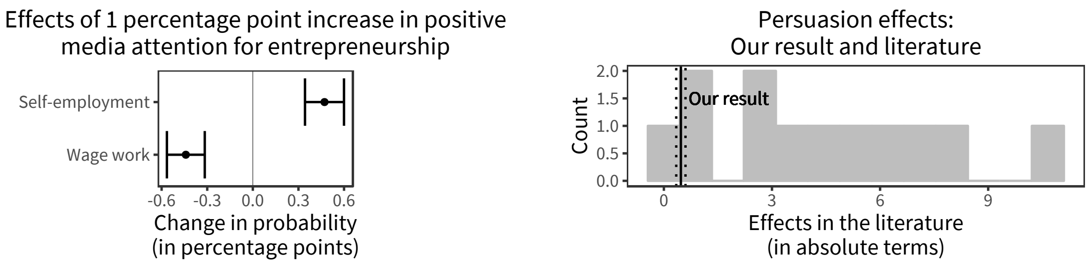
4.5.3.5 *Effect sizes*

In the micro panel, effect sizes cannot be properly interpreted, because of the construction of the Google Trends variable. However, the macro panel allows for a simple interpretation. To approximate effect sizes, we estimate linear probability models, where Table 4.4 is the first stage.

Figure 4.4a shows effects of a 1 percentage point increase in media attention for entrepreneurship, including 95% confidence intervals (full results are in Table 4.C.4 in Appendix C). The probability to select self-employment increases by 0.47 percentage points and the probability to select wage work decreases by 0.44 percentage points.

In Figure 4.4b, we compare the persuasion effect established by us—the 0.5 percentage point increase in the probability to select self-employment—to persuasion effects found in the literature[67] on media effects (viz., Ansolabehere & Iyengar 1995; Gerber & Green 2000; Green, Gerber & Nickerson 2003; Kull, Ramsay & Lewis 2003; Gentzkow & Shapiro 2004; Della Vigna & Kaplan 2007; Gerber, Karlan & Bergan 2009). Our result is consistent with previous findings. The standard deviation of the media variable in our sample is

[67] We use data on media effects compiled by Della Vigna & Kaplan (2007, Table IX). The persuasion effect is computed as the absolute difference in the outcome variable between treatment and control group. All outcome variables are shares so that effects are comparable to our results.



(a) Estimated effect sizes

(b) Persuasion

FIGURE 4.4. Effect sizes in macro panel according to linear probability models

approximately 15 percentage points so that even the small effect size leads to substantial effects given the variation of positive media attention for entrepreneurship. The effect of media on the probability of self-employment is comparable to the effect of watching the Fox News channel on the Republican vote share (a gain of 0.4 to 0.7 percentage points) found by Della Vigna & Kaplan (2007).

4.5.4 *Limitations and further research*

Our approach has several limitations. First of all, as already mentioned, we cannot directly test the theoretical model, but only test for consistency of derived hypotheses. Also, as we are not able to compute optimal choice probabilities in our empirical analysis, we cannot determine whether there are too many or too few choices of self-employment. Therefore, our empirical model is restricted to testing whether positive media reports about (famous) entrepreneurs influence choices in the expected direction.

Furthermore, we cannot identify voluntary unemployment in data. However, preliminary results derived from the theoretical model suggest that media reduces voluntary unemployment by reducing ambiguity about employment options. A reduction in voluntary unemployment caused by media could be seen as a positive effect since unemployment generates negative psychological effects (Paul & Moser 2009) and impairs the generation of valuable information about success probabilities leading to lower life-time earnings (Verbruggen et al. 2015).

Last but not least, the levels of aggregation of our media variables are rather high (region and country levels). It would be more preferable to use media consumption at the individual or household level, and observe the household or individual over a sufficiently long period of time. This would allow to determine more directly whether media influences individuals in their occupational choices. Consequently, further research is necessary.

4.6 Concluding summary

There is no question that media wields significant power in modern societies. Our theoretical and empirical analysis adds a new effect to the literature: Media affects occupational choices. In our theoretical model, we show under what conditions media influences occupational choices, and in which circumstances this influence is positive or negative. We demonstrate that ambiguity-averse individuals might not make individually optimal choices, in the sense that they are not selecting those options that yield the highest expected success probabilities. Instead, they might exhibit a bias for, respectively against, some occupational option due to asymmetric ambiguity aversion. We show that sufficiently intensive positive media reports about entrepreneurs, transporting ambiguity-reducing information, increase the probability of selecting self-employment, while the probability of wage work is reduced. In case of asymmetric ambiguity preferences biased against self-employment, when ambiguity aversion related to self-employment is sufficiently higher than to wage work, media reduces a behavioral bias against self-employment. In that way, our approach also complements recent research according to which entrepreneurship can be modeled as an experiment (Daly 2015; Dillon & Stanton 2016). Positive media reports may foster such experimentation with an occupational option.

Given micro-level data and country-level panel data, we test central predictions from our theoretical model, in particular to what extent media affects choice probabilities. For that reason, we estimate multiple instrumental variable regressions to determine the empirical effects of media. In line with our theoretical model, we establish that the consumption of positive media articles and reports about entrepreneurs significantly increases the probability of self-employment and significantly reduces the probability of wage work.

To conclude, media reports can foster self-employment, while reducing wage work, by providing information that changes individual beliefs. Informational shocks can, thus, have a significant impact on career choices. The established effects are sufficiently large to be of interest; allow for a causal interpretation; are based on observations from two data sets using two different empirical models; and are robust to model specification. However, our regression approaches, relying on repeated cross sections or aggregated data, cannot directly evaluate individual decision histories. Yet, our theoretical model can be used to generate further testable predictions with respect to individual short- and long-run decision behavior, opening up avenues for further research.

5 BUSINESS CYCLES AND START- UPS ACROSS INDUSTRIES: AN EMPIRICAL ANALYSIS FOR GERMANY

In this chapter, we analyze whether start-up rates in different industries systematically change with business cycle variables. We mostly find correlations that are consistent with counter-cyclical influences of the business cycle on entries in both innovative and non-innovative industries . Entries into the large-scale industries, including the innovative part of the manufacturing sector, are more strongly influenced by changes in the cyclical component of unemployment, while entries into small-scale industries, like the knowledge intensive services, are merely influenced by changes in the cyclical component of GDP. Business formation may therefore have a stabilizing effect on the economy.

5.1 Introduction

Research postulates that macroeconomic factors, such as the business cycle and unemployment levels, influence the number of business entries (Parker 2012a; Koellinger & Thurik 2012). Theoretical considerations suggest that these two variables may unfold either pro- or counter-cyclical effects (Bernanke & Gertler 1989; Hopenhayn 1992; Francois & Lloyd-Ellis 2003), thus making the relationship between the macroeconomic variables and entrepreneurial entry ambiguous. The direction of the effect is, however, of crucial importance. Pro-cyclical effects may amplify positive and negative economic shocks that may overheat the economy during boom periods and slow down recovery during recessions. Counter-cyclical effects would be beneficial for the economy when the opening of more businesses would spur economic recovery in recessions, while a decline of business entries in boom periods would not further enhance growth.

Empirical analyses of how these two macroeconomic factors influence business entries, report mixed results (Parker 2012a; Sanchis Llopis, Millán, Baptista, Burke, Parker & Thurik 2015). This holds for longer time periods,[68] as well as for recent shocks like the great recession after the financial crisis of 2008.[69] However, it is not just the direction of the effects that is unknown, but also the composition of new businesses started during

[68]Grant (1996) and Lee & Mukoyama (2015) find pro-cyclical effects for the US, while Georgellis & Wall (2000) report counter-cyclical effects for the UK. Fritsch, Kritikos & Pijnenburg (2015) also find counter-cyclical effects for Germany.

[69]Siemer (2014) finds pro-cyclical effects for the US, while Hundt & Sternberg (2014b) find counter-cyclical effects for Germany.

these periods. In what way do the different stages of the cycle affect innovative start-ups with potential for growth and affect marginal businesses with little or no impact on the economy? In this context, Barlevy (2007) claims that radical innovation positively affects further new businesses to be ventured during boom periods, while Ghatak, Morelli & Sjöström (2007), Román, Congregado & Millán (2013), and Koellinger & Thurik (2012) argue that recessions may especially stimulate the formation of marginal businesses because of falling wages and lower opportunity costs of entrepreneurial activity. Could it be that, even in countries where counter-cyclical relations between the cycle and business entries are observed, innovative businesses with growth potential are more likely to be started during boom periods? The answer to this question is crucial in order to assess whether new business formation has a stabilizing or a destabilizing effect over the cycle.

Therefore, this chapter investigates how much of the observed variations of the start-up rates in different industries can be attributed to changes in GDP growth and to the unemployment level. Using the start-up information of the ZEW Enterprise Panel, we distinguish between innovative and non-innovative industries, as well as between large and small scale industries. We perform the analysis at the NUTS 2 region level so that we are able to control for regional differences with regard to a number of further factors that may influence entrepreneurial entry, in particular, knowledge spillovers and employment in small businesses.

We mostly find correlations that are consistent with counter-cyclical effects on both innovative and non-innovative industries. Entries into large-scale industries including the innovative part of the manufacturing sector increase when unemployment is high and *vice versa*, while there is a significantly negative relationship between changes in GDP and entries into small-scaled businesses, including the knowledge intensive services. Exceptions from this general pattern are the energy-and-mining and the credit-and-insurance sectors. Results remain robust when analyzing several sources of a potential bias (detrending technique, unobserved spatial links, endogenous errors, etc.).

In the following, Section 5.2 summarizes the current state of research on how business cycles relate to new business formation, and presents and motivates the research questions. We then introduce the data in Section 5.3. Section 5.4 presents our empirical approaches and describes the results of the analysis. Section 5.5 concludes.

5.2 Start-ups over the business cycle: Theoretical and empirical research

There is considerable variation in the level of new business formation across industries (Audretsch 1995; Falck 2007). Such difference can be attributed to several factors, including an industry-specific minimum efficient size, qualification requirements, the expected development of demand, and the availability of industry-specific inputs (for an overview, see Parker 2009). Due to these differences, the effect of the business cycle may considerably vary across industries. This section reviews previous theoretical and empirical literature on how cyclical changes in GDP (Section 5.2.1) and unemployment (Section 5.2.2) relate to the entry of new businesses and derives our main research questions (Section 5.2.3). In addition, the research questions are motivated with a simple model of entrepreneurial entry (Section 5.2.4).

For clarification, we say that the effect of GDP is pro-cyclical if the number of entries increases when GDP growth is high. The effect of unemployment is said to be pro-cyclical if the number of new businesses increases when unemployment is low. If the number of entries increases when GDP growth is low or unemployment is high, we speak of a counter-cyclical effect. The effect is a-cyclical if no (statistically significant) relationship between the business cycle variables and new business formation can be established.

5.2.1 *Effects of changes in GDP on start-ups*

Various theories were developed to understand how start-ups, including innovative start-ups, react to changes in business cycles (for an overview, see Parker 2012a). Caballero & Hammour (1994) propose a Schumpeterian model of creative destruction where new businesses entering the market are more productive than old businesses, such that firm entry drives less productive incumbents out of the market. Given that in their model total current demand drives entry and exit, they show that we should expect increasing entry rates during economic upswings, while there should be relatively low levels of new business formation during recessions. Thus, the model proposes a pro-cyclical relationship between variations in GDP and business entries.[70] In a similar direction, Clementi & Palazzo (2016) argue that in response to boom periods, where firms usually realize higher profits, more entrepreneurs may feel attracted to enter markets.

This pro-cyclical effect of GDP changes on start-ups may, however, be limited if we also consider supply side effects. Also, the development of resource prices is usually related to changes in GDP development. Production costs, rents, wages, and other relevant costs for business entries are typically lower in recessions, while boom periods have higher costs (Lewis 2009). Hence, low entry costs might make investments into new businesses more attractive during recessions pointing to a counter-cyclical effect on entry.

Francois & Lloyd-Ellis (2003) and Barlevy (2007) present models where firms are explicitly considered to invest into invention activities. In both models, demand expectations play a central role, but the two models differ in one crucial aspect, namely in the timing when firms turn their inventions into innovations. While Francois & Lloyd-Ellis

[70] As Caballero & Hammour (1994) focus on the relative number of entries in relation to business closures, they also clarify that in case of increasing business entries during recessions, increasing entry rates need to be overcompensated by even higher numbers of business exits given the reduced overall demand in such times.

(2003) propose that more firms start to commercialize their ideas during recessions, thus leading to a counter-cyclical influence of business cycles on entrepreneurial entries, Barlevy (2007) makes a case for the opposite, namely that more innovation takes place during boom periods leading to a pro-cyclical effect.

Francois & Lloyd-Ellis (2003) argue in their model that since resource costs are lower during recessions, entrepreneurs starting firms sensitive to such entry costs might prefer to launch their businesses during recessions even if they made their invention in a boom period. Barlevy (2007) explains that firms make more investments into inventions during boom periods and would not risk any delay, even if such a delay would be efficient from a cost point of view. He argues that entrepreneurs will not risk that potential competitors might "take their ideas away," i.e., commercializing the same idea at a higher speed. Thus, the main reasons for the contradicting expectations of the two models are timing and opportunity costs.

Koellinger & Thurik (2012) provide a different reason for why there may be more start-ups in recessions that might be particularly relevant for innovative new businesses. By applying the prospect theory to start-up decisions, they argue that "innovative business ideas that entail high risk are more likely to be pursued by individuals who suddenly have lower opportunity costs of self-employment than before, for example, as result of a salary cut or of unemployment in a recession" (Koellinger & Thurik 2012, p. 1153). This is because in such situations of a loss position in relation to the prospect theory's reference point, individuals might be more willing to take risk and act less uncertainty averse. According to this argument there should be more innovative start-ups during recessions.

Empirical evidence is inconclusive. Lee & Mukoyama (2015), who restrict their analysis to manufacturing plants, and Clementi & Palazzo (2016), who analyze the overall entry rates for the US, report pro-cyclical effects of output growth. In turn, Glaeser, Kerr & Ponzetto (2010) show that more entrepreneurs enter the market when fixed costs are lower, which usually holds for recessions. Fritsch et al. (2015), who analyze the overall entry rates for Germany, find that new business formation is higher during recessions than during boom periods. Diverging observations can also be found for the great recession in the aftermath of the financial crisis of 2008. Siemer (2014, p. 1) observes an "unprecedented decline in firm entry" in the US, while Hundt & Sternberg (2014b, p. 740) conclude that "the crisis had the effect of supporting entrepreneurial activities in Germany in general." Hence, changes in GDP or output levels can have both pro-cyclical and counter-cyclical effects on business entries (see Hundt & Sternberg 2014a). Even more so, it remains unclear which effect prevails in what type of industry.

5.2.2 *Effects of variations in unemployment on start-ups*

The second business cycle variable, the level of unemployment, may unfold effects that are different from those of the development of GDP or output. When unemployment is high, it seems plausible to assume that more individuals will set up a business out of unemployment (Caliendo & Kritikos 2010). However, these kind of new businesses may favor industries where starting a firm requires relatively few resources and where the minimum efficient size is comparatively low, such as in small-scale services (Román et al. 2013). Consequently, high levels of unemployment may particularly induce entries by the extensive margin. During times of low unemployment, entry into such industries can be expected to be lower because it is easier to find a job in dependent employment.

In contrast, the prosperity-pull hypothesis argues that during times of low unemployment, newly ventured businesses face higher consumer and firm demand for their products and services as more people have jobs, thus increasing their potential profits, and *vice versa*. If that influence should prevail, fewer businesses would be expected to be opened during times of high unemployment and there should be more start-ups when unemployment is low (Parker 2012b).

Both the unemployment-push and the prosperity-pull hypothesis do, however, not account for a further important issue: the availability of resources that may be particularly relevant for firm formation at the intensive margin, i.e., when entrepreneurs aim to establish larger businesses with significant numbers of dependent employees. One may argue that such more ambitious start-ups find more favorable conditions during periods of high unemployment not primarily because the business founders themselves are unemployed, but because high unemployment improves the availability of labor at relatively low wages. This may be particularly relevant for the venturing of innovative businesses with demand for labor (Francois & Lloyd-Ellis 2003). Thus, high unemployment may also have a positive effect on the number of entries into industries, where business founders aim to create innovative and large-scale businesses.

Examining the empirical evidence, Robson (1998) finds no support for a recession-push effect for Great Britain, while Georgellis & Wall (2000), in their analysis for Great Britain, report a positive relationship between a rising level of unemployment and entrepreneurial entry. Foti & Vivarelli (1994), Fairlie (2013), and Fritsch et al. (2015) arrive at similar results, for Italy, the US, and Germany.[71] Hence, the overall picture regarding the effect of unemployment on new business formation is more conclusive, as nearly all studies point to a counter-cyclical influence, i.e., more businesses are opened when unemployment is high. We are not aware of any evidence supporting the prosperity-pull hypothesis.

This brief outline of the possible relationships between cyclical variations in macroe-

[71] There is more empirical evidence for other countries: see Parker (2012b) or Fritsch et al. (2015).

conomic variables and new business formation shows that it is far from clear how these factors, in particular changes in GDP, influence start-up behavior. Since the different effects may only apply to certain types of entry, it is important to distinguish between such types in terms of industries, potential firm sizes, and innovativeness. During times of recessions and of economic prosperity, the entry costs, costs of resources as well as opportunity cost, on the one hand, and current profits and expectations about future profits on the other, can be assumed to unfold different effects on different industries.

5.2.3 *Industry-specific effects of macroeconomic variables: Research questions*

Our central concern is how cyclical variables relate to new business formation conditional on different kinds and sizes of businesses. To distinguish between different kinds of business entries, we follow Holmes & Schmitz (1990) by assuming that individuals with low skills will start firms that have a low potential for innovation, while individuals with high skills will start firms that have a higher potential to introduce innovations. In addition to considering two different types of entrepreneurs with different potentials for innovation, we also assume that businesses can be of two different sizes. Many businesses will require no or only a small number of employees, so that they are basically run by the entrepreneurs themselves. Other businesses may, however, demand a larger number of paid employees. Thus, overall we consider four types of entrepreneurs and their businesses that combine innovation potentials and scale:

Type 1 entrepreneurs are low skilled and have a low potential for innovation, while they expect to employ no or only a small number of workers. Examples include trade or consumer services.

Type 2 entrepreneurs may also have a low potential for innovation, but they attempt to create large-scale businesses. Businesses of this type include, for instance, start-ups in the non-innovative manufacturing industry.

Type 3 entrepreneurs are highly skilled and have a potential to introduce innovations but their business is expected to be small-scale, e.g., firms in knowledge intensive services.

Type 4 consists of highly skilled entrepreneurs with an innovation potential running large-scale businesses, for example innovative manufacturing.

Given these four types we may speculate about type-specific effects. Our review of previous research (Section 5.2.1) demonstrated that variations in the GDP level can on the one hand unfold demand side effects (i.e., current or future profit opportunities), and on the other hand have supply side effects (i.e., changes of entry cost, early stage production cost, as well as resource cost for labor and capital). Which of these effects prevails is an empirical question. We expect that the more innovative a start-up is, the less current demand and the more demand expectations do play a role. Moreover, the larger the expected size of a newly ventured business, the less will the entrepreneur's opportunity costs matter and the larger the role of other cost categories. Thus, correlations between GDP and business entries will differ by size and innovativeness of the firms.

As to the second business cycle variable, unemployment, we expect that variations in the unemployment rate may influence potential entrepreneurs in different ways depending on their skill level and the planned size of their business. For individuals with low skills, an increase in unemployment makes it more difficult to find a job, such that avoiding unemployment leads to a stronger incentive to start an own business. Individuals with high skills, who may also have a higher potential for introducing innovations, can be assumed to be less influenced by changes in the general level of unemployment as they may move more easily between alternative job opportunities. For this type, there are less incentives of venturing a business in order to avoid unemployment. Entrepreneurs who plan to venture businesses with a larger number of employees may be more likely to start their ventures in times of high unemployment. Given that there is no evidence of a prosperity-pull effect of a low level of unemployment, the unemployment rate can be expected to generate counter-cyclical effects. Based on these considerations, we investigate the following three questions.

Question 1
Size- and innovation-potential-specific differences. To what extent are there differences in the influence of cyclical variables on new business formation, when distinguishing between (a) large and small-scale industries and (b) innovative and non-innovative industries?

Question 2
Cyclical variations in GDP. (a) Are demand side effects such as expectations about current or future profit opportunities more or less important for new business formation than supply side effects (e.g., opportunity costs, production and resource costs, other entry-related costs), leading to pro-, a-, or counter-cyclical influences of GDP on business entry? (b) To what extent does the innovativeness and the expected size of a ventured business matter in the cyclical relationship between changes in GDP and business entries?

Question 3
Cyclical variations in unemployment. (a) To what extent do changes in the unemployment level unfold counter-cyclical influences on entries of firms with a low innovation potential that do not expect to employ a large number of personnel (Type 1)? (b) To what extent do changes in the unemployment level unfold any influence on entries of firms with a high innovation potential that do not intend to employ a large amount of personnel, but where the entrepreneurs themselves are less affected by unemployment risks (Type 3)? (c) To what extent do changes in the unemployment level unfold counter-cyclical influences on

entries of firms that expect to employ a larger number of employees (Type 2 and 4) at times of high unemployment when labor is more easily available?

5.2.4 *Motivating the research questions: A simple model of entrepreneurial entry*

To provide a theoretical background for cyclical effects of GDP and unemployment on new business formation, conditional on different kinds of businesses, we construct and solve a simple model of entrepreneurial entry. As demonstrated in the previous sections, there are two types of variables driving the entry decision: current and future profit opportunities and entry costs, which also include opportunity costs (the potential costs of not working for a wage). Also, a potential entrepreneur has to solve the problem of funding, i.e., acquiring capital.

We combine all these factors by constructing and solving a basic matching model, based on Mortensen & Pissarides (1994), with three types of agents. In the model, entrepreneurs and investors are paired such that they can start a business, while entrepreneur-investor pairs can also be separated—generating new business formation as well as market exits. In contrast to models in the literature, such as Bernanke & Gertler (1989) and Rampini (2004), which generally predict that entrepreneurship is pro-cyclical, we take innovation potential, resource costs, and different sizes of potential businesses into account. Consequently, we establish conditions when start-ups are pro-, counter-, or a-cyclically related to GDP.

Agents

All model agents are business-cycle takers, i.e., they react to changes in effective wages and sales opportunities, which are, in turn, determined by exogenous shocks, but cannot influence them. Furthermore, all agents are price takers. The first two types are investors, denoted by $\mathfrak{i}$, and potential entrepreneurs, denoted by $\mathfrak{l}$. Potential entrepreneurs have entrepreneurial ideas. To realize an idea, a potential entrepreneur must be paired with an investor, as the investor is endowed with resources that can be used to start a business—investors cannot start a business on their own because they lack entrepreneurial ideas, while potential entrepreneurs lack the necessary resources. Thus, potential entrepreneurs can be of two different subtypes. They can be actual entrepreneurs, denoted by $\mathfrak{e}$, if they are paired with an investor; or they can be wage workers, denoted by $\mathfrak{w}$, if a match did not take place or a match was destroyed. The third type of agent are pure workers who work for entrepreneurs, but who do not have entrepreneurial ideas or lack the willingness or ability to put such an idea into practice.

For the sake of simplicity, we assume that the supply of investors and pure wage workers is potentially infinite and that entry for investors is free. The number of potential entrepreneurs is normalized such that $\mathfrak{l}(t) = \mathfrak{l} = 1$ for all t, where time is continuous and the type variable denotes the number of agents. Given a fixed number of potential entrepreneurs, the number of actual entrepreneurs in period t is $\mathfrak{e}(t) = 1 - \mathfrak{v}(t)$.

Investors and potential entrepreneurs are risk-neutral and discount future income with rate $\delta \in (0, 1)$. Pure wage workers have only one source of income such that their risk preferences and discounting factor do not matter. Pure wage workers and potential entrepreneurs in wage work receive a utility flow corresponding to the expected wage. More specifically, pure wage workers receive $(1 - \vartheta)w(y)$, where the real wage $w(y) > 0$ increases in sales opportunities $y > 0$ (increasing sale opportunities represent increasing demand) and $\vartheta \in (0, 1)$ is the general probability of unemployment. An increase in unemployment makes it harder to find a job such that the expected wage is reduced. Potential entrepreneurs in wage work receive utility flow $\beta(1 - \vartheta)w(y)$, where $\beta \in \{1, (1 - \vartheta)^{-1}\}$ describes how strongly potential entrepreneurs are directly influenced by the general unemployment rate. We assume (as in Holmes & Schmitz 1990) that individuals with low skills have a low potential for innovation (corresponding to necessity entrepreneurs), and are strongly influenced by movements in the unemployment rate such that $\beta = 1$. For this type of individual, avoiding unemployment—in the model, higher unemployment reduces the expected wage—is a strong incentive to start an own business. Individuals with high skills who have a high potential for introducing innovations (corresponding to opportunity entrepreneurs) are assumed to be barely influenced by changes in the general level of unemployment such that $\beta = (1 - \vartheta)^{-1}$. For this type, there is no incentive to avoid unemployment by venturing a business.

An entrepreneur produces and sells y, pays $\gamma(1 - \vartheta)w(y)$ to pure wage workers,[72] where $\gamma \geqslant 0$ is the number of workers required to produce the product, and receives entrepreneurial income π. An investor, matched with an entrepreneur, receives an income flow of $y - \gamma(1 - \vartheta)w(y) - \pi$. In addition to considering potentials for innovation, we also assume that businesses can be of two different sizes: First, businesses may be driven by entrepreneurs, not requiring a significant number of employees, such that $\gamma = 0$. An example of (solo-) entrepreneur driven businesses are firms providing services that mostly require specialized knowledge but not extensive amounts of labor (such as in knowledge intensive services). Second, other businesses may require a larger number of paid employees (non-entrepreneurial labor), $\gamma \gg 0$, for example, in manufacturing.

Matching

To generate a match, a combination of an investor and an entrepreneur constituting a firm, investors post project vacancies $\mathfrak{p}$. An investor pays a per-period cost given by $c > 0$ to post a project vacancy and to be ready for a business start-up. The matching rate of investors

[72] The model only accounts for resource costs, where resources are pure wage workers needed to produce y. Resource availability (the number of wage workers available for employment) has similar effects, as increasing scarcity of resources leads to higher resource costs.

and potential entrepreneurs is $\theta(t)$. The matching rate of project vacancies is $\phi(t)$. New matches are formed according to a Poisson process. To create a match, individuals rely on a matching technology

$$M(\mathfrak{p}, \mathfrak{v}) = \mathfrak{p}^{1-\varepsilon}\mathfrak{v}^{\varepsilon}$$

where $\varepsilon \in (0, 1)$. Hence, the matching rate of investors and potential entrepreneurs is

$$\theta(t) = \frac{\mathfrak{p}(t)^{1-\varepsilon}\mathfrak{v}(t)^{\varepsilon}}{\mathfrak{v}(t)} = \lambda(t)^{1-\varepsilon}$$

where $\lambda(t) \equiv \mathfrak{p}(t)\mathfrak{v}(t)^{-1}$. The matching rate of project vacancies is

$$\phi(t) = \frac{\mathfrak{p}(t)^{1-\varepsilon}\mathfrak{v}(t)^{\varepsilon}}{\mathfrak{p}(t)} = \lambda(t)^{-\varepsilon}$$

Matches are separated according to a Poisson process with a time-invariant, exogenous, separation rate σ.

Steady state

The model is a standard matching model and the steady state can be easily determined (for details, see, for instance, Hornstein, Krusell & Violante 2005). It is easy to show that the number of potential entrepreneurs in wage work develops according to

$$\dot{\mathfrak{v}}(t) = \sigma - [\theta(t) + \sigma]\mathfrak{v}(t)$$

Using the steady state condition $\dot{\mathfrak{v}}(t) = 0$, we obtain

$$\mathfrak{v}^* = \frac{\sigma}{\theta^* + \sigma} \tag{5.1}$$

where values with an asterisk are steady-state values. To fully specify the steady-state number of entrepreneurs, we must determine θ^* in (5.1). Let η_+ denote the steady-state asset value of matching and η_- the steady-state asset value of not matching. The discounted value of a match for an investor is

$$\delta\eta^{\mathfrak{i}}_+ = y - \gamma(1 - \vartheta)w(y) - \pi + \sigma[\eta^{\mathfrak{i}}_- - \eta^{\mathfrak{i}}_+]$$

and the discounted value of not matching is

$$\delta\eta^{\mathfrak{i}}_- = \beta(1 - \vartheta)w(y) + \theta^*[\eta^{\mathfrak{i}}_+ - \eta^{\mathfrak{i}}_-]$$

However, only $\eta^{\mathrm{i}}_{-} = 0$ is consistent with free entry of investors, such that we obtain

$$\eta^{\mathrm{i}}_{+} = c(\phi^*)^{-1} = (y - \gamma(1 - \vartheta)w(y) - \pi)(\delta + \sigma)^{-1}$$

The discounted asset value of a match for potential entrepreneurs is

$$\delta\eta^{\mathrm{I}}_{+} = \pi + \sigma[\eta^{\mathrm{I}}_{-} - \eta^{\mathrm{I}}_{+}]$$

and the discounted asset value of not matching is

$$\delta\eta^{\mathrm{I}}_{-} = \beta(1 - \vartheta)w(y) + \theta^*[\eta^{\mathrm{I}}_{+} - \eta^{\mathrm{I}}_{-}]$$

Solving the linear system given by $\delta\eta^{\mathrm{I}}_{+}$ and $\delta\eta^{\mathrm{I}}_{-}$ yields

$$\eta^{\mathrm{I}}_{+} = (\beta[1 - \vartheta]\sigma w(y) + [\theta^* + \delta]\pi)(\delta[\theta^* + \delta + \sigma])^{-1}$$

respectively

$$\eta^{\mathrm{I}}_{-} = (\beta[1 - \vartheta]\sigma w(y) + \beta[1 - \vartheta]\delta w(y) + \theta^*\pi)(\delta[\theta^* + \delta + \sigma])^{-1}$$

Investors will accept a match if $\eta^{\mathrm{i}}_{+} > \eta^{\mathrm{i}}_{-} = 0$ and potential entrepreneurs will accept a match if $\eta^{\mathrm{I}}_{+} > \eta^{\mathrm{I}}_{-}$ resulting in the following conditions for entrepreneurial income:

(C.5.1) $\quad y - \gamma(1 - \vartheta)w(y) > \beta(1 - \vartheta)w(y)$

such that a contract, acceptable to both sides, exists and

(C.5.2) $\quad \pi \in \{\pi \in \mathbb{R} : \beta(1 - \vartheta)w(y) < \pi < y - \gamma(1 - \vartheta)w(y)\}$

such that both parties agree. We assume that the income contract realized depends on the bargaining power of potential entrepreneurs $\rho \in (0, 1)$:

$$\pi = [1 - \rho]\beta(1 - \vartheta)w(y) + \rho[y - \gamma(1 - \vartheta)w(y)] \tag{5.2}$$

By plugging π in (5.2) into η^{i}_{+} and using $\lambda^* = (\phi^*)^{-1/\varepsilon}$, we obtain λ^*. Using $\theta^* = (\lambda^*)^{1-\varepsilon}$, we get

$$\theta^* = (1 - \rho)^{\frac{1-\varepsilon}{\varepsilon}} \left[\frac{y - (\gamma + \beta)(1 - \vartheta)w(y)}{c[\delta + \sigma]}\right]^{\frac{1-\varepsilon}{\varepsilon}} \tag{5.3}$$

Consequently, using $\mathfrak{e}^* = 1 - \mathfrak{v}^*$, (5.1), and (5.3), the steady-state number of entrepreneurs is

$$\mathfrak{e}^* = \begin{cases} \frac{\theta^*}{\sigma+\theta^*} & \text{if } y > (\gamma+\beta)(1-\vartheta)w(y) \\ 0 & \text{else} \end{cases} \tag{5.4}$$

Comparative statics

We, now, can determine how the parameters GDP, i.e., sales opportunities, y, and unemployment, ϑ, influence the steady-state number of entrepreneurs. Assume that entrepreneurs have a sufficiently high productivity level, $y > (\gamma + \beta)(1 - \vartheta)w(y)$, such that the steady-state number of entrepreneurs is strictly positive. We consider four types of businesses, which we introduced in Section 5.2.3. GDP effects are obtained by computing first-order derivatives of the steady-state number of entrepreneurs (given in [5.4]) with respect to sales opportunities and then checking for sign conditions. Unemployment effects are computed in a similar way, by evaluating $\partial\mathfrak{e}^*/\partial\vartheta$.

Type 1. As type-1 entrepreneurs have a low potential for innovations and their businesses are mostly driven by entrepreneurial labor, it is reasonable to assume that $\beta = 1$ and $\gamma = 0$ holds. Also, as type-1 entrepreneurs might not face substantial entry costs, it might be reasonable to assume that c is small. The effect of GDP is pro-cyclical if $w'(y) < (1-\vartheta)^{-1}$, a-cyclical if $w'(y) = (1-\vartheta)^{-1}$, and counter-cyclical if $w'(y) > (1-\vartheta)^{-1}$. Thus, if opportunity costs are sufficiently high in a boom, GDP can generate counter-cyclical effects. Effects of unemployment are counter-cyclical.

Type 2. Type-2 entrepreneurs have a low potential for innovations but they create large scale businesses, such that we can assume that $\beta = 1$ but $\gamma > 0$. The effect of GDP is pro-cyclical if $w'(y) < (1 + \gamma)^{-1}(1 - \vartheta)^{-1}$, a-cyclical if $w'(y) = (1 + \gamma)^{-1}(1 - \vartheta)^{-1}$, and counter-cyclical if $w'(y) > (1 + \gamma)^{-1}(1 - \vartheta)^{-1}$. The term $(1 + \gamma)^{-1}(1 - \vartheta)^{-1}$ represents a critical value for the increase in opportunity cost and resource costs during a boom. If opportunity and resource costs are too high when GDP is high, GDP effects can become counter-cyclical. Changes in unemployment generate counter-cyclical effects.

Type 3. Type-3 entrepreneurs are high-skilled with a high innovation potential and run small scale businesses such that $\beta = (1-\vartheta)^{-1}$ and $\gamma = 0$. The impact of GDP is pro-cyclical if $w'(y) < 1$, a-cyclical if $w'(y) = 1$, and counter-cyclical if $w'(y) > 1$. A counter-cyclical GDP effect is generated if opportunity costs increase more than sales opportunities. The effect of unemployment is a-cyclical.

Type 4. Type 4 are high-skilled entrepreneurs with a high innovation potential who are running large scale businesses. A corresponding parameter combination is $\beta = (1 - \vartheta)^{-1}$ and $\gamma > 0$. The effect of GDP is pro-cyclical if $w'(y) < (1 + [1 - \vartheta]\gamma)^{-1}$, a-cyclical if $w'(y) = (1 + [1 - \vartheta]\gamma)^{-1}$, and counter-cyclical if $w'(y) > (1 + [1 - \vartheta]\gamma)^{-1}$. Here, $(1 + [1 - \vartheta]\gamma)^{-1}$ is a critical value for the increase in opportunity cost and resource costs during a boom. If these costs are too high, GDP produces counter-cyclical effects. Changes in unemployment produce counter-cyclical effects.

Summary of model predictions

The model's prediction for the effects of unemployment is that the variable usually generates counter-cyclical effects, except for high-skilled entrepreneurs running small businesses. On the one hand, these types remain themselves mostly unaffected by the risk of unemployment; on the other hand, they do not employ others in large numbers, which is why we might expect that Type 3 does not react to changes in unemployment.

The effect of GDP depends on the balance between optimism, based on better sales opportunities, and the opportunity cost and resource costs of entrepreneurship (the alternative wage income and the costs of hiring labor), since an increase in GDP increases production cost. In line with Rampini (2004) and Barlevy (2007), it might be expected that optimism will dominate and that GDP will generate pro-cyclical effects if we consider Types 1 and 3, as these types are mostly based on entrepreneurial labor. For Types 2 and 4, relying on paid employees, the cost aspect will dominate if the number of wage workers employed is sufficiently large, as

$$\lim_{\gamma\to\infty}(1+\gamma)^{-1}(1-\vartheta)^{-1} = \lim_{\gamma\to\infty}(1+[1-\vartheta]\gamma)^{-1} = 0 \text{ and } w'(y) > 0$$

and the GDP cycle will produce counter-cyclical effects.

With respect to the research questions, our model predicts that size and innovation-potential differences in correlations between start-ups and business cycles (Question 1) are likely to occur. Furthermore, GDP effects are likely to mostly depend on size and not necessarily on the innovation potential (Question 2), as the need to employ a large amount of labor can produce a counter-cyclical GDP effect independent of the innovation potential of an entrepreneur—compare the GDP effects of Type 2 (non-innovative) and Type 4 (innovative). Last, but not least, unemployment is likely to be counter-cyclical if entrepreneurs have a low innovation potential or (plan to) employ other individuals in their companies (Question 3).

5.3 Data and classification of industries

5.3.1 Data description

For our analysis, we use data on start-ups from the Founder Panel of the Center for European Economic Research (ZEW-Mannheim). This source includes nearly every firm established between 1995 and 2013. These data are based on information from Creditreform, the largest German credit rating agency, and allow for identifying innovative start-ups based on their affiliation with certain industries. Like many other data sources on start-ups, these data may not have complete coverage of solo-entrepreneurs. However, once the firm is either registered, hires employees, asks for a bank loan, or unfolds reasonable economic activities, even as a solo entrepreneur, it is included and information is collected on the date when the firm was established, meaning that many solo-entrepreneurs are captured, along with the business founding date (for details see Bersch, Gottschalk, Müller & Niefert 2014).

In our analysis, we distinguish between a number of sectors and industries covering all parts of the private economy in Germany, except for agriculture.[73] We differentiate between industries according to their innovativeness and knowledge intensity. Within the manufacturing sector, we apply the common classification of industries according to their presumed innovativeness. In our first group, we include high-technology manufacturing industries, investing more than 8.5 percent of their annual turnover into Research and Development (R&D), as well as technologically advanced manufacturing industries with R&D intensities between 3.5 and 8.5 percent (OECD 2005; Gehrke, Schasse, Rammer, Frietsch, Neuhäusler & Leidmann 2010). These two types of industries are separated from the non-technology oriented manufacturing industries that invest less in R&D, as well as from the sectors of construction, and energy and mining.

Since the service sector is heterogeneous, we differentiate between a number of different sub-sectors, starting with the traditional services such as trade, transport and postal services. The business oriented services comprise the "other" business services and the knowledge intensive services, which again consist of technology oriented services (such as architectural and engineering activities, technical consultancy, and technical testing and analysis), as well as other non-technology oriented services. We further separate credit and insurance from the business services. Lastly, we include the consumer oriented services comprising hospitality, financial services, real estate services, health care, culture, sports and entertainment, social services, and education.

The analysis is performed at the level of NUTS 2 regions, to account for region-specific influences. The NUTS 2 level is chosen as it is the most spatially disaggregated level for which the sample size of the data on knowledge-intensive business formation is sufficiently large. Data on GDP at a spatially more disaggregated level, such as planning regions or NUTS 3, are considerably less reliable. Thus, the cross-sectional dimension of this analysis is 38.

Moreover, many investigations of the relationship between entrepreneurship and the business cycle use changes in the stock of entrepreneurs as the dependent variable (net entry). Our analysis is based on transitions into self-employment (gross entry). Gross

[73] Agriculture is excluded because new business formation in this sector represents a rather special case that is hardly comparable to other sectors.

entry is better suited to identify how macro-variables influence entrepreneurship at the industry level, as this variable provides information about the dynamics of the economy (see Caballero & Hammour 1994; Nickell 1996). Net entries conceal changes in the gross flows (for an extended discussion, see Fritsch et al. 2015). Hence, we expect to reveal the relevant relationships more reliably than an analysis based on net entry.

The information on the number of unemployed persons and unemployment rates are provided by the Federal Employment Agency (Bundesagentur für Arbeit). The unemployment rate is defined as the share of registered unemployed over the entire working population that comprises all employed plus the registered unemployed. The nominal GDP at the NUTS 2 level is annually provided by Volkswirtschaftliche Gesamtrechnung der Länder (Macroeconomic Accounting of the Federal States; Statistisches Bundesamt a, various volumes) for the period 1994-2012. In order to obtain real values of GDP, the nominal figures are divided by the annual Consumer Price Index (CPI) of the Federal Statistical Office (Statistisches Bundesamt b, various volumes). Since information on the CPI is only available for Federal States, but not for NUTS 2 regions, we divide the nominal GDP at the NUTS 2 level by the regional CPI of the corresponding Federal State.[74]

The cyclical component of the unemployment rate and of real GDP is generated applying the Hodrick-Prescott filter (Hodrick & Prescott 1997). This filter is a statistical tool widely used in empirical analyses for separating the cyclical component of economic development from the corresponding trend (Montoya & de Haan 2008). The two components are estimated in a way that, over long time periods, the sum of the deviations of the cyclical component from the trend is close to zero, thereby minimizing the variability of the growth component. The so-called smoothing or HP filter parameter determines the allowed variability of the growth component. The larger the HP filter parameter, the smoother the trend component is. Following Ravn & Uhlig (2002), we use a HP filter parameter of 6.25 for annual data.[75]

When focusing on start-ups of innovative businesses requiring significant investments into R&D, as well as those needing highly skilled employees, it is important to control for other macroeconomic factors; in particular, those related to knowledge spillovers that also influence entries in such industries. In our approach, for each NUTS 2 region we consider four factors that are consistent with the Knowledge Spillover Theory of Entrepreneurship (see Audretsch & Keilbach 2007b; Acs, Braunerhjelm, Audretsch & Carlsson 2009),

[74] We mostly concentrate on GDP and unemployment as there is data on these two variables varying across regions. Other variables providing information on the business cycle, such as interest rates, are not included because there is no data available at the region level.

[75] Using other plausible values of the HP filter parameter does not lead to any fundamental change in the results. We also discuss the use of a different filter in Section 5.4.3. The deviation from trend of GDP is computed on the basis of 18 observations per region, whereas to filter unemployment shares we have 17–19 observations per region.

namely the share of employees in small businesses, the share of workforce with a tertiary degree, the number of professors, and the number of patents as an R&D output measure.

Previous research shows that employees who have worked in small firms have a higher propensity to start an own business (Almeida & Kogut 1999; Elfenbein et al. 2010) and that the venturing of new businesses is higher in regions where the number of individuals holding a tertiary degree is high (see Audretsch & Keilbach 2007a).[76] We account for both variables by using information on the regional employment share in establishments with up to 20 employees and the share of highly qualified employees, both coming from the Establishment History File of the German Employment Statistics, which covers all employees subject to compulsory social insurance contributions (Spengler 2008).

It is also known that the proximity to universities matters, as start-ups seek to exploit the regional knowledge stock for spillovers from these institutions (Audretsch, Keilbach & Lehmann 2006; Fritsch & Aamoucke 2013). We approximate for the knowledge stock by providing information on the number of professors per 1,000 workforce population (see German Federal Statistical Office – Hochschulstatistik des Statistisches Bundesamt).[77]

Finally, a higher level of research output may also induce higher start-up rates (Shane 2001). To account for this influence, we include the yearly number of patent applications per 1,000 workforce population as a measure of a region's knowledge capital; this is provided by the PatStat data base. The low correlation between the number of professors and the number of patents (see Table 5.A.6 in the Appendix) suggests that the two variables represent unrelated types of knowledge that show considerable divergence in their importance across industries. Since a Breitung (2000) panel unit root test reveals that the levels of these independent variables are not stationary, they are included as growth rates. Given restrictions due to data availability and necessary transformations of independent variables[78] (like the computation of growth rates), our time dimension covers 13 years, from 1996 to 2008.

5.3.2 *Classification of industries*

In our conceptual analysis and the resulting research questions (Section 5.2.3), we distinguish between four types of start-ups according to the innovativeness and the potential size of their ventures. To answer these research questions with data for industries, we

[76] According to the Knowledge Spillover Theory of Entrepreneurship, such knowledge should be conducive to start-ups, particularly in innovative and knowledge-intensive industries (Acs et al. 2009). Fritsch & Aamoucke (2013) find that such effects of regional knowledge on new business formation in innovative and knowledge-intensive industries are highly localized and hardly spill over to adjacent regions.

[77] The analyses of the role of public research institutions for innovative start-ups in Germany by Fritsch & Aamoucke (2015) show that the number of professors can be regarded a good representation of the respective knowledge stock.

[78] The independent variables enter the estimation with a time lag of one year.

need to assign these types of entrepreneurs to different industries. Table 5.1 presents such a classification of industries that is based on average values of innovativeness and minimum efficient size of the industries. According to the classification, the typical entries of Type 1 are in consumer oriented services, construction, and trade. The entries of Type 2 are in non-innovative large-scale industries such as non-innovative manufacturing, transport, but also energy and mining. Examples for innovative industries with low minimum efficient size, Type 3, are knowledge intensive and technology oriented services, as well as credit and insurance. Finally, innovative businesses at a larger scale, Type 4, are the high-tech and technologically advanced manufacturing industries. This classification can of course not fully account for all heterogeneity of start-ups within industries. However, we argue that our analysis is sufficiently valid for an empirical test of our research questions.

Table 5.1. Classification of industries according to innovativeness and minimum efficient size

Innovation/scale	Small scale	Large scale
Non-innovative	Type 1: Consumer oriented services, construction, trade	Type 2: Energy and mining, non-innovative manufacturing, transport and postal services
Innovative	Type 3: Credit and insurance, knowledge intensive and technology oriented services	Type 4: High-tech and technologically advanced manufacturing

A debatable case in this classification of industries, according to their minimum efficient size, is energy and mining. This sector traditionally consists of large-scale power plants and includes comprehensive infrastructures for energy distribution that make entry rather difficult. However, it also comprises a growing share of firms that produce energy based on wind, solar power and water on a relatively small scale. Entries of such small-scale energy producers played a considerable role during our period of analysis. This sector may also represent a rather special case because it was subject to a high level of state intervention such as changes in the regulatory framework and subsidization of energy production from renewable resources.

5.3.3 *Descriptive statistics*

Table 5.2 reports descriptive statistics of the start-up rates in different sectors. In the entire private sector (except agriculture), there are an average of 44 start-ups per 10,000 workforce per year. The largest numbers of these new businesses are in consumer oriented services, trade, and business oriented services. Relatively low rates can be found in manufacturing and in energy and mining. The start-up rates in high-tech manufacturing and technologically advanced manufacturing industries are lower than those in

TABLE 5.2. Descriptive statistics for start-up rates in different sectors

Start-up rate for	Mean	Minimum	Maximum	SD
All private sectors	44.14	30.02	52.19	6.03
– Energy and mining	0.52	0.28	1.02	0.24
– Construction	5.34	3.30	7.40	1.17
– Trade	10.53	6.19	13.36	2.28
– Transport and postal services	1.84	1.17	2.24	0.27
– Credit and insurance	1.67	1.04	2.06	0.31
– Consumer oriented services	12.55	9.04	13.97	1.22
– Manufacturing	2.13	1.58	2.71	0.29
* Non-innovative manufacturing	1.73	1.30	2.13	0.22
* Innovative manufacturing	0.40	0.27	0.58	0.09
· High-tech manufacturing	0.11	0.07	0.16	0.06
· Technologically advanced manufacturing	0.29	0.19	0.41	0.06
– Business oriented services	9.57	7.21	11.41	1.08
* Knowledge intensive services	5.61	3.78	7.08	0.86
· Technology oriented services	2.79	1.95	3.69	0.48

Notes: Yearly number of start-ups per 10,000 workforce in Germany 1995–2013.

non-innovative manufacturing industries. In contrast, the number of new businesses in knowledge intensive services is higher (5.6 start-ups per 10,000 workforce).

Overall, the summary statistics show considerable differences in the magnitude and the variation of business dynamics across sectors. Together with the fact that there may be substantial differences in the economic significance of start-ups across sectors, our observations indicate that an analysis of the influence of the business cycle on start-ups must distinguish between sectors. For instance, a high number of entries into consumer oriented services may overcompensate for considerably smaller numbers of entries in the manufacturing sector or the technology oriented services but these fewer start-ups may have stronger effects on future developments.

Looking at entries over time, Figure 5.1 shows declining start-up rates in most economic sectors. This decline is stronger for Type 1, which is mostly driven by the trade sector, partly explaining the high coefficient of variation for this sector (Table 5.2), but the start-up rates in the innovative industries also show a negative trend over the observation period (see Types 3 and 4 in Figure 5.1).

