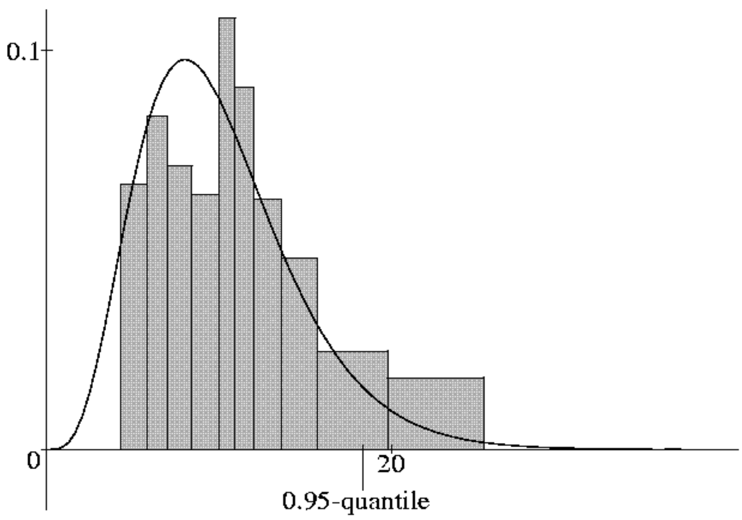
UNIVERSITÄT POTSDAM
Institut für Mathematik

Empirical Minimax Linear Estimates

Henning Läuter



Mathematische Statistik und
Wahrscheinlichkeitstheorie

Universität Potsdam – Institut für Mathematik

Mathematische Statistik und Wahrscheinlichkeitstheorie

Empirical Minimax Linear Estimates

Henning Läuter

Institute of Mathematics, University of Potsdam
e-mail: laeuter@uni-potsdam.de

Preprint 2008/06

November 2008

Impressum

Herausgeber: Mathematische Statistik und Wahrscheinlichkeitstheorie
am Institut für Mathematik

Adresse: Universität Potsdam
Am Neuen Palais 10
14469 Potsdam
Telefon: +49-331-977 1500
Fax: +49-331-977 1578
E-mail: neisse@math.uni-potsdam.de

ISSN 1613-3307

Empirical Minimax Linear Estimates

Henning Läuter
Institute of Mathematics, University of Potsdam, Germany
e-mail: laeuter@uni-potsdam.de

Abstract: We give the explicit solution for the minimax linear estimate. For scale dependent models an empirical minimax linear estimates is defined and we prove that these estimates are Stein's estimates.

Key words and phrases: Minimax linear estimates, admissibility, empirical estimates
AMS subject classification: Primary 62F30; Secondary 62F15,62F10

1 Introduction

In the past years some progress has been made with the study of minimax linear estimates in convex linear models. Considering linear models

$$Y = X\beta + \varepsilon, \tag{1.1}$$

$$\mathsf{E}\varepsilon = 0, \quad \mathsf{Var}\, Y = \sigma^2 I$$

with the condition

$$(\beta, \sigma^2) \in \mathcal{B} = \{(\beta, \sigma^2) : (\beta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}^+, \; \beta^t X^t M X \beta \le k^2 \sigma^2\}, \tag{1.2}$$

where the nonrandom design matrix X is an $n \times p$ matrix, i.e. $X \in \mathfrak{M}_{n\times p}$. M is a known positive semidefinite matrix such that $H = X^t M X$ has a full rank. $\beta \in \mathbb{R}^p$ and σ^2 are unknown parameters. Our aim is to estimate the parameter $\gamma = CX\beta + c$ with the $s \times n$ matrix C and $c \in \mathbb{R}^s$. In the class of linear estimators $\hat{\gamma}$ with

$$\hat{\gamma}(y) = Ty + t, \quad T \in \mathfrak{M}_{s\times n}, \quad t \in \mathbb{R}^s$$

we look for a minimax linear estimate. With the risk

$$R(T,t;\beta,\sigma^2) = \frac{1}{\sigma^2}\mathsf{E}_{(\beta,\sigma^2)}\,(\hat{\gamma}(Y)-\gamma)^t\,Z\,(\hat{\gamma}(Y)-\gamma),$$

where Z is a positive semidefinite $s\times s$ matrix, an estimate γ^* and $\gamma^*(y) = T^*y + t^*$ is called a minimax linear estimate (MILE) if

$$\sup_{\mathcal{B}} R(T^*,t^*;\beta,\sigma^2) = \min_{T,t}\sup_{\mathcal{B}} R(T,t;\beta,\sigma^2).$$