The high correlation values of start-up rates within certain sectors, reported in Tables 5.A.1 and 5.A.2 in Appendix 5, indicate correspondence between new business formations within different fields of economic activity. This is particularly true for start-ups in the

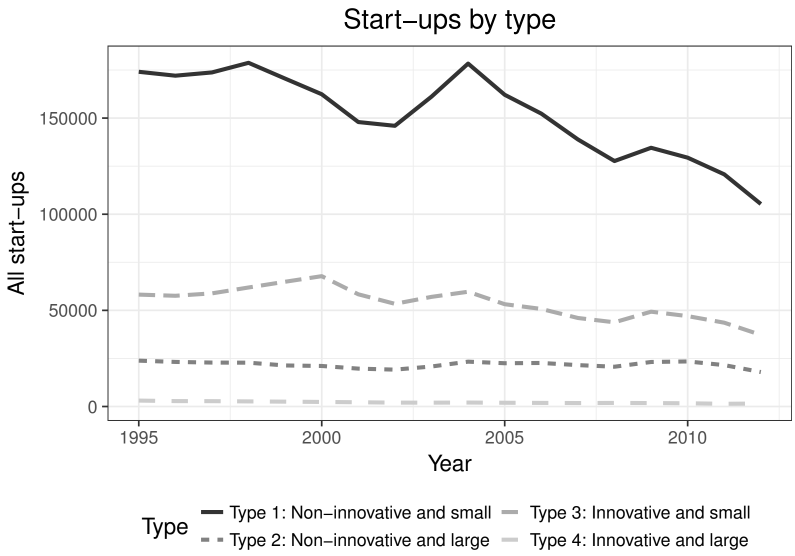


FIGURE 5.1. Start-ups conditional on type

different parts of the service sector suggesting that these start-ups are triggered by similar factors.

The relationship of new business formation in manufacturing with the start-up activity in the service sector is less pronounced, while the correlation of new business formations within the innovative parts of manufacturing is high (Tables 5.A.1 and 5.A.2 in the Appendix). An exception is the negative correlation of the start-up rates in energy and mining with the level of new business formation in other sectors (Table 5.A.1 in the Appendix).

Table 5.3 reports the correlation between start-up rates in different sectors with the cyclical components of the unemployment rate and of GDP. Correlation is strongest for the cycle indicator that is lagged by one year ($t-1$). While the relationship with the unemployment rate is nearly always positive, it tends to be negative for the cyclical component of GDP, particularly in years t and $t-1$. The trade sector correlates most with the cyclical component of the unemployment rate, while correlations seem to be stronger for manufacturing than for services. When it comes to the cyclical component of GDP, the contemporaneous and one year lagged correlation is almost always negative, while the two-year lagged correlation is mostly positive, probably due to the sinusoidal wave pattern of this variable over time. These correlations suggest that the influence of the cycle on new business formation differs quite considerably across industries. According to these correlations it takes, on average, about one year for the business cycle to exert its main influence on the formation of new businesses.

The correlation between the GDP cycle and the unemployment cycle is –0.39 (Table

TABLE 5.3. Correlation between start-up rates in different sectors and business cycle variables

Start-up rate for	Unemployment rate—cyclical component			Real GDP—cyclical component		
	t	$t-1$	$t-2$	t	$t-1$	$t-2$
All private sectors	0.12	0.17	0.04	–0.08	–0.09	0.06
Energy and mining	0.00	–0.04	–0.09	–0.05	–0.11	–0.03
Manufacturing	0.16	0.19	0.01	–0.16	–0.12	0.10
Non-innovative manufacturing	0.16	0.18	–0.02	–0.18	–0.14	0.08
Construction	0.18	0.17	0.04	–0.05	–0.06	0.03
Trade	0.19	0.25	0.05	–0.09	–0.10	0.08
Transport and postal services	0.06	0.15	0.08	0.00	–0.06	0.02
Credit and insurance	0.02	0.08	0.00	–0.02	–0.06	0.06
Business oriented services	0.00	0.04	0.02	–0.04	–0.03	0.04
Consumer oriented services	0.07	0.11	0.02	–0.06	–0.08	0.04
Innovative manufacturing	0.08	0.13	0.07	–0.05	0.00	0.10
High tech manufacturing	0.05	0.14	0.12	0.00	–0.04	0.02
Technologically advanced manufacturing	0.07	0.09	0.03	–0.08	0.03	0.12
Knowledge intensive services	0.00	0.06	0.03	–0.05	–0.04	0.05
Technology-oriented services	0.00	0.09	0.06	–0.06	–0.06	0.05

5.A.6 in the Appendix) such that a multicollinearity problem due to a strong correlation between the cyclical components is unlikely.[79] Therefore, we include the cyclical component of the unemployment rate and the GDP in all models. There is enough variation in the GDP and unemployment cycle across regions, which is demonstrated in Figure 5.A.1 (in the Appendix) depicting bi-regional correlations between the two cyclical components.[80] The median bi-regional correlation of the GDP cycle is 0.55 and of the unemployment cycle 0.88. Thus, the GDP cycle is less synchronized across regions than the unemployment cycle.

[79] We also computed variance inflation factors, which did not indicate any serious multicollinearity problems.
[80] The bi-regional correlation is computed as $\mathrm{Cor}(v_i, v_j)$ where v is the cyclical component (of GDP or unemployment) and i and j are two regions with $i \neq j$.

5.4 Empirical approaches and results

This section presents our empirical results. We start by analyzing short-term business cycle effects in all industries available in our data set. Next, we construct the four types of businesses discussed in Section 5.2.2 and 5.3.2 and answer our research questions. We also provide sensitivity checks, including tests for endogeneity.

5.4.1 Separate analysis for each industry

We aim to assess whether there is evidence that start-up rates in different industries systematically change with business-cycle variables. Therefore, we regress start-up rates on the cyclical components of the unemployment rate and GDP, as well as on a set of control variables related to business entries. The same set of independent variables is included in all models in order to identify differences between industries. We apply a fixed effects panel approach to capture region-specific influences that are invariant over time. To reduce endogeneity problems, all explanatory variables are included with a time lag of one year (Astebro et al. 2013; Buch, Koch & Koetter 2013). As we are mostly interested in short-term correlations, we concentrate on the cyclical components from the previous year. For robustness checks we examine effects using longer lags. Still, we are aware that the fixed-effects estimations do not reflect causal relationships. However, we provide tests for endogeneity in Section 5.4.3, when testing the robustness of results. To allow for an assessment of the relative influence of the estimated coefficients, all variables are standardized with a mean value of zero and a standard deviation of 1. The estimated equation has the form:

$$\text{Start-up rate}^{k}_{i,t} = \mu_i + \alpha_1 \text{BC unemployment}_{i,t-1} + \alpha_2 \text{BC GDP}_{i,t-1} + \boldsymbol{\eta}^{\top}\mathbf{x}_{i,t-1} + u_{i,t} \quad (5.5)$$

The $i = 1, \ldots, N$ is the number of NUTS 2 regions; $t = 1, \ldots, T$ denotes the years, $k = 1, \ldots, 15$ stands for the sector (including aggregates such as all private sectors), μ_i represents the fixed effects and $\mathbf{x}_{i,t}$ is the four-dimensional vector of control variables. The control variables (in $\mathbf{x}$) are, as discussed in Section 5.3.1, the share of employees in businesses with up to 20 employees, the share of employees with a tertiary degree, the number of university professors, and of patent applications per 1,000 workforce.

Table 5.4 (mixed types), 5.5 (non-innovative types), and 5.6 (innovative types) present the results of our fixed effects estimations for start-up rates in different sectors. We start by analyzing how macroeconomic variables affect all start-ups in the overall private sector (Model I). We find positive influences of the cyclical component of unemployment and negative influences of the cyclical component of GDP on new business formation. Thus, the two variables unfold counter-cyclical influences: Business formation is higher

TABLE 5.4. Influence of cyclical variables on start-ups in different sectors with detailed results for every industry in data and aggregates: Mixed types

Scale	All types	Small scale	Large scale
Model	I	II	III
Start-up rates in:	All private sectors	Business oriented services	Manufacturing
Unemployment rate—cyclical component ($t-1$)	0.05***	−0.02	0.11***
GDP—cyclical component ($t-1$)	−0.17***	−0.10***	−0.06
Share of employees in small businesses ($t-1$)	0.02	0.04	0.02
Share of employees with tertiary education ($t-1$)	−0.08**	−0.02	0.02
Number of professors per 1,000 workforce ($t-1$)	0.03	0.02	0.06**
Number of patent applications per 1,000 workforce ($t-1$)	0.09***	0.02**	0.10**
Constant	−0.27***	0.29***	0.32
R^2	0.71	0.82	0.62

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models.

with high unemployment or when GDP is low. These results fully confirm earlier findings for Germany that used two different data sources (the micro-census and the business registration statistics), but could not distinguish between new businesses in different sectors (Fritsch et al. 2015). Among the control variables related to knowledge spillovers, the number of patent applications, has the expected positive sign.

To investigate whether the cyclical macroeconomic variables have differing influence on new business formation in innovative and non-innovative industries, we estimate separate models for these industries (Table 5.5 and 5.6). Starting with the cyclical component of the unemployment rate, we observe mostly counter-cyclical influences: high unemployment levels are positively related to entries into manufacturing, both non-innovative and innovative (Models VII and XV), as well as to entries into construction (Model IV) and into the traditional parts of the service sector, i.e., trade and transport services (Models VI and VIII).

In contrast, high unemployment levels do not significantly influence the start-up rates in knowledge intensive services (Model XI) or, against expectations, in consumer oriented

TABLE 5.5. Influence of cyclical variables on start-ups in different sectors with detailed results for every industry in data and aggregates: Non-innovative types

Scale		Small scale			Large scale	
Model	IV	V	VI	VII	VIII	IX
Start-up rates in:	Construction	Consumer oriented services	Trade	Non-innovative manufacturing	Transport and postal services	Energy and mining
Unemployment rate—cyclical component ($t-1$)	0.06***	0.03	0.09***	0.10***	0.08***	−0.02
GDP—cyclical component ($t-1$)	−0.15***	−0.10***	−0.20***	−0.04	−0.08***	0.13***
Share of employees in small businesses ($t-1$)	0.00	0.02	−0.02	−0.02	−0.01	−0.08***
Share of employees with tertiary education ($t-1$)	−0.15***	−0.07**	−0.06	0.04	−0.03	0.04
Number of professors per 1,000 workforce ($t-1$)	0.03	0.01	0.03	0.07**	0.01	−0.02
Number of patent applications per 1,000 workforce ($t-1$)	0.10***	0.01	0.20***	0.07	0.04	−0.12**
Constant	−0.41***	−0.60***	−0.42***	0.14***	−0.03***	−0.57***
R^2	0.75	0.79	0.56	0.62	0.87	0.36

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models.

services (Model V). Only entries into credit and insurance (Model X) react pro-cyclically to employment rates.

Typically, there is a countercyclical relationship between the cyclical component of

TABLE 5.6. Influence of cyclical variables on start-ups in different sectors with detailed results for every industry in data and aggregates: Innovative types

Scale		Small scale			Large scale	
Model	X	XI	XII	XIII	XIV	XV
Start-up rates in:	Credit and insurance	Knowledge intensive services	Technology oriented services	Techno-logically advanced manufac-turing	High-tech manufac-turing	Innovative manufac-turing
Unemployment rate—cyclical component ($t-1$)	-0.08^{***}	-0.01	0.01	0.07^{**}	0.09^{***}	0.09^{***}
GDP—cyclical component ($t-1$)	-0.30^{***}	-0.12^{***}	-0.18^{***}	-0.04	-0.14^{***}	-0.09
Share of employees in small businesses ($t-1$)	0.05	0.08^{***}	0.17^{***}	0.09^{***}	0.15^{***}	0.13^{***}
Share of employees with tertiary education ($t-1$)	-0.13^{**}	-0.04	-0.06	-0.03	-0.01	-0.02
Number of professors per 1,000 workforce ($t-1$)	0.01	-0.03^{**}	0.03	0.04^{**}	0.02	0.03
Number of patent applications per 1,000 workforce ($t-1$)	0.05	0.06^{***}	0.10^{***}	0.16^{***}	0.12^{**}	0.16^{***}
Constant	0.00	0.29^{***}	0.62^{***}	0.67^{***}	0.47^{***}	0.68^{***}
R^2	0.40	0.77	0.72	0.32	0.48	0.43

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models.

GDP and start-up rates in many innovative and non-innovative sectors. This holds true for all kinds of services, knowledge intensive and technology oriented (Models XI and XII), as well as all traditional services (Models V and VIII), but also the construction sector

(Model IV). Quite remarkably, except for high-tech manufacturing (Model XIV), entries in most parts of the manufacturing sector remain unaffected by GDP. Only in the 'energy and mining' sector (Model IX)—about one percent of all start-ups—does the cyclical component of GDP have a positive influence on business entries.

When focusing on business entries into the innovative parts of the industries, an important difference occurs: The influence of cyclical deviations from unemployment and from GDP levels differ systematically between the manufacturing and the service sector. Changes in unemployment levels lead to counter-cyclical correlations with the innovative part of the manufacturing sector but not with the innovative part of the service sector. While changes in GDP levels are correlated with business entries into knowledge intensive and technology oriented services (Models XI and XII representing all innovative parts of the service sector), in the manufacturing sector it only correlates with start-up rates in high-tech manufacturing (Model XIV). Table 5.7 summarizes all effects conditional on

Table 5.7. Business cycle effects on new business formation

Type	Industry	GDP	Unemployment
Non-innovative and small scale	Construction	–	–
	Consumer oriented services	–	0
	Trade	–	–
Non-innovative and large scale	Non-innovative manufacturing	0	–
	Transport and postal services	–	–
	Energy and mining	+	0
Innovative and small scale	Credit and insurance	–	+
	Knowledge intensive services	–	0
	Technology oriented services	–	0
Innovative and large scale	Technologically advanced manufacturing	0	–
	High-tech manufacturing	–	–

Notes: "+" indicates a pro-cyclical effect, "–" a counter-cyclical effect, and "0" an a-cyclical effect

type (combinations of innovation potential and scale).

At the industry level, the results for the control variables are mostly in accordance with results of previous studies. The share of employees in small businesses has the expected positive sign if statistically significant (except for energy and mining). The variation of the results for the three variables representing distinctive facets of the regional knowledge stock—share of employees with a tertiary degree, number of professors per workforce, and number of patent applications per 1,000 employees—demonstrates differences in the relevant knowledge base for start-ups across industries. Regional knowledge, in particular patents, has a positive influence when focusing on the innovative industries, but also on the non-innovative part of the secondary sector, while it does not seem to play a role for new business formation in the non-innovative service sectors. The findings are consistent with the Knowledge Spillover Theory of Entrepreneurship (Audretsch & Keilbach 2007b; Acs et al. 2009).

5.4.2 Type-specific analysis

To further analyze our research questions proposed in Section 5.2.3, we use our data on entries into different industries to construct the four types, as proposed in Table 5.1. Instead of Equation (5.5), we, now, estimate the following model:

$$\text{Start-up rate}^{\tau}_{i,t} = \mu^{\tau}_{i} + \alpha^{\tau}_{1}\text{BC unemployment}_{i,t-1} + \alpha^{\tau}_{2}\text{BC GDP}_{i,t-1} + \boldsymbol{\eta}^{\top}_{\tau}\mathbf{x}_{i,t-1} + u^{\tau}_{i,t} \quad (5.6)$$

where $\tau \in \{\text{Type 1, Type 2, Type 3, Type 4}\}$ captures the industry type and all other variables in (5.6) also depend on the type. Table 5.8 provides results conditional on type. Using Table 5.8, we obtain the following results with respect to our research questions.

In relation to Question 1, we find that the two cyclical macroeconomic variables influence new business formation in a different way. Hence, the industry context that stands for demand conditions, technologies, production methods, and cost structures matters quite significantly. This result is in line with the predictions of our theoretical model in Section 5.2.4.

In relation to Question 2, we find virtually no correlation that is consistent with a procyclical effects of GDP, as proposed by Rampini (2004), Barlevy (2007) and others. Rather to the contrary, we observe correlations that are consistent with a counter-cyclical influence of GDP on business entries, but only for entries into all small-scale industries (Type 1 and Type 3), while entries into large-scale industries remain mostly uncorrelated to GDP, which also contradicts the predictions from our theoretical model of entrepreneurial entry with respect to large-scale industries. However, the importance of size is consistent with the theoretical model. So, it is the size that matters for the relationship between the cycle and business entries.

TABLE 5.8. Influence of cyclical variables on start-ups conditional on the four types

Start-up rates in:	Type 1: non-innovative and small	Type 2: non-innovative and large	Type 3: innovative and small	Type 4: innovative and large
Unemployment rate—cyclical component ($t-1$)	0.08**	0.10***	–0.01	0.09***
GDP—cyclical component ($t-1$)	–0.19***	–0.03	–0.17***	–0.09
Share of employees in small businesses ($t-1$)	0.00	–0.05	0.11***	0.13***
Share of employees with tertiary education ($t-1$)	–0.11**	0.01	–0.06**	–0.02
Number of professors per 1,000 workforce ($t-1$)	0.03	0.02	0.03	0.03
Number of patent applications per 1,000 workforce ($t-1$)	0.13***	0.01	0.07***	0.16***
Constant	–0.62***	–0.21***	0.36***	0.68***
R^2	0.63	0.75	0.73	0.43

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models.

In relation to Question 3, the correlations between unemployment and business entries point to counter-cyclical influences on entries by Type 1 (the small-scaled non-innovative businesses, Question 3a), while, in line with our theoretical model, the influence of unemployment is a-cyclical in sectors with a high innovation potential that have small average sizes (Type 3, Question 3b). Moreover, correlations between unemployment and business entries also point to counter-cyclical influences on entries into Types 2 and 4, the large-scale industries (Question 3c). In summary, there is more of a "size effect," with changes in GDP being related to business entries into small-scale industries, while changes in unemployment relate more strongly to entries into large-scale industries. Both relations apply for innovative industries.

5.4.3 Robustness tests

We performed several sensitivity tests to examine the robustness of our findings.

5.4.3.1 *Effect dynamics*

As a robustness test, we compare the results from the model using one lag (one year), given in (5.6), to the following distributed lag model capturing short-run dynamics in two prior years:

$$\text{Start-up rate}^{\tau}_{i,t} = \tilde{\mu}^{\tau}_{i} + \alpha^{\tau}_{1,1}\text{BC unemployment}_{i,t-1} + \alpha^{\tau}_{1,2}\text{BC unemployment}_{i,t-2} \tag{5.7}$$
$$+\alpha^{\tau}_{2,1}\text{BC GDP}_{i,t-1} + \alpha^{\tau}_{2,2}\text{BC GDP}_{i,t-2} + \boldsymbol{\eta}^{\top}_{\tau,1}\mathbf{x}_{i,t-1} + \boldsymbol{\eta}^{\top}_{\tau,2}\mathbf{x}_{i,t-2} + \tilde{u}^{\tau}_{i,t}$$

where all variables are included with a lag of two periods. The effects of interest are the cumulative effects of unemployment, $\alpha^{\tau}_{1,1} + \alpha^{\tau}_{1,2}$, and the GDP cycle, $\alpha^{\tau}_{2,1} + \alpha^{\tau}_{2,2}$. Results are provided in Table 5.9. We report the effects of unemployment and GDP.[81]

TABLE 5.9. Short-run dynamics of the influence of cyclical variables on start-ups conditional on the four types

Start-up rates in:	Type 1: non-innovative and small	Type 2: non-innovative and large	Type 3: innovative and small	Type 4: innovative and large
Unemployment rate—cyclical component $t-1$	0.06**	0.10***	−0.01	0.07**
Unemployment rate—cyclical component $t-2$	−0.07***	0.08***	−0.04	0.02
GDP—cyclical component $t-1$	−0.20***	−0.06	−0.17***	−0.07
GDP—cyclical component $t-2$	0.02	0.01	0.01	0.03
Cumulative effect of unemployment	−0.01	0.18***	−0.05	0.09***
Cumulative effect of GDP	−0.18**	−0.05	−0.16***	−0.04

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one or two periods. The number of observations is 418 (38 cross sections, 11 years) in all models.

The results of the dynamic analysis in Table 5.9 are consistent with the static analysis in Table 5.8. The only difference is that the cumulative effect of unemployment on start-ups of Type 1 is a-cyclical in the distributed lag model, which is due to the fact that unemployment effects with a lag of one and two periods cancel each other out.

[81] Additional results are available on request.

5.4.3.2 *Unobserved spatial links between regions*

Studies examining the connection between the business cycle and business entries usually do not account for unobserved spatial links. For instance, Koellinger & Thurik (2012) use a country level panel and control for country fixed effects, but do not control for unobserved dependencies between countries. However, previous studies demonstrate that German municipalities compete on taxes (Büttner 2001), spending (Borck, Caliendo & Steiner 2007), and debt (Borck, Fossen, Freier & Martin 2015). This type of competition is unobserved in our model and would enter it through the error term. To test whether our results are sensitive to including unobserved spatial dependencies, we estimate the following spatial error model with spatial fixed effects:

$$\text{Start-up rate}_{i,t}^{\tau} = \mu_i^{\tau} + \alpha_1^{\tau}\text{BC unemployment}_{i,t-1} + \alpha_2^{\tau}\text{BC GDP}_{i,t-1} + \boldsymbol{\eta}_{\tau}^{\top}\mathbf{x}_{i,t-1} + su_{i,t}^{\tau} \quad (5.8)$$

The main difference between the model in (5.6) and (5.8) is that instead of a simple error term $u_{i,t}^{\tau}$, we now have a spatially lagged error $su_{i,t}^{\tau}$. Suppressing the type indicator for notational convenience, the spatial error is constructed as follows:

$$su_{i,t} = \zeta \sum_{j=1}^{N} \mathrm{w}_{i,j} su_{i,j} + \varepsilon_{i,t}$$

where $\zeta \in (0, 1)$. Thus, the effect in region i depends on the weighted effects from all the other regions $\sum_{j=1}^{N} \mathrm{w}_{i,j} su_{i,j}$ and a region-specific effect $\varepsilon_{i,t}$. The weights are constructed in two steps. In the first step, two regions with a common border are assigned a weight of 1 such that

$$\tilde{\mathrm{w}}_{i,j} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ have a common border} \\ 0 & \text{if no common border or if } i = j \end{cases}$$

while regions without a common border are assigned a weight of 0. In the second step, weights are normalized such that they sum up to 1:

$$\mathrm{w}_{i,j} = \frac{\tilde{\mathrm{w}}_{i,j}}{\sum_{i=1}^{N} \tilde{\mathrm{w}}_{i,j}}$$

Note that this simple way to model spatial effects can accommodate rather complex structures.

The model in (5.8) is estimated with maximum likelihood, as suggested by Elhorst (2003). Results, presented in Table 5.10, are consistent with previous results. All statistically significant effects are counter-cyclical. The only difference to Table 5.8 is that the

TABLE 5.10. Influence of cyclical variables on start-ups conditional on the four types with spatially lagged errors

Start-up rates in:	Type 1: non-innovative and small	Type 2: non-innovative and large	Type 3: innovative and small	Type 4: innovative and large
Unemployment rate—cyclical component ($t-1$)	0.05	0.08**	0.01	0.09**
GDP—cyclical component ($t-1$)	−0.07**	−0.05	−0.10***	−0.01
Share of employees in small businesses ($t-1$)	−0.02	−0.03	0.03	0.09**
Share of employees with tertiary education ($t-1$)	0.00	0.01	−0.01	−0.04
Number of professors per 1,000 workforce ($t-1$)	−0.01	0.02	0.01	0.02
Number of patent applications per 1,000 workforce ($t-1$)	0.03	0.01	0.04	0.09***
Spatial error coefficient (ζ)	0.67***	0.46***	0.44***	0.49***

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation with spatial errors; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models.

effect of unemployment on Type-1 entries becomes insignificant after introducing spatial dependencies. Thus, with spatial dependencies, the difference between small and large businesses becomes even more pronounced. According to Table 5.10, entries into small scale businesses react counter-cyclically to GDP and a-cyclically to unemployment, while entries into large scale businesses mostly react counter-cyclically to unemployment and a-cyclically to GDP.

An additional and important insight provided in Table 5.10 is that the estimate of the spatial error coefficient, ζ, is relatively large and highly significant for all four types. Hence, an analysis of the effects of the business cycle on entry should account for spatial dependencies as these might play an important role.

5.4.3.3 *Potential endogeneity bias*

Consider the following model for business entry:

$$\text{Entry} = \nu \text{Business cycle variable} + bY + \epsilon_1 \tag{5.9}$$

where indices (such as the time index) are dropped for notational convenience. The variable Y includes $\mathbf{x}$ (control variables), a constant, region fixed effects, and potentially additional variables. As we are interested in the effect of the business cycle, ν is the coefficient of interest. The OLS estimate of ν is unbiased if the business cycle variable and the error ϵ_1 are not correlated. Since we include all business cycle variables in lags, the likelihood of such a correlation is reduced. However, to examine whether results are robust to relaxing the assumption of exogeneity, we perform several instrumental-variable-based checks.

To examine a plausible mechanism generating a correlation between the business cycle variable and the error, consider the following model for the business cycle variable:

$$\text{Business cycle variable} = cY + \epsilon_2 \tag{5.10}$$

If ϵ_1 and ϵ_2 are correlated, the OLS estimate of ν will be biased. This might happen if, for instance, there is an unobserved variable U driving business cycle and entry at the same time, i.e., $\epsilon_1 = U + \breve{\epsilon}_1$ and $\epsilon_2 = dU + \breve{\epsilon}_2$, where $\breve{\epsilon}_1$ and $\breve{\epsilon}_2$ are idiosyncratic errors. We try to account for potential endogeneity by applying an instrumental-variable method. In the given setting, we need an instrument for the unemployment and GDP cycle, which enters (5.10) but not (5.9).

A reasonable instrument for unemployment is the implementation of one of the so-called Hartz labor market reforms at the beginning of the year 2005. Before this reform, non-employed individuals receiving social welfare were not required to be available to the labor market. After the implementation of the reform, individuals received a new form of unemployment benefits (instead of social welfare) and were required to be available to the labor market. While the reform was implemented on January 1, 2005, it generated an incentive for individuals receiving social welfare to register as unemployed already in 2004, as the unemployment benefits were higher than the social welfare payments.[82] As registering as unemployed required availability to the labor market, the reform resulted in an increase in unemployment without directly affecting start-up incentives in 2004. A fact that might weaken the exclusion restriction is that an earlier part of the reform was a start-up subsidy for entrepreneurship out of unemployment, the so-called "Ich-AG." However, the subsidy started already two years earlier on January 1, 2003. Furthermore, we tested whether the subsidy might influence results by removing all types of services from small scale businesses, as this is the most common type of start-ups from unemployment (Caliendo & Kritikos 2010), and obtained similar results.[83]

An instrument for the GDP cycle is the pre-crisis peak in 2007. As the German economy is highly export-oriented, the German economy was affected by the pre-crisis boom in the US through trade links. As in the pre-crisis period US GDP and, through the trade

[82] See German newspaper reports from this period (for instance, Rosenfeld 2005 or Spiegel 2005).
[83] Results are available upon request.

channel, the German GDP was pushed above the trend by forces not directly related to the German economy, the pre-crisis boom is a candidate for a valid instrument. The financial crisis might have affected start-ups through the finance channel. To control for this channel, in an additional regression we remove the start-ups in the banking industry and include them as a covariate, approximating the health of this sector.

We estimate separate models for the effects of unemployment and GDP by two-stage least squares. Since the approach removes considerable amounts of variance from the cyclical variables, as the instruments are effectively time dummies, and there is not much variance in case of Type-4 entries (Figure 5.1), we aggregate the types and consider only effects on small- and large-scale industries. First stages are given in Table 5.A.7 (unemployment) and 5.A.8 (GDP) in Appendix 5.A. Table 5.11 shows results of second stage

TABLE 5.11. Influence of the unemployment and GDP cycle on start-ups conditional on size given that the unemployment or GDP cycles are instrumented

Start-up rates in:	Small scale industries		Large scale industries	
Unemployment rate—cyclical component ($t-1$) (instrumented by labor market reform)	−0.01		0.10**	
GDP—cyclical component ($t-1$) (instrumented by pre-crisis boom)		−0.55***		−0.09
Share of employees in small businesses ($t-1$)	−0.03	0.14***	−0.05	−0.05**
Share of employees with tertiary education ($t-1$)	−0.12	−0.08***	0.00	−0.02
Number of professors per 1,000 workforce ($t-1$)	0.02	0.04	0.03	0.03
Number of patent applications per 1,000 workforce ($t-1$)	0.17***	0.06**	0.04	0.05
Constant	−0.31***	−0.19***	−0.12	−0.09
F-test for weak IV	23.96	134.38	23.96	134.38

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; IV estimation with fixed effects; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models. Standard errors are clustered at the region level.

regressions. The instruments have sufficient strength (see the F-test in Table 5.11). In line with previous results, entry into small-scale industries is mostly counter-cyclically influenced by GDP, whereas entry into large-scale industries is mostly counter-cyclically influenced by unemployment. Table 5.A.9 in the Appendix provides results where banking is removed from small-scale industries and entries into banking are included as an additional covariate. Results are consistent with Table 5.11, although the size of the GDP cycle effect on entries into small-scale industries is smaller.

Models using the two instruments are exactly identified such that instrument validity cannot be tested. However, if certain conditions are met, it is possible to construct additional instruments, to be able to perform a Sargan-Hansen test. Let the structural system be given by (5.9) and (5.10). In this case, a recently developed instrumental variable approach proposed by Lewbel (2012) demonstrates that, by exploiting potential heteroskedasticity in the error term in (5.10), it is possible to consistently estimate effects without an available exclusion restriction if certain assumptions hold. Let Z denote a set of variables that are exogenous and that affect start-ups and the business cycle variable, i.e., they are part of (5.9) and (5.10). Z can include some or all elements of Y. In addition to the standard conditions, viz., $\mathbb{E}[Y\epsilon_1] = 0$, $\mathbb{E}[Y\epsilon_2] = 0$, and $\mathbb{E}[YY^\top]$ is non-singular, the approach of Lewbel (2012), which we refer to as Lewbel IV, requires that $\mathrm{Cov}(Z, \epsilon_1\epsilon_2) = 0$ and $\mathrm{Cov}(Z, \epsilon_2^2) \neq 0$. If the conditions hold, $(Z - \bar{Z})\epsilon_2$ is a valid instrument for the business cycle variable, where $\bar{Z}$ is the sample average of Z. If $\mathrm{Cov}(Z, \epsilon_2^2) \neq 0$ does not hold the instrument will be weak (Lewbel 2012) such that testing for instrument strength indirectly tests for the assumption.

To construct the instrument, we assume that Z includes the share of employees in small businesses, the share of employees with tertiary education, and the number of patent applications. The selection is based on the criterion of sufficient instrument strength. Using Z constructed in such a way, we tested if the errors in Equation (5.10) are heteroskedastic, a test if $\mathrm{Cov}(Z, \epsilon_2^2) \neq 0$ holds, and could reject homoskedasticity at the 5% level using a Breusch-Pagan test. Thus, we could not find violations of the Lewbel (2012) conditions for instrument construction.

Lewbel (2012) instruments are combined with the instruments we already used for unemployment and GDP and the model, which is now overidentified, is estimated with IV-GMM. Results are presented in Table 5.12. The F-test results at the first stage (see Table 5.12) suggest that the generated instruments are not weak. Results show that there is no significant counter-cyclical influence of unemployment on entries into small-scale industries but entries into this type of industries are still counter-cyclically influenced by GDP. In case of large-scale industries, the only statistically significant business cycle effect is a counter-cyclical influence of unemployment. Hence, results are consistent with previous estimation results.

As the system is now over-identified, we can perform a Sargan-Hansen test to exam-

TABLE 5.12. Influence of the unemployment and GDP cycle on start-ups conditional on size given that effects are estimated with a combination of Lewbel IV and traditional IV

Start-up rates in:	Small scale industries		Large scale industries	
Unemployment rate—cyclical component ($t-1$) (instrumented by labor market reform and Lewbel instrument)	0.15***		0.26**	
GDP—cyclical component ($t-1$) (instrumented by pre-crisis boom and Lewbel instrument)		−0.54***		0.01
Share of employees in small businesses ($t-1$)	0.01	0.14***	0.04	−0.08**
Share of employees with tertiary education ($t-1$)	−0.09***	−0.08***	−0.00	−0.02
Number of professors per 1,000 workforce ($t-1$)	0.03	0.04	0.03	0.02
Number of patent applications per 1,000 workforce ($t-1$)	0.15***	0.05	−0.01	0.07**
Constant	−0.32***	−0.19	−0.32***	−0.11
Overidentification test (p-value)	0.07	0.74	0.20	0.70
F-test for weak IV	12.40	43.57	12.40	43.57

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; IV estimation with fixed effects; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models. Standard errors are heteroskedasticity robust.

ine whether there is indication that the exclusion restriction for the 2004-reform and the 2007-pre-crisis-peak instruments does not hold. The hypothesis that overidentifying restrictions are valid cannot be rejected at the 5% level (Table 5.12) suggesting that the instruments are valid, conditional on validity of the Lewbel instruments.

5.4.3.4 Detrending technique

An open concern is that results might be driven by the detrending technique. The Hodrick-Prescott filter has a free parameter, set by the user of the procedure and not driven by data that might influence results (Schlicht 2005). Furthermore, the filter has an end-point bias

(Mise, Kim & Newbold 2005). Therefore, to check the robustness of the results, we use an alternative filter, developed by Baxter & King (1999). The Baxter-King filter, isolating the cyclical component from the trend, was explicitly designed to overcome the drawbacks of the Hodrick-Prescott filter.

In line with recommendations of Baxter & King (1999) for annual data, the minimum period of oscillation is set to 2 and the maximum period to 8, while the order of the filter is 3. Results are given in Table 5.A.10 in the Appendix. Nearly all results under the Baxter-King filter are consistent with the ones obtained with the Hodrick-Prescott filter. The only exception is the unemployment effect for Type 4, which is a-cyclical, instead of counter-cyclical, when using the Baxter-King filter. This could be a result of the end-point bias of the Hodrick-Prescott filter. However, as results obtained using the Baxter-King filter are numerically similar to results in Table 5.8 for all the remaining three types, not indicating any difference in predicted business cycle effects, it is more likely that the statistically insignificant result is generated because the Baxter-King filter removes around 80 data points (data at the beginning and end of each time series in each region).

5.4.3.5 Additional tests

As further robustness checks we ran the models for low-density and high-density regions as well as for East and West Germany separately. However, we did not find any significant differences between these spatial categories.

5.5 Summary and conclusions

While nearly all previous empirical analyses of the effect of the business cycle on new business venturing examine start-ups in the overall private sector, our study systematically distinguishes between different industries. We find distinct variations in how the variables that represent the business cycle are correlated with new firm formation across industries. Overall, our analysis with data for German NUTS 2 regions indicates correlations that are consistent with mostly counter-cyclical effects of the business cycle on entries into the market: Either more businesses are started when unemployment is high or when GDP is low. Moreover, correlations differ more between large- and small-scale industries than between innovative and non-innovative industries in the sense that entries into large-scale industries are mainly influenced by variations in unemployment, while entries into small scale industries—all kinds of services—are affected by variations in GDP. Both results are robust to applying different model specifications.

Overall, in virtually no industry is a GDP level above the trend correlated with increasing entries. Although we are not able to make causal interpretations, this observation points to the conclusion that favorable conditions in terms of high GDP might not be germane for start-ups. In fact, according to our results, it is the other way around: An economic downturn, maybe in the sense of lower production or lower entry costs or better future profit opportunities, seems to be a more favorable environment for innovative businesses than boom periods. This holds at least for entries of small-scale businesses, while entries of large-scale businesses are uncorrelated to changes in GDP. We further interpret our second result—the positive correlations between unemployment and entries of large-scaled businesses—as an indicator that these entries are influenced by easier availability of labor when unemployment is high. Moreover, we should also emphasize that we find no correlation between unemployment and entries into consumer services that make more than 25% of all entries. This industry is commonly assumed to provide a well-suited environment for an unemployment-push effect for necessity entrepreneurs. All together, these results are in strong contrast to earlier research (see Ghatak et al. 2007; Koellinger & Thurik 2012; Román et al. 2013) that did not distinguish between entries into different industries but assumed that opportunity driven start-ups should be expected more frequently during boom periods.

Our findings clearly demonstrate the importance of accounting for industry-specific characteristics when analyzing the relationship between the business cycle and new business formation. Since there is little other empirical evidence on industry differences of the effects of the business cycle on new business formation available, more research on this topic is needed. It would be desirable to have similar studies for other countries, particularly for countries like the US where overall pro-cyclical effects seem to prevail. It would be important to understand why in some countries pro-cyclical and in other countries counter-cyclical influences predominate. Is it that in some countries supply-side effects such as production costs influence entry decisions, while in other countries it is the demand side? Or do profit expectations during boom periods and recession differ between countries?

Since new businesses, particularly innovative start-ups, may have a positive effect on economic development (Acs et al. 2009; Fritsch 2013; Kritikos 2014), they might play a crucial role in reducing effects of aggregate economic shocks and supporting economic recovery. Hence, it would be important to analyze how cycle-induced entries affect economic development, for instance, in terms of employment. Empirical analysis for the US points again to opposing results. While Lee & Mukoyama (2015) find, for the US manufacturing sector, that firms opened during recessions start with about 30% more employees than firms opened during boom periods, Sedlácek & Sterk (2017), analyzing entries into all industries, find the opposite, namely more job creation in firms opened during boom periods.

Generally, it would be important to have data that comprise more information about the characteristics of the new businesses in each of these industries, such as their size, the timing, and the amount of innovation efforts, as well as the timing of the commercialization of innovative ideas. Such data could allow for a more precise identification of different types of new ventures and, thus, for a more causal interpretation of the results. Moreover, it would be interesting to identify industry-specific cycles and their effects on new business formation. This would particularly allow for comparing the effects of global conditions such as the nationwide levels of unemployment and GDP with industry-specific developments. To these ends, more empirical research on the effect of the business cycle on new business formation is needed.

6 ARE PERSONALITY PROFILES USEFUL FOR PREDICTION AND ADVICE? A MONTE CARLO STUDY

The human personality predicts a wide range of variables—from musical sophistication to occupational choice. However, which method should be applied if information on personality traits is planned to be used for predictions and advice? In psychological research, group profiles are widely employed. This chapter examines the performance of profiles using the example of career prediction and advice, involving a comparison of average trait scores of successful entrepreneurs with the traits of potential entrepreneurs. Based on a simple theoretical model estimated with GSOEP data and analyzed with Monte Carlo methods, we show, for the first time, that the choice of the comparison method matters substantially. We reveal that under certain conditions the performance of profiles is inferior to the tossing of a coin. Alternative methods deliver better performance and are more robust.

6.1 Motivation

> Based on the comparison with personality profiles of top-entrepreneurs, I selected six out of 1,500 applicants and decided to heavily invest into their entrepreneurial ideas. *(The CEO of an American investment company, Nov. 2016)*

Personality predicts a wide range of variables, including cadet performance (Mayer & Skimmyhorn 2017), musical sophistication (Green, Müllensiefen, Lamb & Rentfrow 2015), migration (Jokela 2009), human values (Fischer & Boer 2014), job satisfaction (Heller, Ferris, Brown & Watson 2009), conflict in interpersonal relations (Bono, Boles, Judge & Lauver 2002), and entrepreneurial success (Zhao & Seibert 2006; Rauch & Frese 2007; Caliendo et al. 2014). However, are observations on personality traits also useful for predictions and advice? More specifically, under what conditions can information on personality traits of individuals being successful in a certain kind of occupation be exploited for generating predictions and recommendations?

In this contribution, we examine the recommendation performance of personality-based advice approaches, given the case of occupational choices. Although we focus on a specific type of choices, our approach is general enough to discuss any variable correlated with personality. Using the case of career recommendations, we demonstrate that the common approach of using group personality profiles, for instance, comparing the average personality of entrepreneurs and workers, results in a very weak recommendation performance, which can be outperformed by a simple coin.

Career decisions are of particular interest because personality is an essential determinant of occupational choices in general (Holland 1997; Heckman, Stixrud & Urzua 2006; Borghans et al. 2008) and matters when individuals decide to start a business, seeking to maintain it successfully (Zhao & Seibert 2006; Rauch & Frese 2007; Caliendo et al. 2014; Frese & Gielnik 2014). Personality is a predictor of entrepreneurial income (Levine & Rubinstein 2016; Manso 2016). Several researchers argue that entrepreneurs may fail even when they have a convincing idea, access to finance, and possess high education but do not have the "necessary" personality traits (Kalleberg & Leicht 1991; Shaver & Scott 1991) to successfully run an own business.

At the same time, individual decisions to become an entrepreneur are often influenced by others' advice (Bosma et al. 2012). In many countries, there is a huge consulting industry offering personality checks to individuals who plan to become entrepreneurs. Products range from online questionnaires (some of them free of charge) to offline offers by chambers of commerce, by psychologists, consultants, coaches, mentors, and other practitioners who suggest comparisons of the personality of the potential entrepreneur with the average personality of successful entrepreneurs, charging substantial fees for their advice (Caliendo, Kritikos, Künn, Loersch, Schröder & Schütz 2014). Similarly, banks and investors are periodically tempted to implement personality inventories as part of their decision process of whether individuals should get loans or equity for their start-ups—with banks and investors aiming to apply deterministic thresholds levels where individuals would get access to capital only if the evaluation of their personality reveals a score above a benchmark (see the introductory quote, but also Rodionova 2015).

The benefits of a helpful advice would be numerous, while wrong advice can be very costly. If proper advice is provided, individuals would make better occupational choices; those who are not suited to become entrepreneurs would avoid costly misallocations; those who are suited to become entrepreneurs would be encouraged to do so. Banks would avoid credit defaults, investors avoid losses, and consumers would largely benefit from entrepreneurial entries creating better or cheaper products. Proper advice for or against entrepreneurship would also greatly benefit society in general, as half of nascent entrepreneurs fail in the first five years (Helmers & Rogers 2010; Quatraro & Vivarelli 2015). Economic and psychological costs generated by entrepreneurial failure, like the loss of the own savings, over-indebtedness, or unemployment in the aftermath of failure, could be reduced, if individuals being unfit for entrepreneurship are correctly advised to remain or become paid employees.

Therefore, to understand whether such an advice is indeed able to achieve the main goal—encouraging individuals with entrepreneurship-prone abilities to start an own business and discouraging those who do not have such abilities—the first step in making recommendations work is to ask a crucial question: What kind of method should be used when processing information for the advice of potential entrepreneurs?

A large part of research in this field (e.g., Begley & Boyd 1987; Stewart & Roth 2001; Zhao & Seibert 2006; Rauch & Frese 2007) employs the method of comparisons of average values of personality traits of managers, individuals in wage employment, or of unsuccessful entrepreneurs, with the average traits of successful entrepreneurs who are at the time of the survey in the market. We refer to this method as the *average-scores* approach, as a profile of a successful entrepreneur is created based on differences in average trait scores in inventories such as the Big Five. This method entails that advice to potential entrepreneurs should be given based on a comparison of their individual scores with the average metric, capturing information about the prototype successful entrepreneur (Zhao & Seibert 2006; Rauch & Frese 2007; Obschonka, Silbereisen & Schmitt-Rodermund 2010).

Although widely applied, to the best of our knowledge, no study has analyzed whether the average-scores or an alternative method produce valuable predictions. We close this gap by approaching this problem in the following way. First, we propose a conceptual framework allowing us to judge whether a methods is able to achieve the main goal, namely to induce better choices. For this reason, we develop a simple recommendation problem where giving an advice is based on personality traits correlated with a variable of interest—in our case, entrepreneurial fitness (for recent insights, see Levine & Rubinstein 2016; Manso 2016). We employ a performance measure that compares the outcome of optimal self-selection, a choice given perfect information, with the outcome produced by an adviser who uses different recommendation approaches. Secondly, based on our conceptual framework, we conduct a number of recommendation experiments using Monte Carlo methods, which numerically examine recommendation performance distributions. The numerical experiments are conducted in two steps. In the first step, we estimate the parameters of our data generating process with data from the German Socio-economic Panel, such that the Monte Carlo experiment is performed under realistic conditions. In the second step, we check the robustness of our results by conducting additional experiments using a multitude of different parameter combinations. We find that average scores' performance substantially lacks robustness.

The remainder of the chapter is organized as follows. We present our conceptual framework in Section 6.2. In Section 6.3, we analyze and compare performance distributions in different situations. Section 6.4 concludes.

6.2 Model and approaches

To analyze the performance of personality profiles and potential alternatives, we, first, introduce a simple model. The model's setup involves two types of individual characteristics and a prediction, respectively recommendation, problem to solve. Second, we develop an intuitive background regarding the model's environment by relating it to the person- and the variable-oriented approach, usually used to represent an entrepreneurial personality in psychology. Third, we introduce and discuss a measure of recommendation performance. Finally, we define and formally implement average scores and alternatives.

6.2.1 Linking two types of individual characteristics: The problem

Consider the following situation: A client of a consultant or mentor, plans to start a business and she wants to find out whether she has the necessary entrepreneurial aptitudes. To find an answer, she turns to an adviser, who is paid upfront and is only interested in producing a good recommendation result by maximizing the utility of the client. The task of the adviser is to give an informed occupational choice recommendation based on several variables, which, by the nature of a forecast, should be immediately observable. We assume that there are only two occupations: entrepreneur and non-entrepreneur.

Individual variables related to entrepreneurial aptitudes consist only of two types. The first type is a personality trait, $\Gamma \in \mathbb{R}$, that can be immediately observed with reasonable effort.[84] The second type is a measure of entrepreneurial fitness, $\Pi \in \mathbb{R}$, that can, in principle, be observed with some, again reasonable, effort, but only after the specific individual has been sufficiently long exposed to the market as an entrepreneur. Personality trait and fitness are assumed to be stochastic variables. We denote a realization of the personality trait by γ and a realization of fitness by π.

π measures fitness in relation to some reference point. The most straightforward method to construct π would be to use monetary entrepreneurial income in relation to the monetary income from alternative sources like wage income (for instance, Carter 2011):

$$\pi = \text{Income from entrepreneurship} - \text{Income from wage work}$$

Other reference points such as non-monetary values—for instance, independence resulting from being your own boss[85]—are also possible.[86]

[84] This simplifying assumption can be relaxed by using a vector of personality variables instead of univariate Γ. However, our simple model already allows for the most important aspects to be captured, such that using more traits would only increase complexity without providing a benefit.

[85] Blanchflower (2000), Hundley (2001), and Benz & Frey (2008a,b) show that entrepreneurs experience higher job satisfaction than wage workers such that non-monetary benefits of entrepreneurship might play a significant role with respect to occupational choices.

[86] There are two possible interpretations of π with respect to risk consistent with the model setting. First,

An individual i with $\pi_i = 0$ is neither fit nor unfit for entrepreneurship. Whether an individual is fit for entrepreneurship is determined by a cutoff approach. Let $\tau \in \mathbb{R}_0^+$ denote a cutoff agreed upon by client and adviser such that an individual is suited to become an entrepreneur if $\pi_i > \tau$. An individual is not suited to become an entrepreneur if $\pi_i \leqslant \tau$. Being perfectly aware of the entrepreneurial fitness of the individual asking for an advice would make the task of giving a good advice trivial to accomplish. However, Π is a typical ex-post variable, not known before the client becomes an entrepreneur, and the client's fitness cannot be directly used as the basis for the advice.

To fully characterize the population of clients, we assume the existence of three more parameters: μ_Γ is the population mean of the personality trait. μ_Π is the population mean of entrepreneurial fitness. ρ is the correlation between personality trait and fitness, measuring the strength and direction of the relationship between both variables. We only consider positive correlations such that $\rho \in (0, 1)$. Without loss of generality, the population variance of the personality trait and entrepreneurial fitness is normalized to 1. Let $\mathbf{v} = [\Gamma, \Pi]^\top$ denote a vector combining personality trait and fitness. We assume that $\mathbf{v}$ is bivariate normal with mean $\mathbf{m} = [\mu_\Gamma, \mu_\Pi]^\top$ and covariance $\mathbf{Q}$, where the covariance between the two variables is ρ, as variances are both 1.

To complete the problem setting, the adviser is assumed to have historical data, Θ, on Γ and Π. Θ is a finite index set where indices represent available historical observations.[87] Furthermore, it is assumed that clients and the historical sample are drawn from the same normal distribution. Note that the latter assumptions might make an advice easier than in reality, as in a real-world situation the characteristics of individuals turning to an adviser can differ from the overall population.

The historical sample consists of $n(\Theta)$ individuals, where $n(\cdot)$ denotes the cardinality of a set, the overall number of its elements. The adviser observes γ_i and π_i for every individual $i \in \Theta$ and can use the common measure of acceptable minimal entrepreneurial fitness, τ, to decompose historical observations into two groups: historical entrepreneurs, $E_\Theta = \{i \in \Theta : \pi_i > \tau\}$ and historical non-entrepreneurs, $E_\Theta^c = \{i \in \Theta : \pi_i \leqslant \tau\}$.

Theoretically, the same can be done for the client group denoted by Ω (a finite index set of clients) and consisting of $n(\Omega)$ individuals. In practice, since individuals in Θ and Ω are sampled from the same distribution, but as we do not know π_i for all $i \in \Omega$, we can only argue that the same association between personality trait and fitness must hold for Θ and Ω; yet, we can observe this association only for individuals in the historical sample.

Thus, the problem is:

entrepreneurship is not associated with any risk such that π is the deterministic relative income (relative to alternative income). Second, entrepreneurship is associated with risk and π is the average relative income but the client is risk-neutral and only cares about averages.

[87] Typically, even if Π can be measured without any problem, we would only have reliable historical data on it for a subset of individuals because it is a counterfactual for those who were never entrepreneurs.

Problem. Given the setting above, advice clients $i \in \Omega$ with respect to their relative entrepreneurial fitness, $\pi_i - \tau$, given that only the client's personality trait γ_i can be observed, such that fitness must be predicted from personality, using historical data Θ on trait and fitness.

6.2.2 The model and psychological research

Our model relates to the two perspectives most dominant in psychological research on how to properly assess the impact of personality on some variable of interest. Magnusson & Törestad (1993) identify the personality- and the variable-oriented approach. In entrepreneurship research, personality- and variable-oriented approaches correspond to the following concepts:

PERSONALITY-PROFILE PERSPECTIVE IN ENTREPRENEURSHIP RESEARCH
Conceptually following the ideas of Schumpeter (1934), a researcher using the personality-oriented approach constructs an entrepreneurship-prone reference personality profile (for examples, see Schmitt-Rodermund 2004; Obschonka et al. 2013). Deviations from the reference profile, measuring goodness of fit of the profile, can be quantified (for a method to measure differences between reference profiles and an observed set of personality traits, see Cronbach & Gleser 1953) and used to predict entrepreneurial aptitudes. Examples of profile-driven entrepreneurship research are Obschonka et al. (2013) and Stuetzer et al. (2012).

VARIABLE-ORIENTED PERSPECTIVE IN ENTREPRENEURSHIP RESEARCH
Variable-oriented approaches "focus on the effects of isolated variables on behavior" (Obschonka et al. 2013, p. 106). They assess the impact of, for instance, a personality trait on some variable of interest—usually, with the help of a linear regression such that trait effects are derived under a *ceteris paribus* condition. A typical empirical conclusion from the variable-oriented viewpoint is, for instance, as follows:

> Evidence suggests that entrepreneurship is associated with higher levels of extraversion, conscientiousness, and openness and lower levels of agreeableness and neuroticism … . (Obschonka et al. 2013, p. 106)

Examples of variable-oriented research include Costa, McCrae & Holland (1984), Zhao & Seibert (2006), Zhao et al. (2010), Manso (2016), and Levine & Rubinstein (2016).

With respect to the compatibility of both perspectives and our model, we establish the following result:

Proposition 6.1. *Our recommendation model is consistent with the personality-profile and the variable-oriented approach to model the entrepreneurial personality.*

Proof. To establish consistency with the profile-based approach, we must essentially answer the following question: What happens to the distribution of the individual trait if we condition on entrepreneurial fitness? Let $\mu_{\Gamma|E}$ denote the mean of Γ for entrepreneurs and let $\mu_{\Gamma|E^c}$ denote the mean of the personality trait for non-entrepreneurs. Similarly, denote the variance of personality trait, Γ, by $\sigma^2_{\Gamma|E}$, respectively $\sigma^2_{\Gamma|E^c}$. It is straightforward to derive that

$$\mu_{\Gamma|E} = \mu_\Gamma + \rho W(\kappa), \qquad \mu_{\Gamma|E^c} = \mu_\Gamma + \rho V(\kappa) \tag{6.1}$$

$$\sigma^2_{\Gamma|E} = 1 - \rho^2 w(\kappa), \qquad \sigma^2_{\Gamma|E^c} = 1 - \rho^2 v(\kappa) \tag{6.2}$$

where $\kappa = \tau - \mu_\Pi$, $W(\kappa) = \phi(\kappa)/[1 - \Phi(\kappa)] > 0$ where $\phi(\cdot)$ is the density and $\Phi(\kappa)$ the distribution function of the standard normal distribution; $V(\kappa) = -\phi(\kappa)/\Phi(\kappa) < 0$; $w(\kappa) = W(\kappa)[1 - W(\kappa)]$; and $v(\kappa) = V(\kappa)[1 - V(\kappa)]$.

As the correlation, ρ, determines how strong the connection is between the personality trait and entrepreneurial fitness, we focus on the role of this parameter. If trait and fitness are independent, the correlation between them is zero such that $\mu_{\Gamma|E} = \mu_{\Gamma|E^c} = \mu_\Gamma$ and $\sigma^2_{\Gamma|E} = \sigma^2_{\Gamma|E^c} = 1$. In such a setting, we cannot construct a distinct personality profile of an entrepreneur. However, if traits and fitness depend on each other with non-zero correlation, there will be a personality profile of an entrepreneur given by $\mu_{\Gamma|E}$. To see this, note that $\rho > 0$ implies

$$|\mu_{\Gamma|E} - \mu_{\Gamma|E^c}| = \rho[W(\kappa) - V(\kappa)] > 0 \tag{6.3}$$

such that there is a difference between the average trait of an entrepreneur and the average trait of a non-entrepreneur. The difference in (6.3) increases in the correlation between trait and entrepreneurial fitness. Furthermore, the variance of the personality trait conditional on being an entrepreneur, $\sigma^2_{\Gamma|E}$, decreases if the correlation between trait and fitness increases, as can be clearly seen in (6.2).

A personality- or profile-oriented approach has the following strategy. It takes the client's personality trait, γ, and compares it to the typical trait, $\mu_{\Gamma|E}$, of an entrepreneur. If Γ and Π are sufficiently correlated, the Γ-values of entrepreneurs will be concentrated in one place and Γ-values of non-entrepreneurs in another. Hence, similarity between the client's γ and profile $\mu_{\Gamma|E}$ is an indication that the client is an entrepreneur. If the correlation is weak, all Γ-values will be located in roughly one place independent from Π such that similarity between the client's trait, γ, and profile $\mu_{\Gamma|E}$ has little meaning.

To show consistency with the variable-oriented approach, let $\Psi \in \mathbb{R}$ denote a normally distributed variable with mean μ_Ψ and variance σ^2_Ψ. Ψ is assumed to capture all factors affecting entrepreneurial fitness that are not related to the personality trait, represented by Γ, such that we can assume that Ψ and personality trait, Γ, are independent. The variable-oriented approach is consistent with the following model of entrepreneurial fitness:

$$\Pi = a\Gamma + b\Psi \tag{6.4}$$

where a and b are nonzero constant scalars. For instance, let Γ represent extraversion (one of the Big Five personality traits). If $a > 0$, more extraversion will increase entrepreneurial fitness, which is in line with previous research (Costa et al. 1984; Zhao & Seibert 2006; Zhao et al. 2010). The difference between the variable-oriented perspective and entrepreneurship-prone profiles is that in the model in (6.4) there is no specific reference profile of an entrepreneur. The assumption underlying (6.4) is that, given $a > 0$, the trait Γ simply positively relates to entrepreneurial fitness, i.e., a higher score in Γ is associated with higher fitness. The model in (6.4) generates a joint distribution of personality trait and fitness that is consistent with our recommendation model. Notice that Π in (6.4) is normal (it is the sum of two normal independent variables) with the following parameters:

$$\mu_\Pi = a\mu_\Gamma + b\mu_\Psi, \qquad \sigma_\Pi^2 = a^2 + b^2\sigma_\Psi^2$$

The covariance between Π and personality trait Γ is given by

$$\sigma(\Pi, \Gamma) = a(1 + 2\mu_\Gamma^2) + 2b\mu_\Psi\mu_\Gamma$$

Furthermore, it can be demonstrated that Π and Γ are jointly normal according to the model in (6.4). The joint distribution of Π in model (6.4) and personality trait Γ is bivariate normal if and only if $Y = \alpha\Pi + \beta\Gamma$ is normal for any constant $\alpha, \beta \in \mathbb{R}$. It is obvious that Y is normal if either $\alpha = 0$ or $\beta = 0$, as Π and Γ are both normal. If $\alpha = \beta = 0$, $Y = 0$ with probability 1, which corresponds to a normal distribution with mean and variance zero. Hence, we must demonstrate that Y is normal if α and β are both nonzero. Note that Π and Γ are dependent and correlated. Furthermore, note that

$$Y = \alpha\Pi + \beta\Gamma = \alpha(a\Gamma + b\Psi) + \beta\Gamma = \delta_\Gamma\Gamma + \delta_\Psi\Psi \tag{6.5}$$

where $\delta_\Gamma \equiv \alpha a + \beta$ and $\delta_\Psi \equiv \alpha b$. Using independence and normality of Γ and Ψ, the moment-generating function of Y is given by

$$\begin{aligned} M_Y(t) = M_\Gamma(\delta_\Gamma t)M_\Psi(\delta_\Psi t) = \exp\left\{t\delta_\Gamma\mu_\Gamma + \frac{1}{2}\delta_\Gamma^2 t^2\right\} \\ \times \exp\left\{t\delta_\Psi\mu_\Psi + \frac{1}{2}\delta_\Psi^2\sigma_\Psi^2 t^2\right\} \end{aligned} \tag{6.6}$$

such that

$$M_Y(t) = \exp\left\{t\left[\delta_\Gamma\mu_\Gamma + \delta_\Psi\mu_\Psi\right] + \frac{1}{2}\left[\delta_\Gamma^2 + \delta_\Psi^2\sigma_\Psi^2\right]t^2\right\}$$

(6.6) is the moment-generating function of a normal distribution with mean $\delta_\Gamma\mu_\Gamma + \delta_\Psi\mu_\Psi$ and variance $\delta_\Gamma^2 + \delta_\Psi^2\sigma_\Psi^2$. As Y is normal for any constant α and β, Π and Γ must be bivariate normal. Without loss of generality, we can normalize a such that $a^2 = 1 - b^2\sigma_\Psi^2$ obtaining $\sigma_\Pi^2 = 1$ and $\rho = \sigma(\Pi, \Gamma)$. Hence, as our recommendation model, the model in (6.4) generates a bivariate normal distribution for $[\Gamma, \Pi]^\top$ with mean $\mathbf{m}$ and covariance $\mathbf{Q}$. ■

6.2.3 *Measuring adviser performance*

A straightforward way to introduce a measure of adviser performance is as follows. Let $\mathbf{a}_i \in \{0, 1\}$ denote an indicator function where $\mathbf{a}_i = 1$ if client $i \in \Omega$ is recommended to become an entrepreneur and $\mathbf{a}_i = 0$ if not. Note that function $\mathbf{a}_i$ depends on the approach applied. Furthermore, let $\mathbf{t}_i = \mathbb{1}\{\pi_i > \tau\} \in \{0, 1\}$, where $\mathbf{t}_i = 1$ if $i \in \Omega$ is actually fit for entrepreneurship and $\mathbf{t}_i = 0$ if not. If clients knew their true entrepreneurial fitness, if information would be perfect, they would self-select into entrepreneurship and non-entrepreneurship according to $\mathbf{t}_i$.

An easy to interpret performance measure can be constructed by comparing recommendations and the actual state of affairs. Thus, let

$$S = n(\Omega)^{-1} \sum_{i \in \Omega} \mathbb{1}\{\mathbf{a}_i = \mathbf{t}_i\} \tag{6.7}$$

The indicator in (6.7) can be interpreted as follows. If $\mathbb{1}\{\mathbf{a}_i = \mathbf{t}_i\} = 1$, recommendation and the actual state of affairs are the same such that the client was recommended entrepreneurship and the client was suitable to become an entrepreneur; or, alternatively, the client was recommended non-entrepreneurship and the client was not an entrepreneur. Put differently, $\mathbb{1}\{\mathbf{a}_1 = \mathbf{t}_1\} + \mathbb{1}\{\mathbf{a}_2 = \mathbf{t}_2\} + \cdots + \mathbb{1}\{\mathbf{a}_{n(\Omega)} = \mathbf{t}_{n(\Omega)}\}$ determines the number of correct recommendations. Consequently, $S \in [0, 1]$ is the relative number of correct recommendations, the recommendation success rate, which is linked to the adviser's reputation—the best adviser delivers recommendations with the highest success rate.

As every individual recommendation success indicator $\mathbb{1}\{\mathbf{a}_i = \mathbf{t}_i\}$ obeys a Bernoulli distribution with success probability p_i and, by construction of recommendation trials, we have $p_1 = p_2 = \cdots = p_{n(\Omega)} = p$, we must have

$$\tilde{S} = \sum_{i \in \Omega} \mathbb{1}\{\mathbf{a}_i = \mathbf{t}_i\} \sim \text{Binomial}(n(\Omega), p) \tag{6.8}$$

where p is the probability of a recommendation success given an arbitrary client from set Ω.

Note that $\mathbb{E}[\tilde{S}] = n(\Omega)p$ and $\mathbb{V}[\tilde{S}] = n(\Omega)p(1-p)$. The success probability p will depend on the distribution of historical and client data, and on the approach used to generate recommendations. Using (6.8), it is easy to establish that

$$\mathbb{E}[S] = n(\Omega)^{-1}\mathbb{E}\left[\sum_{i\in\Omega} \mathbb{1}\{\mathbf{a}_i = \mathbf{t}_i\}\right] = n(\Omega)^{-1}\mathbb{E}[\tilde{S}] = p$$

$$\mathbb{V}[S] = n(\Omega)^{-2}\mathbb{V}\left[\sum_{i\in\Omega} \mathbb{1}\{\mathbf{a}_i = \mathbf{t}_i\}\right] = n(\Omega)^{-2}\mathbb{V}[\tilde{S}] = n(\Omega)^{-1}p(1-p)$$

The variance of recommendation success probability S is largest at $p = 1/2$ and decreases to both sides such that $\mathbb{V}[S]$ takes a minimum at $p = 0$ and $p = 1$. Hence, in our problem setting, a high success probability p automatically implies a low recommendation success variance and an increase in the success rate, given that $p \geqslant 1/2$, reduces variance. Therefore, p allows for the derivation of conclusions with respect to the recommendation success probability *and* success variance at the same time.

6.2.4 *Profile-based recommendations*

First, we provide a general version of the average-scores approach. Then, we suggest some ideas on how to optimize recommendations based on average scores.

General average scores

A straightforward algorithm to implement the average-scores approach consists of four steps. The first step is to take historical data, Θ, and divide individuals into two groups by using the cutoff, τ. One group consists of historical entrepreneurs, E_Θ, and the second group consists of historical non-entrepreneurs, E^c_Θ. Second, construct a personality profile for each group by taking averages over the personality trait γ:

$$\hat{\gamma}_E = n(E_\Theta)^{-1}\sum_{j\in E_\Theta}\gamma_j, \qquad \hat{\gamma}_{E^c} = n(E^c_\Theta)^{-1}\sum_{j\in E^c_\Theta}\gamma_j \tag{6.9}$$

Third, the adviser should, by employing some statistical test, verify that $\hat{\gamma}_E$ and $\hat{\gamma}_{E^c}$ are significantly different from each other. If there is no significant difference, distinct personality profiles do not exist. Lastly, the adviser recommends entrepreneurship if the client is sufficiently similar to the personality profile of an entrepreneur or

$$\mathbf{a}_i^{AS} = \begin{cases} 1 & \gamma_i \in h(\epsilon) \\ 0 & \gamma_i \notin h(\epsilon) \end{cases} \quad \text{for } i \in \Omega \tag{6.10}$$

where $h(\epsilon)$ is a similarity interval given by $h(\epsilon) = (\hat{\gamma}_E - \epsilon, \hat{\gamma}_E + \epsilon)$. The similarity criterion $\epsilon \in \mathbb{R}^+$ is set by the adviser. Internet services specializing in occupational choice recommendations, essentially, employ the general version of the average-scores approach where the similarity criterion is set according to some, rather non-transparent, considerations.

Optimized average scores

As the similarity criterion, ϵ, is a parameter freely set by the adviser, it is plausible to assume that the adviser might try to systematically optimize his recommendation performance by setting an appropriate ϵ, based on, for instance, prior recommendation experience. A proper objective to optimize is the expected recommendation success rate, $p_{AS} = \mathbb{E}[S_{AS}]$. Hence, the task of the adviser is to find an ϵ^* such that

$$p_{AS} = n(\Omega)^{-1}\mathbb{E}\left[\sum_{i\in\Omega} \mathbb{1}\{\mathbf{a}_i^{AS} = \mathbf{t}_i\}\right] = \mathbb{P}(\mathbf{a}^{AS} = 1 \wedge \mathbf{t} = 1) + \mathbb{P}(\mathbf{a}^{AS} = 0 \wedge \mathbf{t} = 0) \quad (6.11)$$

where

$$\mathbb{P}(\mathbf{a}^{AS} = 1 \wedge \mathbf{t} = 1) = \int_{\gamma_E-\epsilon}^{\gamma_E+\epsilon} \int_{\tau}^{\infty} \phi_{\mathbf{m},\mathbf{Q}}(\gamma, \pi)\mathrm{d}\pi\mathrm{d}\gamma \quad (6.12)$$

$$\mathbb{P}(\mathbf{a}^{AS} = 0 \wedge \mathbf{t} = 0) = \int_{-\infty}^{\gamma_E-\epsilon} \int_{-\infty}^{\tau} \phi_{\mathbf{m},\mathbf{Q}}(\gamma, \pi)\mathrm{d}\pi\mathrm{d}\gamma + \int_{\gamma_E+\epsilon}^{\infty} \int_{-\infty}^{\tau} \phi_{\mathbf{m},\mathbf{Q}}(\gamma, \pi)\mathrm{d}\pi\mathrm{d}\gamma \quad (6.13)$$

is maximized. $\phi_{\mathbf{m},\mathbf{Q}}$ denotes the joint distribution of the personality trait and entrepreneurial fitness. To see that p_{AS} is the probability of a successful recommendation if the adviser uses the average-scores approach, note that there are two ways to generate a correct recommendation. The adviser recommends entrepreneurship and the client is an entrepreneur, which occurs with probability $\mathbb{P}(\mathbf{a}^{AS} = 1 \wedge \mathbf{t} = 1)$. Alternatively, the adviser does not recommend entrepreneurship and the client is not an entrepreneur occurring with probability $\mathbb{P}(\mathbf{a}^{AS} = 0 \wedge \mathbf{t} = 0)$. Under the assumption that the joint distribution of the personality trait and entrepreneurial fitness is known and there is some known average personality trait of entrepreneurs γ_E, probabilities are computed by (6.11) and (6.12).

On the basis of ϵ^*, the adviser can construct an optimized similarity interval $h(\epsilon^*)$ such that, in addition to the general average-scores approach, there exists an optimized version. If the optimized version is used, recommendations are given according to

$$\mathbf{a}_i^{OAS} = \begin{cases} 1 & \gamma_i \in h(\epsilon^*) \\ 0 & \gamma_i \notin h(\epsilon^*) \end{cases} \quad \text{for } i \in \Omega \quad (6.14)$$

Optimizing (6.11) requires that the joint distribution of the personality trait and entrepreneurial fitness is perfectly known. In a realistic scenario, this will not hold and parameters must be estimated with historical data. Thus, even if the adviser systematically optimizes, his similarity criterion might deviate from the optimal criterion ϵ^* due to estimation errors.

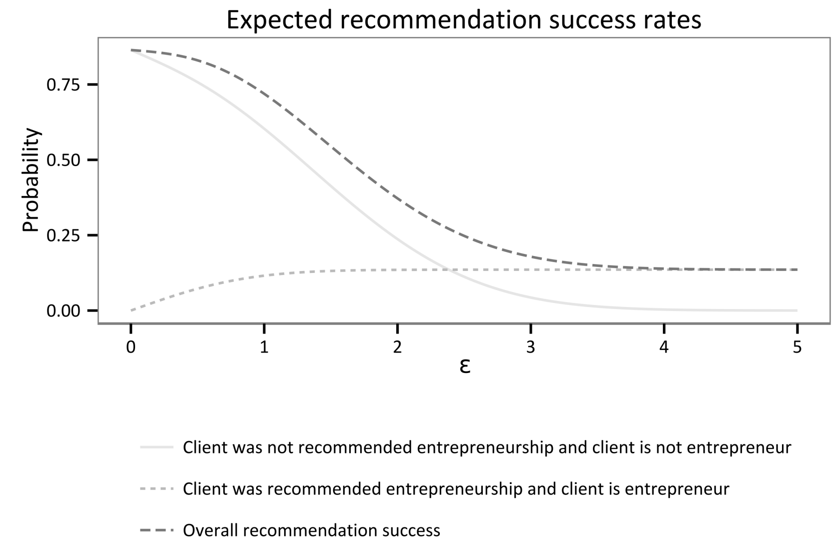


FIGURE 6.1. An example of changes in expected recommendation success when using the average scores approach with different similarity criteria

Example. To provide an example on how to optimize average-scores performance and pitfalls associated with optimization, consider a numerical scenario. Let $\mu_\Gamma = 1$, $\mu_\Pi = -1$, $\rho = 0.8$, and $\tau = 0.1$. For the average personality trait of entrepreneurs, we assume $\gamma_E = \mathbb{E}[\Gamma|\Pi > \tau]$ ($\hat{\gamma}_E$ is just an estimator of $\mathbb{E}[\Gamma|\Pi > \tau]$). The share of entrepreneurs in the population is $1-\Phi(\tau-\mu_\Pi) \approx 14\%$, where Φ is the distribution function of the standard normal. Thus, most clients are not entrepreneurs.

According to (6.11), the probability of a successful recommendation is the sum of the probability that a client recommended entrepreneurship is an entrepreneur and that a client not recommended entrepreneurship is not an entrepreneur. In Figure 6.1, we plot all three probabilities as a function of the similarity criterion, ϵ. The adviser would achieve most recommendation successes if he sets a very strict similarity criterion such that only clients very similar to the entrepreneurship-prone profile are recommended entrepreneurship. This strategy will not necessarily result in an identification of clients suited to become entrepreneurs—in fact, the probability that a client recommended entrepreneurship is an entrepreneur is almost zero—but non-entrepreneurs are identified with a very high probability. The high probability to identify non-entrepreneurs results in a high recommendation success probability. As the similarity criterion increases, the interval $h(\epsilon)$ widens resulting in a higher probability to recommend entrepreneurship. As a consequence, the probability to identify entrepreneurs improves but the probability to identify non-entrepreneurs decreases. What is more important, the probability to identify clients unfit for entrepreneurship decreases more strongly than the probability to identify entrepreneurs such that the overall recommendation success probability decreases.

6.2.5 *Feasible performance bounds*

To benchmark average scores, we establish some plausible performance bounds. The first benchmark, which we call the probability-based approach, directly estimates the probability of being suited to become an entrepreneur, and constitutes a feasible upper performance bound. The lower performance bound is given by the toss of an unbiased coin.

The upper bound, the probability-based approach, exploits the well-known result that every conditional distribution of a multivariate normal distribution is normal itself. Conditional on an observation of personality trait, γ_i for $i \in \Omega$, the probability of being suited to become an entrepreneur is given by

$$\mathbb{P}(\Pi > \tau | \Gamma = \gamma_i) = 1 - \int_{-\infty}^{\tau} \phi_{\tilde{\mu}_\Pi, \tilde{\sigma}_\Pi}(\pi) \mathrm{d}\pi \tag{6.15}$$

where

$$\tilde{\mu}_\Pi = \mu_\Pi + \sigma_\Pi \rho \left(\frac{\gamma_i - \mu_\Gamma}{\sigma_\Gamma} \right), \qquad \tilde{\sigma}_\Pi = (1 - \rho^2)\sigma_\Pi^2$$

The parameters $\mathbf{m}$ and $\mathbf{Q}$ are unknown. Therefore, $\mathbb{P}(\Pi > \tau | \Gamma = \gamma_i)$ is also unknown. However, parameters can be estimated with historical data, to get an estimate of $\mathbb{P}(\Pi > \tau | \Gamma = \gamma_i)$. Let $\mathbf{w}_1, \ldots, \mathbf{w}_{n(\Theta)}$ denote joint observations of personality trait and entrepreneurial fitness. Instead of $\mathbf{m}$, we use

$$\hat{\mathbf{m}} = n(\Theta)^{-1} \sum_{i \in \Theta} \mathbf{w}_i$$

Instead of covariance matrix $\mathbf{Q}$, we use the sample covariance

$$\hat{\mathbf{Q}} = \frac{1}{n(\Theta) - 1} \sum_{i \in \Theta} (\mathbf{w}_i - \hat{\mathbf{m}})(\mathbf{w}_i - \hat{\mathbf{m}})^\top$$

Let $\hat{p}_i^E$ denote the estimated probability that client $i \in \Omega$ is fit for entrepreneurship, an estimate of (6.15) using estimated parameters $\hat{\mathbf{m}}$ and $\hat{\mathbf{Q}}$. Note that the estimated probability that i is not fit for entrepreneurship is $1 - \hat{p}_i^E$. The adviser can recommend entrepreneurship to if the probability that i is an entrepreneur exceeds the probability that i

is a non-entrepreneur:

$$\mathbf{a}_i^{PBA} = \begin{cases} 1 & \hat{p}_i^E > 1 - \hat{p}_i^E \\ 0 & \text{else} \end{cases} \quad \text{for } i \in \Omega \tag{6.16}$$

The probability-based approach can only be applied if the joint distribution of the personality trait and entrepreneurial fitness is (approximately) bivariate normal—if this is not the case, deriving conditional distributions is more difficult.

The lower bound, the coin, generates the following recommendation:

$$\mathbf{a}_i^{COIN} = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2} \end{cases} \tag{6.17}$$

where recommendation probabilities do not depend on data.

6.3 Performance of average scores

In this section, we present the main results of our analysis. In preparation, we introduce two desirable properties of a recommendation approach derived from the intuitive ideas that an approach should perform substantially better than a coin toss and that recommendation success rates should be sufficiently stable. Second, to perform our analysis under realistic conditions, we calibrate our model setting with data from the German Socio-economic Panel. Then, we evaluate the recommendation performance of average scores and test whether this approach has the two desirable properties.

6.3.1 Properties

By introducing the toss of a coin as a benchmark, we introduced an absolute lower performance boundary. Every approach suitable for occupational choice advice should at least outperform the coin. Thus, the first requirement is that the average recommendation success rate of an approach should be substantially larger than 50%, as 50% is the average success rate of an unbiased coin.[88]

[88] Note that $p_{\text{COIN}} = \mathbb{E}[S_{COIN}] = \mathbb{P}(\mathbf{a}_{COIN} = 1 \land \mathbf{t} = 1) + \mathbb{P}(\mathbf{a}_{COIN} = 0 \land \mathbf{t} = 0)$. The coin completely ignores historical and client data such that $\mathbb{P}(\mathbf{a}_{COIN} = a)$ and $\mathbb{P}(\mathbf{t} = t)$ are independent. Hence, we get

$$p_{COIN} = \mathbb{P}(\mathbf{a}_{COIN} = 1)\mathbb{P}(\mathbf{t} = 1) + \mathbb{P}(\mathbf{a}_{COIN} = 0)\mathbb{P}(\mathbf{t} = 0)$$

Given that the probability that an arbitrary individual is an entrepreneur is $\mathbb{P}(\mathbf{t} = 1) = 1 - \Phi(\tau - \mu_\Pi)$, we get

$$p_{COIN} = \frac{1}{2}[1 - \Phi(\tau - \mu_\Pi)] + \frac{1}{2}\Phi(\tau - \mu_\Pi) = \frac{1}{2}$$

As average scores rely on a similarity criterion, a second sensible requirement is that changing the similarity criterion does not significantly affect recommendation success, such that success is reasonably stable. To put it more formally, let ε and ε' denote two different similarity criteria. Let $S(\varepsilon)$ denote the recommendation success rate of an approach depending on some similarity criterion ε. Given the same distribution parameters $\mathbf{m}$ and $\mathbf{Q}$, we can always find a $\Delta(\varepsilon, \varepsilon') = \Delta(\varepsilon', \varepsilon) \in [0, 1]$ such that

$$\Delta(\varepsilon, \varepsilon') = |\mathbb{E}[S(\varepsilon) - S(\varepsilon')]| \leqslant \bar{\Delta}(\varepsilon, \varepsilon')$$

We require that $\Delta(\varepsilon, \varepsilon')$ has a reasonably small upper bound $\bar{\Delta}(\varepsilon, \varepsilon')$.[89]

6.3.2 Model calibration

Our goal is to use a data generating process (DGP) that is as realistic as possible. Therefore, we estimate model parameters with data. The application of Monte Carlo methods requires the specification of the parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_{DGP}^\top, \boldsymbol{\theta}_{Approach}^\top]^\top$ where

$$\boldsymbol{\theta}_{DGP} = \begin{bmatrix} \mu_\Gamma \\ \mu_\Pi \\ \sigma_\Gamma \\ \sigma_\Pi \\ \rho \end{bmatrix}, \qquad \boldsymbol{\theta}_{Approach} = \begin{bmatrix} \tau \\ \epsilon \end{bmatrix}$$

$\boldsymbol{\theta}_{DGP}$, including variances of trait and fitness that might not be 1 in data, fully determines the data generating process. Parameters in $\boldsymbol{\theta}_{Approach}$ are either idiosyncratic to the recommendation approach (similarity criteria) or set by the client (minimal fitness to be an entrepreneur). Consequently, we, first, estimate $\boldsymbol{\theta}_{DGP}$ from data and, then, use the calibrated model to analyze performance distributions.

6.3.2.1 Estimation of model parameters

To estimate the model's parameter $\boldsymbol{\theta}_{DGP}$, we use data from the German Socio-economic Panel (GSOEP). The GSOEP is a longitudinal survey of a large representative sample

The variance of recommendation successes rates is easy to derive and given by

$$\mathbb{V}[S_{COIN}] = \frac{p_{COIN}(1 - p_{COIN})}{n(\Omega)} = \frac{1}{4n(\Omega)}$$

[89]For instance, $\Delta(\varepsilon, \varepsilon') = 1/2$ implies that changing the similarity criterion by a certain amount would change the expected recommendation success rate by 50 percentage points in the same setting indicating that approach performance is unstable.

of German individuals and households with coverage from 1984 to 2015 and provides, among other variables, information on personality, occupational status, and earnings. We restrict our attention to individuals who provided a self-reported measure of the personal willingness to take risk, and reported monthly income from wage work *and* self-employment.[90]

The relation between the willingness to take risk and entrepreneurial outcomes is well established in the literature (see Kihlstrom & Laffont 1979; Caliendo, Fossen & Kritikos 2009, 2010). By focusing on individuals who were both wage workers and self-employed during their careers, we avoid the problem of unknown counterfactuals. We take averages over time of entrepreneurial income and wage to reduce noise. The willingness to take risk is measured on a scale from zero (not willing to take risk) to 10 (very high willingness to take risk). In line with Obschonka et al. (2013), we use the highest reported measure of the willingness to take risk as our trait related to entrepreneurship. In sum, there are 87 available observations, each observation corresponding to a different individual with a certain risk attitude, and experience in entrepreneurship and wage work.

Our model requires that the joint distribution of trait and fitness is bivariate normal. Therefore, data must be transformed, as nominal income differences and risk attitudes are not normally distributed. Sample entrepreneurial fitness is, in line with suggestions in Section 6.2.1, approximated by average entrepreneurial income (AEI) relative to the average wage (AW). To ensure that the condition of joint normality holds, we take the logarithm such that sample fitness is given by

$$\pi_{Sample} = \log\left(\frac{\text{AEI}}{\text{AW}}\right) = \log(\text{AEI}) - \log(\text{AW})$$

The transformation resulting in π_{Sample} has the downside that the minimal fitness requirement must be applied to the difference in log incomes and not nominal incomes. Consequently, $\pi_i > \tau$, indicating that the individual is fit for entrepreneurship, has no simple interpretation. However, using $\tau = 0$, such that $\log(\text{AEI}) - \log(\text{AW}) > \tau = 0$ and transforming the log difference back to nominal incomes yields the condition

$$\exp(\log(\text{AEI}) - \log(\text{AW})) > \exp(\tau) = \exp(0) = 1$$

which is equivalent to AEI > AW. Hence, we use $\tau = 0$ as the minimal fitness requirement since this reduces to the simple condition that entrepreneurial income must be larger than the wage.

The original measure of the willingness to take risk γ_* is transformed by applying the

[90] Self-employment and wage work can take place at the same time or at different time points.

Box-Cox transformation:

$$\gamma_{Sample} = \frac{\gamma_*^{\lambda} - 1}{\lambda}$$

where $\lambda \neq 0$. Given that γ_* is measured on a scale from zero to 10 and we have no observations where $\gamma_* = 0$, and, thus, $d\gamma_{Sample}/d\gamma_* = \gamma_*^{\lambda-1} > 0$, an increase in γ_{Sample} can be interpreted as an increase in the willingness to take risk. We set $\lambda = 1.5$, as this value provides a sufficiently good transformation ensuring the bivariate normality of trait and entrepreneurial fitness.

Without loss of generality, we normalize π_{Sample} and γ_{Sample} such that they have a variance of 1. Table 6.1 depicts data characteristics including skewness and kurtosis, which

TABLE 6.1. Data characteristics

Variable	Min	Max	Mean	Skewness	Kurtosis	p-value for skewness = 0	p-value for kurtosis = 3
Normalized trait	0.28	4.65	2.61	−0.32	2.81	0.19	0.10
Normalized fitness	−3.00	2.12	−0.31	−0.41	3.24	0.93	0.43

Notes: The test of skewness is performed in line with D'Agostino (1970) and the test of kurtosis in line with Anscombe & Glynn (1983).

are tested against the skewness and kurtosis of a univariate normal distribution, where skewness is zero and kurtosis is 3. The hypothesis that skewness and kurtosis are similar to the normal distribution cannot be rejected at the 5%-level. In addition, we perform the multivariate normality test of Mardia (1974), which cannot reject the hypothesis of multivariate normality at the 5%-level (p-value for multivariate skewness is 0.12 and for multivariate kurtosis 0.09).

Using the assumption of bivariate normality, we estimate the model's parameters by maximum likelihood resulting in

$$\hat{\boldsymbol{\theta}}_{DGP} = \begin{bmatrix} \hat{\mu}_{\Gamma} \\ \hat{\mu}_{\Pi} \\ \hat{\sigma}_{\Gamma} \\ \hat{\sigma}_{\Pi} \\ \hat{\rho} \end{bmatrix} = \begin{bmatrix} 2.61 \\ -0.31 \\ 0.99 \\ 0.99 \\ 0.21 \end{bmatrix}$$

The estimated correlation, $\hat{\rho} = 0.21$, is positive and significantly different from zero at the 5%-level (p-value $= 0.04$). Thus, more willingness to take risk is associated with higher entrepreneurial fitness. Put differently, willingness to take risk predicts entrepreneurial outcomes.

6.3.3 Performance analysis

The analysis is performed using $\hat{\theta}_{DGP}$ to generate draws from the historical and the client sample. The historical sample is assumed to include 1,000 individuals, whereas the client sample includes 100. Given parameters $\hat{\theta}_{DGP}$, there are two trait profiles. The average trait value for individuals with sufficient entrepreneurial fitness is 2.8, whereas individuals who are not fit for entrepreneurship have an average trait of 2.5.

As τ is fixed ($\tau = 0$), $\theta_{Approach}$ has one parameter that must be set: the similarity criterion ϵ. We assume that $\epsilon \in [0.01, 4.99]$, as performance does not change much for $\epsilon > 4$, and vary the parameter in steps of 0.02, resulting in 250 values. For one parameter value, we simulate performance $M = 10,000$ times and compute expected performance by taking the simulation average

$$M^{-1} \sum_{m=1}^{M} S_m$$

The lower performance boundary is 0.5 and the upper boundary, based on the probability-based approach (PBA) and simulated with $\hat{\theta}_{DGP}$, is 0.67.

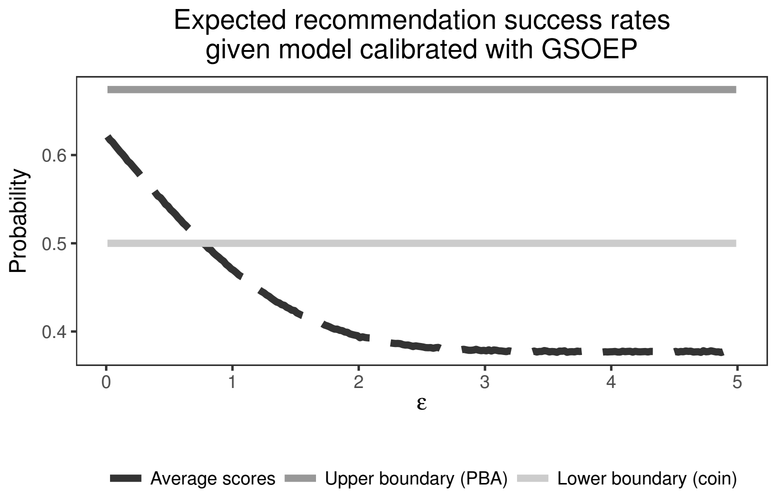


FIGURE 6.2. Expected recommendation success using model calibrated with GSOEP data

Figure 6.2 presents simulation results. Expected average-scores performance is depicted as a function of the approach-specific similarity criterion. We also show the two feasible performance bounds. The figure reveals that average-scores performance can be optimized by using a very strict similarity criterion ($\epsilon = 0.01$) such that the maximal recommendation success rate is 62%, which is 5 percentage points less than the feasible upper performance bound (probability-based approach). However, average-scores performance is very sensitive to the similarity criterion.

We postulated two intuitive requirements an approach should fulfill to be considered as a prediction and recommendation method. The first requirement is that the average recommendation success rate should be substantially larger than the average success rate resulting from the toss of a coin. Table 6.2 shows the minimal performance of average scores. The minimal performance of average scores is only 38%, which is below the coin. Thus, average scores violate the first requirement.

TABLE 6.2. Summary of performance analysis of average scores based on GSOEP

Min	Max	Max − min	Maximal performance as percentage of upper boundary	Minimal performance as percentage of lower boundary
0.38	0.62	0.24	0.92	0.75

The second requirement demands that an approach depending on a similarity criterion is sufficiently robust, in the sense that changing the similarity criterion should not have a large effect on average recommendation success rates. Sufficient recommendation success stability is an important property since data will be plagued by measurement errors and estimations also produce errors. Our robustness measure, $\Delta(\varepsilon, \varepsilon')$, determines how much potential adviser mistakes—setting an inappropriate similarity criterion—cost in terms of average recommendation success rates. In Table 6.2, we depict the difference between maximal and minimal performance. Using an inappropriate similarity criterion can result in a performance loss of 24 percentage points.

Hence, average scores, respectively profiles, might not exhibit sufficient performance and are not sufficiently stable to be considered as a method for predictions and recommendations.

6.3.4 Robustness of results

Although the model calibrated with data is close to real-world conditions, it only reflects a particular setting, as parameters are estimated with German data. To check the robustness of our results, we repeated a numerical recommendation experiment using different combinations of parameters resulting in different recommendation settings. We, especially, varied the assumption on the share of entrepreneurs in the population. The simulation setup and results are described in Appendix 6.A. Results are consistent with those obtained with German data.

6.4 Conclusion

The major aim of this contribution is to provide an answer to the question of whether using information on personality scores of successful individuals provides a helpful method to predict the success of an arbitrary individual. While predictions and advice based on a comparison of a prototype with the scores of the individual seeking such advice has been established in the recent literature and in everyday business, it has not been discussed whether this is a proper approach and whether alternative approaches might deliver better results.

Using career advice for illustration purposes, we design a simple framework involving two stochastic variables, having various degrees of correlation, generated from a bivariate normal distribution. One variable is interpreted as entrepreneurial fitness. The other variable is an individual personality trait. Our model's setting is consistent with a holistic and a variable-oriented view of the entrepreneurial personality. The problem to solve is to give a recommendation regarding an individual's entrepreneurial fitness by examining only the individual's personality variable, while having historical data on both variables. The data generating process of the problem setting is calibrated with German data on personality and entrepreneurial fitness.

We demonstrate that, predictions and recommendations based on the average-scores approach sometimes provide even worse predictions and recommendations than the toss of a coin. For instance, if there are many entrepreneurs in the client group of an adviser, using a too strict average-scores similarity criterion will result in too few recommendations for entrepreneurship because many clients with a high probability to be an entrepreneur are recommended against. At the same time, as many entrepreneurs are not properly identified and, due to the strictness of the similarity criterion, recommended non-entrepreneurship, the probability to identify a non-entrepreneur is also low. In such a situation, a coin, always generating 50% correct recommendations, does better. In a calibrated model approximating real-world conditions, we find that average scores have a maximal success probability of around 60% and that performance is highly unstable.

From a policy perspective, if individuals seek external (career) advice, they should, rather avoid following a consultant or internet-based questionnaire comparing clients' scores with average scores of groups (such as entrepreneurs), as they would risk an advice inferior to the toss of a coin. Furthermore, banks and investors should be aware that personality-score evaluations based on average reference personalities, which are relatively cheap and easy to generate, are unlikely to significantly reduce the risk of their loan or investment portfolio.

7 APPENDICES

Appendix 2.A

This appendix provides the proof of Lemma 2.1.

Proof of Lemma 2.1. Let s_1 be an optimally constructed firm with $M_1 > 0$ employees where the entrepreneur j is not counted as employee. m_1 individuals are in the entrepreneur's unit where $m_1 > 0$. $M_1 - m_1 = m_1'$ individuals are in other units where $m_1' > 0$. Let s_2 be the same firm structure but with no individuals in other units other than the entrepreneur's unit (individuals in the entrepreneur's unit are the same in s_1 and s_2), i.e., we have $m_2 = M_2 = m_1$ and $m_2' = 0$. Worker's wage bills W are

$$\begin{aligned} W_{s_1} &= \sum_{i=1}^{m_1}[v_{\{2\}i} + \Delta v_{\mathcal{U}_i^{(0)}}] + \sum_{i=m_1+1}^{m_1'}[v_{\{2\}i} + \Delta v_{\mathcal{U}_i^{(0)}}] \\ W_{s_2} &= \sum_{i=1}^{m_1} v_{\{2\}i} \end{aligned}$$

where $\Delta v_{\mathcal{U}_i^{(0)}}$ is the managerial advantage of i over second-best managers of her unit. Entrepreneurial income in structure 1 is $\pi_{s_1} = v_{\{1\}j} + \sum_{i=1}^{M_1} v_{\{1\}i} - W_{s_1}$ or

$$\pi_{s_1} = v_{\{1\}j} + \sum_{i=1}^{m_1}[v_{\{1\}i} - v_{\{2\}i} - \Delta v_{\mathcal{U}_i^{(0)}}] + \sum_{i=m_1+1}^{m_1'}[v_{\{1\}i} - v_{\{2\}i} - \Delta v_{\mathcal{U}_i^{(0)}}]$$

However, since the aggregate managerial advantage of a worker is fully rewarded to this worker, we must have $\sum_{i=1}^{m_1} \Delta v_{\mathcal{U}_i^{(0)}} + \sum_{i=m_1+1}^{m_1'} \Delta v_{\mathcal{U}_i^{(0)}} = \sum_{i=m_1+1}^{m_1'}[v_{\{1\}i} - v_{\{2\}i}]$ and, therefore

$$\pi_{s_1} = v_{\{1\}j} + \sum_{i=1}^{m_1}[v_{\{1\}i} - v_{\{2\}i}]$$

Entrepreneurial income in structure 2 is $\pi_{s_2} = v_{\{1\}j} + \sum_{i=1}^{M_2} v_{\{1\}i} - W_{s_2}$ or

$$\pi_{s_2} = v_{\{1\}j} + \sum_{i=1}^{m_1}[v_{\{1\}i} - v_{\{2\}i}]$$

where we used $M_2 = m_1$. We, thus, can verify that $\pi_{s_2} = \pi_{s_1}$, i.e., entrepreneurial income is the same across structures s_1 and s_2. ■

Appendix 2.B

This appendix provides the proof of Lemma 2.3.

Proof of Lemma 2.3. Use iteration to obtain the following equations. The attraction of wage work is

$$\mathscr{W}_{i,t} = \frac{\alpha_i^t}{t+1}\bar{\delta} w_{i,0} + \frac{1-\alpha_i^t}{(t+1)(1-\alpha_i)} w_{i,0} + \frac{1}{t+1}\sum_{k=1}^{t} \alpha_i^{t-k} \tilde{w}_{i,k} \tag{2.B.1}$$

The attraction of entrepreneurship without market feedback is

$$\mathscr{E}_{i,t} = \frac{\alpha_i^t}{t+1}\bar{\delta}(1-c)\hat{\pi}_{i,0} + \frac{1-\alpha_i^t}{(t+1)(1-\alpha_i)}(1-c)\hat{\pi}_{i,0} + \frac{1-c}{t+1}\sum_{k=1}^{t} \alpha_i^{t-k} \tilde{\pi}_{i,k} \tag{2.B.2}$$

and the attraction for an individual becoming entrepreneur in period 1 with feedback is

$$\mathscr{E}_{i,t} = \frac{\alpha_i^t}{t+1}\bar{\delta}(1-c)\hat{\pi}_{i,0} + \frac{1-\alpha_i^t}{(t+1)(1-\alpha_i)}(1-c)\pi_{i,0} \tag{2.B.3}$$

$$+\frac{1-c}{t+1}\sum_{k=1}^{t} \alpha_i^{t-k} u_{i,k} + \frac{1-c}{t+1}\sum_{k=1}^{t} \alpha_i^{t-k} \tilde{\pi}_{i,k}$$

Using Assumption 2.1.1, restricting labor market interactions, we obtain $\tilde{w}_{i,t} = \tilde{\pi}_{i,t} = 0$ for all t.

As a side note, consider what would happen if feedback was noise-free such that $u_{i,k} = 0$. The relative attraction of wage work for an individual who is entrepreneur in period 1 would be

$$\mathscr{W}_{i,t} - \mathscr{E}_{i,t} = \frac{\alpha_i^t}{t+1}\hat{\varkappa} + \frac{1-\alpha_i^t}{(t+1)(1-\alpha_i)}\varkappa$$

$$\hat{\varkappa} \equiv \bar{\delta}[w_{i,0} - (1-c)\hat{\pi}_{i,0}], \qquad \varkappa \equiv w_{i,0} - (1-c)\pi_{i,0}$$

where $\hat{\varkappa} < 0$, as the individual started as entrepreneur. There are two interesting cases. Case 1: The individual starts as entrepreneur and is right about it such that $\varkappa < 0$. Then, $\mathscr{W}_{i,t} - \mathscr{E}_{i,t} < 0$ for all t and i never selects an option different from the full-information equilibrium, which is entrepreneurship. Case 2: The individual starts as entrepreneur and is wrong about it such that $\varkappa > 0$. Then, the individual will switch to wage work at $t^* + 1$

where

$$t^* = \frac{\log(-\varkappa[\hat{\varkappa}(1-\alpha_i) - \varkappa]^{-1})}{\log(\alpha_i)}$$

and remain a wage worker for the rest of her career. Hence, given a noise-free feedback, the best option is always identified.

Given Assumption 2.1.1, the relative attraction of wage work for an employee in period 1, based on (2.B.1) and (2.B.2), is

$$\mathscr{W}_{i,t} - \mathscr{E}_{i,t} = \frac{\alpha_i^t}{t+1}\bar{\delta}[w_{i,0} - (1-c)\hat{\pi}_{i,0}] + \frac{1-\alpha_i^t}{(t+1)(1-\alpha_i)}[w_{i,0} - (1-c)\hat{\pi}_{i,0}] > 0 \quad \text{for all } t$$

where the relation is given because selecting wage work in period 1 implies $w_{i,0} - (1 - c)\hat{\pi}_{i,0} > 0$. However, switching to entrepreneurship requires that $\mathscr{W}_{i,t} - \mathscr{E}_{i,t} \leqslant 0$ for some t, which cannot hold. Without restrictions on interactions, $\mathscr{W}_{i,t} - \mathscr{E}_{i,t}$ can become non-positive as

$$\sum_{k=1}^{t} \alpha_i^{t-k}\tilde{w}_{i,k} < (1-c)\sum_{k=1}^{t} \alpha_i^{t-k}\tilde{\pi}_{i,k} \tag{2.B.4}$$

holds if workers are denied a sufficient number of opportunities to earn more. The same argument applies to entrepreneurs who exit and consider returning to entrepreneurship, as this would also require that (2.B.4) holds, which is ruled out by Assumption 2.1.1. ■

Appendix 2.C

This appendix provides the proofs of Proposition 2.2, 2.3, 2.4, and 2.5.