This problem was discussed and solved by J. Kuks & V. Olman (1972) for $rank\, Z = 1$. For general Z a characterization of the MILE was given by H. Läuter (1975) and K. Hoffmann (1978). The equivalence with a spectral characterization was discussed in H. Drygas &H. Läuter (1994). H. Drygas and J. Pilz (1996) and also V.L. Girko (1996) showed the equivalence of spectral theory and bayesian analysis in minimax estimation problems. In N. Gaffke & B. Heiligers (1989) and J. Pilz (1991) several special cases are discussed. B.F. Arnold & P. Stahlecker (2000) determined for another objective function the minimax solution. An explicit representation of the MILE was not known up to now. The present paper is organized as follows. In the next chapter we formulate the explicit general solution for the MILE. As a consequence we give an empirical MILE for scale dependent models (1.2). Here the term "empirical" is used in the similar sense as in empirical Bayes estimates. The risk of such estimates is calculated and we see that under normal distributions some of these estimates are Stein estimates.

2 Representation of the MILE

Let us consider the scale invariant model (1.1) with (1.2) and the parameter $\gamma = CX\beta + c$ is to be estimated. In Läuter (1975) it was remarked that

$$\chi(T) := \sup_{\mathcal{B}_1} R(T^*,t^*;\beta,\sigma^2) = \operatorname{tr} ZTT^t + k^2\,\lambda_{\max}[Z^{\frac{1}{2}}(TX-CX)H^{-1}(TX-CX)^tZ^{\frac{1}{2}}],$$

where $\lambda_{\max}(A)$ is the largest eigenvalue of A. Let $\mathcal{R}(A)$ be the space spanned by the column vectors of the matrix A and $P_{\mathcal{R}(A)}$ denotes the orthogonal projection on $\mathcal{R}(A)$. For a symmetric matrix A with the spectral representation

$$A = \sum_{i=1}^{k}\mu_i a_i {a_i}^t, \quad \mu_1 \ge \mu_2 \ge ... \ge \mu_l > 0 \ge \mu_{l+1} \ge ... \ge \mu_k$$

we denote by $A^>$ the positive part

$$A^> = \sum_{i=1}^{l}\mu_i a_i {a_i}^t.$$

With

$$V = k^2 H^{-\frac{1}{2}} X^t X H^{-\frac{1}{2}} =: k^2 \tilde{V}, \tag{2.3}$$

$$G = \frac{1}{k^2} H^{\frac{1}{2}} (X^t X)^{-1} X^t C^t Z C X (X^t X)^{-1} H^{\frac{1}{2}} =: \frac{1}{k^2} \tilde{G} \tag{2.4}$$

we denote

$$\tilde{A}_\eta = \tilde{V}^{-1/2} (\eta \tilde{V}^{-1} - \tilde{V}^{1/2} \tilde{G} \tilde{V}^{1/2})^{>} \tilde{V}^{-1/2} \tag{2.5}$$

and

$$A_\rho := \frac{1}{k^2} \tilde{A}_{\frac{\rho}{k^2}}. \tag{2.6}$$

Obviously $A_\rho \in \mathfrak{M}_{p \times p}$ is positive semidefinite (denoted by $A_\rho \geq 0$). Then the following theorem gives the general representation of the MILE.

Theorem 1 *1. There exists a $\rho_0 \geq 0$ such that*

$$\text{tr}\,[(G + A_{\rho_0})^{1/2}] = \sqrt{\rho_0}(1 + \text{tr}\,(V^{-1})). \tag{2.7}$$

2. With

$$D_{\rho_0} := \frac{1}{\sqrt{\rho_0}} (G + A_{\rho_0})^{1/2} - V^{-1}$$

and

$$T^* = k^2\, C X H^{-\frac{1}{2}} D_{\rho_0} H^{-\frac{1}{2}} X^t (I + k^2\, X H^{-\frac{1}{2}} D_{\rho_0} H^{-\frac{1}{2}} X^t)^{-1} \tag{2.8}$$

the estimator γ^* with

$$\gamma^*(y) = T^* y + c$$

is a MILE for γ.

Now the problem is reduced to find a solution of (2.7) for the given A_ρ and this is easy to determine. The restriction to linear estimates is not really so restrictive. This is justified by the maximality of the normal distribution in the sense of the risk for the best unbiased estimate (BUE) being maximal under normality and the BUE being linear.