2.C.1 Overconfidence

Proof of Proposition 2.2. Use the results on truncated normal and bivariate normal distributions provided by Greene (2002)[91] to obtain the following averages. The mean of estimated self-management productivities of entrepreneurs is given by

$$\mathbb{E}[\nu_E] = \mathbb{E}[\nu|\nu \geqslant \xi] = \mathbb{E}[\nu] + \mathbb{V}[\nu]^{\frac{1}{2}}\lambda(\kappa)$$

where

$$\lambda(\kappa) \equiv \frac{f(\kappa)}{1-F(\kappa)} > 0, \quad \kappa \equiv \frac{\xi - \mathbb{E}[\nu]}{\mathbb{V}[\nu]^{\frac{1}{2}}}$$

[91] Greene, W. (2002). *Econometric Analysis* (5. ed.). New Jersey: Prentice Hall International.

$f(\cdot)$ is the density and $F(\cdot)$ the distribution function of the standard normal. The mean of actual self-management productivities of entrepreneurs is

$$\mathbb{E}[v_E] = \mathbb{E}[v|\nu \geqslant \xi] = \mathbb{E}[v] + \gamma\lambda(\kappa)$$

The difference in means is

$$\mathbb{E}[\nu_E] - \mathbb{E}[v_E] = (\mathbb{V}[\nu]^{\frac{1}{2}} - \gamma)\lambda(\kappa)$$

Hence, entrants overestimate their entrepreneurial abilities, such that $\mathbb{E}[\nu_E] - \mathbb{E}[v_E] > 0$, as long as $\mathbb{V}[\nu]^{\frac{1}{2}} - \gamma > 0$. Given $\mathbb{V}[\nu]^{\frac{1}{2}} - \gamma > 0$, overestimation becomes more prominent if γ decreases. Note that $\lambda'(\kappa) > 0$ and $\partial\kappa/\partial\xi > 0$. For the cut-off, we have

$$\begin{aligned}
\xi &= \frac{1}{1-c}v_{\{2\}1} + \frac{c}{1-c}\sum_{j\in\mathcal{U}_1^{(0)}}[v_{\{1\}j} - v_{\{2\}j}] \\
\xi &= \frac{1}{1-c}v_{\{2\}2} + \frac{c}{1-c}\sum_{j\in\mathcal{U}_2^{(0)}}[v_{\{1\}j} - v_{\{2\}j}] \\
&\vdots \\
\xi &= \frac{1}{1-c}v_{\{2\}(I-1)} + \frac{c}{1-c}\sum_{j\in\mathcal{U}_{I-1}^{(0)}}[v_{\{1\}j} - v_{\{2\}j}] \\
\xi &= \frac{1}{1-c}v_{\{2\}I} + \frac{c}{1-c}\sum_{j\in\mathcal{U}_I^{(0)}}[v_{\{1\}j} - v_{\{2\}j}]
\end{aligned}$$

where

$$\frac{\partial\xi}{\partial v_{\{2\}i}} = \frac{1}{1-c} > 0, \qquad \frac{\partial\xi}{\partial\sum_{j\in\mathcal{U}_i^{(0)}}[v_{\{1\}j} - v_{\{2\}j}]} = \frac{c}{1-c} > 0$$

Thus, the cut-off increases in costs, c. Furthermore, the cut-off increases if the average productivity under the second-best manager $I^{-1}\sum_{i=1}^{I} v_{\{2\}i}$ increases given constant managerial advantages for all i. Last, but not least, the cut-off increases if the average managerial advantage $I^{-1}\sum_{i=1}^{I}\sum_{j\in\mathcal{U}_i^{(0)}}[v_{\{1\}j} - v_{\{2\}j}]$ increases given a constant productivity under the second-best manager for all i. ∎

2.C.2 Survival

Proof of Proposition 2.3. Using Assumption 2.1.1, (2.B.1), and (2.B.3), the relative attraction of wage work is given by

$$\mathscr{W}_{i,t} - \mathscr{E}_{i,t} = \mathscr{Y}_{i,t} - \frac{1-c}{t+1}U_{i,t}$$

$$\mathscr{Y}_{i,t} = \frac{\alpha_i^t}{t+1}\bar{\delta}[w_{i,0} - (1-c)\hat{\pi}_{i,0}] + \frac{1-\alpha_i^t}{(t+1)(1-\alpha_i)}[w_{i,0} - (1-c)\pi_{i,0}]$$

where $w_{i,0} - (1-c)\hat{\pi}_{i,0} < 0$ as individual i strictly preferred to select entrepreneurship in period 1. For the probability to remain entrepreneur, we have

$$\mathbb{P}(\mathbb{d}_{i,t+1} = 1|\mathbb{d}_{i,1} = 1) = \mathbb{P}(\mathscr{X}_{i,t} \leqslant 0) = \mathbb{P}\left(U_{i,t} \geqslant \frac{t+1}{1-c}\mathscr{Y}_{i,t}\right)$$

Thus, we have to compute

$$\mathbb{P}(\mathbb{d}_{i,t+1} = 1|\mathbb{d}_{i,1} = 1) = 1 - \mathbb{P}\left(U_{i,t} < \mathscr{Q}_{i,t}\right), \quad \mathscr{Q}_{i,t} = \frac{t+1}{1-c}\mathscr{Y}_{i,t}$$

Given Assumption 2.2.1, $U_{i,t}$ follows a uniform distribution with support in $[U_{i,t}^-, U_{i,t}^+]$. Hence, we get

$$1 - \mathbb{P}\left(U_{i,t} < \mathscr{Q}_{i,t}\right) = 1 - \frac{\mathscr{Q}_{i,t} - U_{i,t}^-}{U_{i,t}^+ - U_{i,t}^-} = \frac{3}{2} - \frac{1}{\frac{2(1-\alpha_i^t)}{1-\alpha_i}u_i^+}\mathscr{Q}_{i,t} \tag{2.C.1}$$

Taking the first derivative with respect to time yields

$$\dot{\mathbb{P}}(\mathbb{d}_{i,t+1} = 1|\mathbb{d}_{i,1} = 1) = \frac{\bar{\delta}(\alpha_i - 1)\alpha_i^t \log(\alpha_i)}{2(1-c)(\alpha_i - 1)^2 u_i^+}[w_{i,0} - (1-c)\hat{\pi}_{i,0}]$$

where $\log(\alpha_i) < 0$ such that $\bar{\delta}(\alpha_i - 1)\alpha_i^t \log(\alpha_i)\left\{(1-c)(\alpha_i - 1)^2 2u_i^+\right\}^{-1} > 0$. Consequently, we get $\dot{\mathbb{P}}(\mathbb{d}_{i,t+1} = 1|\mathbb{d}_{i,1} = 1) < 0$, as $w_{i,0} < (1-c)\hat{\pi}_{i,0}$. Taking the second derivative with respect to time yields

$$\ddot{\mathbb{P}}(\mathbb{d}_{i,t+1} = 1|\mathbb{d}_{i,1} = 1) = -\frac{\bar{\delta}(\alpha_i - 1)\alpha_i^t \log(\alpha_i)^2(\alpha_i^t + 1)}{2(1-c)(\alpha_i^t - 1)u_i^+}[w_{i,0} - (1-c)\hat{\pi}_{i,0}]$$

where $2(1-c)(\alpha_i^t - 1)u_i^+ < 0$ and $-\bar{\delta}(\alpha_i - 1)\alpha_i^t \log(\alpha_i)^2(\alpha_i^t + 1) > 0$ such that $\ddot{\mathbb{P}}(\mathbb{d}_{i,t+1} = 1|\mathbb{d}_{i,1} = 1) > 0$. Finally, taking the limit, with respect to time, yields $\lim_{t\to\infty} \mathbb{P}(\mathbb{d}_{i,t+1} =$

$1|\mathbb{d}_{i,1} = 1) = \psi_i$ where

$$\psi_i = \frac{w_{i,0} - (1-c)u_i^+ - (1-c)\pi_{i,0}}{2(c-1)u_i^+}$$

■

2.C.3 Non-linear impact of traits

Proof of Proposition 2.4. The proposition can be verified by taking the first derivative of (2.C.1) with respect to the relevant trait-related variable. The derivative with respect to the Stability trait parameter is

$$\eta_{\alpha_i} = \frac{\bar{\delta}\alpha_i^{t-1}[(\alpha_i - 1)t - \alpha_i^{t+1} + \alpha_i]}{2(1-c)(\alpha_i^t - 1)^2 u_i^+}[w_{i,0} - (1-c)\hat{\pi}_{i,0}]$$

We have $2(1-c)(\alpha_i^t - 1)^2 u_i^+, \bar{\delta}\alpha_i^{t-1} > 0$. $\omega_t \equiv (\alpha_i - 1)t - \alpha_i^{t+1} + \alpha_i < 0$ can be demonstrated by induction. It is obvious that $\omega_1 = (\alpha_i - 1) + (\alpha_i - \alpha_i^2) < 0$. Assume that $\omega_t < 0$ holds for some t (induction hypothesis). Consider $\omega_{t+1} = (\alpha_i - 1)(t+1) - \alpha_i^{t+2} + \alpha_i = (\alpha_i - 1)t - \alpha_i^{t+1}\alpha_i + \alpha_i^{t+1} + \alpha_i - 1 - \alpha_i\alpha_i^{t+1} = \omega_t + (1-\alpha_i)(\alpha_i^{t+1} - 1)$. Since $(1-\alpha_i)(\alpha_i^{t+1} - 1) < 0$ and, by induction hypothesis, $\omega_t < 0$, $\omega_{t+1} < 0$. Thus, if $\omega_t < 0$, than $\omega_{t+1} < 0$. Consequently, we have $\omega_t < 0$ for all $t > 0$. Combining all arguments yields $\eta_{\alpha_i} > 0$. As $\hat{\pi}_{i,0} = \varepsilon_i + \pi_{i,0}$, we also get

$$\frac{\partial \eta_{\alpha_i}}{\partial \varepsilon_i} = -(1-c)\frac{\bar{\delta}\alpha_i^{t-1}[(\alpha_i - 1)t - \alpha_i^{t+1} + \alpha_i]}{2(1-c)(\alpha_i^t - 1)^2 u_i^+} > 0$$

The derivative with respect to estimated entrepreneurial income is

$$\eta_{\hat{\pi}_{i,0}} = -\frac{\alpha_i^t \bar{\delta}(\alpha_i - 1)}{2(1-\alpha_i^t)u_i^+} > 0$$

and does not depend on error ε_i. ■

2.C.4 Earning differentials

Proof of Proposition 2.5. Using Assumption 2.1.1 and assumptions on model parameters, we get

$$\mathscr{W}_{k,t} - \mathscr{E}_{k,t} = \mathscr{Y}_{k,t} - \frac{1}{t+1}U_{k,t}, \quad \mathscr{Y}_{k,t} = \frac{1}{t+1}[w_{k,0} - \pi_{k,0}]$$

$$\mathscr{W}_{l,t} - \mathscr{E}_{l,t} = \mathscr{Y}_{l,t} - \frac{1}{t+1}U_{l,t}, \quad \mathscr{Y}_{l,t} = \frac{\alpha_l^t}{t+1}\bar{\delta}[w_{l,0} - \hat{\pi}_{l,0}] + \frac{1-\alpha_l^t}{(t+1)(1-\alpha_l)}[w_{l,0} - \pi_{l,0}]$$

As noise is independent, the probability of a positive earning differential is

$$\mathbb{P}(\Delta_{t+1} > 0) = \mathbb{P}\left(\mathscr{Y}_{k,t} - \frac{1}{t+1}U_{k,t} \leqslant 0\right)\mathbb{P}\left(\mathscr{Y}_{l,t} - \frac{1}{t+1}U_{l,t} > 0\right)$$

Hence:

$$\mathbb{P}(\Delta_{t+1} > 0) = \left[1 - \mathbb{P}\left(U_{k,t} < (t+1)\mathscr{Y}_{k,t}\right)\right]\mathbb{P}\left(U_{l,t} < (t+1)\mathscr{Y}_{l,t}\right) \tag{2.C.2}$$

The probability that entrepreneur k remains is

$$1 - \mathbb{P}\left(U_{k,t} < (t+1)\mathscr{Y}_{k,t}\right) = 1 - \mathbb{P}\left(U_k < w_{k,0} - \pi_{k,0}\right) = q_k$$

where

$$q_k = \frac{w_{k,0} - \pi_{k,0} - u_k^-}{2u_k^-} > 0$$

For the probability that entrepreneur l exits, we have

$$\dot{\mathbb{P}}\left(U_{l,t} < (t+1)\mathscr{Y}_{l,t}\right) = \frac{\bar{\delta}(\alpha_l - 1)\alpha_l^t \log(\alpha_l)}{2(\alpha_l^t - 1)^2 u_l^-}[w_{l,0} - \hat{\pi}_{l,0}] > 0$$

Thus, if $q_k \neq 0$, we get $\dot{\mathbb{P}}(\Delta_{t+1} > 0) > 0$. Taking the limit of the probability of entrepreneurial exit of l, we obtain

$$\overline{q} = \lim_{t\to\infty} \mathbb{P}\left(U_{l,t} < (t+1)\mathscr{Y}_{l,t}\right) = \frac{w_{l,0} - \pi_{l,0} + u_l^-}{2u_l^+}$$

where

$$\frac{\partial \overline{q}}{\partial \pi_{l,0}} = -\frac{1}{2u_l^+} < 0$$

Consequently, $\mathbb{P}\left(U_{l,t} < (t+1)\mathscr{Y}_{l,t}\right)$ is bounded between $\underline{q} = \mathbb{P}\left(U_{l,1} < 2\mathscr{Y}_{l,1}\right)$ and $\overline{q}$, where the latter is never reached. Multiplication by $q_k > 0$ yields

$$q_k\underline{q} \leqslant q_k\mathbb{P}\left(U_{l,t} < (t+1)\mathscr{Y}_{l,t}\right) < q_k\overline{q}$$

Since $q_k = 1 - \mathbb{P}\left(U_{k,t} < (t+1)\mathscr{Y}_{k,t}\right)$ and (2.C.2), we obtain

$$q_k\underline{q} \leqslant \mathbb{P}(\Delta_{t+1} > 0) < q_k\overline{q} \quad \text{for } t > 0$$

where the upper boundary decreases in $\pi_{l,0}$. ∎

Appendix 2.D

This appendix describes how we simulate the full version of our model. Simulations proceed in three steps.

STEP 1. Given numerical and distributional assumptions, generate one draw of model variables.

STEP 2. Test whether $\mathfrak{A}$ has a tolerable level of circularity. A tolerable level of circularity is given if management is transitive in general and up to 5 (chain-link) restructurings affecting initial entrepreneurs are not problematic.

STEP 3. Simulate the behavior of individuals, while testing for additional, relevant, non-circularity issues. If a relevant issue appears, return to the first step and re-simulate.

We use the concept of chain restructuring, which can be described as follows. Assume that an entrepreneur i abandons her firm and goes to her best alternative manager j. This is one chain link. Entrepreneur i is said to survive the restructuring if she has a real alternative if she decides to abandon entrepreneurship, i.e., the second-best manager j is not part of i's firm. The matrix $\mathfrak{A}$ is said to survive a restructuring with one chain link if the above holds for every entrepreneur and her best alternative manager in $\mathfrak{A}$. A n chain-link restructuring is defined in a similar way. Individual i abandons her firm and goes to entrepreneur h, entrepreneur h abandons her firm and goes to entrepreneur k, and so on. We test whether matrix $\mathfrak{A}$ survives a 5 chain-link restructuring.

Note that the simulation does not require general non-circularity. The real-world counterpart of $\mathfrak{A}$, respectively $\mathfrak{P}$, will also not be perfectly non-circular. But, being not perfectly non-circular is usually not an obstacle if the number of individuals involved is sufficiently large. Two individuals without well-defined interactions do not pose an issue if they never interact. In our simulation exercise, we let the number of individuals be large enough and impose circularity only where interactions are likely, but not as a general condition.

TABLE 2.C.1. Numerical assumptions

Parameter	Value
I	300
T	60
c	0.01
m	0.10
δ	0.96
$\tilde{u}$	0.80
Average Stability parameter α	0.10
Standard deviation of Stability parameter	0.10
Average non-self-management productivities	1.40
Standard deviation of non-self-management productivities	0.50
Correlation between estimated and true self-management productivities	{0, 0.9}
Average of estimated and true self-management productivities	1.30
Standard deviation of estimated self-management productivities	1.70
Standard deviation of true self-management productivities	1.00

Appendix 3.A

This appendix demonstrates that ranking distributions according to the highest expected values is equivalent to ranking distributions according to success probabilities given a certain success criterion, and shows the properties of the mean-variance rule.

3.A.1 Relation between success probabilities and average incomes

Using a large cross-country income distribution data set with approximately 800 country-year observations, Lopez & Servén (2006) demonstrate that incomes are well approximated by the log-normal distribution. Thus, a reasonable assumption is that Ψ is log-normal with unknown parameters.

Lemma 3.A1. *Let f_1 and f_2 denote two log-normal income distributions with corresponding payoffs Ψ_1 and Ψ_2. Let $\mathbb{E}[\log\Psi] = \zeta$ and $\mathbb{V}[\log\Psi] = \nu$ such that $\mathbb{E}[\Psi] = \exp(\zeta + \nu/2)$. Furthermore, let $\succ_{\mathbb{E}}$ and $\succ_{SP}$ denote two rankings such that*

$$f_1 \succ_{\mathbb{E}} f_2 \textit{ if } \mathbb{E}[\Psi_1] > \mathbb{E}[\Psi_2]$$

and

$$f_1 \succ_{SP} f_2 \textit{ if } \mathbb{P}(\Psi_1 > \bar{\Pi}) > \mathbb{P}(\Psi_2 > \bar{\Pi})$$

If $\bar{\Pi} = \exp([\zeta_2 + \nu_2 + \sqrt{\nu_1}\sqrt{\nu_2}]/2)$, then $f_1 \succ_{\mathbb{E}} f_2$ and $f_1 \succ_{SP} f_2$ are equivalent, i.e., if $f_1 \succ_{\mathbb{E}} f_2$, we must also have $f_1 \succ_{SP} f_2$ and vice versa.

Proof. The ranking $f_1 \succ_{\mathbb{E}} f_2$ requires that $\exp(\zeta_1 + \nu_1/2) > \exp(\zeta_2 + \nu_2/2)$. Using the properties of $\exp(\cdot)$, this condition is equivalent to

$$\zeta_1 > \zeta_1^* = \zeta_2 + \frac{1}{2}(\nu_2 - \nu_1)$$

The ranking $f_1 \succ_{\mathrm{SP}} f_2$ requires that

$$F_{\mathrm{Norm}}\left(\frac{\log\bar{\Pi} - \zeta_2}{\sqrt{\nu_2}}\right) > F_{\mathrm{Norm}}\left(\frac{\log\bar{\Pi} - \zeta_1}{\sqrt{\nu_1}}\right)$$

where F_{Norm} is the distribution function of the standard normal. By the properties of the distribution function, we must have

$$\sqrt{\nu_1}(\log\bar{\Pi} - \zeta_2) > \sqrt{\nu_2}(\log\bar{\Pi} - \zeta_1)$$

for $f_1 >_{\mathrm{SP}} f_2$ to hold. This condition is equivalent to $\zeta_1 > \zeta_1^{**}$ where

$$\zeta_1^{**} = \frac{\sqrt{v_1}}{\sqrt{v_2}}\zeta_2 + \frac{\sqrt{v_2} - \sqrt{v_1}}{\sqrt{v_2}} \log \bar{\Pi}$$

It is easy to see that $\zeta_1^{*} = \zeta_1^{**}$ if $v_1 = v_2$. For $v_1 \neq v_2$, we must have $\bar{\Pi} = \bar{\Pi}_{\mathbb{E}}$ where

$$\bar{\Pi}_{\mathbb{E}} = \exp\left(\frac{1}{2}[\zeta_2 + v_2 + \sqrt{v_1}\sqrt{v_2}]\right)$$

■

It is also possible to show consistency of rankings using less assumptions (but slightly different conditions):

Lemma 3.A2. *Let 1 and 2 denote two options with existing expected values. Let* $\phi_1 = \mathbb{P}(\Psi_1 > \bar{\Pi})$, $\phi_2 = \mathbb{P}(\Psi_2 > \bar{\Pi})$, $m_1 = \mathbb{E}[\Psi_1]$, *and* $m_2 = \mathbb{E}[\Psi_2]$. *Furthermore, assume that* $\bar{\Pi} \gg m_2$. *If the expected reward of option 1 is sufficiently greater than of option 2 such that*

$$m_1 > \frac{L - \bar{\Pi}}{\bar{\Pi}} m_2 + \bar{\Pi}$$

we have $\phi_1 > \phi_2$.

Proof. If option 1 has a higher success probability than option 2, we have

$$\mathbb{P}(\Psi_1 > \bar{\Pi}) > \mathbb{P}(\Psi_2 > \bar{\Pi}) \tag{3.A.1}$$

Using Markov's inequality, the right hand side of (3.A.1) is bounded as follows:

$$\mathbb{P}(\Psi_2 > \bar{\Pi}) \leqslant \frac{m_2}{\bar{\Pi}}$$

Note that $\mathbb{P}(\Psi_1 > \bar{\Pi}) = 1 - \mathbb{P}(\Psi_1 \leqslant \bar{\Pi})$. Using the reverse Markov's inequality, we obtain

$$\mathbb{P}(\Psi_1 \leqslant \bar{\Pi}) \leqslant \frac{L - m_1}{L - \bar{\Pi}}$$

Hence, we have

$$\mathbb{P}(\Psi_1 > \bar{\Pi}) \geqslant \frac{m_1 - \bar{\Pi}}{L - \bar{\Pi}}$$

As $L - \bar{\Pi} > 0$, (3.A.1) holds if

$$m_1 > \frac{L - \bar{\Pi}}{\bar{\Pi}} m_2 + \bar{\Pi}$$

Since $\bar{\Pi} > m_2$, we have $m_1 > m_2$. ■

3.A.2 Properties of mean-variance rule

The proofs of properties S1, S2, and S3 in Proposition 3.1 are as follows.

Proof of S1. Let $\boldsymbol{\sigma}_{i,t} = \boldsymbol{\sigma}(\mathbb{x}_{i,t}; \lambda)$. Furthermore, let $\boldsymbol{\sigma}_{j,t} = \max\{\boldsymbol{\sigma}_{\mathfrak{w},t}, \boldsymbol{\sigma}_{\mathfrak{e},t}\}$. Assume that in period t we are required to select a safe option $\mathfrak{u}$ over a non-safe option $j \neq \mathfrak{u}$ because $\boldsymbol{\sigma}_{j,t-1} \leqslant 0$. Then, $\boldsymbol{\sigma}_{i,t} = \boldsymbol{\sigma}_{i,t-1}$ for all $i \in \mathbb{O}$ because we do not obtain any new information by selecting unemployment. Consequently, $\boldsymbol{\sigma}_{j,t} \leqslant 0$. Hence: If $\boldsymbol{\sigma}_{j,t} \leqslant 0$, then $\boldsymbol{\sigma}_{j,h} \leqslant 0$ for all $h \in \{t+1, t+2, \ldots\}$ and $j \neq \mathfrak{u}$ by induction. There are two interesting starting points. First, we will start with $\mathfrak{u}$ in period 1 if $\boldsymbol{\sigma}_{j,0} \leqslant 0$ and there will be no transitions at all. Second, we will start with option $j \neq \mathfrak{u}$ if $\boldsymbol{\sigma}_{j,0} > 0$ and there will be no changes of occupational status after t^* if for some $t^* > 1$ we have $\boldsymbol{\sigma}_{j,t^*-1} \leqslant 0$. ■

Proof of S2. The mean-variance rule does not always stay on a winner. A success can make an option more ambiguous if not expected and an ambiguity-averse gambler might switch from a winner.

Let i be an option producing a success and j a second option. Assume that

$$\boldsymbol{\sigma}(\alpha_i, \beta_i; \lambda) > \boldsymbol{\sigma}(\alpha_j, \beta_j; \lambda)$$

such that option i is selected. After option i is selected, it is assumed to generate a success. Consequently, if $\boldsymbol{\sigma}(\alpha_i + 1, \beta_i; \lambda) \geqslant \boldsymbol{\sigma}(\alpha_i, \beta_i; \lambda)$, the gambler would always stay on a winner. If $\boldsymbol{\sigma}(\alpha_i + 1, \beta_i; \lambda) < \boldsymbol{\sigma}(\alpha_i, \beta_i; \lambda)$, the gambler does not necessarily stay on a winner since we might have $\boldsymbol{\sigma}(\alpha_i + 1, \beta_i; \lambda) < \boldsymbol{\sigma}(\alpha_j, \beta_j; \lambda)$. Put differently, if we have

$$\frac{\lambda}{2} u(\alpha, \beta) > \mu(\alpha + 1, \beta) - \mu(\alpha, \beta), \qquad u(\alpha, \beta) = \mathsf{v}(\alpha + 1, \beta) - \mathsf{v}(\alpha, \beta)$$

the gambler might not stay on a winner. Note that $\mu(\alpha+1, \beta) - \mu(\alpha, \beta) > 0$. In most cases, we have $u(\alpha, \beta) < 0$ and not staying on a winner requires

$$\lambda < 2\frac{\mu(\alpha + 1, \beta) - \mu(\alpha, \beta)}{u(\alpha, \beta)}$$

where

$$\frac{\mu(\alpha + 1, \beta) - \mu(\alpha, \beta)}{u(\alpha, \beta)} < 0$$

such that ambiguity-averse individuals, with $\lambda > 0$, will always stay on a winner. However, if a success was not expected, we could have $u(\alpha, \beta) > 0$. For instance, if $\alpha = 1$ and $\beta = 10$, which is an event with a positive probability if $\phi \in (0, 1)$, a success is not expected such that $u(1, 10) > 0$. In such a case, the gambler will not stay on a winner and switch from option i to j if ambiguity aversion is sufficiently high or, more precisely, if

$$\lambda > 2\frac{\mu(\alpha_j, \beta_j) - \mu(\alpha_i + 1, \beta_i)}{\mathsf{v}(\alpha_j, \beta_j) - \mathsf{v}(\alpha_i + 1, \beta_i)}$$

■

Proof of S3. The property is easy to verify if we consider a choice between unemployment and a non-unemployment option. We can always find a $\lambda > 0$ such that $\boldsymbol{\sigma}_{i,h} \leqslant 0$ for all $i \in \mathbb{O}_{-\mathfrak{u}}$ and some period $h > 0$. Given that unemployment is selected in one period, unemployment is selected in all consecutive periods (see property S1). This is independent from true success probabilities $\phi_{\mathfrak{e}}$ and $\phi_{\mathfrak{w}}$ such that the best option might not be identified. However, the property also holds if we assume that unemployment is never weakly better than a non-unemployment option:

$$q_{\mathfrak{u}} = \mathbb{P}(\boldsymbol{\sigma}_{i,t} > 0 \text{ for all } i \in \mathbb{O}_{-\mathfrak{u}} \text{ and all } t) = 1$$

Note that $q_{\mathfrak{u}} = 1$ requires

$$\lambda < 2\frac{\mu_{i,t}}{\mathsf{v}_{i,t}} \qquad \text{for all } i \in \mathbb{O}_{-\mathfrak{u}} \text{ and all } t$$

The following argument is a modified version of Theorem 2 in Brezzi & Lai (2000). To see that the property holds in a situation where $q_{\mathfrak{u}} = 1$, first, consider the following two conditions:

$$\mathbb{P}\left(\lim_{t\to\infty} \mu_{i,t} = \phi_i\right) = 1 \tag{3.A.2}$$

$$\mathbb{P}\left(\lim_{t\to\infty} \mathsf{v}_{i,t} = 0\right) = 1 \tag{3.A.3}$$

Note that t is, here, the number of times option i is sampled from. To get (3.A.2), use the assumption that the initial prior $\mathbb{x}_0 = (\alpha_0, \beta_0)$ is sampled from a distribution with the right success probability ϕ, and note that

$$\mu_t = \frac{\alpha_t}{\gamma_t} = \frac{1}{t}B_t, \qquad B_t = \sum_{h=1}^{t} b_h, \qquad b_h \sim \text{Bernoulli}(\phi), \qquad \mathbb{E}[b_h] = \phi$$

where notation is slightly abused to extend the number of observed periods to pseudo observations, $\mathbb{x}_0$. By applying the Strong Law of Large Numbers, we obtain (3.A.2). To show (3.A.3), note that if pseudo observations are sampled from the correct distribution, we have $\alpha_t = B_t$ and $\beta_t = t - B_t$, such that

$$\mathsf{v}_t = \frac{B_t(t - B_t)}{t^2(t+1)}$$

where $0 < B_t \leqslant t$. Now, notice that for a fixed t the expression $B_t(t - B_t)$ takes a maximum at $B_t = t/2$ such that we get

$$\mathsf{v}_t = \frac{B_t(t - B_t)}{t^2(t+1)} \leqslant \frac{1}{4(t+1)} \qquad \text{for all } t$$

By letting $t \to \infty$, we obtain (3.A.3). If (3.A.2) and (3.A.3) are combined, we get

$$\mathbb{P}\left(\lim_{t\to\infty} \boldsymbol{\sigma}_{i,t} = \lim_{t\to\infty} \mu_{i,t} - \frac{\lambda}{2}\lim_{t\to\infty} \mathsf{v}_{i,t} = \phi_i\right) = 1 \tag{3.A.4}$$

According to (3.A.4), an infinite amount of information is equivalent to perfect information. Hence, if the gambler selects an occupational option for an infinite amount of periods, she will almost surely learn the true success probability of this occupational option.

Yet, only one occupation will be selected for an infinite number of periods. To see this, define the number of periods option $i \in \mathbb{O}_{-\mathfrak{u}}$ is selected by $T_{i,t}$ and let $T_i^\infty \equiv \lim_{t\to\infty} T_{i,t}$. On the event $T_\mathfrak{e}^\infty = \infty$, $\boldsymbol{\sigma}_{\mathfrak{e},t} \to \phi_\mathfrak{e}$ almost surely. On the event $T_\mathfrak{w}^\infty = \infty$, $\boldsymbol{\sigma}_{\mathfrak{w},t} \to \phi_\mathfrak{w}$ almost surely. But, since only the occupation with the highest index is selected, we cannot simultaneously have $T_\mathfrak{e}^\infty = \infty$ and $T_\mathfrak{w}^\infty = \infty$ if $\phi_\mathfrak{e} \neq \phi_\mathfrak{w}$ or

$$\mathbb{P}\left(T_i^\infty < \infty \text{ for all } i \in \mathbb{O}_{-\mathfrak{u}} \text{ except one } i\right) = 1$$

It can be demonstrated that the best option i^* is not necessarily the option selected for an infinite number of periods or

$$\mathbb{P}\left(T_{i^*}^\infty < \infty\right) > 0 \tag{3.A.5}$$

To obtain (3.A.5), assume, without loss of generality, $i^* = \mathfrak{w}$ such that $\phi_\mathfrak{w} > \phi_\mathfrak{e}$. Let

$$q_\mathfrak{e} \equiv \mathbb{P}\left(\boldsymbol{\sigma}_{\mathfrak{e},t} \geqslant \varepsilon \text{ for all } t > 0\right) \tag{3.A.6}$$

for some ε and let

$$q_\mathfrak{w} \equiv \mathbb{P}\left(\boldsymbol{\sigma}_{\mathfrak{w},k} < \varepsilon\right) \tag{3.A.7}$$

for some k. Consider the joint probability of the event in (3.A.6) and (3.A.7) given by

$$q^* \equiv \mathbb{P}\left(\boldsymbol{\sigma}_{\mathfrak{w},k} < \varepsilon \leqslant \boldsymbol{\sigma}_{\mathfrak{e},t} \text{ for all } t > 0\right)$$

If $q^* > 0$, (3.A.5) holds.

Note that $\boldsymbol{\sigma}_{\mathfrak{w}}$ and $\boldsymbol{\sigma}_{\mathfrak{e}}$ are independent as the underlying options we sample from are independent. By independence, we get $q^* = q_{\mathfrak{e}} q_{\mathfrak{w}}$. Let $\varepsilon \in (0, \phi_{\mathfrak{e}})$. Given $q_{\mathfrak{u}} = 1$ and (3.A.4), we can always find a ε such that $q_{\mathfrak{e}} > 0$. It is also obvious that $q_{\mathfrak{w}} > 0$. For instance: Let $\mathbb{x}_{\mathfrak{w},0} = (2, 2)$, which occurs with a positive probability. Assume that we obtain k nonsuccesses. The probability to obtain k nonsuccesses is given by $[1 - \phi_{\mathfrak{w}}]^k > 0$. In such a setting, we get

$$\boldsymbol{\sigma}_{\mathfrak{w},k} = \frac{2k^2 + 18k + 40 - (k+2)\lambda}{(k+4)^2(k+5)}$$

such that

$$\frac{\mathrm{d}\boldsymbol{\sigma}_{\mathfrak{w},k}}{\mathrm{d}k} = \frac{(2k^2 + 11k + 8)\lambda - 2k^3 - 28k^2 - 130k - 200}{(k+4)^3(k+5)^2} < 0$$

if

$$\lambda < \lambda_q \equiv \frac{2k^3 + 28k^2 + 130k + 200}{2k^2 + 11k + 8}, \qquad \lambda_q \in (15, \infty)$$

As $\boldsymbol{\sigma}_{\mathfrak{w},k}$ monotonically approaches zero as k increases, it is always possible to generate a sequence of events with a positive probability such that $\boldsymbol{\sigma}_{\mathfrak{w},k} < \varepsilon$. Hence: $q^* > 0$. ■

Appendix 3.B

This appendix compares choice patterns produced by the mean-variance rule to patterns obtained in Bernoulli bandit experiments. Most experiments analyze the explanatory performance of the optimal decision rule and try to classify deviations from optimal behavior. Deviations are quite common (e.g., Meyer & Shi 1995; Anderson 2001; Gans et al. 2007).

Some results are strongly consistent with the mean-variance rule. For instance, Anderson (2001) finds that individuals experiment less than optimal and that this phenomenon is best explained by ambiguity aversion. According to Anderson (2001), individuals follow a rule with a lower than the optimal index. Brezzi & Lai (2000) demonstrated that the optimal index has the lower bound $\mathbf{g} \geqslant \mu$. An ambiguity-averse gambler deciding according to the mean-variance rule has an index $\boldsymbol{\sigma} < \mu$ implying $\boldsymbol{\sigma} < \mathbf{g}$. Put differently, an ambiguity-averse gambler following the mean-variance rule operates with an index that is always lower than the optimal Gittins index.

Two empirical findings that are not obviously consistent with the mean-variance rule are as follows.

1. Steyvers et al. (2009) calculate behavioral characteristics (denoted henceforth by BCHAR) for individuals experimenting with Bernoulli bandits with varying success distributions. A BCHAR is, for instance, the number of times individuals selected

an option with fewer successes and nonsuccesses than an alternative option, which is interpreted as a measure of exploratory behavior. Steyvers et al. (2009) find no significant correlation between any of the BCHARS and personality traits.

2. Gans et al. (2007) apply different rules to data generated by individuals who played with Bernoulli bandits. Gans et al. (2007) establish a ranking of the explanatory performance of different rules.

The first result appears to contradict my assumption that decisions are affected by personality through the parameter λ. The ranking of Gans et al. (2007) should be reproducible with data generated by the mean-variance rule. I show that both empirical findings are consistent with the mean-variance rule.

3.B.1 *Personality traits and characteristics of behavior*

The rationale behind the first test is based on the idea that personality can determine individual levels of ambiguity aversion but aspects of ambiguity-averse behavior can look exactly the same across different levels of aversion. In other words, ambiguity aversion does not have a linear effect on behavior. A BCHAR might capture the difference between an ambiguity-averse and an ambiguity-neutral individual but BCHARS may not capture the difference between two ambiguity-averse individuals.

Let $\upsilon \in \mathbb{R}$ denote an observable personality trait. Let $\mathbf{w}_{\upsilon,\lambda}$ denote a draw from the joint distribution of the observable personality trait and $\lambda \in \mathbb{R}$ representing ambiguity preferences, where ambiguity affinity is possible. The joint distribution is bivariate normal. I assume that personality almost completely determines the parameter λ such that the correlation between υ and λ is almost perfect (0.99). The personality trait has a variance of 1; the ambiguity-preferences parameter has a variance of 2; and the mean of the personality trait is zero.

There are four groups of individuals. Each group corresponds to 100 draws of $\mathbf{w}_{\upsilon,\lambda}$.

Group 1 — For this group, I assume that the mean of λ is 50. Individuals from group 1 are ambiguity averse with a probability of almost 100%.

Group 2 — For this group, the mean of λ is 10. Individuals from group 2 are ambiguity averse with a very high probability but the probability is slightly smaller than for individuals from group 1.

Group 3 — The mean of λ is zero. A member of group 3 is ambiguity averse with a probability of 50%.

Group 4 — For the last group, I assume that the mean of λ is −10 such that group-4 individuals are ambiguity averse with a probability of around 0%.

The Monte Carlo experiment proceeds as follows.

STEP 1. Groups are drawn.

STEP 2. Individuals in each group use the mean-variance rule to play with a two-armed bandit over 100 periods. The bandit setting corresponds to the setting used by Steyvers et al. (2009).

STEP 3. After each individual finishes the game, five BCHARS, which are described below, are calculated.

STEP 4. The algorithm computes the correlation between BCHARS and the observable personality trait.

This procedure is repeated 10,000 times for each combination of priors and success probabilities. The five BCHARS, which are also calculated by Steyvers et al. (2009), are as follows. EXPLOITATION is the number of times the gambler selected an option with more successes and more nonsuccesses than the alternative option. EXPLORATION is the number of times the gambler selected an option with fewer successes and fewer nonsuccesses than the alternative option. BETTER is the number of times the gambler selected an option with more success and fewer nonsuccesses than the alternative option. WORSE is the number of times the gambler selected an option with fewer successes and more nonsuccesses than the alternative option. UNTRIED is the number of times the gambler selected an option never selected before.

Decisions can be influenced by decision noise. Noise is defined as follows.

Definition 3.5. A rule-based behavior is said to be affected by decision noise $\epsilon \in (0, 0.5]$ if, in some situation, given two different options 1 and 2, a decision rule prescribes to select (a) option 1 with probability 1 and option 2 with probability zero or (b) option 1 with probability zero and option 2 with probability 1 but the gambler selects (a) option 1 with probability $1 - \epsilon$ and option 2 with probability ϵ or (b) option 1 with probability ϵ and option 2 with probability $1 - \epsilon$.

The parameter ϵ determines the probability to make a mistake by deviating from the prescribed choice probability. I set $\epsilon = 0.01$ such that there are only very few mistakes.

Figure 3.B.1 presents median correlations between the personality trait and BCHARS conditional on group affiliation. WORSE does not capture any trait differences inside groups. BETTER, EXPLOITATION, EXPLORATION, and UNTRIED reflect trait differences inside groups if there are some ambiguity-neutral or ambiguity-seeking individuals in the group. The correlations are plausible. For instance, an increase in the personality trait, which goes along with an increase in ambiguity aversion, is associated with an increase in EXPLOITATION behavior and a decrease in EXPLORATION behavior. However, given

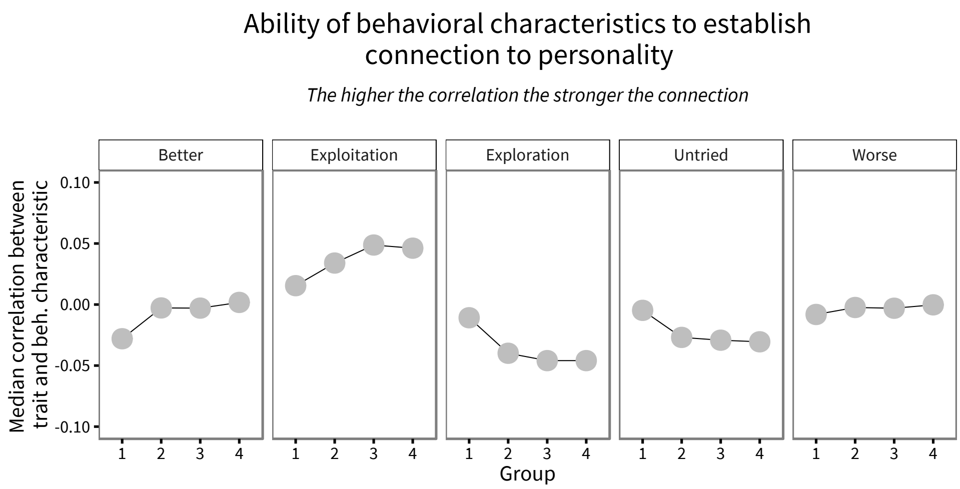


FIGURE 3.B.1. BCHARS and personality

that there is an almost perfect correlation between the personality trait and the ambiguity-aversion parameter λ determining behavior, correlations between the personality trait and BCHARS are rather small (they do not exceed ± 0.05).

The correlations also tend to decrease with the share of ambiguity-averse individuals in groups. For instance, in the first group, which is completely ambiguity-averse, almost no BCHAR can establish a meaningful connection between personality and behavior—every individual in the totally ambiguity-averse group, who does not make a mistake, exploits but does not explore and never tries an untried option (besides the first period). Since a large share of ambiguity-averse individuals (in an arbitrary group) is a realistic scenario, there is a simple explanation for the results of Steyvers et al. (2009).

3.B.2 Explanatory performance

Gans et al. (2007) use a Bernoulli bandit experiment to estimate the explanatory performance of different decision rules. Among others, Gans et al. (2007) consider the following three rules: the myopic rule, the simple rule, and exponential smoothing. Explanatory performance is measured by the Bayesian information criterion (BIC). According to Gans et al. (2007), exponential smoothing performs better than the simple rule, while the simple rule performs better than the myopic rule. In what follows, I will describe the three rules and show that applying the rules to data generated by the mean-variance rule results in the ranking of Gans et al. (2007).

The rules are interesting because they are derived from different assumptions. Alternative rules are as follows.

MYOPIC RULE. The myopic rule is a special case of the Gittins index with a zero probability to continue. In the context of the mean-variance rule, myopia corresponds to the behavior of ambiguity-neutral individuals. The myopic index is

$$\mathbf{m}_{i,t} = \mu(\mathbb{x}_{i,t}) \tag{3.B.1}$$

Simple rule. According to the simple rule, an option can have two qualities: It is either good or bad but never both. The index of the simple rule is (see Gans et al. 2007):

$$\mathbf{c}_{i,t} = \mathbf{c}_{i,t-1} + (a_0 + a)\pi_t - a \tag{3.B.2}$$

which corresponds to a random walk. Gans et al. (2007) set $a_0 = 1$ and $a > 0$ is a free parameter.

Exponential smoothing rule. Exponential smoothing is a purely descriptive rule. The smoothing index is a weighted average of the smoothing index from the previous period and new information:

$$\mathbf{e}_{i,t} = (1 - \zeta)\mathbf{e}_{i,t-1} + \zeta\pi_t \tag{3.B.3}$$

where $\zeta \in [0, 1]$ is a weight. The weight is a free parameter. Starting values, i.e., $\mathbf{e}_{i,0}$, are also free parameters. In line with Gans et al. (2007), I assume that the initial index of one option is 2/3.

The Monte Carlo experiment proceeds as follows.

Step 1. The algorithm generates a decision history based on the mean-variance rule and some combination of ambiguity preferences, probabilities to succeed, and noise levels. Decision histories are generated for 100 periods, while the choice setting is adapted from Gans et al. (2007). There are two noise levels such that $\epsilon \in \{0.03, 0.08\}$. Given both noise levels, decisions are mostly determined by the mean-variance rule. Ambiguity preferences are given by $\lambda \in \{-20, -10, 1, 20, 40, 50\}$ such that gamblers can be ambiguity seekers ($\lambda \in \{-20, -10\}$), can be almost ambiguity neutral ($\lambda = 1$), or can be ambiguity-averse ($\lambda \in \{20, 40, 50\}$), where ambiguity preferences include high levels of aversion.

Step 2. BIC values are estimated by maximum likelihood with a probit model. Each model has a sensitivity parameter. The myopic model has no additional free parameters. The model based on the simple rule has one additional parameter. The exponential smoothing model has two additional parameters.

The two steps are sequentially repeated 10,000 times. Note that, here, a smaller BIC value indicates higher explanatory performance.

Table 3.B.1 presents simulation results by showing the median BIC value conditional on ambiguity preferences, noise level, and decision rule. If we consider a setting with suffi-

TABLE 3.B.1. Simulated explanatory performance of myopic, simple, and exponential-smoothing rule on the basis of data generated by the mean-variance rule; rows that are consistent with the Gans et al. (2007) ranking are in bold

Noise level	Ambiguity preferences	Median of BIC		
		Myopic rule	Simple rule	Smoothing rule
$\epsilon = 0.03$	$\lambda = -20$	63.291	65.884	62.300
	$\lambda = -10$	45.497	54.040	59.040
	$\lambda = 1$	34.529	41.555	60.352
	$\lambda = 20$	143.235	42.652	45.882
	$\lambda = 40$	**143.235**	**44.571**	**40.764**
	$\lambda = 50$	**143.235**	**44.907**	**40.764**
$\epsilon = 0.08$	$\lambda = -20$	100.124	101.808	96.738
	$\lambda = -10$	86.180	93.503	94.888
	$\lambda = 1$	76.992	83.459	95.681
	$\lambda = 20$	143.235	70.316	72.391
	$\lambda = 40$	**143.235**	**69.542**	**64.543**
	$\lambda = 50$	**143.235**	**69.972**	**64.543**

ciently high ambiguity aversion, such that $\lambda \geqslant 40$, we can observe the ranking established by Gans et al. (2007). Exponential smoothing is better than the simple rule, while the simple rule is better than the myopic rule. The ranking cannot be reproduced if ambiguity aversion is not sufficiently high or gamblers are ambiguity seekers. Note, furthermore, that increasing noise reduces the explanatory performance of the rules, i.e., the BIC value increases. To summarize, if most decisions can be derived from the mean-variance rule with sufficiently high ambiguity aversion and are not random, theoretical behavior is similar to the behavior of actual gamblers.

Appendix 3.C

This appendix provides proofs of Proposition 3.1, 3.2, 3.3, and 3.4. Furthermore, I also isolate the effect of myopia (from ambiguity aversion) and show that myopia is not the driving force behind the results.

3.C.1 Propensity to select unemployment

Proof of Proposition 3.1. Let $\mu_{i,t} = \mu(\mathbb{x}_{i,t})$, $\mathsf{v}_{i,t} = \mathsf{v}(\mathbb{x}_{i,t})$, and $\mathbf{r}_{i,t} = \mathbf{r}(\mathbb{x}_{i,t}; \dots)$, where $\mathbf{r}(\mathbb{x}_{i,t}; \dots)$ is either the mean-variance or the optimal index.

Using the boundaries from G2, we obtain $\mathbb{P}(\mathbf{g}_{\mathfrak{e},t} > 0) = \mathbb{P}(\mathbf{g}_{\mathfrak{w},t} > 0) = 1$ for all t. Hence, we have $\mathbb{P}(d_t^{[\mathbf{g}]} = \mathfrak{u}) = 0$ for all t. Consider the first problem, i.e., $p_{\mathfrak{u},1} > 0$. By independence of sampling of pseudo observations, we get

$$p_{\mathfrak{u},1} = \mathbb{P}(\boldsymbol{\sigma}_{\mathfrak{e},0} \leqslant 0)\mathbb{P}(\boldsymbol{\sigma}_{\mathfrak{w},0} \leqslant 0)$$

Let $X_i \in [1, \gamma_{i,0}]$ denote the number of pseudo successes in occupation $i \in \mathbb{O}_{-\mathfrak{u}}$. After straightforward transformations, we get

$$p_{\mathfrak{u},1} = L_{\mathfrak{e}}(\lambda)L_{\mathfrak{w}}(\lambda)$$

where L_i is the distribution function of the random variable

$$Y_i = 2\frac{\gamma_{i,0}(\gamma_{i,0}+1)}{\gamma_{i,0} - X_i}$$

The random variable Y_i has support in $[\underline{Y_i}, \overline{Y_i}]$ where

$$\underline{Y_i} = 2\frac{\gamma_{i,0}(\gamma_{i,0}+1)}{\gamma_{i,0} - 1}, \quad \overline{Y_i} = 2\gamma_{i,0}(\gamma_{i,0}+1)$$

As Y_i is a strictly increasing function of the binomially distributed random variable X_i, we must have

$$L_i(\lambda) = \begin{cases} 0 & \text{if } \lambda < \underline{Y_i} \\ G([\lambda\gamma_{i,0} - 2\gamma_{i,0}(\gamma_{i,0}+1)]\lambda^{-1}; \phi_i, \gamma_{i,0}) & \text{if } \lambda \in [\underline{Y_i}, \overline{Y_i}] \\ 1 & \text{if } \lambda > \overline{Y_i} \end{cases}$$

where G is the distribution function of the binomial distribution. Hence, it is always possible to find a large enough $\lambda \geqslant \underline{Y_i}$ such that $L_{\mathfrak{w}}(\lambda), L_{\mathfrak{e}}(\lambda) > 0$.

To show that $p_{\mathfrak{u},1} = 1$ for some λ, simply note that if

$$\lambda > \max\{\overline{Y_{\mathfrak{e}}}, \overline{Y_{\mathfrak{w}}}\}$$

we must have $L_{\mathfrak{e}}(\lambda) = L_{\mathfrak{w}}(\lambda) = 1$. ■

3.C.2 Overconfidence and ambiguity aversion

Proof of Proposition 3.2. Let $\mu_{i,t} = \mu(\mathbb{x}_{i,t})$, $\mathrm{v}_{i,t} = \mathrm{v}(\mathbb{x}_{i,t})$, and $\mathbf{r}_{i,t} = \mathbf{r}(\mathbb{x}_{i,t}; \ldots)$, where $\mathbf{r}(\mathbb{x}_{i,t}; \ldots)$ is either the mean-variance or the optimal index. Using C1 and the boundaries in G2, we can establish $\mathbf{g}_{\mathfrak{e},0} > \mathbf{g}_{\mathfrak{w},0}$ such that $d_1^{[\mathbf{g}]} = \mathfrak{e}$. By sufficiently high overconfidence,

$\boldsymbol{\sigma}_{\mathfrak{e},0} > \boldsymbol{\sigma}_{\mathfrak{w},0}$ and, thus, $d_1^{[\sigma]} = \mathfrak{e}$. Given C3, $\boldsymbol{\sigma}_{\mathfrak{w},0} > 0$. Note that, given signal $\mathbf{s}_{\mathfrak{e},t}$, $\vartheta_{\mathfrak{e}} = \mathsf{v}(\alpha_{\mathfrak{e},0}, \beta_{\mathfrak{e},0} + t_{\mathsf{v}})$ where

$$t_{\mathsf{v}} = -\frac{4\beta_{\mathfrak{e},0} - \sqrt{9\alpha_{\mathfrak{e},0}^2 + 10\alpha_{\mathfrak{e},0} + 1} + \alpha_{\mathfrak{e},0} + 1}{4}$$

Note that $\mu_{\mathfrak{e},t} + \delta(1-\delta)^{-1}\mathsf{v}_{\mathfrak{e},t} \leqslant \mu_{\mathfrak{e},t} + \delta(1-\delta)^{-1}\vartheta_{\mathfrak{e}}$. Using the boundaries in G2, we can establish that a transition incentive will definitely occur for the first time when $\mu_{\mathfrak{e},t} + \delta(1-\delta)^{-1}\vartheta_{\mathfrak{e}} = \mu_{\mathfrak{w},0}$ but not later. Given signal $\mathbf{s}_{\mathfrak{e},t}$, we obtain $\mu_{\mathfrak{e},t} = \alpha_{\mathfrak{e},0}(\gamma_{\mathfrak{e},0} + t)^{-1}$. Hence, we have $\mathcal{T}_{\mathbf{g}} \leqslant \mathcal{T}_{\mathbf{g}}^{+}$, where

$$\mathcal{T}_{\mathbf{g}}^{+} = \frac{(1-\delta)\alpha_{\mathfrak{e},0} - w_0\gamma_{\mathfrak{e},0}}{w_0}, \qquad w_0 = (1-\delta)\mu_{\mathfrak{w},0} - \delta\vartheta_{\mathfrak{e}} \tag{3.C.1}$$

Using C2, we get $w_0 > 0$. Furthermore, using $\mu = \alpha\gamma^{-1}$ and $\mu_{\mathfrak{e},0} > \mu_{\mathfrak{w},0}$ (by C1), we obtain $(1-\delta)\alpha_{\mathfrak{e},0} - w_0\gamma_{\mathfrak{e},0} > 0$ such that $\mathcal{T}_{\mathbf{g}}^{+} > 0$. Hence, we get

$$0 < \mathcal{T}_{\mathbf{g}} \leqslant \mathcal{T}_{\mathbf{g}}^{+}$$

Now consider the mean-variance rule. The mean-variance rule has an incentive to switch to wage work when $\boldsymbol{\sigma}_{\mathfrak{e},t} \leqslant \boldsymbol{\sigma}_{\mathfrak{w},0}$. As $\mu_{\mathfrak{e},t} - (\lambda/2)\vartheta_{\mathfrak{e}} \leqslant \mu_{\mathfrak{e},t} - (\lambda/2)\mathsf{v}_{\mathfrak{e},t}$, a transition incentive will occur later than period $t = \mathcal{T}_{\boldsymbol{\sigma}}^{-}$, where $\mathcal{T}_{\boldsymbol{\sigma}}^{-}$ solves $\mu_{\mathfrak{e},t} - (\lambda/2)\vartheta_{\mathfrak{e}} = \mu_{\mathfrak{w},0} - (\lambda/2)\mathsf{v}_{\mathfrak{w},0}$. Hence, we have $\mathcal{T}_{\boldsymbol{\sigma}} \geqslant \mathcal{T}_{\boldsymbol{\sigma}}^{-}$, where

$$\mathcal{T}_{\boldsymbol{\sigma}}^{-} = \frac{w_1\lambda - 2\gamma_{\mathfrak{e},0}\mu_{\mathfrak{w},0} + 2\alpha_{\mathfrak{e},0}}{-(\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}})\lambda + 2\mu_{\mathfrak{w},0}}$$

$$w_1 = (\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}})\gamma_{\mathfrak{e},0}$$

It is obvious that $w_1 > 0$, as, according to condition C2, $\vartheta_{\mathfrak{e}} < \mathsf{v}_{\mathfrak{w},0}$. Furthermore, $-(\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}})\lambda + 2\mu_{\mathfrak{w},0} > 0$, as, according to C3, $\lambda < 2\mu_{\mathfrak{w},0}\mathsf{v}_{\mathfrak{w},0}^{-1}$ implying $\lambda < 2\mu_{\mathfrak{w},0}(\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}})^{-1}$. Also, we have

$$w_1\lambda - 2\gamma_{\mathfrak{e},0}\mu_{\mathfrak{w},0} + 2\alpha_{\mathfrak{e},0} = ([\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}}]\lambda + 2[\mu_{\mathfrak{e},0} - \mu_{\mathfrak{w},0}])\gamma_{\mathfrak{e},0} > 0$$

since $\mu_{\mathfrak{e},0} > \mu_{\mathfrak{w},0}$ (C1). Thus, we have $\mathcal{T}_{\boldsymbol{\sigma}}^{-} > 0$ and

$$\mathcal{T}_{\boldsymbol{\sigma}}^{-} \leqslant \mathcal{T}_{\boldsymbol{\sigma}} < \infty$$

Now consider

$$\mathcal{T}_{\boldsymbol{\sigma}}^{-} - \mathcal{T}_{\mathbf{g}}^{+} = \frac{\lambda(1-\delta)(\vartheta_{\mathfrak{e}} - \mathsf{v}_{\mathfrak{w},0}) + 2\delta\sqrt{\vartheta_{\mathfrak{e}}}}{([1-\delta]\mu_{\mathfrak{w},0} - \delta\vartheta_{\mathfrak{e}})([\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}}]\lambda - 2\mu_{\mathfrak{w},0})}$$

It is obvious that $(\mathsf{v}_{\mathfrak{w},0}-\vartheta_{\mathfrak{e}})\lambda-2\mu_{\mathfrak{w},0} < 0$ (C3). Using C2, we must have $(1-\delta)\mu_{\mathfrak{w},0}-\delta\vartheta_{\mathfrak{e}} > 0$ such that

$$([1-\delta]\mu_{\mathfrak{w},0}-\delta\vartheta_{\mathfrak{e}})([\mathsf{v}_{\mathfrak{w},0}-\vartheta_{\mathfrak{e}}]\lambda-2\mu_{\mathfrak{w},0}) < 0$$

Finally, if $\lambda > \lambda_{\boldsymbol{\sigma}} \equiv 2\delta(1-\delta)^{-1}\sqrt{\vartheta_{\mathfrak{e}}}(\mathsf{v}_{\mathfrak{w},0}-\vartheta_{\mathfrak{e}})^{-1}$, we have $\lambda(1-\delta)(\vartheta_{\mathfrak{e}}-\mathsf{v}_{\mathfrak{w},0})+2\delta\sqrt{\vartheta_{\mathfrak{e}}} < 0$ such that

$$\mathcal{T}_{\boldsymbol{\sigma}}^{-} > \mathcal{T}_{\mathbf{g}}^{+}$$

implying

$$\mathcal{T}_{\boldsymbol{\sigma}} > \mathcal{T}_{\mathbf{g}}$$

If there are enough pseudo observations with respect to entrepreneurship, such that C2 holds, we have $\lambda_{\boldsymbol{\sigma}} < 2\mu_{\mathfrak{w},0}\mathsf{v}_{\mathfrak{w},0}^{-1}$, such that $\lambda > \lambda_{\boldsymbol{\sigma}}$ is consistent with C3.

To demonstrate the second part of the proposition, note that either $\mathcal{T}_{\boldsymbol{\sigma}} < t_{\mathsf{v}}$, or $\mathcal{T}_{\boldsymbol{\sigma}} = t_{\mathsf{v}}$, or $\mathcal{T}_{\boldsymbol{\sigma}} > t_{\mathsf{v}}$. $\mu_{\mathfrak{e},t}$ always decreases in t (given signal $\mathbf{s}_{\mathfrak{e},t}$) such that $\mu_{\mathfrak{e},t+1} < \mu_{\mathfrak{e},t}$. $\gamma_{\mathfrak{e},t}$ always increases in t such that $\gamma_{\mathfrak{e},t+1} > \gamma_{\mathfrak{e},t}$. If $\mathcal{T}_{\boldsymbol{\sigma}} < t_{\mathsf{v}}$, ambiguity increases in t such that $\mathsf{v}_{\mathfrak{e},t+\Delta_t} > \mathsf{v}_{\mathfrak{e},t}$, where $\Delta_t > 0$ is sufficiently small. If $\mathcal{T}_{\boldsymbol{\sigma}} = t_{\mathsf{v}}$, $\mathsf{v}_{\mathfrak{e},t_{\mathsf{v}}-\Delta_t} < \mathsf{v}_{\mathfrak{e},t_{\mathsf{v}}}$ and $\mathsf{v}_{\mathfrak{e},t_{\mathsf{v}}+\Delta_t} < \mathsf{v}_{\mathfrak{e},t_{\mathsf{v}}}$. If $\mathcal{T}_{\boldsymbol{\sigma}} > t_{\mathsf{v}}$, ambiguity decreases in t such that $\mathsf{v}_{\mathfrak{e},t+\Delta_t} < \mathsf{v}_{\mathfrak{e},t}$. Let $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)$ such that $\boldsymbol{\sigma}_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)} = \boldsymbol{\sigma}_{\mathfrak{w},0}$ or

$$\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)}-\mu_{\mathfrak{w},0} = \frac{\lambda}{2}(\mathsf{v}_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)}-\mathsf{v}_{\mathfrak{w},0}) \tag{3.C.2}$$

The right hand side of (3.C.2) is negative by C2. Hence, we must have $\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)} < \mu_{\mathfrak{w},0}$. Let $\lambda' > \lambda$ and consider

$$\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)}-\mu_{\mathfrak{w},0} > \frac{\lambda'}{2}(\mathsf{v}_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)}-\mathsf{v}_{\mathfrak{w},0}) \tag{3.C.3}$$

Consider how $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda)$ must change if we want to establish an equality instead of the inequality in (3.C.3), i.e., if we want a $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda')$ solving

$$\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda')}-\mu_{\mathfrak{w},0} = \frac{\lambda'}{2}(\mathsf{v}_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}(\lambda')}-\mathsf{v}_{\mathfrak{w},0})$$

Case 1: Assume that $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda) < t_{\mathsf{v}}$. Here, it is obvious that it does not make sense to decrease $\mathcal{T}_{\boldsymbol{\sigma}}$, as, in such a case, $\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mu_{\mathfrak{w},0}$ would increase, while $(\lambda'/2)(\mathsf{v}_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mathsf{v}_{\mathfrak{w},0})$ would decrease. Hence, we must have $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda') > \mathcal{T}_{\boldsymbol{\sigma}}(\lambda)$ for $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda) < t_{\mathsf{v}}$.

Case 2: Assume that $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda) > t_{\mathsf{v}}$. In such a case, it is required that $\mu_{\mathfrak{e},t}$ decreases faster over time than $(\lambda/2)\mathsf{v}_{\mathfrak{e},t}$ if $t = \mathcal{T}_{\boldsymbol{\sigma}}(\lambda)$ is the *first* period where $\boldsymbol{\sigma}_{\mathfrak{e},t} = \boldsymbol{\sigma}_{\mathfrak{w},0}$. In the contrary case, $t = \mathcal{T}_{\boldsymbol{\sigma}}(\lambda)$ cannot be the first period solving $\boldsymbol{\sigma}_{\mathfrak{e},t} = \boldsymbol{\sigma}_{\mathfrak{w},0}$. Hence, here again, we must increase $\mathcal{T}_{\boldsymbol{\sigma}}$ such that $\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mu_{\mathfrak{w},0}$ decreases, while $(\lambda'/2)(\mathsf{v}_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mathsf{v}_{\mathfrak{w},0})$ also decreases, however, not as fast as $\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mu_{\mathfrak{w},0}$. (If we decrease $\mathcal{T}_{\boldsymbol{\sigma}}$, $\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mu_{\mathfrak{w},0}$ increases, whereas $(\lambda'/2)(\mathsf{v}_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mathsf{v}_{\mathfrak{w},0})$ also increases but slower than $\mu_{\mathfrak{e},\mathcal{T}_{\boldsymbol{\sigma}}}-\mu_{\mathfrak{w},0}$.) Consequently, we must have $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda') > \mathcal{T}_{\boldsymbol{\sigma}}(\lambda)$ for $\mathcal{T}_{\boldsymbol{\sigma}}(\lambda) > t_{\mathsf{v}}$.

Case 3: If we assume that $\mathcal{T}_\sigma(\lambda) = t_v$ and decrease $\mathcal{T}_\sigma$, we have the case $\mathcal{T}_\sigma(\lambda) < t_v$, where decreasing $\mathcal{T}_\sigma$ cannot balance (3.C.3) but an increase in $\mathcal{T}_\sigma$ can; and if we assume that $\mathcal{T}_\sigma(\lambda) = t_v$ and increase $\mathcal{T}_\sigma$, we have the case $\mathcal{T}_\sigma(\lambda) > t_v$, where increasing $\mathcal{T}_\sigma$ can balance (3.C.3) but not decreasing. To summarize, in all considered cases, we must have $\mathcal{T}_\sigma(\lambda') > \mathcal{T}_\sigma(\lambda)$ for $\lambda' > \lambda$. ∎

3.C.3 Transitions

Proof of Lemma 3.3. Notice that the maximal number of plausible transitions is 2 (see Table 3.1). The Gittins index generates the following transition numbers:

$$\mathcal{M}^{[g]} = \{0, 1, 2\}$$

Hence, we have $\max \mathcal{M}^{[g]} = 2$.

Let

$$\lambda^* \equiv 2 \min\{\mu_{\mathfrak{e}} v_{\mathfrak{e}}^{-1}, \mu_{\mathfrak{w}} v_{\mathfrak{w}}^{-1}\}$$

$$\lambda^{**} \equiv 2 \max\{\mu_{\mathfrak{e}} v_{\mathfrak{e}}^{-1}, \mu_{\mathfrak{w}} v_{\mathfrak{w}}^{-1}\}$$

Note that $\lambda^* > 0$ and $\lambda^{**} > 0$ are unique and $\lambda^{**} > \lambda^*$, since $\mu_{\mathfrak{e}} v_{\mathfrak{e}}^{-1} \neq \mu_{\mathfrak{w}} v_{\mathfrak{w}}^{-1}$ (according to Definition 3).

Assume that $\lambda \in (0, \lambda^*)$. If the mean-variance rule starts with option $\mathfrak{e}$, it will generate either sequence DS3, or DS5, or DS9. If the mean-variance rule starts with option $\mathfrak{w}$, it will generate either sequence DS2, or DS4, or DS8. The mean-variance rule will not start with option $\mathfrak{u}$ or select $\mathfrak{u}$ in the second period if $\lambda < \lambda^*$. Consequently, we have $\mathcal{M}^{[\sigma]}_{\lambda<\lambda^*} = \{0, 1, 2\}$ such that $\max \mathcal{M}^{[\sigma]}_{\lambda<\lambda^*} = 2$.

Assume that $\lambda \in [\lambda^*, \lambda^{**})$. If the mean-variance rule starts with option $\mathfrak{e}$, it will generate either sequence DS3 or DS7. If the mean-variance rule starts with option $\mathfrak{w}$, it will generate either sequence DS2 or DS6. The rule will not start with option $\mathfrak{u}$ if $\lambda < \lambda^{**}$. As a consequence, we get $\mathcal{M}^{[\sigma]}_{\lambda^*\leqslant\lambda<\lambda^{**}} = \{0, 1\}$ such that $\max \mathcal{M}^{[\sigma]}_{\lambda^*\leqslant\lambda<\lambda^{**}} = 1$.

Assume that $\lambda \in [\lambda^{**}, \infty)$. The only decision sequence the mean-variance rule can generate is DS1. Consequently, we get $\mathcal{M}^{[\sigma]}_{\lambda\geqslant\lambda^{**}} = \{0\}$ and $\max \mathcal{M}^{[\sigma]}_{\lambda\geqslant\lambda^{**}} = 0$.

Now, let $\lambda' \in (0, \lambda^*)$, $\lambda'' \in [\lambda^*, \lambda^{**})$, and $\lambda''' \in [\lambda^{**}, \infty)$. It is obvious that $\lambda''' > \lambda'' > \lambda'$. Only λ' generates transition numbers such that $\mathcal{M}^{[\sigma]}_{\lambda'} = \mathcal{M}^{[g]}$. For the remaining levels of ambiguity aversion, we get $\max \mathcal{M}^{[\sigma]}_{\lambda''} < \max \mathcal{M}^{[g]}$ and $\max \mathcal{M}^{[\sigma]}_{\lambda'''} < \max \mathcal{M}^{[g]}$. Furthermore:

$$\max \mathcal{M}^{[\sigma]}_{\lambda'} > \max \mathcal{M}^{[\sigma]}_{\lambda''} > \max \mathcal{M}^{[\sigma]}_{\lambda'''}$$

∎

3.C.4 Sampling

Proof of Lemma 3.4. Assume that $\mu_{\mathfrak{e}} > \mu_{\mathfrak{w}}$. As a success in entrepreneurship is generated with probability $\phi_{\mathfrak{e}}$ and a success in wage work with probability $\phi_{\mathfrak{w}}$, we obtain

$$\mathcal{F}_{\mathfrak{w}}^{[\mathbf{g}]} = \begin{cases} 0 & \text{with probability } \phi_{\mathfrak{e}} \\ 1 & \text{with probability } (1-\phi_{\mathfrak{e}})(1-\phi_{\mathfrak{w}}) \\ \infty & \text{with probability } (1-\phi_{\mathfrak{e}})\phi_{\mathfrak{w}} \end{cases} \tag{3.C.4}$$

To derive (3.C.4), note that the Gittins index starts with option $\mathfrak{e}$ if $\mu_{\mathfrak{e}} > \mu_{\mathfrak{w}}$. After selecting option $\mathfrak{e}$, the gambler learns everything about the option. With probability $\phi_{\mathfrak{e}}$, the reward of entrepreneurship is 1 and the gambler generates decision sequence DS3 in Table 3.1. In DS3 only entrepreneurship is selected such that $\mathcal{F}_{\mathfrak{w}}^{[\mathbf{g}]} = 0$. With probability $1-\phi_{\mathfrak{e}}$, the reward of entrepreneurship is zero and the gambler selects wage work. With probability $(1-\phi_{\mathfrak{e}})(1-\phi_{\mathfrak{w}})$, the reward of entrepreneurship *and* wage work is zero and the gambler generates sequence DS9 in Table 3.1. In DS9 wage work is selected once such that $\mathcal{F}_{\mathfrak{w}}^{[\mathbf{g}]} = 1$. With probability $(1-\phi_{\mathfrak{e}})\phi_{\mathfrak{w}}$, the reward of entrepreneurship is zero but the reward of wage work is 1. In such a case, the gambler generates decision sequence DS5 in Table 3.1. In DS5 wage work is selected for an infinite number of periods such that $\mathcal{F}_{\mathfrak{w}}^{[\mathbf{g}]} = \infty$. All other sampling frequencies can be obtained in a similar way.

If we assume $\mu_{\mathfrak{w}} > \mu_{\mathfrak{e}}$, we obtain

$$\mathcal{F}_{\mathfrak{w}}^{[\mathbf{g}]} = \begin{cases} 1 & \text{with probability } 1-\phi_{\mathfrak{w}} \\ \infty & \text{with probability } \phi_{\mathfrak{w}} \end{cases} \tag{3.C.5}$$

For the sampling frequency of unemployment, we have

$$\mathcal{F}_{\mathfrak{u}}^{[\mathbf{g}]} = \begin{cases} 0 & \text{with probability } 1-(1-\phi_{\mathfrak{e}})(1-\phi_{\mathfrak{w}}) \\ \infty & \text{with probability } (1-\phi_{\mathfrak{e}})(1-\phi_{\mathfrak{w}}) \end{cases} \tag{3.C.6}$$

Equation (3.C.4), (3.C.5), and (3.C.6) combined yield

$$\mathbb{P}(\mathcal{F}_{\mathfrak{w}}^{[\mathbf{g}]} = 0) > 0, \quad \mathbb{P}(\mathcal{F}_{\mathfrak{u}}^{[\mathbf{g}]} < \infty) > 0$$

Consider two cases. Case 1: Let $\lambda \geqslant 2\max\{\mu_{\mathfrak{e}}\mathrm{v}_{\mathfrak{e},0}^{-1}, \mu_{\mathfrak{w}}\mathrm{v}_{\mathfrak{w},0}^{-1}\}$. Given this particular level of ambiguity aversion, we get $\boldsymbol{\sigma}_{\mathfrak{e},0}, \boldsymbol{\sigma}_{\mathfrak{w},0} \leqslant 0$. Hence:

$$\mathbb{P}(\mathcal{F}_{\mathfrak{u}}^{[\boldsymbol{\sigma}]} = \infty) = 1$$

such that we get

$$\mathbb{P}(\mathcal{F}_{\mathfrak{u}}^{[\sigma]} > \mathcal{F}_{\mathfrak{u}}^{[\mathbf{g}]}) > 0$$

Case 2: Let $\lambda \in (2\mu_{\mathfrak{e}}\mathrm{v}_{\mathfrak{e},0}^{-1}, 2\mu_{\mathfrak{w}}\mathrm{v}_{\mathfrak{w},0}^{-1})$. Given λ, we have $\boldsymbol{\sigma}_{\mathfrak{e},0} < 0$ and $\boldsymbol{\sigma}_{\mathfrak{w},0} > 0$ such that the mean-variance rule will always start with wage work. With probability $\phi_{\mathfrak{w}}$, the mean-variance rule samples only from wage work. With probability $1 - \phi_{\mathfrak{w}}$, the mean-variance rule only samples once from wage work. However, the mean-variance rule will always sample from wage work at least once, while the Gittins index does not sample from wage work when $\mu_{\mathfrak{e}} > \mu_{\mathfrak{w}}$ and entrepreneurship delivers a success, which occurs with probability $\phi_{\mathfrak{e}}$. ■

3.C.5 *Myopia* versus *ambiguity aversion*

A major drawback of using myopic utility is that results might be mostly driven by myopia and not by ambiguity aversion. However, it is relatively easy to show that this is not the case. The purely myopic index is

$$\mathbf{m}(\mathbb{x}_{i,t}) = \mu(\mathbb{x}_{i,t}) \tag{3.C.7}$$

corresponding to a mean-variance rule with ambiguity neutrality (and risk neutrality). Furthermore, $\mathbf{m}$ generates the optimal strategy if $\delta = 0$. The result stems from the fact that if $\delta = 0$, there will be no incentive to explore (the game ends in the next period). As only the incentive to exploit remains, it is optimal to select the currently most promising option (for a formal argument, see Gittins et al. 2011).

3.C.5.1 *Myopia and probability to select unemployment*

The myopic index will never select unemployment. To see, note that $\mu(\mathbb{x}_{i,t}) > 0$ for all $i \in \mathbb{O}_{-\mathfrak{u}}$ and t, whereas $\mu(\mathbb{x}_{\mathfrak{u},t}) = 0$. Hence:

$$\mathbb{P}(d_t^{[\mathbf{m}]} = \mathfrak{u}) = 0$$

Consequently results in Proposition 3.1 are driven by ambiguity aversion and not by myopia.

3.C.5.2 *Myopia and correction of overconfidence*

Given the myopic rule and conditions in Proposition 3.2, a transition from entrepreneurship to wage work occurs in period

$$\mathcal{T}_{\mathbf{m}} = \frac{\alpha_{\mathfrak{e},0} - \gamma_{\mathfrak{e},0}\mu_{\mathfrak{w},0}}{\mu_{\mathfrak{w},0}}$$

Using results established in the proof of Proposition 3.2, it is easy to show that

$$\mathcal{T}_{\boldsymbol{\sigma}}^{-} - \mathcal{T}_{\mathbf{m}} = \frac{\alpha_{\mathfrak{e},0}(\vartheta_{\mathfrak{e}} - \mathsf{v}_{\mathfrak{w},0})\lambda}{\mu_{\mathfrak{w},0}([\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}}]\lambda - 2\mu_{\mathfrak{w},0})}$$

Using C2, we must have $\alpha_{\mathfrak{e},0}(\vartheta_{\mathfrak{e}} - \mathsf{v}_{\mathfrak{w},0})\lambda < 0$. Using C3, we must have $[\mathsf{v}_{\mathfrak{w},0} - \vartheta_{\mathfrak{e}}]\lambda - 2\mu_{\mathfrak{w},0} < 0$. Hence, we get

$$\mathcal{T}_{\boldsymbol{\sigma}}^{-} - \mathcal{T}_{\mathbf{m}} > 0$$

and the mean-variance rule transitions later than the myopic rule. Also, note that increasing ambiguity aversion will have no effect on $\mathcal{T}_{\mathbf{m}}$ but $\mathcal{T}_{\boldsymbol{\sigma}}$ will increase.

3.C.5.3 Myopia and fast learning

In an environment with fast learning, the purely myopic rule and the Gittins index induce the same decisions. Thus, deviations from optimal behavior, established in Proposition 3.3 and 3.4, must be generated by ambiguity aversion.

Furthermore, it can be demonstrated that given sufficiently high ambiguity aversion and fast learning, the myopic mean-variance rule maximizes both the myopic and the global mean-variance criterion:

Proposition 3.5. *Assume that unemployment is absorbing for any potential decision rule such that*

$$\mathbb{P}(d_t^{[\mathbf{r}]} = \mathfrak{u} | d_{t-1}^{[\mathbf{r}]} = \mathfrak{u}) = 1 \qquad \textit{for all } \mathbf{r}$$

qua assumption.[92] *There always exists a sufficiently high level of ambiguity aversion, λ, such that a decision rule maximizing the myopic mean-variance criterion $\boldsymbol{\sigma}_t$ will also maximize the global mean-variance criterion $\sum_{t=1}^{\infty} \delta^{t-1}\boldsymbol{\sigma}_t$ in expectations given initial prior information $\mathbb{X}$.*

Proof. Given that unemployment is absorbing such that $\mathbb{P}(d_t = \mathfrak{u} | d_{t-1} = \mathfrak{u}) = 1$, the only plausible decision sequences for any rule are DS1–DS9. Let

$$\Theta_{\mathrm{DS}} = \sum_{t=1}^{\infty} \delta^{t-1} \boldsymbol{\sigma}_t^{[\mathrm{DS}]}$$

denote the global mean-variance criterion generated by decision sequence DS. Furthermore, let $\tilde{\Theta}$ denote the expected global mean-variance outcome generated by the myopic mean-variance rule. There are three setting of interest: (1) $\boldsymbol{\sigma}_{\mathfrak{e}}, \boldsymbol{\sigma}_{\mathfrak{w}} < 0$, (2) $\boldsymbol{\sigma}_i > 0$ and $\boldsymbol{\sigma}_j < 0$ where $i \neq j$ and $i, j \in \mathbb{O}_{-\mathfrak{u}}$, (3) $\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j > 0$ and $\boldsymbol{\sigma}_i > \boldsymbol{\sigma}_j$ where $i \neq j$ and $i, j \in \mathbb{O}_{-\mathfrak{u}}$.

[92] Note that $\mathbb{P}(d_t^{[\mathbf{r}]} = \mathfrak{u} | d_{t-1}^{[\mathbf{r}]} = \mathfrak{u}) = 1$ holds for any index rule but not necessarily for non-index rules.

Setting 1: Assume that $\boldsymbol{\sigma}_{\mathfrak{e}}, \boldsymbol{\sigma}_{\mathfrak{w}} < 0$. The myopic mean-variance rule generates

$$\tilde{\Theta}_1 = 0$$

For decision sequences 6–9, we obtain $\Theta_6, \Theta_7, \Theta_8, \Theta_9 < 0$. For decision sequences 2–5, we get $\Theta_2, \Theta_3, \Theta_4, \Theta_5 < 0$ if

$$\lambda > \max\left\{\frac{2}{\mathsf{v}_{\mathfrak{w}}}\left(\mu_{\mathfrak{w}} + \frac{\delta}{1-\delta}\right), \frac{2}{\mathsf{v}_{\mathfrak{e}}}\left(\mu_{\mathfrak{e}} + \frac{\delta}{1-\delta}\right),\right.$$

$$\left.\frac{2}{\mathsf{v}_{\mathfrak{w}} + \delta \mathsf{v}_{\mathfrak{e}}}\left(\mu_{\mathfrak{w}} + \delta\mu_{\mathfrak{e}} + \frac{\delta^2}{1-\delta}\right), \frac{2}{\mathsf{v}_{\mathfrak{e}} + \delta \mathsf{v}_{\mathfrak{w}}}\left(\mu_{\mathfrak{e}} + \delta\mu_{\mathfrak{w}} + \frac{\delta^2}{1-\delta}\right)\right\}$$

Setting 2: Assume, wlog., that $\boldsymbol{\sigma}_{\mathfrak{e}} > 0$ and $\boldsymbol{\sigma}_{\mathfrak{w}} < 0$. The myopic mean-variance rule generates

$$\tilde{\Theta}_2 = \mu_{\mathfrak{e}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{e}} + \mu_{\mathfrak{e}}\frac{\delta}{1-\delta} > 0$$

It is obvious that starting with $\mathfrak{u}$ generates zero. There are two alternatives that are not obviously inferior to $\tilde{\Theta}_2$.[93] (i) Starting with $\mathfrak{w}$ yields at best

$$\tilde{\Theta}_2' = \mu_{\mathfrak{w}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{w}} + \mu_{\mathfrak{w}}\frac{\delta}{1-\delta} + (1-\mu_{\mathfrak{w}})\left(\delta\left[\mu_{\mathfrak{e}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{e}}\right] + \mu_{\mathfrak{e}}\frac{\delta^2}{1-\delta}\right)$$

If the level of ambiguity aversion is sufficiently large such that

$$\lambda > \frac{2[(\delta\mu_{\mathfrak{e}} - 1)\mu_{\mathfrak{w}} + (1-\delta)\mu_{\mathfrak{e}}]}{(\delta-1)(\mathsf{v}_{\mathfrak{w}} + [\delta - \delta\mu_{\mathfrak{w}} - 1]\mathsf{v}_{\mathfrak{e}})}$$

we have $\tilde{\Theta}_2 > \tilde{\Theta}_2'$. (ii) The second alternative is to start with entrepreneurship but to switch to wage work, instead of unemployment, given a nonsuccess in entrepreneurship. In such a case, the expected outcome is at best

$$\tilde{\Theta}_2'' = \mu_{\mathfrak{e}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{e}} + \mu_{\mathfrak{e}}\frac{\delta}{1-\delta} + (1-\mu_{\mathfrak{e}})\left(\delta\left[\mu_{\mathfrak{w}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{w}}\right] + \mu_{\mathfrak{w}}\frac{\delta^2}{1-\delta}\right)$$

Here, we get $\tilde{\Theta}_2 > \tilde{\Theta}_2''$ if

$$\lambda > 2\frac{\mu_{\mathfrak{w}}}{(1-\delta)\mathsf{v}_{\mathfrak{w}}}$$

Consequently, no alternative decision rule provides a better expected global mean-

[93] An obviously inferior alternative is, for instance, selecting unemployment before trying entrepreneurship.

variance outcome than the myopic mean-variance rule when ambiguity aversion is sufficiently large.

Setting 3: Assume that $\boldsymbol{\sigma}_{\mathfrak{e}}, \boldsymbol{\sigma}_{\mathfrak{w}} > 0$ and, wlog., $\boldsymbol{\sigma}_{\mathfrak{e}} > \boldsymbol{\sigma}_{\mathfrak{w}}$. The myopic mean-variance rule generates

$$\tilde{\Theta}_3 = \mu_{\mathfrak{e}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{e}} + \mu_{\mathfrak{e}}\frac{\delta}{1-\delta} + (1-\mu_{\mathfrak{e}})\left(\delta\left[\mu_{\mathfrak{w}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{w}}\right] + \mu_{\mathfrak{w}}\frac{\delta^2}{1-\delta}\right) > 0$$

Starting with $\mathfrak{u}$ generates zero, which is inferior to $\tilde{\Theta}_3$. Starting with $\mathfrak{w}$, instead of $\mathfrak{e}$, yields

$$\tilde{\Theta}_3' = \mu_{\mathfrak{w}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{w}} + \mu_{\mathfrak{w}}\frac{\delta}{1-\delta} + (1-\mu_{\mathfrak{w}})\left(\delta\left[\mu_{\mathfrak{e}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{e}}\right] + \mu_{\mathfrak{e}}\frac{\delta^2}{1-\delta}\right)$$

at best. Let

$$K \equiv (1-\delta)(\mathsf{v}_{\mathfrak{w}} - \mathsf{v}_{\mathfrak{e}}) + \delta(\mu_{\mathfrak{e}}\mathsf{v}_{\mathfrak{w}} - \mu_{\mathfrak{w}}\mathsf{v}_{\mathfrak{e}})$$

If entrepreneurship is sufficiently less ambiguous than wage work, such that $\mathsf{v}_{\mathfrak{w}} > \mathsf{v}_{\mathfrak{e}}$ and $\mu_{\mathfrak{e}}\mathsf{v}_{\mathfrak{w}}^{-1} > \mu_{\mathfrak{w}}\mathsf{v}_{\mathfrak{e}}^{-1}$, we have $K > 0$, which is consistent with $\boldsymbol{\sigma}_{\mathfrak{e}} > \boldsymbol{\sigma}_{\mathfrak{w}}$. Then, $\tilde{\Theta}_3'$ is inferior to $\tilde{\Theta}_3$, such that $\tilde{\Theta}_3 > \tilde{\Theta}_3'$, if

$$\lambda > 2(\mu_{\mathfrak{w}} + \mu_{\mathfrak{e}})K^{-1}$$

Not switching to wage work in case of a nonsuccess in entrepreneurship (either staying in entrepreneurship or switching to unemployment) generates

$$\tilde{\Theta}_3'' = \mu_{\mathfrak{e}} - \frac{\lambda}{2}\mathsf{v}_{\mathfrak{e}} + \mu_{\mathfrak{e}}\frac{\delta}{1-\delta} < \tilde{\Theta}_3$$

Hence, given a sufficiently high level of ambiguity aversion, no alternative decision rule generates a higher expected global mean-variance outcome than the myopic mean-variance rule. ■

Appendix 3.D

In this appendix, I discuss the effect of risk aversion. Assuming that $\lambda = 0$ (ambiguity neutrality) and using the mean-variance formulation of Maccheroni et al. (2013), we obtain

$$\boldsymbol{\sigma}_{i,t} = \mu_{i,t} - \frac{\theta}{2} \times \frac{\alpha_{i,t}\beta_{i,t}}{\gamma_{i,t}^2}$$

where the variance of π_i is computed according to a beta-Bernoulli model. Assume that an individual starts with observations $(\alpha_{i,0}, \beta_{i,0})$ and selects entrepreneurship in the first period, such that

$$\mu_{i,0} - \frac{\theta}{2} \times \frac{\alpha_{i,0}\beta_{i,0}}{\gamma_{i,0}^2} - \boldsymbol{\sigma}_{\mathfrak{w},0} > 0$$

In line with the setup of Proposition 3.2, assume that there are only nonsuccesses in en-

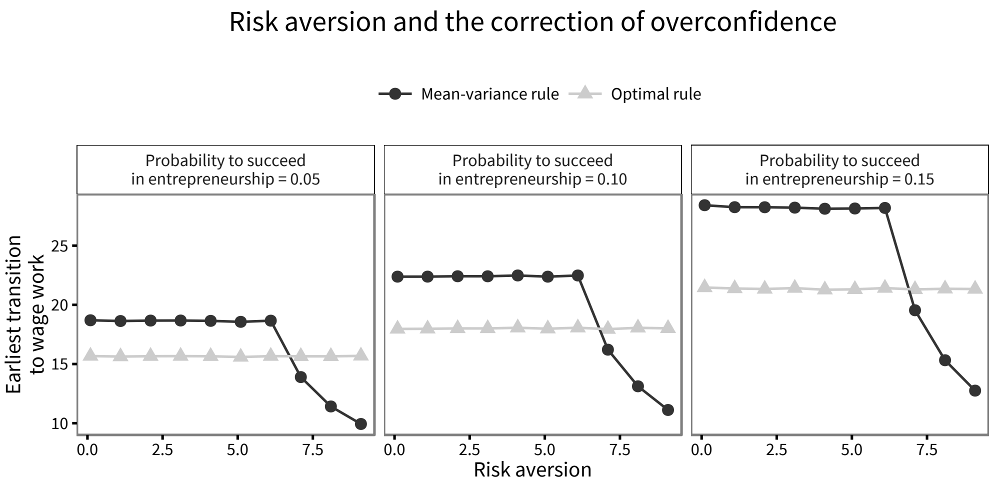


FIGURE 3.D.1. Effect of risk aversion on correction of overconfidence

trepreneurship, which we denote by ϖ, such that

$$\mu_{i,t} - \frac{\theta}{2} \times \frac{\alpha_{i,t}\beta_{i,t}}{\gamma_{i,t}^2} - \boldsymbol{\sigma}_{\mathfrak{w},0} = \frac{\alpha_{\mathrm{e},0}(\gamma_{\mathrm{e},0} + \varpi) - \theta\alpha_{\mathrm{e},0}(\beta_{\mathrm{e},0} + \varpi) - 2(\gamma_{\mathrm{e},0} + \varpi)^2\boldsymbol{\sigma}_{\mathfrak{w},0}}{2(\gamma_{\mathrm{e},0} + \varpi)^2}$$

where $\varpi = t$ by construction. The minimum level of risk aversion that would make the individual willing to switch to wage work in period t is

$$\theta^*(\varpi) = \frac{\alpha_{\mathrm{e},0} - 2(\gamma_{\mathrm{e},0} + \varpi)\boldsymbol{\sigma}_{\mathfrak{w},0}}{\alpha_{\mathrm{e},0}}$$

The individual who is willing to endure one nonsuccess more, before switching to wage work, has a risk aversion level of

$$\theta^*(\varpi + 1) = \frac{\alpha_{\mathrm{e},0} - 2(\gamma_{\mathrm{e},0} + \varpi + 1)\boldsymbol{\sigma}_{\mathfrak{w},0}}{\alpha_{\mathrm{e},0}}$$

Now, it is obvious that

$$\theta^*(\varpi + 1) - \theta^*(\varpi) = -2\boldsymbol{\sigma}_{\mathfrak{w},0}\alpha_{\mathrm{e},0}^{-1} < 0$$

if wage work is considered to be better than unemployment. As in case of a high level of overconfidence, the prior number of successes, $\alpha_{e,0}$, will also be high, different levels of risk aversion will only have a small effect on the willingness to transition to wage work.

As an example, consider the second scenario given in Table 3.2. Let $\lambda = 0$ but $\theta \in \{0.1, 1.1, \ldots, 8.1, 9.1\}$, such that instead of considering different levels of ambiguity aversion, we consider different levels of risk aversion. Simulating the scenario generates Figure 3.D.1. Risk aversion generates a pattern different from ambiguity aversion. First, the transition time to wage work is only slightly delayed. Second, the delay does not increase in aversion. Only, if risk aversion is high, individuals change their transition behavior—they transition too early compared to optimal behavior. Hence, it is unlikely that earnings-puzzle-like observations can be explained by risk aversion.

Appendix 3.E

In this appendix, I test two central predictions of the bandit model and demonstrate that both predictions are consistent with available data.

3.E.1 First hypothesis: Propensity to avoid ambiguous options

A testable version of the prediction that increasing ambiguity aversion will amplify the tendency to select a "sure" option, based on Proposition 3.2 and 3.5, is as follows.

Hypothesis 3.1. Compared to a low level of ambiguity aversion, a higher aversion level is associated with a lower probability to select entrepreneurship because the propensity to select an option with low ambiguity increases.

In the bandit choice model, the low-ambiguity option is represented by unemployment. Some types of wage work can also be a low-ambiguity option. An example is Japan's lifetime employment system where workers are hired immediately after their graduation and, on the basis of a gentleman's agreement, are not laid off until they retire (Sullivan & Peterson 1991). However, there is no indication that entrepreneurship generally represents the most certain option as entrepreneurs are usually confronted with new problems that cannot be fully anticipated, whereas wage workers and the unemployed solve rather well-known problems. Furthermore, information about entrepreneurship is generally scarcer than information about non-entrepreneurial options (Kelley, Brush, Greene & Litovsky 2013).

3.E.1.1 Modeling ambiguity aversion and choice of entrepreneurship

I assess the first prediction with a clustering model. Assume that we can observe the ambiguity aversion and the probability to select entrepreneurship of individual (country) $j = 1, \ldots, N_j$. Let $\mathbb{P}e_j$ denote j's probability to select entrepreneurship and let, furthermore, λ_j^e denote j's ambiguity-aversion type. Using a parsimonious model, I simply consider the joint distribution of $\mathbb{P}e_j$ and λ_j^e, where $\mathbb{P}e_j$ and λ_j^e are combined into vector $\mathbf{x}_j = [\lambda_j^e, \mathbb{P}e_j]^\top$. I assume that $\mathbf{x}_j$ can be modeled with a Gaussian mixture model of the following form:

$$\prod_{j=1}^{N_j} \left\{ \sum_{c=1}^{N_c} p_c f_c(\mathbf{x}_j; \mathbf{m}_c, \Sigma_c) \right\} \tag{3.E.1}$$

$$f_c(\mathbf{x}_j; \mathbf{m}_c, \Sigma_c) = (2\pi)^{-1} |\Sigma_c|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} [\mathbf{x}_j - \mathbf{m}_c]^\top \Sigma_c^{-1} [\mathbf{x}_j - \mathbf{m}_c] \right)$$

N_j is the overall number of individuals or observations. N_c is the number of clusters. $p_c > 0$ is the probability that an observation belongs to cluster c. Note that $\sum_{c=1}^{N_c} p_c = 1$. $\mathbf{m}_c = [\bar{\lambda}_c, \overline{\mathbb{P}e_c}]^\top$ is a cluster-specific mean and Σ_c is a cluster-specific covariance matrix.

The mixture model in (3.E.1) allows data to be clustered such that different clusters of observations might be generated by different distributions. The number of clusters is finite. Σ_c can be written in terms of its Eigen decomposition such that

$$\Sigma_c = \kappa_c \mathbf{A}_c \mathbf{S}_c \mathbf{A}_c^\top$$

where κ_c is the volume, $\mathbf{A}_c$ the orientation, and $\mathbf{S}_c$ the shape (Banfield & Raftery 1993). I use 10 different specifications of Σ_c that differ with respect to how they treat group differences. For instance, it can be assumed that volume and shape can be equal across clusters but not orientation. For the number of clusters, I assume a maximum of 9 clusters, i.e., $N_c \in \{1, 2, \ldots, 9\}$.

In its most parsimonious version, Hypothesis 3.1 requires only two clusters: a cluster with moderate or low average ambiguity aversion and a cluster with high average ambiguity aversion. According to Hypothesis 3.1, the cluster with high average ambiguity aversion should exhibit a lower average probability to select entrepreneurship than the other cluster. I will show that given a number of alternative empirical models the best-performing model is the one suggested by Hypothesis 3.1.

3.E.1.2 Data on ambiguity aversion and choice of entrepreneurship

I use Hofstede's uncertainty avoidance index, available from Hofstede (2015) and described in Hofstede et al. (2010), as a proxy for ambiguity-aversion type. Hofstede's index is never negative and more ambiguity aversion is represented by higher index values.

To measure the probability to select entrepreneurship, I rely on established business ownership rates, representing the long-term tendency of a society to select entrepreneurship, provided by the Global Entrepreneurship Monitor and available from GEM (2015).

Business ownership data is annual and the range of periods is 2001–2013 but not all periods are available for all countries. However, since Hofstede's index is time-invariant, I average over all available periods.[94] As a result, I have 67 joint observation of average business ownership rate and ambiguity-aversion type.

3.E.1.3 Ambiguity aversion and the probability of entrepreneurship

Empirical models are estimated with the expectation-maximization algorithm.[95] Table 3.E.1 depicts the Bayesian information criterion (BIC) for estimated models. A higher

TABLE 3.E.1. Explanatory performance of different model specifications, where the best model is marked by an asterisk

Number of clusters	BIC of best covariance specification
1	− 1006.12
2	− 997.92*
3	− 1003.24
4	− 1010.01
5	− 1019.69
> 5	⩽ −1025.16

BIC indicates better explanatory performance given a penalty for model complexity.

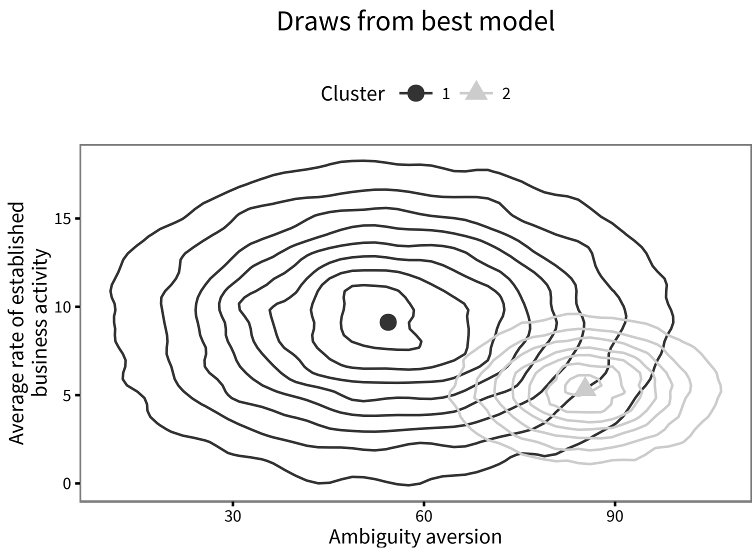
According to Table 3.E.1, the best model has two clusters. The model suggested by Table 3.E.1 has the following cluster means:

$$\mathbf{m}_1 = \begin{bmatrix} \bar{\lambda}_1 & = 54.379 \\ \overline{\mathbb{P}e}_1 = & 9.124 \end{bmatrix}, \quad \mathbf{m}_2 = \begin{bmatrix} \bar{\lambda}_2 & = 85.299 \\ \overline{\mathbb{P}e}_2 = & 5.325 \end{bmatrix}$$

where the first element is average ambiguity-aversion type and the second the average rate of established business activity. Figure 3.E.1 shows 100,000 draws from the best model's density. The average business ownership rate of the high-aversion cluster (the gray cluster with mean $\mathbf{m}_2$) is lower than the corresponding rate of the cluster with low or moderate

[94] Data characteristics are provided in Appendix 3.F.