From (2.8) it becomes clear that the MILE is a type of shrinkage estimate. With the orthonormal eigenvectors $e_1, ...e_p \in \mathcal{R}(X)$ of $k^2\, X H^{-\frac{1}{2}} D_{\rho_0} H^{-\frac{1}{2}} X^t$ and corresponding eigenvalues $\lambda_1, ..., \lambda_p$ we have

$$T^* = C P_{\mathcal{R}(X)} - C \sum_{i=1}^{p} \frac{1}{1 + \lambda_i} e_i e_i^t. \tag{2.9}$$

From the representation

$$D_{\rho_0} = \frac{1}{\sqrt{\rho_0}} \frac{1}{k} (\tilde{G} + \tilde{A}_{\frac{\rho_0}{k^2}})^{1/2} - \frac{1}{k^2} \tilde{V}^{-1}$$

we see the explicit dependence on k and ρ_0. This can be used for the calculation of the λ_i for different k and ρ_0.

2.1 MILE for commutable $\tilde{G}, \tilde{V}$

Assuming that $\tilde{G}, \tilde{V}$ are commutable, i.e. they have the same eigenvectors, say $h_1, ..., h_p$. $\tilde{G}$ has the eigenvalues $\gamma, ..., \gamma_p$ and $\tilde{V}$ has the eigenvalues $\nu_1, ..., \nu_p$. Then A_ρ from (2.6) has the eigenvalues

$$\frac{1}{k^2\nu_i}(\frac{\rho}{k^2\nu_i} - \gamma_i\nu_i)_+, \quad i = 1, ..., p$$

with $a_+ := \max(0, a)$. Consequently D_ρ has the eigenvalues

$$\frac{1}{\sqrt{\rho}\, k}\sqrt{\gamma_i + \frac{1}{\nu_i}(\frac{\rho}{k^2\nu_i} - \gamma_i\nu_i)_+} - \frac{1}{k^2\nu_i} = \frac{1}{k\nu_i}\Big(\frac{\nu_i\sqrt{\gamma_i}}{\sqrt{\rho}} - \frac{1}{k}\Big)_+. \tag{2.10}$$

Hence we have the spectral representation

$$k^2\, XH^{-\frac{1}{2}}D_\rho H^{-\frac{1}{2}}X^t = \sum_{i=1}^{p} \lambda_i e_i\, e_i^t \tag{2.11}$$

with

$$\lambda_i = \frac{k\nu_i}{\sqrt{\rho}}\sqrt{\gamma_i + (\frac{\rho}{k^2\nu_i^2} - \gamma_i)_+} - 1 = \Big(\frac{k\nu_i\sqrt{\gamma_i}}{\sqrt{\rho}} - 1\Big)_+. \tag{2.12}$$

Theorem 2 *For the unique ρ_0 with $trD_{\rho_0} = 1$ we have*

$$T^* = CP_{\mathcal{R}(X)} - C\sum_{i=1}^{p} \frac{1}{1 + \Big(\frac{k\nu_i\sqrt{\gamma_i}}{\sqrt{\rho_0}} - 1\Big)_+}\, e_i\, e_i^t.$$

Proof : From (2.10) follows that

$$trD_\rho = \sum_{i=1}^{p} \frac{1}{k\nu_i}\Big(\frac{\sqrt{\gamma_i\nu_i^2}}{\sqrt{\rho}} - \frac{1}{k}\Big)_+$$

is a monotone decreasing function in ρ and for $\rho \leq \max\{1/(\gamma_i k^2\nu_i^2), i = 1, ..., p\}$ it is continuous and strictly monotone from ∞ to 0. Hence, ρ_0 with $trD_\rho = 1$ exists and it is unique. Now from (2.9) the assertion follows.

□

3 Empirical minimax linear estimate

By construction it follows from (2.7) that ρ_0 depends on k, consequently $\rho_0 = \rho_0(k)$. If k is unknown then the MILE is also unknown. If one substitutes an estimation $\hat{\kappa}$ for k we get empirical minimax linear estimates. Using (2.9) then λ_i and e_i depend on k. If we substitute $\hat{\kappa}$ instead of k directly in (2.8) then we get

$$\hat{T}^*y = \hat{\kappa}^2\, CXH^{-\frac{1}{2}}D_{\rho_0(\hat{\kappa})}H^{-\frac{1}{2}}X'(I + \hat{\kappa}^2 XH^{-\frac{1}{2}}D_{\rho_0(\hat{\kappa})}H^{-\frac{1}{2}}X')^{-1}y. \tag{3.13}$$

More generally

$$\hat{T}^* = CP_{\mathcal{R}(X)} - C\sum_{i=1}^{p} \frac{\mu_i}{1+\hat{\lambda}_i}\, \hat{e}_i\, \hat{e}_i^t \tag{3.14}$$

is an empirical MILE if the μ_i's are positive constants ≤ 1 and $\hat{\lambda}_i$, $\hat{e}_i$ are estimates for the corresponding λ_i, e_i. An appropriate estimation $\hat{\kappa}$ is those with

$$\hat{\kappa}^2 = \eta||M^{1/2}y||^2 = \eta\, y^t\, M\, y \tag{3.15}$$

for a positive constant η. $||z||$ denotes the Euclidean norm of z. We calculate the risk for an empirical MILE.