[95] To estimate the model, I use the mclust package in R. An alternative suggested by Fraley & Raftery (2007) is to replace the maximum-likelihood estimator, in the expectation-maximization algorithm, by a maximum-posterior estimator. This modified approach can substantially reduce the number of non-fitted models. However, the general result (provided upon request) does not change. An additional concern is that established business activity may underestimate the amount of information about entrepreneurship (those who abandon entrepreneurship also provide information) such that the assumption that entrepreneurship is the most ambiguous option available becomes problematic. To address this issue, I estimate the model only with countries that have an average business activity rate not above 10%. Results (provided upon request) are still consistent with Hypothesis 3.1.



Note: Point and triangle indicate cluster means.

FIGURE 3.E.1. Density of the best model according to Bayesian information criterion

ambiguity aversion (the black cluster with mean $\mathbf{m}_1$). Hence, I obtain the following result consistent with Hypothesis 3.1:

Empirical result 3.1. Higher ambiguity aversion is associated with a lower level of long-term entrepreneurial activity.

3.E.2 Second hypothesis: Learning efficiency of entrepreneurs

Consider two individuals whose only difference is how strong they dislike ambiguity. Both individuals exhibit the same level of overconfidence with respect to entrepreneurship. Assume that both individuals start an own company. According to the theoretical model, the more ambiguity-averse individual will learn less efficiently than the less ambiguity-averse individual. Consequently, the more ambiguity-averse individual has a *higher* probability to survive in entrepreneurship than the less ambiguity-averse individual given the same conditions because the less ambiguity-averse individual will correct her false beliefs (she is overconfident) more efficiently than will the more ambiguity-averse individual (who is also overconfident). Hence, the bandit model makes the following testable prediction, on the basis of Proposition 3.3:

Hypothesis 3.2. Given that overconfidence of entrepreneurs is a common phenomenon, the probability of firm death should be lower for individuals with relatively high ambiguity aversion than for individuals with relatively low ambiguity aversion.

3.E.2.1 *Modeling ambiguity aversion and entrepreneurial survival*

To test Hypothesis 3.2, I examine the impact of ambiguity aversion on firm death rates. Let $D \in (0,1)$ denote the death rate of firms. Assume that death rates are individual- and time-specific, where individuals (countries in our setting) are indexed by j and time by t. Rates assume values in $(0,1)$; the distribution of rates is not necessarily symmetric; and there also might be heteroskedasticity. Hence, a simple linear regression might not be appropriate. In such a setting, Ferrari & Cribari-Neto (2004) propose to use a beta regression. Ferrari & Cribari-Neto (2004) suggest to parameterize the beta distribution in terms of its mean and precision.[96]

Using the parameterization suggested by Ferrari & Cribari-Neto (2004), I assume the following:

$$D_{j,t} \sim \mathfrak{B}(\mu^D_{j,t}, \gamma^D), \quad \mu^D_{j,t} = \text{logit}^{-1}(\eta_{j,t}), \quad \gamma^D \text{ is constant} \tag{3.E.2}$$

such that death rates $D_{j,t}$ follow a beta distribution with mean $\mu^D_{j,t}$ and constant precision γ^D. $\eta_{j,t}$ is a linear function of ambiguity aversion type and a number of additional covariates that is connected to the mean by a logit-link function. Mean and variance both vary across time and individuals such that we do not have to assume homoskedasticity:

$$\mathbb{E}[D_{j,t}] = \mu^D_{j,t}, \quad \mathbb{V}[D_{j,t}] = \frac{\mu^D_{j,t}(1-\mu^D_{j,t})}{1+\gamma^D}$$

Data is used to predict $\eta_{j,t}$ given by

$$\eta_{j,t} = \text{constant} + e_\lambda \mathbb{1}\{j \in H\} + \mathbf{c}_{j,t}^\top \mathbf{e} \tag{3.E.3}$$

The indicator $\mathbb{1}\{j \in H\}$, capturing ambiguity aversion types, is 1 if country j has relatively high ambiguity aversion (i.e., j belongs to the high-ambiguity-aversion group [type] H) and zero else. $\mathbf{c}_{j,t}$ is a vector of additional covariates. e_λ and $\mathbf{e}$ are coefficients to be estimated. According to Hypothesis 3.2, I expect $e_\lambda < 0$ such that belonging to the high-ambiguity-aversion type decreases (average) firm death rates.

As a robustness test, I also estimate a linear regression

$$\log\left(\frac{D_{j,t}}{1-D_{j,t}}\right) = \eta_{j,t} + \varepsilon_D \tag{3.E.4}$$

where ε_D is an error term with the usual OLS properties. The dependent variable (firm

---

[96] Let μ^D denote the mean and γ^D the precision. A beta distribution is traditionally parameterized in terms of α^D and β^D such that density is $f(y;\alpha^D,\beta^D) = \mathrm{B}(\alpha^D,\beta^D)^{-1} y^{\alpha^D-1}(1-y)^{\beta^D-1}$. A parameterization in terms of $\mu^D = \alpha^D(\alpha^D+\beta^D)^{-1}$ and $\gamma^D = \alpha^D + \beta^D$ yields density $f(y;\mu^D,\gamma^D) = \mathrm{B}(\mu^D\gamma^D, [1-\mu^D]\gamma^D)^{-1} y^{\mu^D\gamma^D-1}(1-y)^{[1-\mu^D]\gamma^D-1}$.

death rates, $D_{j,t}$) is transformed such that it takes values on the real line. However, the model in (3.E.4) does not allow for the identification of effect sizes—due to the transformation of the dependent variable, only effect directions can be identified.

Ambiguity-aversion type might be endogenous, $\mathbb{1}\{j \in H\}$ and ε_D might be correlated. Therefore, I also test whether there might be an endogeneity problem by using two different instruments for $\mathbb{1}\{j \in H\}$ in the linear model: an indicator that is 1 if the religious majority in the country under examination is Roman Catholic and zero otherwise; and the exposure of a country to natural disasters, measured by the number of disasters between 1900 and 2000. Religion is exogenous to firm survival in the period under consideration but it is linked to ambiguity aversion (see, for instance, Huang 2008). Natural disasters are also exogenous to firm survival but can have an impact on a society's preferences toward ambiguity.

3.E.2.2 Data on entrepreneurial survival

I use the same measure of ambiguity aversion as before, i.e., Hofstede's index. Figure

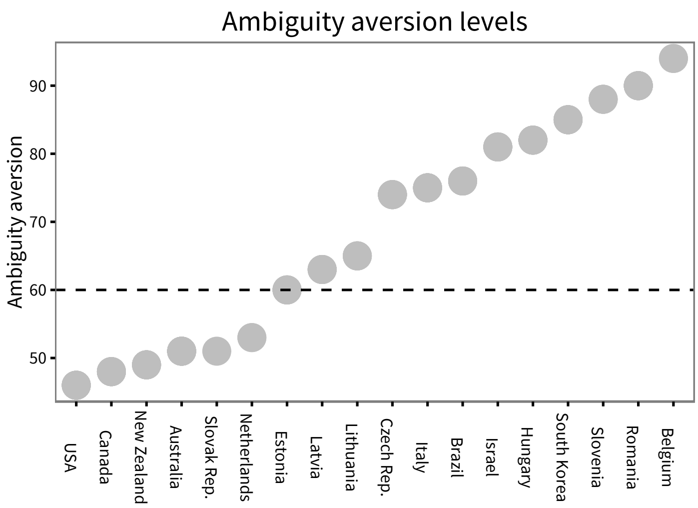


FIGURE 3.E.2. Levels of ambiguity aversion (smaller values indicate less aversion)

3.E.2 shows country-specific ambiguity aversion levels. I assume that ambiguity aversion is relatively high if the level of uncertainty aversion is above 60, which is indicated with a horizontal line in Figure 3.E.2.

OECD (2015) provides a small unbalanced panel of annual enterprise death rates for a number of countries. Death rates are reported for the period from 2006 to 2012 but some periods are missing for some countries.[97]

[97] Data characteristics of enterprise death rates are provided in Appendix 3.F.

The binary variable indicating whether the religious majority is Roman Catholic is constructed on the basis of population surveys. I classify the following countries as Roman Catholic: Australia, Belgium, Brazil, Canada, Czech Republic, Hungary, Italy, Lithuania, Netherlands, New Zealand, Slovakia, and Slovenia. Exposure to natural disasters is constructed on the basis of data provided by Guha-Sapir et al. (2016). Exposure to natural disasters is significantly negatively correlated with high ambiguity aversion (correlation: –0.409; p-value: ≈ 0).

Additional covariates (captured by vector $\mathbf{c}$) are the following: GDP growth, inflation, real interest rates, start-up costs, the unemployment rate, and average ease of doing business. All covariates are provided by the World Bank and available from World Bank (2015).

Average ease of doing business is constructed on the basis of annual indicators, where a higher indicator value corresponds to a situation where doing business is easier, and is used to measure between-country differences in ease of doing business.[98] It should be noted that average ease of doing business and country-specific ambiguity aversion are highly correlated (correlation: –0.645; p-value: ≈ 0), implying that in countries where ambiguity aversion is lower doing business is easier (or *vice versa*). Hence, by including average ease of doing business indicators, it is possible to capture country differences that are related to the business environment but not directly related to ambiguity preferences.

The 2006–2012 period includes the financial crisis, which affected entrepreneurial survival. Therefore, I construct a crisis dummy that is 1 in year 2008, 2009, and 2010, but zero else.[99] A total of 89 joint observations are available for analysis.

3.E.2.3 *Effects of ambiguity aversion on the death of firms*

The beta regression model, given by (3.E.2) and (3.E.3), is estimated by maximum likelihood with the bias correction proposed by Kosmidis & Firth (2010).[100] The linear model, given by (3.E.3) and (3.E.4), is estimated using OLS with heteroskedasticity-robust standard errors. Estimation results are presented in Table 3.E.2. With the exception of significance levels and effect sizes, there is no substantial difference between the beta regression and OLS.

Significant effects not related to ambiguity aversion are all plausible. For instance, an increase in GDP growth significantly reduces death rates, which is a plausible effect since improved demand conditions should foster entrepreneurial survival. An increase in real interest rates, corresponding to tighter restrictions on borrowing money, signifi-

[98] In Appendix 3.F, I provide characteristics of the ease of doing business indicator.
[99] The crisis dummy is, as it might be expected, positively correlated with entrepreneurial death rates (correlation: 0.266) and the correlation is significant at a 5%-level (p-value: 0.012).
[100] Maximum-likelihood based beta regressions are implemented in the R-package betareg.

TABLE 3.E.2. Determinants of firm death rates

Variable	Beta regression		OLS (transformed death rates)	
	Coefficient	Standard error	Coefficient	Standard error
Belonging to high-ambiguity-aversion group[a]	−0.256**	(0.120)	−0.325**	(0.159)
GDP growth	−0.025***	(0.009)	−0.026**	(0.011)
Inflation	0.036**	(0.014)	0.033**	(0.014)
Real interest rate	0.022***	(0.006)	0.023***	(0.005)
Start-up costs	0.041***	(0.008)	0.046***	(0.008)
Unemployment rate	0.007	(0.014)	0.006	(0.013)
Average ease of doing business[b]	0.019***	(0.007)	0.016**	(0.008)
Crisis dummy	0.139*	(0.083)	0.172**	(0.086)
Constant	−4.094***	(0.639)	−3.940***	(0.788)
Precision	99.77***	(12.910)		
	Pseudo R^2 = 0.352		R^2 = 0.290	

Note: [a]Indicator is 1 if country belongs to high-ambiguity-aversion group and zero else; [b]an increase in the indicator signifies that doing business becomes easier; ***significant at 1%-level; **significant at 5%-level; *significant at 10%-level; standard errors are in parentheses; in case of OLS, standard errors are robust; note that it is not possible to identify effect sizes in OLS model, while effect directions can be identified.

cantly increases death rates. An increase in start-up costs, reducing available resources for entrepreneurial survival by imposing higher entry barriers, also significantly increases death rates. The financial crisis significantly increases death rates. An interesting finding is that if doing business becomes easier, death rates are increased, which is a plausible result if ease of doing business reflects competition effects: If doing business becomes easier, competition is intensified such that survival probabilities decrease.

What is most important, belonging to the high-ambiguity-aversion type significantly decreases death rates in the beta and the OLS model: The estimated coefficient of ambiguity aversion, in Table 3.E.2, is negative and significant at the 5%-level in case of the beta regression and the OLS model.

To assess whether there is an endogeneity problem, I estimate the linear model by two-stage least squares using the Roman Catholic indicator and exposure to natural disasters as instruments. For each of the two instrumental variable regressions, I run two tests: an F-test for strength of instrument and a Wu-Hausman test for endogeneity. Test results are presented in Table 3.E.3. Both instruments have sufficient strength. An endogeneity problem cannot be detected at the 5%-level in case of both instruments. Thus, OLS is preferred to IV.

TABLE 3.E.3. Endogeneity problem

Instrument	H0: Instrument is weak	H0: There is no endogeneity problem
Roman Catholic	Rejected at 5% (p-value: 0.001)	Not rejected at 5% (p-value: 0.980)
Exposure to natural disasters	Rejected at 5% (p-value: 0.001)	Not rejected at 5% (p-value: 0.566)

To summarize, I establish the following result consistent with Hypothesis 3.2:

Empirical result 3.2. Higher ambiguity aversion is associated with a higher probability of firm survival.

I conclude that two central choice patterns predicted by the bandit model are consistent with available data.

Appendix 3.F

This appendix provides data characteristics. The data set features three important variables: ambiguity preferences, approximated by Hofstede's uncertainty avoidance index; average established business activity; and enterprise death rates. Furthermore, I use additional covariates (viz., GDP growth, inflation, real interest rate, start-up costs, unemployment rates, and average ease of doing business).

Countries in data sets used to assess the two predictions are shown in Table 3.F.1. Table 3.F.2 presents a detailed analysis of Hofstede's index. The table is the result of the application of k-means clustering with three clusters, where I used the Hartigan & Wong (1979) algorithm to minimize within-cluster sums of squares. According to Table 3.F.2**,** cluster means explain ca. 87% of the overall variation (100% is perfect fit). The high-aversion group consists of approx. 48% of all countries with an available index. The moderate- and high-aversion group accounts for ca. 84% of all countries, while the low-aversion group accounts only for 16%. To summarize, most societies have either a moderate or a high level of ambiguity aversion.

The Global Entrepreneurship Monitor defines the established business ownership rate as the percentage of the population aged 18–64 who run or manage a business that paid payments to the owner for more than 42 months. Figure 3.F.1 shows the distribution of *average* established business activity. The distribution is skewed such that low values are more likely than high values. Consequently, most countries have a rather low average rate of established business activity.

OECD (2015) defines firm death as follows.

Table 3.F.1. Countries in data

First prediction	Second prediction
Argentina, Australia, Austria, Bangladesh, Belgium, Brazil, Canada, Chile, China, Colombia, Costa Rica, Croatia, Czech Rep., Denmark, Ecuador, El Salvador, Estonia, Finland, France, Germany, Great Britain, Greece, Guatemala, Hong Kong, Hungary, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Korea South, Latvia, Lithuania, Luxembourg, Malaysia, Mexico, Morocco, Netherlands, New Zealand, Norway, Pakistan, Panama, Peru, Philippines, Poland, Portugal, Romania, Russia, Serbia, Singapore, Slovak Rep., Slovenia, Spain, Suriname, Sweden, Switzerland, Taiwan, Thailand, Trinidad and Tobago, Turkey, USA, Uruguay, Venezuela, Vietnam	Australia, Belgium, Brazil, Canada, Czech Rep., Estonia, Hungary, Israel, Italy, Korea South, Latvia, Lithuania, Netherlands, New Zealand, Romania, Slovak Rep., Slovenia, USA
67 countries	18 countries

> An employer enterprise death occurs either at the death of an enterprise with at least one employee in the year of death or when an enterprise shrinks to below the threshold of one employee for at least two years. … The employer enterprise death rate corresponds to the number of deaths of employer enterprises as a percentage of the population of active enterprises with at least one employee. (OECD 2015, p. 52)

Figure 3.F.2 plots the distribution of firm death rates, which appears fairly normal such that it is not surprising that OLS and the beta regression deliver similar answers. Figure 3.F.3 shows the relation between high ambiguity aversion and the instruments, where exposure to natural disasters is normalized by the maximal number of natural disasters in the sample.

In Figure 3.F.4, I plot the distributions of GDP growth, inflation, real interest rates, start-up costs, unemployment rates, and average ease of doing business.

TABLE 3.F.2. Low-, moderate-, and high-aversion groups

	Low-aversion group	Moderate-aversion group	High-aversion group
Group mean of Hofstede's index	28.00	58.48	88.09
Countries belonging to group	China, Denmark, Great Britain, Hong Kong, India, Ireland, Jamaica, Malaysia, Singapore, Sweden, Vietnam	Australia, Austria, Bangladesh, Canada, Ecuador, Estonia, Finland, Germany, Indonesia, Iran, Latvia, Lithuania, Luxembourg, Morocco, Netherlands, New Zealand, Norway, Pakistan, Philippines, Slovak Rep., Switzerland, Taiwan, Thailand, Trinidad and Tobago, USA	Argentina, Belgium, Brazil, Bulgaria, Chile, Colombia, Costa Rica, Croatia, Czech Rep., El Salvador, France, Greece, Guatemala, Hungary, Israel, Italy, Japan, Korea South, Malta, Mexico, Panama, Peru, Poland, Portugal, Romania, Russia, Serbia, Slovenia, Spain, Suriname, Turkey, Uruguay, Venezuela
Group size	11	25	33
Within sum of squares	966.00	1714.24	2360.73
Total variation accounted for by group means		86.8%	

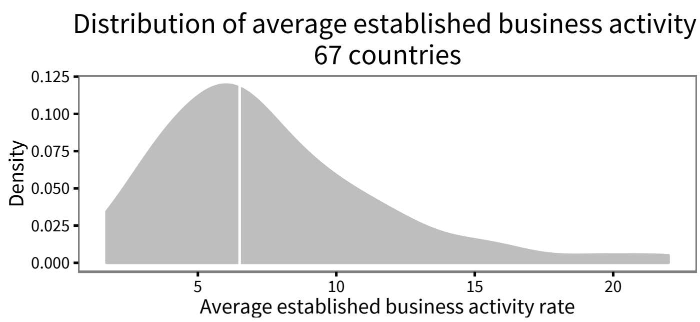


FIGURE 3.F.1. Distribution of average established business activity rates (white line is median)

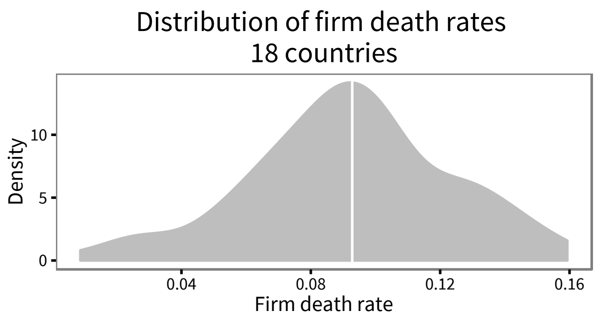


FIGURE 3.F.2. Distribution of firm death rates (white line is median)

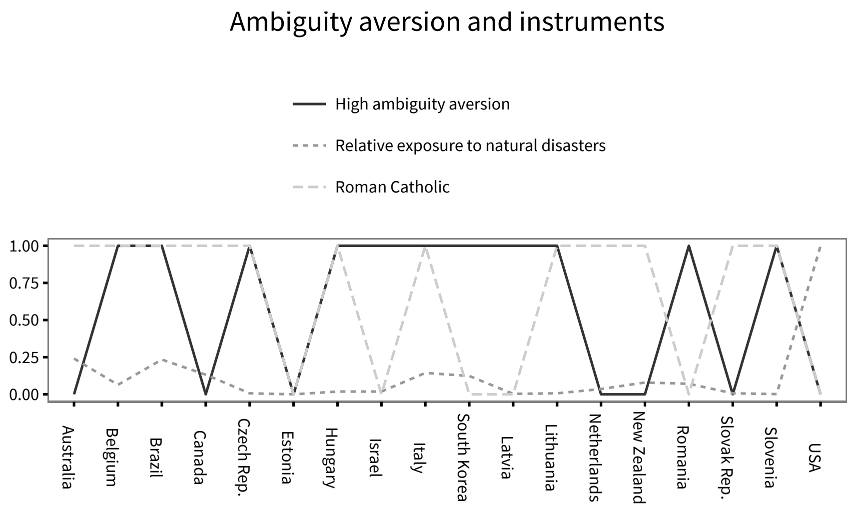


FIGURE 3.F.3. Instruments

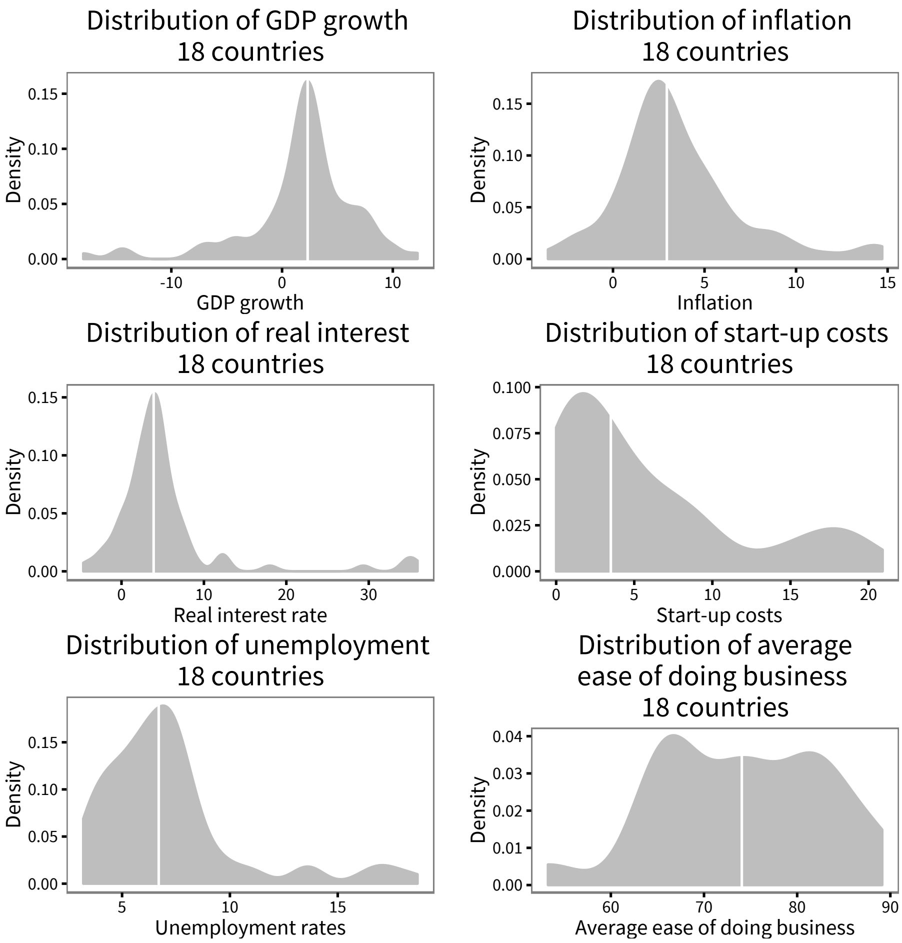


FIGURE 3.F.4. Distributions of GDP growth, inflation, real interest rates, start-up costs, unemployment rates, and average ease of doing business (white vertical lines represent medians)

Appendix 4.A

This appendix presents proofs of Lemma 4.1 and 4.2, and Proposition 4.1, 4.2, and 4.3.

4.A.1 *Choice probabilities*

Proof of Lemma 4.1. Note that $a_n = a_{\mathcal{S},n} + a_{\mathcal{W},n} = a_{\mathcal{S},0} + a_{\mathcal{W},0} + n$ such that a_n is deterministic. The binomial distribution can be constructed out of N i.i.d. draws from a Bernoulli distribution. Put differently, if

$$q = \sum_{n=1}^{N} q_n^*$$

where $q_1^*, \ldots, q_N^*$ are i.i.d. draws from Bernoulli(ϕ), then $q \sim$ Binomial(N, ϕ). Consider an arbitrary period. Assume that we selected and observed self-employment. The payoff is $\omega_{\mathcal{S}} \in \{0, 1\}$ and has distribution Bernoulli($\phi_{\mathcal{S}}$). The parameter $a_{\mathcal{S}}$ will be updated by adding $\omega_{\mathcal{S}}$, while the parameter $a_{\mathcal{W}}$ will be updated by adding $1 - \omega_{\mathcal{S}}$. Assume that we selected and observed wage work. The payoff is $\omega_{\mathcal{W}} \in \{0, 1\}$ with distribution Bernoulli($\phi_{\mathcal{W}}$). $a_{\mathcal{S}}$ is updated by adding $1 - \omega_{\mathcal{W}}$, whereas $a_{\mathcal{W}}$ is updated by adding $\omega_{\mathcal{W}}$. Let $\omega^* = 1 - \omega_{\mathcal{S}}$. It can be shown that $\omega^* \sim$ Bernoulli($\phi_{\mathcal{W}}$). The moment generating function of $\omega_{\mathcal{W}}$ is $M_{\omega_{\mathcal{W}}}(t) = \exp(t)\phi_{\mathcal{W}} + (1 - \phi_{\mathcal{W}})$. Consider the moment generating function of ω^*:

$$\begin{aligned} M_{\omega^*}(t) &= \mathbb{E}\left[\exp(t\omega^*)\right] = \exp(t)\mathbb{E}\left[\exp(-t\omega_{\mathcal{S}})\right] \\ &= \phi_{\mathcal{S}} + \exp(t)(1 - \phi_{\mathcal{S}}) = \exp(t)\phi_{\mathcal{W}} + (1 - \phi_{\mathcal{W}}) \end{aligned}$$

Hence, $\omega_{\mathcal{W}}$ can be replaced by $1 - \omega_{\mathcal{S}}$, as both have the same distribution. Furthermore:

$$\sum_{n=1}^{N} \omega_n^* = \sum_{n=1}^{N} (1 - \omega_{\mathcal{S},n}) \sim \text{Binomial}(N, \phi_{\mathcal{W}})$$

Now, let $q_{\mathcal{S},n} \sim$ Binomial($n, \phi_{\mathcal{S}}$) such that $a_{\mathcal{S},n} = a_{\mathcal{S},0} + q_{\mathcal{S},n}$ and $a_{\mathcal{W},n} = a_{\mathcal{W},0} + n - q_{\mathcal{S},n}$. Rewrite $\mathfrak{u}_n$ and $\mathfrak{b}_n$ as follows:

$$\mathfrak{u}_n(q_{\mathcal{S},n}) = \gamma_n^* + \delta_n^* q_{\mathcal{S},n} \tag{4.A.1}$$

$$\gamma_n^* \equiv -\frac{n + a_{\mathcal{W},0} - a_{\mathcal{S},0}}{a_n}, \quad \delta_n^* \equiv \frac{2}{a_n} > 0$$

$$\mathfrak{b}_n(q_{\mathcal{S},n}) = \gamma_n + \delta_n q_{\mathcal{S},n} \tag{4.A.2}$$

$$\gamma_n \equiv \frac{a_{\mathcal{S},0}\theta_{\mathcal{W}} - (n + a_{\mathcal{W},0})\theta_{\mathcal{S}} - (a_n^2 + a_n)n - (a_n^2 + a_n)a_{\mathcal{W},0} + a_{\mathcal{S},0}a_n^2 + a_{\mathcal{S},0}a_n}{a_n^2(a_n + 1)}$$

$$\delta_n \equiv \frac{\theta_{\mathcal{W}} + \theta_{\mathcal{S}} + 2a_n(a_n + 1)}{a_n^2(a_n + 1)} > 0$$

$\mathfrak{u}_n$ and $\mathfrak{b}_n$ are both strictly increasing in $q_{\mathcal{S},n}$, as $\mathfrak{u}_n'(q_{\mathcal{S},n}) = \delta_n^* > 0$ and $\mathfrak{b}_n'(q_{\mathcal{S},n}) = \delta_n > 0$, and invertible. We are interested in the probabilities to select wage work given by $\mathbb{P}(\mathfrak{u}_n \leqslant 0)$ and $\mathbb{P}(\mathfrak{b}_n \leqslant 0)$. Let $H(x; n, \phi)$ denote the cumulative distribution function of the binomial distribution given parameters n and ϕ. Using the properties $\mathfrak{u}_n$ and $\mathfrak{b}_n$, it is easy to establish that

$$\mathbb{P}(\mathfrak{u}_n \leqslant 0) = H(\tau_n^{\mathfrak{u}}; n, \phi_{\mathcal{S}}), \quad \tau_n^{\mathfrak{u}} \equiv \mathfrak{u}_n^{-1}(0) = -\frac{\gamma_n^*}{\delta_n^*} \tag{4.A.3}$$

and

$$\mathbb{P}(\mathfrak{b}_n \leqslant 0) = H(\tau_n^{\mathfrak{b}}; n, \phi_{\mathcal{S}}), \quad \tau_n^{\mathfrak{b}} \equiv \mathfrak{b}_n^{-1}(0) = -\frac{\gamma_n}{\delta_n} \tag{4.A.4}$$

■

4.A.2 Optimality of unbiased utility

Proof of Lemma 4.2. We can derive the optimal strategy by slightly modifying the argument of Bradt et al. (1956). Let $\mathrm{s}^* = \{\mathfrak{d}_n^*\}_{n=1}^N$ where $\mathfrak{d}_n^* \in \mathbb{O}$ denotes the (unknown) optimal strategy. Furthermore, let

$$\mathcal{V}_N(a_{\mathcal{S},0}, a_{\mathcal{W},0}, \mathrm{s}) = \sum_{n=1}^{N} \mathbb{E}\left[\Omega(\mathfrak{d}_n) | a_{\mathcal{S},0}, a_{\mathcal{W},0}\right]$$

denote the expected payoff associated with some strategy s. Consider an arbitrary period $k > 0$. Let s_{N-k}^* denote the optimal strategy for the remaining $N-k$ periods and let $\mathcal{V}_{N-k}$ denote the corresponding expected payoff. Assume that we consider selecting self-employment in period $k+1$. The expected payoff in period $k+1$ is $\mu_{\mathcal{S},k}$. In case of a success, which occurs with probability $\mu_{\mathcal{S},k}$, $a_{\mathcal{S},k}$ is updated to $a_{\mathcal{S},k}+1$, while $a_{\mathcal{W},k}$ remains the same. In case of a failure in self-employment, which occurs with probability $1-\mu_{\mathcal{S},k}$, $a_{\mathcal{S},k}$ remains the same, whereas $a_{\mathcal{W},k}$ is updated to $a_{\mathcal{W},k}+1$. Hence, the expected payoff from selecting self-employment in period $k+1$ is

$$\begin{aligned}\mathcal{V}_{\mathcal{S}} \equiv \mu_{\mathcal{S},k} + \mu_{\mathcal{S},k}\mathcal{V}_{N-k}(a_{\mathcal{S},k}+1, a_{\mathcal{W},k}, \mathrm{s}_{N-k}^*) \\ +(1-\mu_{\mathcal{S},k})\mathcal{V}_{N-k}(a_{\mathcal{S},k}, a_{\mathcal{W},k}+1, \mathrm{s}_{N-k}^*)\end{aligned} \tag{4.A.5}$$

Given a similar line of reasoning, the expected payoff from selecting wage work in period $k+1$ is

$$\mathcal{V}_{\mathcal{W}} \equiv \mu_{\mathcal{W},k} + \mu_{\mathcal{W},k}\mathcal{V}_{N-k}(a_{\mathcal{W},k}+1, a_{\mathcal{S},k}, \mathrm{s}^*_{N-k}) \qquad (4.A.6)$$
$$+(1-\mu_{\mathcal{W},k})\mathcal{V}_{N-k}(a_{\mathcal{W},k}, a_{\mathcal{S},k}+1, \mathrm{s}^*_{N-k})$$

We should select self-employment if $\mathcal{V}_{\mathcal{S}}$ is strictly larger than $\mathcal{V}_{\mathcal{W}}$; be indifferent if $\mathcal{V}_{\mathcal{S}}$ and $\mathcal{V}_{\mathcal{W}}$ are equal; and select wage work if $\mathcal{V}_{\mathcal{W}}$ is strictly larger than $\mathcal{V}_{\mathcal{S}}$. Note that this holds for an arbitrary period and is, therefore, a general prescription. Furthermore, notice that $1-\mu_{\mathcal{S},k} = \mu_{\mathcal{W},k}$ and $1-\mu_{\mathcal{W},k} = \mu_{\mathcal{S},k}$. Hence: We should strictly prefer self-employment if $\mu_{\mathcal{S},k} > \mu_{\mathcal{W},k}$; be indifferent if $\mu_{\mathcal{S},k} = \mu_{\mathcal{W},k}$; and strictly prefer wage work if $\mu_{\mathcal{W},k} > \mu_{\mathcal{S},k}$. This prescription is equivalent to the prescription made by relative unbiased utility. ■

4.A.3 Non-optimality of unbiased utility

Proof of Proposition 4.1. Note that the distribution function of the binomial is

$$H(\tau; n, \phi) = \sum_{k=0}^{\lfloor\tau\rfloor} \binom{n}{k} \phi^k (1-\phi)^{n-k}$$

where $\lfloor\tau\rfloor$ is the greatest integer less than or equal to τ. Hence, a sufficiently large increase (decrease) in τ increases (decreases) H, by the properties of distribution functions. To assess a potential bias, use Lemma 4.1 to establish:

$$\tau_n^{\mathfrak{u}} - \tau_n^{\mathfrak{b}} = \begin{cases} > 0 & \text{if } \theta_{\mathcal{S}} < \theta_{\mathcal{W}} \\ = 0 & \text{if } \theta_{\mathcal{S}} = \theta_{\mathcal{W}} \\ < 0 & \text{if } \theta_{\mathcal{S}} > \theta_{\mathcal{W}} \end{cases} \quad \text{for all } n \qquad (4.A.7)$$

According to Lemma 4.2, $\tau_n^{\mathfrak{u}}$ induces optimal behavior such that deviations from it constitute a bias reducing career successes. Case 1: If $\theta_{\mathcal{S}} < \theta_{\mathcal{W}}$, $\tau_n^{\mathfrak{u}} > \tau_n^{\mathfrak{b}}$ and biased utility has a potential bias against wage work, as $\mathbb{P}(\mathfrak{u}_n \leqslant 0) \geqslant \mathbb{P}(\mathfrak{b}_n \leqslant 0)$, respectively a potential bias for self-employment, as $1-\mathbb{P}(\mathfrak{u}_n \leqslant 0) \leqslant 1-\mathbb{P}(\mathfrak{b}_n \leqslant 0)$. Case 2: If $\theta_{\mathcal{S}} = \theta_{\mathcal{W}}$, there is no bias. Case 3: If $\theta_{\mathcal{S}} > \theta_{\mathcal{W}}$, $\tau_n^{\mathfrak{u}} < \tau_n^{\mathfrak{b}}$ and biased utility has a potential bias for wage work, as $\mathbb{P}(\mathfrak{u}_n \leqslant 0) \leqslant \mathbb{P}(\mathfrak{b}_n \leqslant 0)$, respectively a potential bias against self-employment, as $1-\mathbb{P}(\mathfrak{u}_n \leqslant 0) \geqslant 1-\mathbb{P}(\mathfrak{b}_n \leqslant 0)$. Note that

$$\frac{\partial}{\partial\theta_{\mathcal{S}}}(\tau_n^{\mathfrak{u}} - \tau_n^{\mathfrak{b}}) < 0, \quad \frac{\partial}{\partial\theta_{\mathcal{W}}}(\tau_n^{\mathfrak{u}} - \tau_n^{\mathfrak{b}}) > 0$$

such that at some point (given a large enough $\theta_{\mathcal{S}}$ or $\theta_{\mathcal{W}}$), we have $\mathbb{P}(\mathfrak{u}_n \leqslant 0) > \mathbb{P}(\mathfrak{b}_n \leqslant 0)$ or $\mathbb{P}(\mathfrak{u}_n \leqslant 0) < \mathbb{P}(\mathfrak{b}_n \leqslant 0)$, i.e., the bias is relevant if either $\theta_{\mathcal{S}}$ or $\theta_{\mathcal{W}}$ is sufficiently large. ■

4.A.4 *Media and behavior*

Proof of Proposition 4.2. Denote $\tau_n^{\mathfrak{u}}$ affected by media $\mathfrak{m} > 0$ by $\mathring{\tau}_n^{\mathfrak{u}}$. The impact of media is given by

$$\frac{\partial}{\partial \mathfrak{m}} \mathring{\tau}_n^{\mathfrak{u}} = -\frac{1}{2} < 0 \tag{4.A.8}$$

Let $\mathring{\tau}_n^{\mathfrak{b}}$ denote $\tau_n^{\mathfrak{b}}$ given that self-employment is affected by $\mathfrak{m} > 0$. It follows that

$$\frac{\partial}{\partial \mathfrak{m}} \mathring{\tau}_n^{\mathfrak{b}} = B B_0^{-1} \tag{4.A.9}$$

$$\begin{aligned}
B \equiv{} & -(\theta_{\mathcal{W}}^2 + (\theta_{\mathcal{S}} + \mathfrak{m}^2 + (2a_{\mathcal{W},0} + 2a_{\mathcal{S},0} + 2 - 2n)\mathfrak{m} - (2a_{\mathcal{W},0} + 2a_{\mathcal{S},0} + 1)n + a_{\mathcal{W},0}^2 + (2a_{\mathcal{S},0} + 2)a_{\mathcal{W},0} + a_{\mathcal{S},0}^2 \\
& + 2a_{\mathcal{S},0})\theta_{\mathcal{W}} + (3\mathfrak{m}^2 + (2n + 6a_{\mathcal{W},0} + 6a_{\mathcal{S},0} + 2)\mathfrak{m} + (2a_{\mathcal{W},0} + 2a_{\mathcal{S},0} + 1)n + 3a_{\mathcal{W},0}^2 + (6a_{\mathcal{S},0} + 2)a_{\mathcal{W},0} \\
& + 3a_{\mathcal{S},0}^2 + 2a_{\mathcal{S},0})\theta_{\mathcal{S}} + 2\mathfrak{m}^4 + (8a_{\mathcal{W},0} + 8a_{\mathcal{S},0} + 4)\mathfrak{m}^3 + (12a_{\mathcal{W},0}^2 + (24a_{\mathcal{S},0} + 12)a_{\mathcal{W},0} + 12a_{\mathcal{S},0}^2 \\
& + 12a_{\mathcal{S},0} + 2)\mathfrak{m}^2 + (8a_{\mathcal{W},0}^3 + (24a_{\mathcal{S},0} + 12)a_{\mathcal{W},0}^2 + (24a_{\mathcal{S},0}^2 + 24a_{\mathcal{S},0} + 4)a_{\mathcal{W},0} + 8a_{\mathcal{S},0}^3 + 12a_{\mathcal{S},0}^2 \\
& + 4a_{\mathcal{S},0})\mathfrak{m} + 2a_{\mathcal{W},0}^4 + (8a_{\mathcal{S},0} + 4)a_{\mathcal{W},0}^3 + (12a_{\mathcal{S},0}^2 + 12a_{\mathcal{S},0} + 2)a_{\mathcal{W},0}^2 + (8a_{\mathcal{S},0}^3 + 12a_{\mathcal{S},0}^2 \\
& + 4a_{\mathcal{S},0})a_{\mathcal{W},0} + 2a_{\mathcal{S},0}^4 + 4a_{\mathcal{S},0}^3 + 2a_{\mathcal{S},0}^2)
\end{aligned}$$

where $B_0 > 0$ but the sign of B is ambiguous. However, it is easy to show that $B < 0$ if $\theta_{\mathcal{S}} > \mathring{\theta}_{\mathcal{S},n}$ where

$$\begin{aligned}
\mathring{\theta}_{\mathcal{S},n} \equiv{} & (-\theta_{\mathcal{W}}^2 - ([-2\mathfrak{m} - 2a_{\mathcal{W},0} - 2a_{\mathcal{S},0} - 1]n + \mathfrak{m}^2 + 2(a_{\mathcal{W},0} + a_{\mathcal{S},0} + 1)\mathfrak{m} + a_{\mathcal{W},0}^2 + 2(a_{\mathcal{S},0} + 1)a_{\mathcal{W},0} \\
& + a_{\mathcal{S},0}^2 + 2a_{\mathcal{S},0})\theta_{\mathcal{W}} - 2\mathfrak{m}^4 - (8a_{\mathcal{W},0} + 8a_{\mathcal{S},0} + 4)\mathfrak{m}^3 - (12a_{\mathcal{W},0}^2 + (24a_{\mathcal{S},0} + 12)a_{\mathcal{W},0} + 12a_{\mathcal{S},0}^2 + 12a_{\mathcal{S},0} + 2)\mathfrak{m}^2 \\
& - (8a_{\mathcal{W},0}^3 + (24a_{\mathcal{S},0} + 12)a_{\mathcal{W},0}^2 + (24a_{\mathcal{S},0}^2 + 24a_{\mathcal{S},0} + 4)a_{\mathcal{W},0} + 8a_{\mathcal{S},0}^3 + 12a_{\mathcal{S},0}^2 + 4a_{\mathcal{S},0})\mathfrak{m} - 2a_{\mathcal{W},0}^4 \\
& - (8a_{\mathcal{S},0} + 4)a_{\mathcal{W},0}^3 - (12a_{\mathcal{S},0}^2 + 12a_{\mathcal{S},0} + 2)a_{\mathcal{W},0}^2 - (8a_{\mathcal{S},0}^3 + 12a_{\mathcal{S},0}^2 + 4a_{\mathcal{S},0})a_{\mathcal{W},0} - 2a_{\mathcal{S},0}^4 - 4a_{\mathcal{S},0}^3 - 2a_{\mathcal{S},0}^2) \\
& [\theta_{\mathcal{W}} + (2\mathfrak{m} + 2a_{\mathcal{W},0} + 2a_{\mathcal{S},0} + 1)n + 3\mathfrak{m}^2 + (6a_{\mathcal{W},0} + 6a_{\mathcal{S},0} + 2)\mathfrak{m} + 3a_{\mathcal{W},0}^2 + (6a_{\mathcal{S},0} + 2)a_{\mathcal{W},0} + 3a_{\mathcal{S},0}^2 + 2a_{\mathcal{S},0}]^{-1}
\end{aligned}$$

It is straightforward to show that $\mathring{\theta}_{\mathcal{S},n}$ is strictly increasing in n. Moreover, it is easy to demonstrate that $\lim_{n\to\infty} \mathring{\theta}_{\mathcal{S},n} = \theta_{\mathcal{W}}$ such that

$$\mathring{\theta}_{\mathcal{S},1} \leqslant \mathring{\theta}_{\mathcal{S},n} < \theta_{\mathcal{W}}$$

Hence, if $\theta_{\mathcal{S}} \geqslant \theta_{\mathcal{W}}$, we have $B < 0$ and, consequently:

$$\frac{\partial}{\partial \mathfrak{m}} \mathring{\tau}_n^{\mathfrak{b}} < 0 \tag{4.A.10}$$

Now, given sufficiently intensive media, we must have

$$H(\mathring{\tau}_n^{\mathfrak{u}}; n, \phi_{\mathcal{S}}) < H(\tau_n^{\mathfrak{u}}; n, \phi_{\mathcal{S}}) \tag{4.A.11}$$

and

$$H(\mathring{\tau}_n^{\mathfrak{b}}; n, \phi_{\mathcal{S}}) < H(\tau_n^{\mathfrak{b}}; n, \phi_{\mathcal{S}}) \tag{4.A.12}$$

if $\theta_{\mathcal{S}} \geqslant \theta_{\mathcal{W}}$. Using Lemma 4.1, (4.A.11) and (4.A.12) imply

$$\mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{u}} = \mathcal{S}) > \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{S}), \qquad \mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{u}} = \mathcal{W}) < \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{W}) \tag{4.A.13}$$

$$\mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}} = \mathcal{S}) > \mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{S}), \qquad \mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}} = \mathcal{W}) < \mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{W}) \tag{4.A.14}$$

where $\mathring{\mathfrak{d}}_n$ is a choice affected by media and $\mathfrak{d}_n$ a choice without the influence of media. ∎

4.A.5 Media and bias against self-employment

Proof of Proposition 4.3. Using Proposition 4.1 and given that $\theta_{\mathcal{S}}$ is sufficiently larger than $\theta_{\mathcal{W}}$, we have

$$H(\tau_n^{\mathfrak{b}}; n, \phi_{\mathcal{S}}) > H(\tau_n^{\mathfrak{u}}; n, \phi_{\mathcal{S}})$$

As $\mathring{\tau}_n^{\mathfrak{b}}$ is strictly decreasing in $\mathfrak{m}$ if $\theta_{\mathcal{S}} > \theta_{\mathcal{W}}$ and $\lim_{\mathfrak{m}\to\infty} \mathring{\tau}_n^{\mathfrak{b}} = -\infty$, there exists only one $\mathfrak{m}_n^*$ solving

$$\mathring{\tau}_n^{\mathfrak{b}}(\mathfrak{m}_n^*) = \tau_n^{\mathfrak{u}}$$

Hence, for every n there exists an $\mathfrak{m}_n \in (0, \mathfrak{m}_n^*]$ such that

$$H(\tau_n^{\mathfrak{u}}; n, \phi_{\mathcal{S}}) \leqslant H(\mathring{\tau}_n^{\mathfrak{b}}(\mathfrak{m}_n); n, \phi_{\mathcal{S}}) < H(\tau_n^{\mathfrak{b}}; n, \phi_{\mathcal{S}})$$

Put differently, for every n there always exists an $\mathfrak{m}_n \in (0, \mathfrak{m}_n^*]$ such that the bias against self-employment and for wage work is reduced:

$$\mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{S}) < \mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}}(\mathfrak{m}_n) = \mathcal{S}) \leqslant \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{S})$$

$$\mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{W}) \leqslant \mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}}(\mathfrak{m}_n) = \mathcal{W}) < \mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{W})$$

Therefore, if $\mathfrak{m} \in (0, \mathfrak{m}^{**}]$ where $\mathfrak{m}^{**} \equiv \min\{\mathfrak{m}_1^*, \mathfrak{m}_2^*, \dots, \mathfrak{m}_N^*\}$, there is at least one period n^* where $\mathbb{P}(\mathring{\mathfrak{d}}_{n^*}^{\mathfrak{b}} = \mathcal{S}) = \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{S})$ and $\mathbb{P}(\mathring{\mathfrak{d}}_{n^*}^{\mathfrak{b}} = \mathcal{W}) = \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{W})$, while for all the remaining periods $\mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{S}) \leqslant \mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}} = \mathcal{S}) \leqslant \mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{S})$ and $\mathbb{P}(\mathfrak{d}_n^{\mathfrak{u}} = \mathcal{W}) \leqslant \mathbb{P}(\mathring{\mathfrak{d}}_n^{\mathfrak{b}} =$

$\mathcal{W}) \leqslant \mathbb{P}(\mathfrak{d}_n^{\mathfrak{b}} = \mathcal{W})$. Note, however, that media might also be too intensive such that $\mathfrak{m}_n > \mathfrak{m}_n^*$ resulting in $H(\mathring{\tau}_n^{\mathfrak{b}}(\mathfrak{m}_n); n, \phi_S) < H(\tau_n^{\mathfrak{u}}; n, \phi_S)$, i.e., a bias against self-employment might be transformed into a bias *for* self-employment. ■

Appendix 4.B

In this appendix, we describe our data.

4.B.1 Characteristics of micro panel

Table 4.B.1 presents variable descriptions.