3.1 Risk of empirical MILE

Assuming that $\tilde{G}, \tilde{V}$ are commutable, i.e. they have the same eigenvectors, say $h_1, ..., h_p$. From section 2.1, it follows that also the h_i's are the eigenvectors from $\tilde{A}_\eta$ and so the eigenvectors $h_1, ..., h_p$ of A_ρ are independent of ρ. These $h_1, ..., h_p$ are also the eigenvectors of D_ρ for any ρ. Consequently $e_1, ...e_p$ with $h_i = H^{-\frac{1}{2}}X^t e_i$ is a system of orthonormal eigenvectors of $XH^{-\frac{1}{2}}D_{\rho_0}H^{-\frac{1}{2}}X^t$. Using (2.10) the empirical MILE (3.13) can be written as

$$\hat{T}^* = CP_{\mathcal{R}(X)} - C\sum_{i=1}^{p} \frac{\mu_i}{1+\left(\frac{\hat{\kappa}\nu_i\sqrt{\gamma_i}}{\sqrt{\rho_0(\hat{\kappa})}} - 1\right)_+}\, e_i\, e_i^t.$$

From (2.10) and the condition $\operatorname{tr} D_{\rho_0} = 1$ it follows that $\sqrt{\rho_0} \approx 1/k$ at least for large k. Then we consider the empirical MILE

$$\hat{T}^* = CP_{\mathcal{R}(X)} - C\sum_{i=1}^{p} \frac{\mu_i}{1+\hat{\lambda}_i}\, e_i\, e_i^t$$

emp3 with

$$\hat{\lambda}_i = \left(\hat{\kappa}^2\, \nu_i\, \sqrt{\gamma_i} - 1\right)_+ \tag{3.16}$$

according to (2.12). This estimation $\hat{T}^*y$ is of the form

$$\hat{T}y = CP_{\mathcal{R}(X)}y - CQy$$

for an $n \times n$ matrix $Q = Q(y^tMy)$ with $Q(y^tMy) \leq P_{\mathcal{R}(X)}$. Denote

$$L = \sum_{i=1}^{p} \gamma_i\, e_i\, e_i^t.$$

Then we get the following representations.

Lemma 1 *We have*

1.

$$MC^tZCQ \;=\; Q^{1/2}LQ^{1/2} \tag{3.17}$$

2.

$$M^{1/2+}QC^tZCM^{1/2} \;=\; \sum_{i=1}^{p} \frac{\gamma_i \mu_i}{1+\hat\lambda_i} M^{1/2+}e_i\, e_i^t M^{1/2+}$$

for orthogonal vectors $M^{1/2+}e_i, \quad i = 1, ..., p.$

3.

$$\operatorname{tr} QC^tZC \;=\; \sum_{i=1}^{p} \frac{\gamma_i \mu_i}{1+\hat\lambda_i}\, e_i^t M^+ e_i \tag{3.18}$$

4.

$$\frac{Y^tQ^{1/2}LQ^{1/2}Y}{Y^tMY} \;\le\; \lambda_{\max}\Big(M^{1/2+}Q^{1/2}LQ^{1/2}M^{1/2+}\Big) = \max_i\{\frac{\gamma_i\mu_i}{1+\hat\lambda_i}\, e_i^tM^+e_i\}$$

5.

$$\frac{Y^tQC^tZCQY}{Y^tMY} \;\le\; \lambda_{\max}\Big(M^{1/2+}QLQM^{1/2+}\Big) = \max_i\{\frac{\gamma_i\mu_i^2}{(1+\hat\lambda_i)^2}\, (e_i^tM^+e_i)^2\}$$

By Stein's identity we get a bound for the risk of the empirical MILE.

Lemma 2 *For* $\mathcal{R}(M) \subseteq \mathcal{R}(X)$, $Y \sim \mathsf{N}_n(X\beta, \sigma^2 I)$ *we have*

$$\begin{aligned}\sigma^2\, R(\theta, \hat T) \le \sigma^2 \operatorname{tr} ZCC^t - 2 \operatorname{tr} \mathsf{E}\, Q(Y^tMY)C^tZC + \\ +4\, \mathsf{E}\, \frac{Y^tQ^{1/2}(Y^tMY)LQ^{1/2}(Y^tMY)Y}{Y^tMY} + \mathsf{E}\, Y^tQ(Y^tMY)C^tZCQ(Y^tMY)Y.\end{aligned} \tag{3.19}$$

The empirical MILE is defined with

$$Q = \sum_{i=1}^{p} \frac{\mu_i}{1+\hat\lambda_i}\, e_i\, e_i^t$$

and $\hat\lambda_i$ as in (3.16). In order to find conditions for an empirical MILE to be minimax it is more convenient to look at

$$\tilde Q(y^tMy) \;=\; Q(y^tMy)\, y^tMy.$$

We see that

$$\tilde Q(y^tMy) \;=\; \sum_{i=1}^{p} \frac{\mu_i\, y^tMy}{1 + (\eta\, y^tMy\, \nu_i \sqrt{\gamma_i} - 1)_+}\, e_i\, \mathrm{e}_i^t$$

is an increasing matrix in y^tMy and it is bounded by

$$\tilde{Q}(y^tMy) \leq \sum_{i=1}^{p} \frac{\mu_i}{\eta\nu_i\sqrt{\gamma_i}}\, e_i\, e_i^t.$$

Then the next theorem gives a condition for the empirical MILE to be minimax.

Theorem 3 *Let be $\mathcal{R}(M) \subseteq \mathcal{R}(X)$ and $Y \sim \mathsf{N}_n(X\beta, \sigma^2 I)$. If*

$$2\mathsf{E}\frac{1}{Y^tMY}\,\mathrm{tr}\, M^{1/2+}\tilde{Q}^{1/2}L\tilde{Q}^{1/2}M^{1/2+} \geq 4\mathsf{E}\frac{1}{Y^tMY}\lambda_{\max}\Big(M^{1/2+}\tilde{Q}^{1/2}L\tilde{Q}^{1/2}M^{1/2+}\Big)+$$
$$+\mathsf{E}\frac{1}{Y^tMY}\lambda_{\max}\Big(M^{1/2+}\tilde{Q}L\tilde{Q}M^{1/2+}\Big) \tag{3.20}$$

then the empirical MILE is minimax.

Here one sees that $p \geq 3$ is necessary because for $p=1$ or $p=2$ we have

$$R := \frac{\mathrm{tr}\, M^{1/2+}\tilde{Q}^{1/2}L\tilde{Q}^{1/2}M^{1/2+}}{\lambda_{\max}\Big(M^{1/2+}\tilde{Q}^{1/2}L\tilde{Q}^{1/2}M^{1/2+}\Big)} \leq 2.$$

And if $R > 2$ then one finds small μ_i's so that (3.20) is fulfilled, because the last term in (3.20) depends on $\tilde{Q}$ in a quadratic way and the others only linearly.
This Theorem 3 gives several possibilities for constructing Stein's estimators. Let $\gamma_1, ..., \gamma_q$ be the eigenvalues larger 0. We put $a_i := e_i^tM^+e_i$ and choose for any positive constant c such values μ_i that

$$\frac{\mu_i\sqrt{\gamma_i}\,a_i}{\nu_i} = c, \quad i = 1, ..., q$$

and we admit only such c that $\mu_i \leq 1$.

Theorem 4 *Let be $\mathcal{R}(M) \subseteq \mathcal{R}(X)$ and $Y \sim \mathsf{N}_n(X\beta, \sigma^2 I)$. If $\nu_1\sqrt{\gamma_1} = \max\{\nu_i\sqrt{\gamma_i}, \quad i = 1, ..., p\}$ and*

$$2\sum_{i=1}^{q}\frac{\nu_i\sqrt{\gamma_i}}{\nu_1\sqrt{\gamma_1}} - 4 \;\geq\; \frac{c}{\eta}\max\{\frac{1}{\gamma_j a_j};\, j = 1, ..., q\} \tag{3.21}$$

then the empirical MILE (3.14) is a Stein's estimator.

One recognizes that only the ratio c/η plays a role. Also we see that under the condition

$$\sum_{i=1}^{q}\frac{\nu_i\sqrt{\gamma_i}}{\nu_1\sqrt{\gamma_1}} \;>\; 2 \tag{3.22}$$

such c/η exist. Obviously this condition can be fulfilled only for $q \geq 3$. This is the same condition as in the other known Stein's estimates.

4 Proofs

Proof of Theorem 1: We have from (2.5) for any ρ

$$A_\rho = V^{-1/2}(\rho V^{-1} - V^{1/2} G V^{1/