TABLE 4.B.1. Variables in micro panel

Variable (source if not IHIS)	Description	Values
Consumption of articles about famous entrepreneurs (Google Trends)	Relative frequency of the search item "famous entrepreneurs" in the US regions	100 for state with highest frequency (SHF); all other values relative to SHF; numbers aggregated over time and normalized by 1,000
Number of disasters (Guha-Sapir et al. 2016)	Number of natural and complex disasters	Numerical
Self-employed	Individual is self-employed	1 = Self-employed; 0 = Not self-employed
Wage worker	Individual is a wage worker	1 = Worker; 0 = Not worker
Age	Individual's age	Numerical
Earnings	Total earnings during the previous calendar year	1 = \$01 to \$4999; 2 = \$5000 to \$9999; 3 = \$10000 to \$14999; 4 = \$15000 to \$19999; 5 = \$20000 to \$24999; 6 = \$25000 to \$34999; 7 = \$35000 to \$44999; 8 = \$45000 to \$54999; 9 = \$55000 to \$64999; 10 = \$65000 to \$74999; 11 = \$75000 and over

Education	Educational attainment	1 = Never attended/kindergarten only; 2 = Grade 1; 3 = Grade 2; 4 = Grade 3; 5 = Grade 4; 6 = Grade 5; 7 = Grade 6; 8 = Grade 7; 9 = Grade 8; 10 = Grade 9; 11 = Grade 10; 12 = Grade 11; 13 = 12th grade, no diploma; 14 = High school graduate; 15 = GED or equivalent; 16 = Some college, no degree; 17 = AA degree: technical/vocational/occupational; 18 = AA degree: academic program; 19 = Bachelor's degree (BA, AB, BS, BBA); 20 = Master's degree (MA, MS, Med, MBA); 21 = Professional (MD, DDS, DVM, JD); 22 = Doctoral degree (PhD, EdD)
Years on job	Years on main or longest or last job	0 = Less than a year; 1, 2, 3, . . . = Numerical value for number of years
Health	Health status	1 = Excellent; 2 = Very Good; 3 = Good; 4 = Fair; 5 = Poor
Effort	Felt everything an effort, past 30 days	0 = None of the time; 1 = A little of the time; 2 = Some of the time; 3 = Most of the time; 4 = All of the time
Feelings interfered with life	Feelings interfered with life, past 30 days	1 = A lot; 2 = Some; 3 = A little; 4 = Not at all
Hopeless	How often felt hopeless, past 30 days	0 = None of the time; 1 = A little of the time; 2 = Some of the time; 3 = Most of the time; 4 = All of the time
Nervous	How often felt nervous, past 30 days	0 = None of the time; 1 = A little of the time; 2 = Some of the time; 3 = Most of the time; 4 = All of the time
Restless	How often felt restless, past 30 days	0 = None of the time; 1 = A little of the time; 2 = Some of the time; 3 = Most of the time; 4 = All of the time
Sad	How often felt sad, past 30 days	0 = None of the time; 1 = A little of the time; 2 = Some of the time; 3 = Most of the time; 4 = All of the time
Worthless	How often felt worthless, past 30 days	0 = None of the time; 1 = A little of the time; 2 = Some of the time; 3 = Most of the time; 4 = All of the time

Born in US	Born in the United States	1 = Born in US; 0 = Not born in US
Female	Gender	1 = Female; 0 = Male
Non-white	Ethnicity	1 = Non-white; 0 = White
Full-time work	Usually work full time	1 = Usually full time; 0 = Usually not full time
Limitations	Has any activity limitation	1 = Limited in any way; 0 = Not limited in any way
Got dividends	Received income from dividends from stocks/funds, previous calendar year	1 = Yes; 0 = No
Got food stamps	Authorized to receive Food Stamps, last calendar year	1 = Yes; 0 = No
Got welfare	Received income from welfare/public assistance, previous calendar year	1 = Yes; 0 = No

In Table 4.B.2, we show descriptive statistics. Correlations are provided in Figure 4.B.1. Figure 4.B.2 shows the shares of self-employed and wage workers conditional on regions and time.

4.B.2 Characteristics of macro panel

In Table 4.B.3, we present descriptive statistics.

In 48% of all countries, doing business is relatively easy. Figure 4.B.3 shows the variation of country averages of the shares of self-employed and wage workers. Table 4.B.4 shows correlations.

There is a strong negative correlation (–1.0) between the share of self-employed and the share of wage workers; i.e., it appears that most self-employed recruit themselves from the wage workers' group. Furthermore, less corruption is strongly positively correlated (0.8) with relative ease of doing business, i.e., doing business is easier in less corrupt societies—it might also be one reason for lower levels of corruption. Ease of doing business is negatively correlated with the share of self-employed (–0.6), but positively correlated with the share of wage workers (0.7).

Figure 4.B.4 shows distributions of the dependent variables (original and transformed

TABLE 4.B.2. Descriptive statistics for micro panel
(a) Non-binary variables

Variable	Min	1st quartile	Median	Mean	3rd quartile	Max	SD
Consumption of articles about famous entrepreneurs	0.14	0.42	0.91	1.12	1.81	3.39	0.80
Number of disasters (countries and regions)	3.00	6.00	10.00	14.80	24.00	38.00	10.99
Age	18.00	25.00	36.00	37.48	48.00	64.00	13.34
Earnings	1.00	2.00	3.00	3.84	5.00	11.00	2.51
Education	1.00	14.00	16.00	15.65	18.00	22.00	3.06
Years on job	0.00	0.00	2.00	4.92	6.00	35.00	6.90
Health	1.00	2.00	2.00	2.40	3.00	5.00	1.04
Effort	0.00	0.00	1.00	1.36	2.00	4.00	1.27
Feelings interfered with life	1.00	2.00	3.00	2.96	4.00	4.00	0.98
Hopeless	0.00	0.00	0.00	0.63	1.00	4.00	0.99
Nervous	0.00	0.00	2.00	1.48	2.00	4.00	1.09
Restless	0.00	0.00	2.00	1.58	2.00	4.00	1.19
Sad	0.00	0.00	1.00	1.01	2.00	4.00	1.07
Worthless	0.00	0.00	0.00	0.48	1.00	4.00	0.9

(b) Binary variables

Variable	Share of individuals with characteristic
Self-employed	0.12
Wage worker	0.88
Born in US	0.85
Female	0.67
Non-white	0.24
Full-time work	0.26
Limitations	0.16
Got dividends	0.09
Got food stamps	0.07
Got welfare	0.02

by the logit transformation) in our data set. Note that the shares of wage workers and self-employed obey an asymmetric distribution and that skewness is not completely removed by the logit transformation (a reason to consider beta regression models).

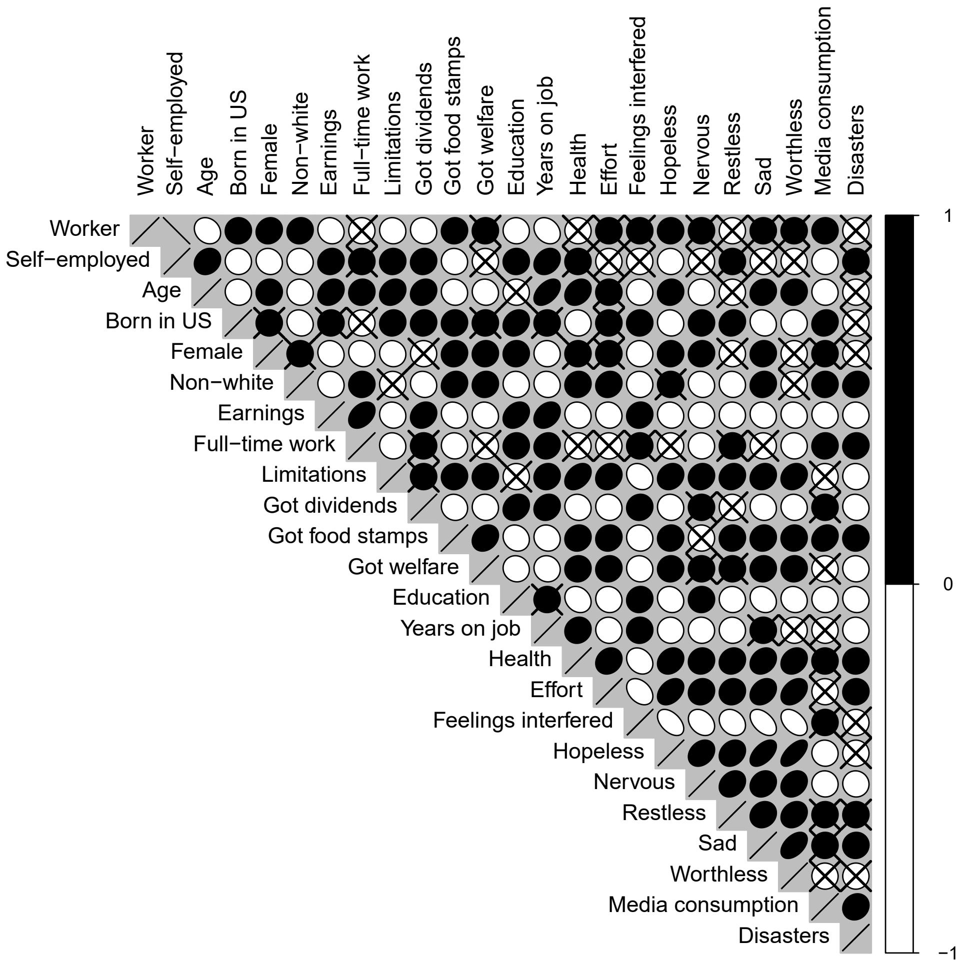


FIGURE 4.B.1. Correlations in micro panel: Crossed out correlations are not significant at the 5%-level, while ellipses indicate strength (diagonal line is perfect correlation, whereas a perfect circle is no correlation) and direction (black is positive and white is negative correlation)

4.B.3 *Stability of ease of doing business*

In our empirical macro model, we assume that ease of doing business is sufficiently stable such that if doing business was relatively easy in 2010–2013, it was also relatively easy in 2003–2009. We, now, examine the stability of relative ease of doing business. We consider only countries where ease of doing business could be observed in all periods 2010–2013. Since the only variable of interest is ease of doing business, we do not have to ensure that all other variables are observed in the same period, meaning that data from a large number of countries (66 countries) is available. For each period $n = 2010, \dots, 2013$, we construct a group, denoted by $\mathcal{E}_n$, consisting of all countries where doing business was relatively easy

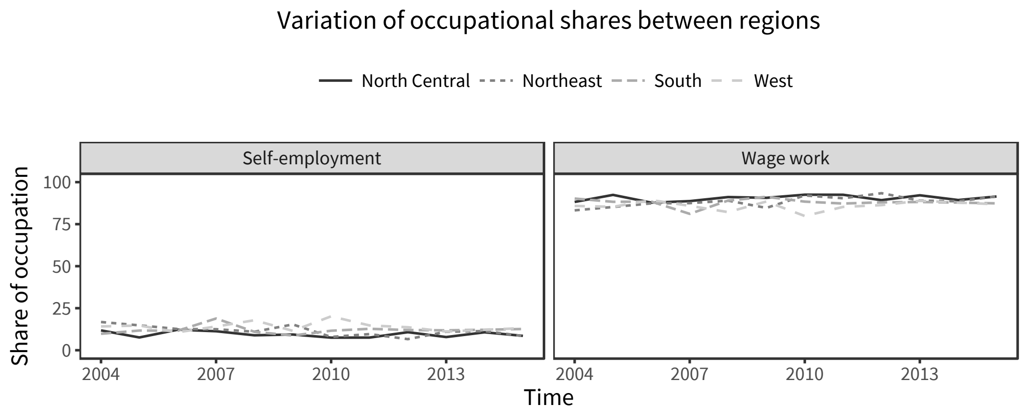


FIGURE 4.B.2. Occupational shares

TABLE 4.B.3. Descriptive statistics for macro panel

Variable	Min	1st quartile	Median	Mean	3rd quartile	Max	SD
Media	19.00	49.00	57.00	57.64	67.00	88.00	14.64
Number of disasters (countries)	1.00	1.25	3.00	4.33	5.00	29.00	4.72
Share of self-employed	6.32	12.29	16.53	20.94	26.89	55.36	11.77
Share of wage workers	39.44	63.60	77.72	71.52	81.04	89.70	12.67
Fear of failure	0.15	0.28	0.32	0.33	0.36	0.58	0.07
Inflation	−6.01	1.88	3.21	4.53	5.52	34.93	5.34
GDP (per capita)	6187	15301	22695	25695	34551	77173	13152.52
GDP growth	−17.96	1.00	3.15	2.76	5.17	10.60	3.94
Real interest rate	−10.89	1.74	3.62	5.39	6.56	46.92	8.93
Lack of corruption	2.10	3.73	5.70	5.86	7.60	9.60	2.18

in period n based on k-means clustering with two clusters. (In all countries not part of $\mathcal{E}_n$ doing business was relatively difficult.)

In Table 4.B.5, we provide two measures of stability. First, the intersection with the previous period is defined as

$$\text{Intersection with previous period}_n \equiv \frac{|\mathcal{E}_{n-1} \cap \mathcal{E}_n|}{|\mathcal{E}_n|} \in [0, 1]$$

and captures the number of countries where doing business was easy in period n and $n-1$

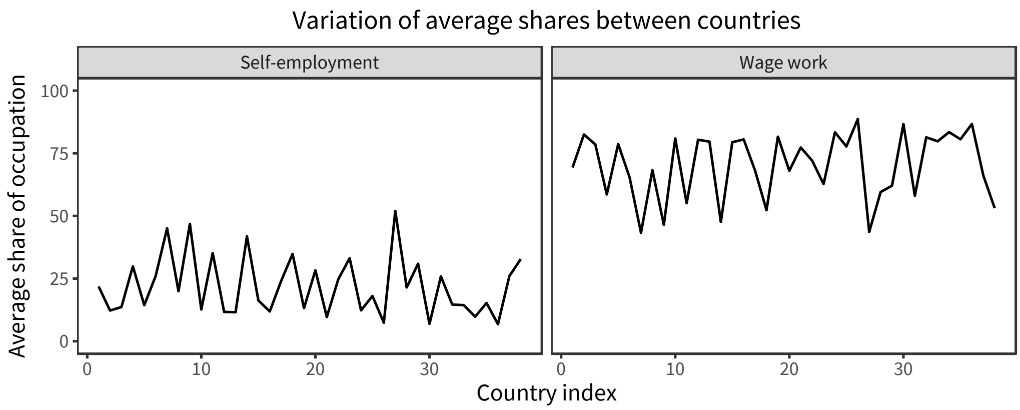


FIGURE 4.B.3. Differences between countries in shares of self-employed and wage workers

TABLE 4.B.4. Correlations in macro panel

	M	FOF	I	GDP	GDPG	RIR	COR	WWS	SES	DBE
FOF	−0.1									
I	0.1	0.1								
GDP	−0.1	−0.0	−0.4**							
GDPG	0.3**	−0.1	0.2**	−0.2**						
RIR	0.2**	0.1	−0.2**	−0.3**	−0.0					
COR	−0.0	−0.2**	−0.5**	0.8**	−0.1	−0.2**				
WWS	−0.3**	0.0	−0.3**	0.7**	−0.2**	−0.4**	0.7**			
SES	0.4**	−0.0	0.3**	−0.7**	0.3**	0.3**	−0.6**	−1.0**		
DBE	0.1	−0.0	−0.4**	0.7**	−0.1	−0.2**	0.8**	0.7**	−0.6**	
DIS	−0.5**	−0.1	−0.0	−0.1	−0.1	0.1	−0.0	0.1	−0.1	−0.2**

Notes: **Correlation is significant at the 5%-level; M = media reports intensity (GEM-based variable); FOF = fear of entrepreneurial failure; I = inflation; GDP = GDP per capita; GDPG = GDP growth; RIR = real interest rate; COR = lack of corruption; WWS = share of wage workers; SES = share of self-employed; DBE = doing business is relatively easy (dummy is 1 if yes); DIS = number of natural disasters in other country

TABLE 4.B.5. Stability of relative ease of doing business

Time	Number of countries where doing business is relatively easy	Intersection with previous period	Intersection with first period
2010	29		100%
2011	31	94%	94%
2012	32	97%	91%
2013	35	91%	83%

relative to the number of countries where doing business was easy in period *n*. Second,

FIGURE 4.B.4. Distributions of dependent variables in macro panel

the intersection with the first period is given by

$$\text{Intersection with first period}_n \equiv \frac{|\mathcal{E}_{2010} \cap \mathcal{E}_n|}{|\mathcal{E}_n|} \in [0, 1]$$

and captures the number of countries where doing business was easy in 2010 and period n relative to the number of countries where doing business was easy in period n.

If relative ease of doing business is stable, we expect both measures to be high. We find that the intersection with the previous period is never below 91%. The intersection with the first period is never below 83%. In particular, in 83% of all countries where doing business was relatively easy in 2013, doing business was also relatively easy in 2010. Consequently, relative ease of doing business is acceptably stable for our purposes.

Appendix 4.C

This appendix provides additional results and further information.

4.C.1 Additional results for micro panel model

Table 4.C.1 presents results for a heteroskedastic probit without IV.

Table 4.C.1. Probit estimates of marginal effects in micro panel model *without* using IV

	Self-employment		Wage work	
Variable	Coefficient	SE	Coefficient	SE
Consumption of articles about famous entrepreneurs	0.155**	(0.068)	−0.155**	(0.068)
Age	0.196***	(0.037)	−0.196***	(0.037)
Age^2	−0.002***	(0.000)	0.002***	(0.000)
Female†	−2.254***	(0.608)	2.254***	(0.608)
Born in US†	−0.182	(0.109)	0.182	(0.109)
Non-white†	−0.267	(0.176)	0.267	(0.176)
Full-time work†	−0.433***	(0.117)	0.433***	(0.117)
Earnings	0.004	(0.015)	−0.004	(0.015)
Got dividends†	0.323**	(0.135)	−0.323**	(0.135)
Got food stamps†	−0.145	(0.169)	0.145	(0.169)
Got welfare†	0.081	(0.278)	−0.081	(0.278)
Education	0.013	(0.019)	−0.013	(0.019)
Years on job	0.082***	(0.017)	−0.082***	(0.017)
Limitations†	0.050	(0.100)	−0.050	(0.100)
Health‡	0.024	(0.065)	−0.024	(0.065)
Effort	0.030	(0.031)	−0.030	(0.031)
Feelings interfered with life	−0.050	(0.045)	0.050	(0.045)
Hopeless	−0.084	(0.052)	0.084	(0.052)
Nervous	0.013	(0.033)	−0.013	(0.033)
Restless	0.079**	(0.034)	−0.079**	(0.034)
Sad	−0.030	(0.042)	0.030	(0.042)
Worthless	−0.074	(0.054)	0.074	(0.054)
North Central†	−0.340	(0.195)	0.340	(0.195)
Northeast†	−0.034	(0.200)	0.034	(0.200)
West†	0.024	(0.184)	−0.024	(0.184)
Constant	−5.341***	(0.707)	5.341***	(0.707)
Variable: Variance model		Coefficient	SE	
North Central†		−0.032	(0.073)	
Northeast†		−0.054	(0.083)	
West†		0.086	(0.072)	
Age		0.014***	(0.003)	
Female†		0.626***	(0.114)	
Non-white†		−0.087	(0.067)	
Health‡		−0.056**	(0.026)	
Education		0.000	(0.001)	
		10,851 obs.		

Notes: † Dummy variable; ‡ increase indicates more health problems; *** significant at the 1%-level; ** significant at the 5%-level; standard errors in parentheses are heteroskedasticity-consistent.

4.C.2 Additional results for macro panel model

In Table 4.C.2, we present results for linear models without the use of an IV. Table 4.C.3

TABLE 4.C.2. Estimates of marginal effects on transformed shares in macro panel model *without* IV

	Self-employment		Wage work	
Variable	Coefficient	SE	Coefficient	SE
Media attention for entrepreneurship	0.015***[**]	(0.003)	−0.011***[**]	(0.003)
Doing business is relatively easy†	−1.349**	(0.611)	0.976**	(0.475)
Ambiguity aversion index	−0.008	(0.007)	0.007	(0.005)
Ambiguity aversion index × doing business is relatively easy	0.010	(0.007)	−0.007	(0.006)
Inflation	−0.014	(0.012)	0.011	(0.010)
GDP	0.000***	(0.000)	0.000***[***]	(0.000)
GDP growth	0.012	(0.011)	0.002	(0.009)
Real interest rate	0.002	(0.003)	−0.003	(0.003)
Lack of corruption‡	−0.028	(0.030)	0.057**	(0.023)
Constant	−0.841	(0.556)	0.071	(0.429)
		171 obs.		

Notes: †Dummy is 1 if yes and zero else; ‡increase indicates less corruption; ***significant at the 1%-level; **significant at the 5%-level; [***]significant at the 1%-level with country-level clustering; [**]significant at the 5%-level with country-level clustering; standard errors in parentheses are heteroskedasticity-consistent.

shows the assignment yielding the strongest instrument in the macro panel model. Table 4.C.4 presents estimation results of linear probability models. A linear probability model corresponds to replacing $\mathfrak{L}(p_{j,n,i})$ with $p_{j,n,i}$, such that the coefficient of media reflects the percentage-points effect on the probability of an occupation given that positive media attention for entrepreneurship increases by 1 percentage point.

Table 4.C.3. Assignment of countries resulting in strongest instrument

Country	Paired country with disasters	Country	Paired country with disasters
Argentina	Croatia	South Korea	Serbia
Australia	Latvia	Latvia	Hong Kong
Belgium	Greece	Malaysia	Jamaica
Brazil	Romania	Mexico	Venezuela
Canada	Trinidad and Tobago	Netherlands	Slovenia
Chile	Colombia	New Zealand	Pakistan
Colombia	Iran	Norway	Uruguay
Croatia	Guatemala	Peru	Italy
Ecuador	Germany	Poland	Australia
UK	Mexico	Romania	Spain
Greece	New Zealand	Russia	Turkey
Hong Kong	Sweden	Serbia	Peru
Hungary	USA	Singapore	Slovak Republic
Iran	Taiwan	Slovenia	Canada
Ireland	Norway	Sweden	Japan
Israel	Poland	Switzerland	France
Italy	Thailand	USA	Belgium
Jamaica	Switzerland	Uruguay	UK
Japan	Ireland	Venezuela	Hungary

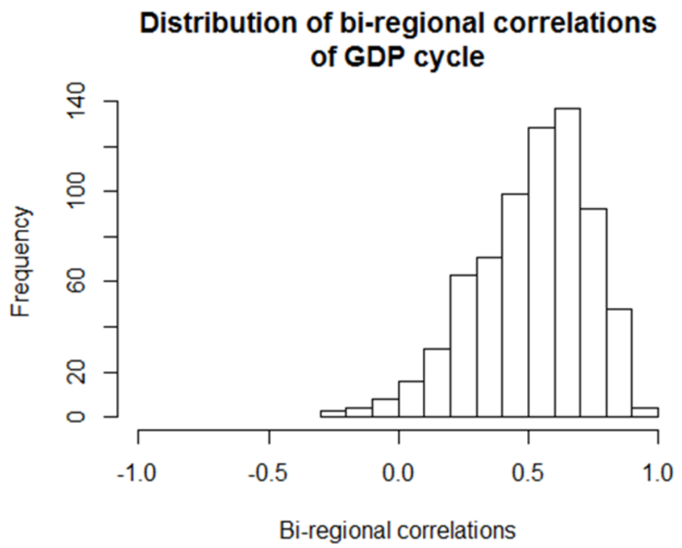
Table 4.C.4. Marginal media effects according to IV linear probability models using macro panel

	Self-employment		Wage work	
Variable	Coefficient	SE	Coefficient	SE
Media attention for entrepreneurship§	0.005***[***]	(0.001)	–0.004***[***]	(0.001)
Doing business is relatively easy†	–0.209**	(0.103)	0.225**	(0.101)
Ambiguity aversion index	–0.001	(0.001)	0.001	(0.001)
Ambiguity aversion index × doing business is relatively easy	0.002	(0.001)	–0.002	(0.001)
Inflation	–0.003	(0.002)	0.003	(0.002)
GDP	0.000**	(0.000)	0.000***[**]	(0.000)
GDP growth	0.000	(0.002)	0.002	(0.002)
Real interest rate	0.000	(0.001)	0.000	(0.001)
Lack of corruption‡	–0.010**	(0.005)	0.013***	(0.005)
Constant	0.181	(0.101)	0.663***[***]	(0.100)
		171 obs.		

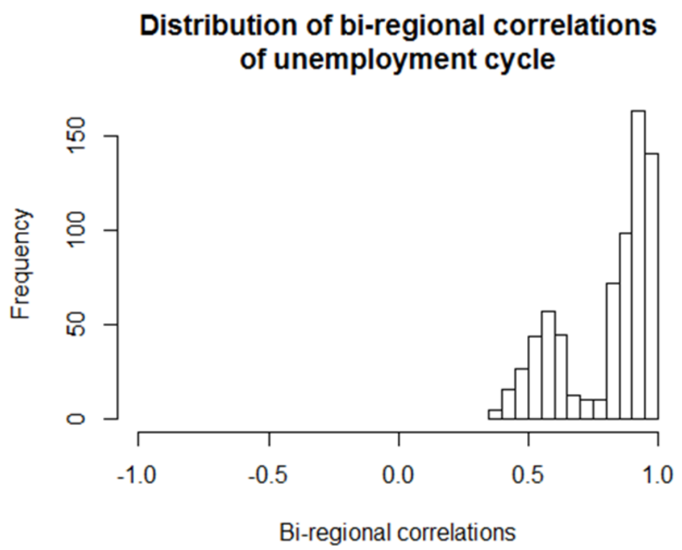
Notes: §Media attention is instrumented by number of natural disasters in other countries; †dummy is 1 if yes and zero else; ‡increase indicates less corruption; ***significant at the 1%-level; **significant at the 5%-level; [***]significant at the 1%-level with country-level clustering; [**]significant at the 5%-level with country-level clustering; standard errors in parentheses are heteroskedasticity-consistent.

Appendix 5.A

Table 5.A.1–5.A.6 provide additional information on data. In Table 5.A.7–5.A.10, we present additional estimation results. Finally, Figure 5.A.1 shows the distributions of bi-regional correlations between the GDP and the unemployment cycle.



(a) GDP cycle



(b) Unemployment cycle

FIGURE 5.A.1. Bi-regional correlations between GDP and unemployment cycle

TABLE 5.A.1. Correlation between start-up rates in different sectors

	Consumer oriented services	Business oriented services	Credit and insurance	Transport and postal services	Trade	Construction	Manufacturing	Energy and mining
All private sectors	0.95	0.8	0.90	0.95	0.96	0.96	0.75	−0.60
Energy and mining	−0.36	−0.48	−0.67	−0.53	−0.72	−0.66	−0.34	1
Manufacturing	0.70	0.46	0.46	0.68	0.80	0.69	1	
Construction	0.86	0.79	0.85	0.95	0.95	1		
Trade	0.84	0.74	0.84	0.91	1			
Transport and postal services	0.91	0.79	0.84	1				
Credit and insurance	0.85	0.92	1					
Business oriented services	0.91	1						

TABLE 5.A.2. Correlation between start-up rates in predominantly innovative industries

	Technology-oriented services	Knowledge intensive services	Non-innovative manufacturing	Technologically advanced manufacturing	High-tech manufacturing	Innovative manufacturing
All private sectors	0.88	0.93	0.68	0.74	0.78	0.76
Innovative manufacturing	0.81	0.74	0.72	0.99	0.98	1
High-tech manufacturing	0.84	0.76	0.75	0.96	1	
Technologically advanced manufacturing	0.79	0.72	0.69	1		
Non-innovative manufacturing	0.60	0.59	1			
Knowledge intensive services	0.98	1				

TABLE 5.A.3. Classification of industries

Industry or sector	Industry code (NACE 1993)
All private sectors	10–93 (without 91)
Energy and mining	10, 11, 12, 13, 14, 40, 41
Manufacturing	15–37
High-tech manufacturing industries	23.30, 24.20, 24.41, 24.61, 29.11, 29.60, 30.02, 31.62, 32.10, 32.20, 33.20, 33.30, 35.30
Technologically advanced manufacturing industries	22.33, 24.11, 24.12, 24.13, 24.14, 24.17, 24.30, 24.42, 24.62, 24.63,24.64, 24.66, 29.12, 29.13, 29.14, 29.31, 29.32, 29.40, 29.52, 29.53, 29.54, 29.55, 29.56, 30.01, 31.10, 31.40, 31.50, 32.30, 33.10, 33.40, 34.10, 34.30, 35.40
Non-technology oriented manufacturing	15–37 without high-tech and technologically advanced manufacturing industries
Construction	45
Trade	50, 51, 52
Transport and postal services	60, 61, 62, 63, 641
Credit and insurance	65, 66, 67
Technology oriented services	642, 72, 731, 742, 743
Non-technology oriented services	73.2, 74.11, 74.12, 74.13, 74.14, 74.4
Knowledge intensive services	Technology and non-technology oriented services
Other business oriented services	71.1, 71.2, 71.3, 74.5, 74.6, 74.7, 74.8 (without 74.87), 90
Business oriented services	Technology oriented, non-technology oriented, and other business oriented services
Consumer oriented services	55, 70 71.4, 80.4, 85, 92, 93

TABLE 5.A.4. Definition of variables

Variable	Definition
Start-up rate	Number of newly founded firms in the industry[a] per 10,000 regional workforce[b].
Unemployment rate—cyclical component	Number of registered unemployed persons over the entire working population.[c]
GDP—cyclical component	Nominal GDP divided by the annual consumer price index (CPI) of the Federal Statistical Office.[d]
Employees in small businesses	Share of employees in establishments with less than 20 employees.[b]
Employees with tertiary education	Share of employees with a university degree.[b]
Number of professors	Number of university professors at universities in the region in the respective year per per 1,000 workforce.[e]
Patent applications	Number of patent applications with an inventor residing in the region[e] per 1,000 workforce.[f]

Data sources: a) ZEW Mannheim Founder Panel. b) Establishment History File of the Social Insurance Statistics. c) Federal Employment Agency. d) Federal Statistical Office, Working Committee "Volkswirtschaftliche Gesamtrechnung der Länder". e) German University Statistics, Federal Statistical Office. f) PATSTAT database. g) Deutsche Bundesbank.

TABLE 5.A.5. Descriptive statistics for independent variables

Variable	Mean	Minimum	Maximum	SD
Unemployment rate—cyclical component (normalized)	–0.363	–2.52e-09	1	0.164
GDP—cyclical component (normalized)	0.607	–1.29e-08	1	0.061
Employees in small businesses	0.307	0.219	0.419	0.044
Employees with tertiary education	0.239	0.139	0.416	0.057
Number of professors	0.431	0.048	1.227	0.208
Patent applications	0.135	0.036	0.353	0.061

TABLE 5.A.6. Correlations among variables

		1	2	3	4	5
1	Unemployment rate—cyclical component	1				
2	GDP—cyclical component	–0.390	1			
3	Employees in small businesses	0.015	0.033	1		
4	Employees with tertiary education	–0.057	–0.026	–0.421	1	
5	Number of professors	–0.003	0.013	–0.398	0.338	1
6	Patent applications	–0.179	0.092	–0.382	0.101	–0.003

Table 5.A.7. First stage for unemployment cycle

2004 labor market reform	0.61***
Share of employees in small businesses $t-1$	–0.27***
Share of employees with tertiary education $t-1$	–0.20***
Number of professors per 1,000 workforce $t-1$	–0.04
Number of patent applications per 1,000 workforce $t-1$	0.30***
Constant	–0.05

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. Besides the instrument, the remaining independent variables are included in growth rates. All variables besides the instrument are lagged by one period, including the dependent variable. The number of observations is 494 (38 cross sections, 13 years) in all models. Standard errors are clustered at the region level.

Table 5.A.8. First stage for GDP cycle

2007 pre-crisis boom	1.29***
Share of employees in small businesses $t-1$	0.24***
Share of employees with tertiary education $t-1$	0.04
Number of professors per 1,000 workforce $t-1$	0.04
Number of patent applications per 1,000 workforce $t-1$	–0.15***
Constant	0.13***

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. Besides the instrument, the remaining independent variables are included in growth rates. All variables besides the instrument are lagged by one period, including the dependent variable. The number of observations is 494 (38 cross sections, 13 years) in all models. Standard errors are clustered at the region level.

TABLE 5.A.9. Influence of the GDP cycle on start-ups conditional on size given that the GDP cycle is instrumented by the 2007 pre-crisis peak

Start-up rates in:	Type 1 + 3 without banking: Small size industries	Type 2 + 4: Large size industries
GDP—cyclical component ($t-1$) (instrumented by pre-crisis boom)	−0.36***	−0.09
Share of employees in small businesses ($t-1$)	0.06**	−0.05
Share of employees with tertiary education ($t-1$)	−0.03	−0.01
Number of professors per 1,000 workforce ($t-1$)	0.02	0.02
Number of patent applications per 1,000 workforce ($t-1$)	0.09***	0.05
Entries into credit and insurance ($t-1$) (proxy for "health" of finance sector)	0.33***	0.04
Constant	−0.22***	−0.09***

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; IV estimation with fixed effects; business cycle is the cyclical component of the Hodrick-Prescott filtered variables. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 494 (38 cross sections, 13 years) in all models. Standard errors are clustered at the region level.

TABLE 5.A.10. Influence of cyclical variables on start-ups conditional on the four types using Hodrick-Prescott and Baxter-King filter in the same data set

Start-up rates in:	Type 1: non-innovative and small		Type 2: non-innovative and large		Type 3: innovative and small		Type 4: innovative and large	
	HP	BK	HP	BK	HP	BK	HP	BK
Unemployment rate—cyclical component ($t-1$)	0.05**	0.06***	0.11***	0.10***	–0.03	–0.03	–0.02	–0.01
GDP—cyclical component ($t-1$)	–0.18***	–0.17***	–0.01	–0.02	–0.19***	–0.20***	–0.05	–0.05
Share of employees in small businesses ($t-1$)	–0.05	–0.06	–0.09***	–0.10***	0.12***	0.11***	0.15***	0.15***
Share of employees with tertiary education ($t-1$)	–0.10**	–0.10**	0.04	0.03	–0.05**	–0.05**	–0.12***	–0.12**
Number of professors per 1,000 workforce ($t-1$)	0.02	0.02	0.02	0.02	0.02	0.03	0.01	0.02
Number of patent applications per 1,000 workforce ($t-1$)	0.04	0.05	–0.03	-0.03	0.06***	0.06***	0.10**	0.10**
Constant	–0.65***	–0.64***	–0.24***	–0.23***	0.32***	0.33***	0.52***	0.52***
R^2	0.63	0.62	0.76	0.76	0.70	0.70	0.46	0.46

Notes: ***, ** statistically significant at 1 percent and 5 percent, respectively; fixed effects estimation; business cycle is the cyclical component of the Hodrick-Prescott filtered (HP) or Baxter-King filtered (BK) variables, where the same observations are used in the estimation procedure. The remaining independent variables are included in growth rates. All variables are lagged by one period. The number of observations is 411 (due to filtering values at the beginning and the end of a time series are lost) in all models.

Appendix 6.A

To check the robustness of the results obtained with GSOEP data, this appendix provides additional simulation results using 1,620 combinations of parameter values.

6.A.1 Simulation setup

Let $l = 1, \ldots, L$ denote all parameter combinations. Let S^l denote the recommendation performance of an arbitrary approach given parameter combination l. We consider $L = 1,620$ combinations. As before, for every parameter combination, we compute $10,000$ simulations with sample sizes $n(\Theta) = 1,000$ and $n(\Omega) = 100$. Given a sample of historical and client data, we apply three approaches to the same simulated data:

- general average scores (GAS);
- average scores with an optimized similarity criterion, given that **m** and **Q** are known (OAS); and
- the probability-based approach (PBA).

Parameters, which are given in Table 6.A.1, are selected in a way such that a high number of different conditions is covered. Correlation between personality trait and entrepreneurial fitness ranges from weak, $\rho = 0.1$, to strong, $\rho = 0.9$.

Table 6.A.1. Variations in parameters

Parameter(s)	Values
μ_Γ, μ_Π	$\in \{-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2\}$
ρ	$\in \{0.1, 0.3, 0.5, 0.7, 0.9\}$
τ, ϵ	$\in \{0.01, 0.15\}$

ϵ is the non-optimized similarity criterion of average scores. Following the general logic of entrepreneurship-prone personality profiles, we assume that to receive a recommendation for entrepreneurship the client's personality trait must be sufficiently similar to the profile, and that sufficient similarity promises good recommendation results. Hence, we use a rather strict (small) similarity criterion for average scores. However, we consider two different similarity criteria to examine the effect of changes in similarity criteria on average success rates (to test the second requirement). In particular, if $\epsilon = 0.01$, we say that the similarity criterion is strict, whereas $\epsilon = 0.15$ is interpreted as a tolerant similarity criterion.

To compute ϵ^* for the optimized version of average scores, we numerically maximize (6.11) for every parameter combination. Given the assumption on μ_Π and τ, we cover a wide range of population shares of entrepreneurs, which is demonstrated in Figure 6.A.1.

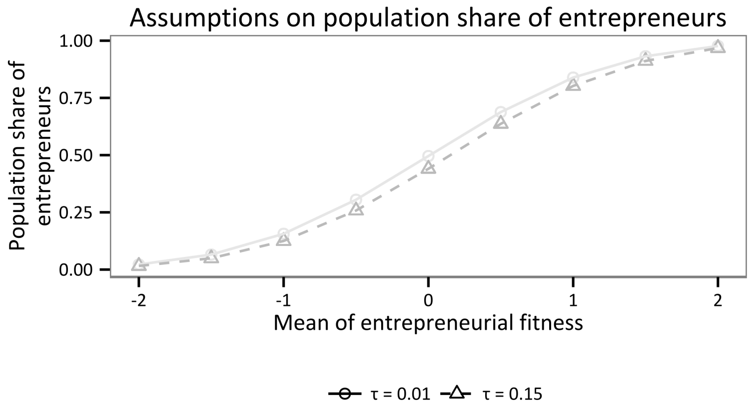


FIGURE 6.A.1. Simulated population shares of entrepreneurs

6.A.2 *Performance analysis using 1,620 parameter combinations*

BENCHMARKING SUCCESS PROBABILITIES

To get an overview over average performance, we compute the simulation average, approximating $\mathbb{E}[S^l]$, for every parameter combination l and every approach. In Figure 6.A.2, we

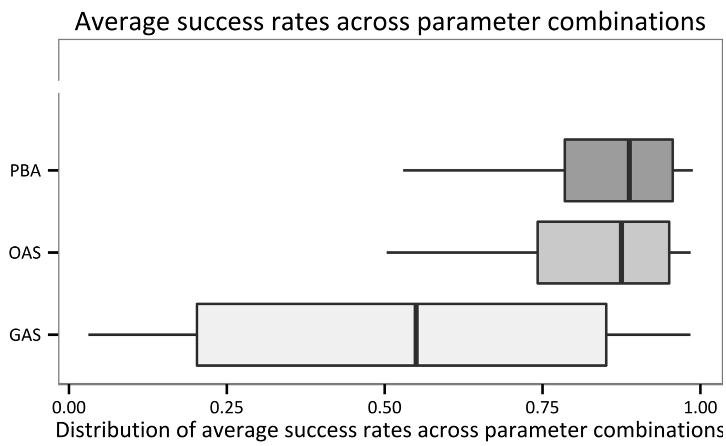


FIGURE 6.A.2. Distribution of average success rates, $p = \mathbb{E}[S]$

plot the distribution of average recommendation success rates across parameter combinations. Figure 6.A.2 reveals that the general average-scores approach (GAS) substantially underperforms compared to all other approaches. In Figure 6.A.3, we only show distributions of average success rates for a high correlation between personality trait and entrepreneurial fitness ($\rho = 0.9$). Still, even when correlation between personality trait and entrepreneurial fitness is high, general average scores underperform in comparison to all other approaches.

In contrast to general average scores, optimized average scores (OAS) exhibit high av-

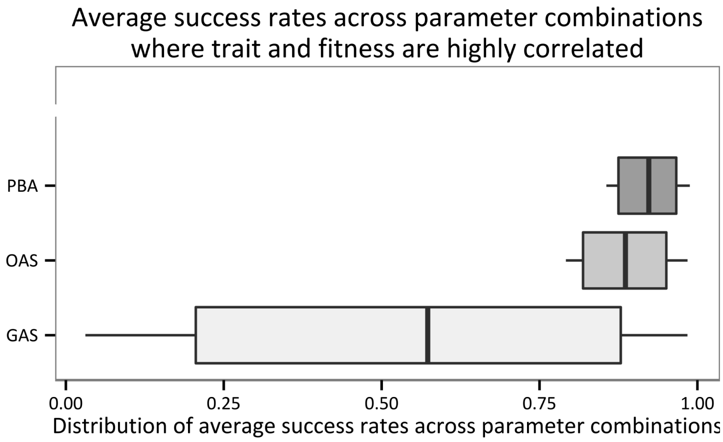
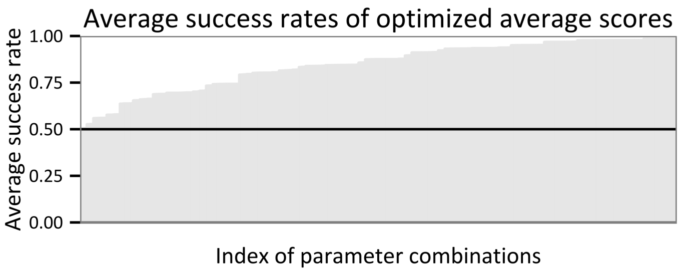


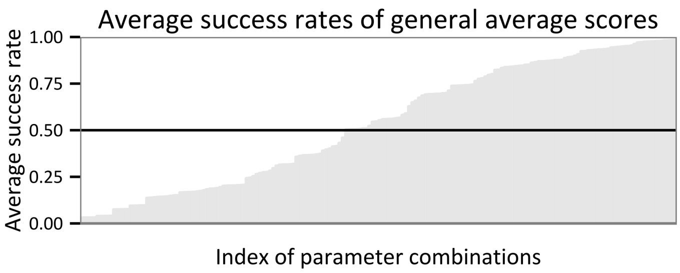
Figure 6.A.3. Distribution of average success rates, $p = \mathbb{E}[S]$, if correlation is high

erage success rates, which are slightly inferior to the upper boundary of recommendation performance represented by the probability-based approach (PBA). The results on relative performance are consistent with those obtained with the GSOEP calibrated model.

Testing requirements



(a) Average success rates of optimized average scores



(b) Average success rates of general average scores

Figure 6.A.4. Recommendation success of average scores

To test the first requirement, in Figure 6.A.4, we plot average recommendation success rates of (a) the optimized average-scores approach and (b) the general average-scores approach as a function of the parameter combinations index, l. Optimized average scores (Figure 6.A.4a) always fulfill the first requirement. In case of general average scores (Figure 6.A.4b), average success rates are smaller than 50%, the approach is inferior to the coin, in about 44% of all parameter combinations. More specific, only if the population share of entrepreneurs is low (about 19% on average, ranging between about 2% and 50), general average scores outperform the coin with respect to average recommendation success rates.

To test the second requirement, let $l_{\varepsilon,\varepsilon'} = (l_\varepsilon, l_{\varepsilon'})$ denote a pair of parameter combination where all parameters besides the similarity criterion are exactly the same. Our simulation-based measure of robustness, the simulation counterpart of $|\mathbb{E}[S(\varepsilon) - S(\varepsilon')]|$, is

$$\tilde{\Delta}(l_{\varepsilon,\varepsilon'}) = M^{-1}\left|\sum_{m=1}^{M} S_m^{l_\varepsilon} - \sum_{m=1}^{M} S_m^{l_{\varepsilon'}}\right|$$

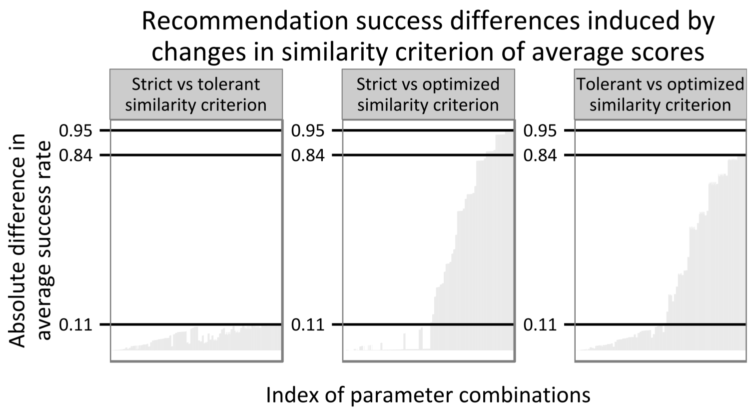


FIGURE 6.A.5. Similarity-criterion-induced changes in average success rates

In Figure 6.A.5, we present robustness measures for average scores. Changing from a strict ($\epsilon = 0.01$) to a tolerant ($\epsilon = 0.15$) similarity criterion, or *vice versa*, changes average recommendation success rates by about 11 percentage points at maximum. The results become more striking when we compare the strict and the tolerant criterion to the optimized similarity criterion ϵ^*. The difference in average success rates between the strict and the optimized criterion is 95 percentage points at maximum, while the success rate difference between the tolerant and the optimized criterion is approx. 84 percentage points at maximum. The results indicate that average scores are not robust—mistakes of the adviser can generate high costs (e.g., a loss in average recommendation success rates of 95 percentage points).

List of tables

List of figures

Bibliography

Acharya, V. & Pedersen, L. (2005). Asset Pricing with Liquidity Risk. *Journal of Financial Economics*, *77*, 375–410.

Acs, Z., Braunerhjelm, P., Audretsch, D., & Carlsson, B. (2009). The Knowledge Spillover Theory of Entrepreneurship. *Small Business Economics*, *32*, 15–30.

Ait-Sahalia, Y. & Brandt, M. (2001). Variable Selection for Portfolio Choice. *Journal of Finance*, *56*, 1297–1351.

Almeida, P. & Kogut, B. (1999). Localisation of Knowledge and the Mobility of Engineers in Regional Networks. *Management Science*, *45*, 905–917.

Almlund, M., Duckworth, A., Heckman, J., & Kautz, T. (2011). Personality Psychology and Economics. *NBER Working Paper*, *16822*.

Alvarez, R. & Brehm, J. (1995). American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values. *American Journal of Political Science*, *41*, 345–374.

Anderson, C. (2001). *Behavioral Models of Strategies in Multi-armed Bandit Problems*. PhD thesis, California Institute of Technology, Pasadena.

Anokhin, S. & Schulze, W. (2009). Entrepreneurship, Innovation, and Corruption. *Journal of Business Venturing*, *24*, 465–476.

Anscombe, F. & Glynn, W. (1983). Distribution of Kurtosis Statistic for Normal Statistics. *Biometrika*, *70*, 227–234.

Ansolabehere, S. & Iyengar, S. (1995). *Going Negative: How Attack Ads Shrink and Polarize the Electorate*. New York: Free Press.

Antonovics, K. & Golan, L. (2012). Experimentation and Job Choice. *Journal of Labor Economics*, *30*, 333–366.

Ardagna, S. & Lusardi, A. (2010). Explaining International Differences in Entrepreneurship: The Role of Individual Characteristics and Regulatory Constraints. In J. Lerner & A. Schoar (Eds.), *International Differences in Entrepreneurship*. Univ. of Chicago Press.

Arthur, M. (1994). The Boundaryless Career: A New Perspective for Organizational Inquiry. *Journal of Organizational Behavior*, *15*, 295–306.

Arum, R. & Müller, W. (2004). Self-employment Dynamics in Advanced Economies. In R. Arum & W. Müller (Eds.), *The Re-emergence of Self-employment* (pp. 1–35). Princeton: Princeton Univ. Press.

Astebro, T., Braunerhjelm, P., & Broström, A. (2013). Does Academic Entrepreneurship Pay? *Industrial and Corporate Change*, *22*, 281–311.

Astebro, T. & Chen, J. (2014). The Entrepreneurial Earnings Puzzle. *Journal of Business Venturing*, *29*, 88–105.

Astebro, T., Chen, J., & Thompson, P. (2011). Stars and Misfits: Self-Employment and Labor Market Frictions. *Management Science*, *57*, 1999–2017.

Audretsch, D. (1995). *Innovation and Industry Evolution*. Cambridge (MA): MIT Press.

Audretsch, D. & Keilbach, M. (2007a). The Location of Entrepreneurship Capital: Evidence from Germany. *Papers in Regional Science*, *86*, 351–365.

Audretsch, D. & Keilbach, M. (2007b). The Theory of Knowledge Spillover Entrepreneurship. *Journal of Management Studies*, *44*, 1242–1254.

Audretsch, D., Keilbach, M., & Lehmann, E. (2006). *Entrepreneurship and Economic Growth*. Oxford: Oxford University Press.

Banerjee, A. & Duflo, E. (2014). Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program. *Review of Economic Studies*, *81*, 572–607.

Banfield, J. & Raftery, A. (1993). Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, *49*, 803–821.

Barlevy, G. (2007). On the Cyclicality of Research and Development. *American Economic Review*, *97*, 1131–1164.

Baum, J. & Locke, E. (2004). The Relation of Entrepreneurial Traits, Skill, and Motivation to Subsequent Venture Growth. *Journal of Applied Psychology*, *89*, 587–598.

Baxter, M. & King, R. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *Review of Economics and Statistics*, *81*, 575–593.

Begley, T. & Boyd, D. (1987). Psychological Characteristics Associated with Performance in Entrepreneurial Firms and Smaller Businesses. *Journal of Business Venturing*, *22*, 147–173.

Bellman, R. (1956). A Problem in the Sequential Design of Experiments. *Sankhyā: The Indian Journal of Statistics*, *16*, 221–229.

Benz, M. & Frey, B. (2008a). Being Independent is a Great Thing. *Economica*, *75*, 362–383.

Benz, M. & Frey, B. (2008b). The Value of Doing What You Like. *Journal of Economic Behavior and Organization*, *68*, 445–455.

Bergemann, D. & Hege, U. (2005). The Financing of Innovation. *RAND Journal of Economics*, *36*, 719–752.

Bernanke, B. & Gertler, M. (1989). Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review*, *79*, 14–31.

Bernardo, A. & Welch, I. (2001). On the Evolution of Overconfidence and Entrepreneurs. *Journal of Economics & Management Strategy*, *10*, 301–330.

Bernhardt, A., Morris, M., Handcock, M., & Scott, M. (2001). *Divergent Paths: Economic Mobility in the New American Labor Market*. New York: Russell Sage.

Berninghaus, S. & Seifert-Vogt, H. (1987). International Migration under Incomplete Information. *Swiss Journal of Economics and Statistics*, *123*, 199–218.

Berridge, K. & Robinson, T. (1998). What is the Role of Dopamine in Reward: Hedonic Impact, Reward Learning, or Incentive Salience? *Brain Research Reviews*, *28*, 309–369.

Bersch, J., Gottschalk, S., Müller, B., & Niefert, M. (2014). The Mannheim Enterprise Panel (MUP) and Firm Statistics for Germany. *ZEW Discussion Paper*, *14-104*.

Besley, T. & Burgess, R. (2002). The Political Economy of Government Responsiveness: Theory and Evidence from India. *Quarterly Journal of Economics*, *117*, 1415–1451.

Bewley, T. (2002). Knightian Decision Theory. Part I. *Decisions in Economics and Finance*, *25*, 79–110.

Biemann, T., Zacher, H., & Feldman, D. (2012). Career Patterns: A Twenty-year Panel Study. *Journal of Vocational Behavior*, *81*, 159–170.

Black, F. (1986). Noise. *Journal of Finance*, *41*, 529–543.

Blanchflower, D. (2000). Self-employment in OECD Countries. *Labour Economics*, *7*, 471–505.

Blanchflower, D. & Oswald, A. (1998). What Makes an Entrepreneur? *Journal of Labor Economics*, *16*, 26–60.

Bono, J., Boles, T., Judge, T., & Lauver, K. (2002). The Role of Personality in Task and Relationship Conflict. *Journal of Personality*, *70*, 311–344.

Borck, R., Caliendo, M., & Steiner, V. (2007). Fiscal Competition and the Composition of Public Spending: Theory and Evidence. *Public Finance Analysis*, *63*, 264–277.

Borck, R., Fossen, F., Freier, R., & Martin, T. (2015). Race to the Debt Trap? *Regional Science and Urban Economics*, *53*, 20–37.

Borghans, L., Duckworth, A., Heckman, J., & ter Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, *43*, 972–1059.

Bosma, N., Hessels, J., Schutjens, V., Van Praag, M., & Verheul, I. (2012). Entrepreneurship and Role Models. *Journal of Economic Psychology*, *33*, 410–424.

Bradt, R., Johnson, S., & Karlin, S. (1956). On Sequential Designs for Maximizing the Sum of n Observations. *Annals of Mathematical Statistics*, *27*, 1060–1074.

Braguinsky, S., Klepper, S., & Ohyama, A. (2012). High-Tech Entrepreneurship. *Journal of Law and Economics*, *55*, 869–900.

Breitung, J. (2000). The Local Power of Some Unit Root Tests for Panel Data. In *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels* (pp. 161–178). Amsterdam: JAI, Press.

Brezzi, M. & Lai, T. (2000). Incomplete Learning from Endogenous Data in Dynamic Allocation. *Econometrica*, *68*, 1511–1516.

Brezzi, M. & Lai, T. (2002). Optimal Learning and Experimentation in Bandit Problems. *Journal of Economic Dynamics and Control*, *27*, 87–108.

Buch, C., Koch, C., & Koetter, M. (2013). Do Banks Benefit from Internationalization? Revisiting the Market Power-risk Nexus. *Review of Finance*, *17*, 1401–1435.

Buntaine, M. (2011). Does the Asian Development Bank Respond to Past Environmental Performance when Allocating Environmentally Risky Financing. *World Development*, *39*, 336–350.

Burke, A., FitzRoy, F., & Nolan, M. (2008). What makes a Die-hard Entrepreneur? Beyond the 'Employee or Entrepreneur' Dichotomy. *Small Business Economics*, *31*, 93–115.

Busenitz, L. & Barney, J. (1997). Differences Between Entrepreneurs and Managers in Large Organizations. *Journal of Business Venturing*, *12*, 9–30.

Büttner, T. (2001). Local Business Taxation and Competition for Capital: The Choice of the Tax Rate. *Regional Science and Urban Economics*, *31*, 215–245.

Caballero, R. & Hammour, M. (1994). The Cleansing Effect of Recessions. *American Economic Review*, *84*, 1350–1368.

Caliendo, M., Fossen, F., & Kritikos, A. (2009). Risk Attitudes of Nascent Entrepreneurs. *Small Business Economics*, *32*, 153–167.

Caliendo, M., Fossen, F., & Kritikos, A. (2010). The Impact of Risk Attitudes on Entrepreneurial Survival. *Journal of Economic Behavior and Organization*, *76*, 45–63.

Caliendo, M., Fossen, F., & Kritikos, A. (2014). Personality Characteristics and The Decision to Become and Stay Self-Employed. *Small Business Economics*, *42*, 787–814.

Caliendo, M. & Kritikos, A. (2009). I Want to, But I also Need to: Start-Ups Resulting from Opportunity and Necessity. *IZA Discussion Paper*, *4661*.

Caliendo, M. & Kritikos, A. (2010). Start-Ups by the Unemployed. *Small Business Economics*, *35*, 71–92.

Caliendo, M., Kritikos, A., Künn, S., Loersch, C., Schröder, H., & Schütz, H. (2014). Evaluation der Programme Gründercoaching Deutschland. *IZA Research Report*, *61*.

Camerer, C., Ho, T., & Chong, J. (2008). *Handbook of Experimental Economics Results*, volume 1, chapter Learning and Equilibrium in Games, (pp. 607–615). Amsterdam: Elsevier.

Camerer, C. & Ho, T.-H. (1999). Experience-weighted Attraction Learning in Normal Form Games. *Econometrica*, *67*, 827–874.

Camerer, C. & Lovallo, D. (1999). Overconfidence and Excess Entry. *American Economic Review*, *89*, 306–318.

Camerer, C. & Weber, M. (1992). Recent Developments in Modeling Preferences: Uncertainty and Ambiguity. *Journal of Risk and Uncertainty*, *5*, 325–370.

Campanale, C. (2010). Private Equity Returns in a Model of Entrepreneurial Choice with Learning. *B.E. Journal of Macroeconomics*, *10*, 1–37.

Carter, S. (2011). The Rewards of Entrepreneurship: Exploring the Incomes, Wealth, and Economic Well-being of Entrepreneurial Households. *Entrepreneurship Theory and Practice*, *35*, 39–55.

Chen, M. & Doane, D. (2008). *Informality in South Asia: A Review*. Cambridge, MA: WIEGO.

Ciavarella, M., Buchholtz, A., Riordan, C., Gatewood, R., & Stokes, G. (2004). The Big Five and Venture Survival. *Journal of Business Venturing*, *19*, 465–483.

Clementi, G. & Palazzo, B. (2016). Entry, Exit, Firm Dynamics and Aggregate Fluctuations. *American Economic Journal: Macroeconomics*, *8*, 1–41.

Cobb-Clark, D. & Schurer, S. (2012). The Stability of Big-Five Personality Traits. *Economics Letters*, *115*, 11–15.

Cooper, A. & Woo, C. (1988). Entrepreneurs' Perceived Chances for Success. *Journal of Business Venturing*, *3*, 97–108.

Costa, P., McCrae, R., & Holland, J. (1984). Personality and Vocational Interests in an Adult Sample. *Journal of Applied Psychology*, *69*, 390–400.

Cowling, M. & Taylor, M. (2001). Entrepreneurial Women and Men: Two Different Species? *Small Business Economics*, *16*, 167–175.

Cronbach, L. & Gleser, G. (1953). Assessing the Similarity Between Profiles. *Psychological Bulletin*, *50*, 456–473.

Cunha, F., Heckman, J., Lochner, L., & Masterov, D. (2005). Interpreting the Evidence on Life Cycle Skill Formation. *NBER Working Paper*, *11331*.

D'Agostino, R. (1970). Transformation to Normality of the Null Distribution of G1. *Biometrika*, *57*, 679–681.

Daly, M. (2015). The Long Term Returns of Attempting Self-employment with Regular Employment as a Fall Back Option. *Labour Economics*, *35*, 26–52.

De Meza, D. & Southey, C. (1996). The Borrower's Curse: Optimism, Finance, and Entrepreneurship. *Economic Journal*, *106*, 375–386.

De Paola, M., Scoppa, V., & Lombardo, R. (2010). Can Gender Quotas Break Down Negative Stereotypes? Evidence from Changes in Electoral Rules. *Journal of Public Economics*, *94*, 344–353.

Della Vigna, S. & Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *Quarterly Journal of Economics*, *122*, 1187–1234.

Della Vigna, S. & La Ferrara, E. (2015). Economic and Social Impacts of the Media. *NBER Working Paper*, *21360*.

Depue, R. & Collins, P. (1999). Neurobiology of the Structure of Personality: Dopamine, Facilitation of Incentive Motivation, and Extraversion. *Behavioral and Brain Sciences*, *22*, 491–517.

DeYoung, C. (2006). Higher-order Factors of the Big Five in a Multi-informant Sample. *Journal of Personality and Social Psychology*, *91*, 1138–1151.

DeYoung, C., Peterson, J., & Higgins, D. (2002). Higher-order Factors of the Big Five Predict Conformity: Are there Neuroses of Health? *Personality and Individual Differences*, *33*, 533–552.

DeYoung, C., Peterson, J., & Higgins, D. (2005). Sources of Openness/Intellect: Cognitive and Neuropsychological Correlates of the Fifth Factor of Personality. *Journal of Personality*, *73*, 825–858.

Digman, J. (1997). Higher-order Factors of the Big Five. *Journal of Personality and Social Psychology*, *73*, 1246–1256.

Dillon, E. & Stanton, C. (2016). Self-employment Dynamics and the Returns to Entrepreneurship. *Working Paper*, *17-022*. Harvard Business School.

Djankov, S., Qian, Y., Roland, G., & Zhuravskaya, E. (2006). Entrepreneurship in China and Russia Compared. *Journal of the European Economic Association*, *4*, 352–365.

Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit. *Journal of Personality and Social Psychology*, *92*, 1087–1101.

Dyck, A., Moss, D., & Zingales, L. (2008). Media versus Special Interest. *NBER Working Paper*.

Dyck, A., Volchkova, N., & Zingales, L. (2008). The Corporate Governance Role of the Media: Evidence from Russia. *The Journal of Finance*, *63*, 1093–1135.

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, *82*, 171–185.

Eisensee, T. & Strömberg, D. (2007). News Droughts, News Floods, and U.S. Disaster Relief. *Quarterly Journal of Economics*, *122*, 693–728.

Elfenbein, D., Hamilton, B., & Zenger, T. (2010). The Small Firm Effect and the Entrepreneurial Spawning of Scientists and Engineers. *Management Science*, *56*, 659–681.

Elhorst, J. (2003). Specification and Estimation of Spatial Panel Data Models. *International Regional Sciences Review*, *26*, 244–268.

Ellsberg, D. (1961). Risk, Ambiguity and Savage Axioms. *Quarterly Journal of Economics*, *75*, 643–679.

Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review*, *101*, 3253–3285.

Evans, D. & Jovanovic, B. (1989). An Estimated Model of Entrepreneurial Choice under Liquidity Constraints. *Journal of Political Economy*, *97*, 808–827.

Evans, D. & Leighton, L. (1989). Some Empirical Aspects of Entrepreneurship. *The American Economic Review*, *79*, 519–535.

Fairlie, R. (2013). Entrepreneurship, Economic Conditions, and the Great Recession. *Journal of Economics and Management Strategy*, *22*, 207–231.

Fairlie, R., Karlan, D., & Zinman, J. (2015). Behind the GATE Experiment: Evidence on Effects of and Rationales for Subsidized Entrepreneurship Training. *American Economic Journal: Economic Policy*, *7*, 125–161.

Falck, O. (2007). *Emergence and Survival of New Businesses*. Heidelberg: Physica.

Fellner, W. (1961). Distortion of Subjective Probabilities as a Reaction to Uncertainty. *Quarterly Journal of Economics*, *75*, 670–689.

Ferber, M. & Waldfogel, J. (1998). The Long-term Consequences of Non-standard Work. *Monthly Labor Review*, *121*, 3–12.

Ferrari, S. & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, *31*, 799–815.

Fifield, A. (2016). Start-ups Encounter Hiccups in the Land of the Rising Sun. *The Washington Post*. 23 June 2016.

Fischer, R. & Boer, D. (2014). Motivational Basis of Personality Traits: A Meta-Analysis of Value-Personality Correlations. *Journal of Personality*, *83*, 491–510.

Foti, A. & Vivarelli, M. (1994). An Econometric Test of the Self-employment Model—the Case of Italy. *Small Business Economics*, *6*, 81–93.

Fraley, C. & Raftery, A. (2007). Bayesian Regularization for Normal Mixture Estimation and Model-Based Clustering. *Journal of Classification*, *24*, 155–181.

Francois, P. & Lloyd-Ellis, H. (2003). Animal Spirits Through Creative Destruction. *American Economic Review*, *93*, 530–550.

Frankfurter Allgemeine Zeitung (2017). Wie Vox Deutschland zum Start-up-Land macht. *Frankfurter Allgemeine Zeitung, 5 September 2017*.

Frese, M. & Gielnik, M. (2014). The Psychology of Entrepreneurship. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*, 413–438.

Frisch, D. & Baron, J. (1988). Ambiguity and Rationality. *Journal of Behavioral Decision Making*, *1*, 149–157.

Fritsch, M. (2013). New Business Formation and Regional Development—A Survey and Assessment of the Evidence. *Foundations and Trends in Entrepreneurship*, *9*, 249–364.

Fritsch, M. & Aamoucke, R. (2013). Regional Public Research, Higher Education, and Innovative Start-ups—An Empirical Investigation. *Small Business Economics*, *41*, 865–885.

Fritsch, M. & Aamoucke, R. (2015). Fields of Knowledge in Higher Education Institutions, and Innovative Start-Ups—An Empirical Investigation. *Papers in Regional Science*. forthcoming.

Fritsch, M., Kritikos, A., & Pijnenburg, K. (2015). Business Cycles, Unemployment and Entrepreneurial Entry—First Evidence from Germany. *International Entrepreneurship and Management Journal*, *11*, 267–286.

Furnham, A. (1992). Personality and Learning Style: A Study of three Instruments. *Personality and Individual Differences*, *13*, 429–438.

Gadzella, B., Ginther, D., Master, W., & Guthrie, D. (1997). Predicting Students as Deep and Shallow Processors of Information. *Perceptual and Motor Skills*, *84*, 875–881.

Gans, H., Knox, G., & Croson, R. (2007). Simple Models of Discrete Choice and Their Performance in Bandit Experiments. *Manufacturing & Service Operations Management*, *9*, 383–408.

Gehrke, B., Schasse, U., Rammer, C., Frietsch, R., Neuhäusler, P., & Leidmann, M. (2010). Listen wissens- und technologieintensiver Güter und Wirtschaftszweige. Studien zum deutschen Innovationssystem. *Frauenhofer ISI, NIW, ZEW*, *19-2010*.

Geisler-Brenstein, E., Schmeck, R., & Hetherington, J. (1996). An Individual Difference Perspective on Student Diversity. *Higher Education*, *31*, 73–96.

GEM (2015). gemconsortium.org.

Gentzkow, M. (2006). Television and Voter Turnout. *Quarterly Journal of Economics*, *121*, 931–972.

Gentzkow, M. & Shapiro, J. (2004). Media, Education, and Anti-Americanism in the Muslim World. *Journal of Economic Perspectives*, *18*, 117–133.

Georgarakos, D. & Tatsiramos, K. (2009). Entrepreneurship and Survival Dynamics of Immigrants to the U.S. and their Descendants. *Labour Economics*, *16*, 161–170.

Georgellis, Y. & Wall, H. (2000). What Makes a Region Entrepreneurial? Evidence from Britain. *Annals of Regional Science*, *34*, 385–403.

Gerber, A. & Green, D. (2000). The Effects of Canvassing, Telephone Calls and Direct Mail on Voter Turnout. *American Political Science Review*, *94*, 653–663.

Gerber, A., Karlan, D., & Bergan, D. (2009). Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions. *American Economic Journal*, *1*, 35–52.

Gervais, S. & Odean, T. (2001). Learning To Be Overconfident. *The Review of Financial Studies*, *14*, 1–27.

Ghatak, M., Morelli, M., & Sjöström, T. (2007). Entrepreneurial Income, Occupational Choice, and Trickle-up Policies. *Journal of Economic Theory*, *137*, 27–48.

Ghirardato, P., Maccheroni, F., & Marinacci, M. (2004). Differentiating Ambiguity and Ambiguity Attitude. *Journal of Economic Theory*, *118*, 133–173.

Gilboa, I. & Schmeidler, D. (1989). Maxmin Expected Utility with Non-unique Prior. *Journal of Mathematical Economics*, *18*, 141–153.

Gill, D. & Prowse, V. (2016). Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-*k* Analysis. *Journal of Political Economy*. forthcoming.

Gimeno, J., Folta, T., Cooper, A., & Woo, C. (1997). Survival of the Fittest? Entrepreneurial Human Capital and the Persistence of Underperforming Firms . *Administrative Science Quarterly*, *42*, 750–783.

Gindling, T. & Newhouse, D. (2012). Self-employment in the Developing World. *World Bank Policy Research Working Paper*, *6201*.

Gittins, J., Glazebrook, K., & Weber, R. (2011). *Multi-armed Bandit Allocation Indices* (2. ed.). Chichester: Wiley.

Gittins, J. & Jones, D. (1974). A Dynamic Allocation Index for the Sequential Design of Experiments. In J. Gani (Ed.), *Progress in Statistics* (pp. 241–266). Amsterdam, NL: North-Holland.

Glaeser, E., Kerr, W., & Ponzetto, G. (2010). Clusters of Entrepreneurship. *Journal of Urban Economics*, *67*, 150–168.

Goedhuys, M. & Sleuwaegen, L. (1999). Barriers to Growth of Firms in Developing Countries, Evidence from Burundi. In D. Audretsch & A. Thurik (Eds.), *Innovation, Industry Evolution and Employment*. Cambridge Univ. Press.

Gompers, P., Lerner, J., & Scharfstein, D. (2005). Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999. *Journal of Finance*, *60*, 577–614.

Granrose, C. & Baccili, P. (2006). Do Psychological Contracts Include Boundaryless or Protean Careers? *Career Development International*, *11*, 163–182.

Grant, D. (1996). The Political Economy of New Business Formation Across the American States 1970-1985. *Social Science Quarterly*, *77*, 28–42.

Green, D., Gerber, A., & Nickerson, D. (2003). Getting Out The Vote in Local Elections. *Journal of Politics*, *65*, 1083–1096.

Green, D., Müllensiefen, D., Lamb, M., & Rentfrow, P. (2015). Personalty Predicts Musical Sophistication. *Journal of Research in Personality*, *58*, 154–158.

Griffin, D. & Varey, C. (1996). Towards a Consensus on Overconfidence. *Organizational Behavior and Human Decision Processes*, *65*, 227–231.

Guha-Sapir, D., Below, R., & Hoyois, P. (2016). EM-DAT: The CRED/OFDA International Disaster Database. Université Catholique de Louvain. Brussels, Belgium. www.emdat.be.

Hall, D. & Moss, J. (1998). The New Protean Career Contract: Helping Organizations and Employees Adapt. *Organizational Dynamics*, *26*, 22–37.

Hamilton, B. (2000). Does Entrepreneurship Pay? *Journal of Political Economy*, *108*, 604–631.

Hamilton, B., Papageorge, N., & Pande, N. (2017). The Right Stuff? Personality and Entrepreneurship. *Discussion Paper*.

Harding, R., Brooksbank, D., Hart, M., Jones-Evans, D., Levie, J., O'Reilly, J., & et al. (2006). *Global Entrepreneurship Monitor United Kingdom 2005*. London: London Business School.

Harris, S., Wright, A., Hayward, C., Starr, J., Whalley, L., & Deary, I. (2005). The Functional COMT Polymorphism, Val158Met, is Associated with Logical Memory and the Personality Trait Intellect/Imagination in a Cohort of Healthy 79 Year Olds. *Neuroscience Letters*, *385*, 1–6.

Hartigan, J. & Wong, M. (1979). A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society*, *28*, 100–108.

Hartog, J., Van Praag, M., & Van der Sluis, J. (2010). If You are so Smart Why are You Not an Entrepreneur? Returns to Cognitive and Social Ability. *Journal of Economics and Management Strategy*, *19*, 947–989.

Heckman, J., Stixrud, J., & Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labour Market Outcomes and Social Behavior. *Journal of Labor Economics*, *24*, 411–482.

Heller, D., Ferris, D., Brown, D., & Watson, D. (2009). The Influence of Work Personality on Job Satisfaction: Incremental Validity and Mediation Effects. *Journal of Personality*, *77*, 1051–1084.

Helmers, C. & Rogers, M. (2010). Innovation and the Survival of New Firms in the UK. *Review of Industrial Organization*, *36*, 227–248.

Hennighausen, T. (2015). Exposure to Television and Individual Beliefs: Evidence from a Natural Experiment. *Journal of Comparative Economics*, *43*, 956–980.

Hintermaier, T. & Steinberger, T. (2002). Occupational Choice and the Private Equity Premium Puzzle. *Journal of Economic Dynamics and Control*, *29*, 1765–1783.

Hirsh, J., DeYoung, C., & Peterson, J. (2009). Metatraits of the Big Five differentially Predict Engagement and Restraint of Behavior. *Journal of Personality*, *77*, 1085–1102.

Hirshleifer, S., McKenzie, D., Almeida, R., & Ridao-Cano, C. (2015). The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey. *The Economic Journal*. forthcoming.

Ho, T.-H., Camerer, C., & Chong, J.-K. (2007). Self-tuning Experience Weighted Attraction Learning in Games. *Journal of Economic Theory*, *133*, 177–198.

Hodrick, R. & Prescott, E. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking*, *29*, 1–16.

Hofstede, G. (2015). www.geerthofstede.nl/research--vsm.

Hofstede, G., Hofstede, G., & Minkov, M. (2010). *Cultures and Organizations* (3. ed.). New York: McGraw-Hill.

Holland, D. & Shepherd, D. (2013). Deciding to Persist: Adversity, Values, and Entrepreneurs' Decision Policies. *Entrepreneurship Theory and Practice*, *37*, 331–358.

Holland, J. (1997). *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments* (3. ed.). Odessa, FL: Psychological Assessment Resources.

Holm, H., Opper, S., & Nee, V. (2013). Entrepreneurs Under Uncertainty: An Economic Experiment in China . *Management Science*, *59*, 1671–1687.

Holmes, T. & Schmitz, J. (1990). A Theory of Entrepreneurship and Its Application to the Study of Business Transfers. *Journal of Political Economy*, *98*, 265–294.

Hong, H., Scheinkman, J., & Xiong, W. (2006). Asset Float and Speculative Bubbles. *Journal of Finance*, *61*, 1073–1117.

Hopenhayn, H. (1992). Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, *60*, 1127–1150.

Hornstein, A., Krusell, P., & Violante, G. (2005). Unemployment and Vacancy Fluctuations in the Matching Model: Inspecting the Mechanism. *Economic Quarterly*, *91*, 19–51.

Huang, R. (2007). Distance and Trade: Disentangling Unfamiliarity Effects and Transport Cost Effects. *European Economic Review*, *51*, 161–181.

Huang, R. (2008). Tolerance for Uncertainty and the Growth of Informationally Opaque Industries. *Journal of Development Economics*, *87*, 333–353.

Hundley, G. (2001). Why and When are the Self-employed More Satisfied with Their Work? *Industrial Relations*, *40*, 293–316.

Hundt, C. & Sternberg, R. (2014a). Explaining New Firm Creation in Europe from a Spatial and Time Perspective: A Multilevel Analysis based upon Data of Individuals, Regions and Countries. *Papers in Regional Science*, *95*, 223–257.

Hundt, C. & Sternberg, R. (2014b). How Did the Economic Crisis Influence New Firm Creation? *Journal of Economics and Statistics*, *234*, 722–756.

Hyytinen, A., Ilmakunnas, P., & Toivanen, O. (2013). The Return-to-entrepreneurship Puzzle. *Labour Economics*, *20*, 57–67.

Hyytinen, A. & Rouvinen, P. (2008). The Labour Market Consequences of Self-employment Spells: European Evidence. *Labour Economics*, *15*, 246–271.

Inklaar, R. & Yang, J. (2012). The Impact of Financial Crises and Tolerance for Uncertainty. *Journal of Development Economics*, *97*, 466–480.

Jagannathan, R. & Ma, T. (2003). Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps. *Journal of Finance*, *58*, 1651–1683.

Jang, K., Livesley, W., Angleitner, A., Riemann, R., Ando, J., Ono, Y., Vernon, P., & Hamer, D. (2001). Covariance Structure of Neuroticism and Agreeableness: A Twin and Molecular Genetic Analysis of the Role of the Serotonin Transporter Gene. *Journal of Personality and Social Psychology*, *81*, 295–304.

John, O. & Srivastava, S. (2001). The Big-Five Trait Taxonomy. In L. Pervin & O. John (Eds.), *Handbook of Personality* (2. ed.). chapter 4, (pp. 102–138). New York: Guilford Press.

Jokela, M. (2009). Personality Predicts Migration within and between U.S. States. *Journal of Research in Personality*, *43*, 79–83.

Jovanovic, B. (1979). Job Search and the Theory of Turnover. *Journal of Political Economy*, *87*, 972–990.

Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica*, *50*, 649–670.

Kahn, B. & Sarin, R. (1988). Modelling Ambiguity in Decisions Under Uncertainty. *Journal of Consumer Research*, *15*, 265–272.

Kalleberg, A. L. & Leicht, K. T. (1991). Gender and Organizational Performance: Determinanst of Small Business Survival and Success. *Academy of Management Journal*, *34*, 136–161.

Karlan, D. & Valdivia, M. (2011). Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions. *Review of Economics and Statistics*, *93*, 510–527.

Kawaguchi, D. (2003). Human Capital Accumulation of Salaried and Self-employed Workers. *Labour Economics*, *10*, 55–71.

Kelley, D., Brush, C., Greene, P., & Litovsky, Y. (2013). *Global Entrepreneurship Monitor: 2012 Women's Report*. London: London Business School.

Kerr, W., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as Experimentation. *Journal of Economic Perspectives*, *28*, 25–48.

Kihlstrom, R. & Laffont, J.-J. (1979). A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion. *Journal of Political Economy*, *87*, 719–748.

Klein, D., Kotov, R., & Bufferd, S. (2011). Personality and Depression: Explanatory Models and Review of the Evidence. *Annual Review of Clinical Psychology*, *7*, 269–295.

Klibanoff, P., Marinacci, M., & Mukerji, S. (2005). A Smooth Model of Decision Making under Ambiguity. *Econometrica*, *73*, 1849–1892.

Knaup, A. (2005). Survival and Longevity in the Business Employment Dynamics Data. *Monthly Labor Review*, *May 2005*, 50–56.

Knaup, A. & Piazza, M. (2007). Business Employment Dynamics. *Monthly Labor Review*, *September 2007*, 3–10.

Knight, F. (1921). *Risk, Uncertainty, and Profit*. New York: Houghton Mifflin.

Koellinger, P., Minniti, M., & Schade, C. (2007). I Think I Can, I Think I Can. *Journal of Economic Psychology*, *28*, 502–527.

Koellinger, P. & Thurik, A. (2012). Entrepreneurship and the Business Cycle. *Review of Economics and Statistics*, *94*, 1143–1156.

Komarraju, M., Karau, S., Schmeck, R., & Avdic, A. (2011). The Big Five Personality Traits, Learning Styles, and Academic Achievement. *Personality and Individual Differences*, *51*, 472–477.

Konon, A. (2016). Career Choice under Ambiguity. Paper presented at the Annual Conference of the German Economic Association (Verein für Socialpolitik) 2016.

Kosmidis, I. & Firth, D. (2010). A Generic Algorithm for Reducing Bias in Parametric Estimation. *Electronic Journal of Statistics*, *4*, 1097–1112.

Kritikos, A. (2014). Entrepreneurs and Their Impact on Jobs and Economic Growth. *IZA World of Labor*, *8*.

Kull, S., Ramsay, C., & Lewis, E. (2003). Misperceptions, the Media, and the Iraq War. *Political Science Quarterly*, *118*, 569–598.

La Ferrara, E., Chong, A., & Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*, *4*, 1–31.

Lazear, E. (2005). Entrepreneurship. *Journal of Labor Economics*, *23*, 649–680.

Lee, Y. & Mukoyama, T. (2015). Entry and Exit of Manufacturing Plants over the Business Cycle. *European Economic Review*, *77*, 20–27.

Levine, R. & Rubinstein, Y. (2016). Smart and Illicit: Who Becomes an Entrepreneur and Do They Earn More? *Quaterly Journal of Economics*. forthcoming.

Lewbel, A. (2012). Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. *Journal of Business and Economic Statistics*, *30*, 67–80.

Lewis, V. (2009). Business Cycle Evidence on Firm Entry. *Macroeconomic Dynamics*, *13*, 605–624.

Lloyd-Ellis, H. & Bernhardt, D. (2000). Enterprise, Inequality and Economic Development . *Review of Economic Studies*, *67*, 147–168.

Lopez, J. & Servén, L. (2006). A Normal Relationship? Poverty, Growth and Inequality. *World Bank Policy Research Working Paper*, *3814*.

Lowe, R. & Ziedonis, A. (2006). Overoptimism and the Performance of Entrepreneurial Firms. *Management Science*, *52*, 173–186.

Lucas, R. (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics*, *9*, 508–523.

Maccheroni, F., Marinacci, M., & Ruffino, D. (2013). Alpha as Ambiguity: Robust Mean-variance Portfolio Analysis. *Econometrica*, *81*, 1075–1113.

Maccheroni, F., Marinacci, M., & Rustichini, A. (2006). Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica*, *74*, 1447–1498.

Maccheroni, F., Marinacci, M., Rustichini, A., & Taboga, M. (2009). Portfolio Selection with Monotone Mean-variance preferences. *Mathematical Finance*, *19*, 487–521.

MacDonald, G. (1988). The Economics of Rising Stars. *American Economic Review*, *78*, 155–166.

Magnusson, D. & Törestad, B. (1993). A Holistic View of Personality: A Model Revisited. *Annual Review of Psychology*, *44*, 427–452.

Manso, G. (2016). Experimentation and the Returns to Entrepreneurship. *Review of Financial Studies*. forthcoming.

Mansour, H. (2012). Does Employer Learning Vary by Occupation? *Journal of Labor Economics*, *30*, 415–444.

Manuck, S., Flory, J., McCaffery, J., Matthews, K., Mann, J., & Muldoon, M. (1998). Aggression, Impulsivity, and Central Nervous System Serotonergic Responsivity in a Nonpatient Sample. *Neuropsychopharmacology*, *19*, 287–299.

Mardia, K. (1974). Applications of Some Measures of Multivariate Skewness and Kurtosis for Testing Normality and Robustness Studies. *Sankhy A*, *36*, 115–128.

Maritz, A. (2004). New Zealand Necessity Entrepreneurs. *International Journal of Entrepreneurship and Small Business*, *1*, 255–264.

Markman, G. & Baron, R. (2003). Person–entrepreneurship Fit: Why Some People are More Successful as Entrepreneurs than Others. *Human Resource Management Review*, *13*, 281–301.

Markman, G., Baron, R., & Balkin, D. (2005). Are Perseverance and Self-efficacy Costless? Assessing Entrepreneurs' Regretful Thinking. *Journal of Organization Behavior*, *26*, 1–19.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance*, *7*, 77–91.

Mayer, J. & Skimmyhorn, W. (2017). Personality Attributes that Predict Cadet Performance at West Point. *Journal of Research in Personality*, *66*, 14–26.

McCrae, R. & John, O. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*, 175–215.

Meyer, R. & Shi, Y. (1995). Sequential Choice under Ambiguity. *Management Science*, *41*, 817–834.

Miller, R. (1984). Job Matching and Occupational Choice. *Journal of Political Economy*, *92*, 1086–1120.

Minnesota Population Center and State Health Access Data Assistance Center (2016). *Integrated Health Interview Series: Version 6.21*. Minneapolis: University of Minnesota. www.ihis.us.

Minniti, M. (2005). Entrepreneurship and Network Externalities. *Journal of Economic Behavior and Organization*, *57*, 1–27.

Mise, E., Kim, T.-H., & Newbold, P. (2005). On Suboptimality of the Hodrick-Prescott Filter at Time Series Endpoints. *Journal of Macroeconomics*, *27*, 53–67.

Montoya, L. & de Haan, J. (2008). Regional Business Cycle Synchronization in Europe? *International Economics and Economic Policy*, *5*, 123–137.

Mortensen, D. & Pissarides, C. (1994). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies*, *61*, 397–415.

Muehlfeld, K., Urbig, D., & Weitzel, U. (2017). Entrepreneurs' Exploratory Perseverance in Learning Settings. *Entrepreneurship Theory and Practice*. forthcoming.

Newey, W. (1987). Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables. *Journal of Econometrics*, *36*, 231–250.

Nickell, S. (1996). Competition and Corporate Performance. *Journal of Political Economy*, *104*, 724–746.

Obschonka, M., Schmitt-Rodermund, E., Silbereisen, R., Goslin, S., & Potter, J. (2013). The Regional Distribution and Correlates of an Entrepreneurship-prone Personality Profile in the United States, Germany, and the United Kingdom: A Socioecological Perspective. *Journal of Personality and Social Psychology*, *105*, 104–122.

Obschonka, M., Silbereisen, R., & Schmitt-Rodermund, E. (2010). Entrepreneurial Intentions as Development Outcome. *Journal of Vocational Behavior*, *77*, 63–72.

OECD (2005). *OECD Handbook on Economic Globalization Indicators*. Paris: OECD Publishing.

OECD (2015). *Entrepreneurship at a Glance 2015*. Paris: OECD Publishing.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.

Parker, C. (2012a). Theories of Entrepreneurship, Innovation, and the Business Cycle. *Journal of Economic Surveys*, *26*, 377–394.

Parker, S. (2009). Why Do Small Firms Produce the Entrepreneurs. *Journal of Socio-Economics*, *38*, 484–494.

Parker, S. (2012b). *The Economics of Entrepreneurship*. Cambridge: Cambridge University Press.

Paul, K. & Moser, K. (2009). Unemployment Impairs Mental Health. *Journal of Vocational Behavior*, *74*, 264–282.

Poschke, M. (2013). Who Becomes an Entrepreneur? Labor Market Prospects and Occupational Choice. *Journal of Economic Dynamics and Control*, *37*, 693–710.

Quatraro, F. & Vivarelli, M. (2015). Drivers of Entrepreneurship and Post-entry Performance of Newborn Firms in Developing Countries. *The World Bank Research Observer*, *30*, 277–305.

Ramírez, A. & Tadesse, S. (2009). Corporate Cash Holdings, Uncertainty Avoidance, and the Multinationality of Firms. *International Business Review*, *18*, 387–403.

Rampini, A. (2004). Entrepreneurial Activity, Risk, and the Business Cycle. *Journal of Monetary Economics*, *51*, 555–573.

Rauch, A. & Frese, M. (2007). Let's Put the Person back into Entrepreneurship Research. *European Journal of Work and Organizational Psychology*, *16*, 353–385.

Ravn, M. & Uhlig, H. (2002). On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. *Review of Economics and Statistics*, *84*, 371–376.

Robson, M. (1998). Self-Employment in the UK Regions. *Applied Economics*, *30*, 313–322.

Rodionova, Z. (2015). The Personality Test that Could One Day Decide your Credit Score. *The Independent*, *2. September*.

Roessler, C. & Koellinger, P. (2012). Entrepreneurship and Organization Design. *European Economic Review*, *56*, 888–902.

Román, C., Congregado, E., & Millán, J. (2013). Start-up Incentives: Entrepreneurship Policy or Active Labour Market Programme? *Journal of Business Venturing*, *28*, 151–175.

Rosenfeld, D. (2005). Höchste Arbeitslosigkeit seit der Wiedervereinigung. *Der Tagesspiegel*. Published on January 5, Retrieved from http://www.tagesspiegel.de.

Rothschild, M. (1974). A Two-Armed Bandit Theory of Market Pricing. *Journal of Economic Theory*, *9*, 185–202.

Sanchis Llopis, J., Millán, J., Baptista, R., Burke, A., Parker, C., & Thurik, R. (2015). Good Times, Bad Times: Entrepreneurship and the Business Cycle. *International Entrepreneurship and Management Journal*, *11*, 243–251.

Savage, L. (1954). *The Foundations of Statistics*. New York: Wiley.

Schlicht, E. (2005). Estimating the Smoothing Parameter in the So-called Hodrick-Prescott Filter. *Journal of Japanese Statistical Society*, *35*, 99–119.

Schmeidler, D. (1989). Subjective Probability and Expected Utility without Additivity. *Econometrica*, *57*, 571–587.

Schmitt-Rodermund, E. (2004). Pathways to Successful Entrepreneurship: Parenting, Personality, Entrepreneurial Competence, and Interest. *Journal of Vocational Behavior*, *65*, 498–518.

Schumpeter, J. (1934). *The Theory of Economic Development*. Cambridge: Harvard Univ. Press.

Sedlácek, P. & Sterk, V. (2017). The Growth Potential of Startups over the Business Cycle. *American Economic Review*. forthcoming.

Shane, S. (2001). Technology Regimes and New Firm Formation. *Management Science*, *47*, 1173–1181.

Shaver, K. & Scott, L. (1991). Person, Process, Choice: The Psychology of New Venture Creation. *Entrepreneurship Theory & Practice*, *16*, 23–45.

Sherman, R. (1974). The Psychological Difference Between Ambiguity and Risk. *Quarterly Journal of Economics*, *88*, 166–169.

Siemer, M. (2014). Firm Entry and Employment Dynamics in the Great Recession. *FEDS Discussion Paper*, *2014-56*.

Simon, H. (1959). Theories of Decision-Making in Economics and Behavioral Science. *American Economics Review*, *9*, 253–283.

Siniscalchi, M. (2009). Vector Expected Utility and Attitudes Toward Variation Authors. *Econometrica*, *77*, 801–855.

Soroka, S. & McAdams, S. (2015). News, Politics, and Negativity. *Political Communication*, *32*, 1–22.

Soto, C., John, O., Gosling, S., & Potter, J. (2011). Age Differences in Personality Traits from 10 to 65. *Journal of Personality and Social Psychology*, *100*, 330–348.

Spengler, A. (2008). The Establishment History Panel. *Journal of Applied Social Science Studies*, *128*, 501–509.

Spiegel (2005). Höchste Jahresarbeitslosigkeit seit der Wiedervereinigung. *Der Spiegel*. Published on January 4, Retrieved from http://www.spiegel.de.

Staiger, D. & Stock, J. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, *65*, 557–586.

Statistisches Bundesamt a (various volumes). *Volkswirtschaftliche Gesamtrechnung der Länder*. Wiesbaden: Statistisches Bundesamt. http://www.vgrdl.de/VGRdL/tbls/home.asp?lang=en-GB.

Statistisches Bundesamt b (various volumes). *Fachserie 11 – Bildung und Kultur*. Wiesbaden: Statistisches Bundesamt.

Stewart, W. & Roth, P. (2001). Risk Propensity Differences between Entrepreneurs and Managers. *Journal of Applied Psychology*, *86*, 145–153.

Steyvers, M., Lee, M., & Wagenmakers, E.-J. (2009). A Bayesian Analysis of Human Decision-making on Bandit Problems. *Journal of Mathematical Psychology*, *53*, 168–179.

Stock, J. H. & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. Andrews & J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). Cambridge: Cambridge University Press.

Strömberg, D. (2004). Radio's Impact on Public Spending. *Quarterly Journal of Economics*, *119*, 189–221.

Stuetzer, M., Goethner, M., & Cantner, U. (2012). Do Balanced Skills Help Nascent Entrepreneurs to Make Progress in the Venture Creation Process? *Economics Letters*, *117*, 186–188.

Sullivan, J. & Peterson, R. (1991). A Test of Theories Underlying the Japanese Lifetime Employment System. *Journal of International Business Studies*, *22*, 79–97.

Sullivan, S. & Baruch, Y. (2009). Advances in Career Theory and Research: A Critical Review and Agenda for Future Exploration. *Journal of Management*, *35*, 1542–1571.

Taylor, M. (1996). Earnings, Independence or Unemployment: Why Become Self-Employed? *Oxford Bulletin of Economics and Statistics*, *58*, 253–266.

Terza, J., Basu, A., & Rathouz, P. (2008). Two-stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling. *Journal of Health Economics*, *27*, 531–543.

The Economist (1997). Adventures with Capital. *The Economist*, *25 January 2016*, p. 17.

Thimme, J. & Völkert, C. (2015). Ambiguity in the Cross-Section of Expected Returns: An Empirical Assessment. *Journal of Business and Economic Statistics*, *33*, 418–429.

Tobin, J. (1958). Liquidity Preference as Behavior Towards Risk. *The Review of Economic Studies*, *25*, 65–86.

Valcour, P. & Tolbert, P. (2003). Gender, Family and Career in the Era of Boundarylessness: Determinants and Effects of Intra- and Inter-organizational Mobility. *International Journal of Human Resource Management*, *14*, 768–787.

Verbruggen, M., van Emmerik, H., Van Gils, A., Meng, C., & de Grip, A. (2015). Does Early-career Underemployment Impact Future Career Success? A Path Dependency Perspective. *Journal of Vocational Behavior*, *90*, 101–110.

Vereshchagina, G. & Hopenhayn, H. (2009). Risk Taking by Entrepreneurs. *American Economic Review*, *99*, 1808–1830.

Wagner, J. (2006). Are Nascent Entrepreneurs 'Jacks-Of-All-Trades'? A Test Of Lazear's Theory of Entrepreneurship with German Data. *Applied Economics*, *38*, 2415–2419.

Wagner, J. (2007). What a Difference a Y Makes: Female and Male Nascent Entrepreneurs in Germany. *Small Business Economics*, *28*, 1–21.

Weber, R. (1992). On the Gittins Index for Multiarmed Bandits. *The Annals of Applied Probability*, *2*, 1024–1033.

White, R., Thornhill, S., & Hampson, E. (2006). Entrepreneurs and Evolutionary Biology. *Organizational Behavior and Human Decision Processes*, *100*, 21–34.

Williams, D. (2000). Consequences of Self-employment for Women and Men in the United States. *Labour Economics*, *7*, 665–687.

Wolfe, M. & Patel, P. (2016). Grit and Self-employment: A Multi-country Study. *Small Business Economics*, *47*, 853–874.

World Bank (2015). http://data.worldbank.org/.

World Bank Group (2015). www.doingbusiness.org/methodology.

Yatchew, A. & Griliches, Z. (1985). Specification Error in Probit Models. *Review of Economics and Statistics*, *67*, 134–139.

Zhang, L. (2003). Does the Big Five Predict Learning Approaches? *Personality and Individual Differences*, *34*, 1431–1446.

Zhao, H. & Seibert, S. (2006). The Big Five Personality Dimensions and Entrepreneurial Status. *Journal of Applied Psychology*, *91*, 259–271.

Zhao, H., Seibert, S., & Lumpkin, G. (2010). The Relationship of Personality to Entrepreneurial Intentions and Performance: A Meta-Analytic Study. *Journal of Management*, *36*, 381–404.

Zingales, L. (2000). In Search of New Foundations. *The Journal of Finance*, *55*, 1623–1653.

Declarations of co-authorship

Declaration of co-authorship for Chapter 2 of Alexander Konon's thesis

I, Alexander Kritikos, co-authored Chapter 2 of this thesis and the corresponding paper. My contribution to the paper consisted of many formal and informal discussions about the topic of entrepreneurial decision processes under imperfect information, the formulation of the research question, the role of personality in entrepreneurial decisions, and the interpretation of the results. We discussed several versions of the chapter and decided upon various changes.

Alexander Konon made important and independent contributions to all parts of the paper including, but not limited to, the construction of the model, deriving analytical results, as well as providing and simulating a version of the theoretical model, which relies on less assumptions. He provided drafts of the manuscript and contributed substantially to its finalization. His overall contribution share is 65%.

Alexander Kritikos

Place/Date: Berlin, 6.12.2017

Declaration of co-authorship for Chapter 4 of Alexander Konon's thesis

I, Alexander Kritikos, co-authored Chapter 4 of this thesis and the corresponding paper. My contribution to the paper consisted of many formal and informal discussions about the topic of the effects of media on occupational choice, the formulation of the research question, the role of risk and ambiguity in occupational choice, and the interpretation of the results. We discussed several versions of the chapter and decided upon various changes.

Alexander Konon made important and independent contributions to all parts of the paper including, but not limited to, the construction of the theoretical model, deriving analytical results, providing a simulation of the model, and conducting an empirical analysis at the micro and macro level. He provided drafts of the manuscript and contributed substantially to its finalization. His overall contribution share is 75%.

Alexander Kritikos

Place/Date: Berlin, 6.12.2017

Declaration of co-authorship for Chapter 5 of Alexander Konon's thesis

We, Michael Fritsch and Alexander Kritikos, co-authored Chapter 5 of this thesis and the corresponding paper. Our contribution to the paper consisted of many formal and informal discussions about the topic of the effect of business cycles on business venturing, the formulation of the research question, the role of different types of businesses, and the interpretation of the results. We discussed several versions of the chapter and decided upon various changes.

Alexander Konon made important and independent contributions to all parts of the paper including, but not limited to, the construction of the theoretical model, deriving analytical results, type-specific empirical analysis, and robustness checks. He provided drafts of the manuscript and contributed substantially to its finalization. His overall contribution share is 40%.

Michael Fritsch

Place/Date: Jena, 5.12.2017

Prof. Dr. Michael Fritsch
Chair of Business Dynamics,
Innovation, and Economic Change
Friedrich-Schiller-University Jena
Faculty of Economics and Business Administration
Carl-Zeiss-Straße 3 · D-07743 Jena · Germany

Alexander Kritikos

Place/Date: Berlin, 6.12.2017

Declaration of co-authorship for Chapter 6 of Alexander Konon's thesis

I, Alexander Kritikos, co-authored Chapter 6 of this thesis and the corresponding paper. My contribution to the paper consisted of many formal and informal discussions about the topic of a proper recommendation approach for individuals intending to venture an own business, the formulation of the research question, the effect of personality on business success, and the interpretation of the results. We discussed several versions of the chapter and decided upon various changes.

Alexander Konon made important and independent contributions to all parts of the paper including, but not limited to, the formulation of the formal problem, deriving analytical results, estimation and calibration of the theoretical model and analysis, and simulations to check the robustness of the results. He provided drafts of the manuscript and contributed substantially to its finalization. His overall contribution share is 60%.

Alexander Kritikos

Place/Date: Berlin, 6.12.2012

Declaration of academic honesty

Eidesstattliche Erklärung und Einverständniserklärung nach § 7 Abs. 2 Nr. 6 der Promotionsordnung der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Potsdam vom 27. August 2002, geändert durch deren erste und zweite Satzung zur Änderung vom 29. Februar 2012 und 20. Juni 2012:

Hiermit versichere ich an Eides statt, dass

- meine hinsichtlich der früheren Teilnahme an Promotionsverfahren gemachten Angaben richtig sind;
- die eingereichte Arbeit oder wesentliche Teile derselben in keinem anderen Verfahren zur Erlangung eines akademischen Grades vorgelegt worden sind;
- bei der Anfertigung der Dissertation die Grundsätze zur Sicherung guter wissenschaftlicher Praxis der DFG eingehalten wurden, die Dissertation selbständig und ohne fremde Hilfe verfasst wurde, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt worden sind und die den benutzten Werken wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht wurden.

Einer Überprüfung der eingereichten Dissertationsschrift bzw. der an deren Stelle eingereichten Schriften mittels einer Plagiatssoftware stimme ich zu.

Ort/Datum:

Unterschrift: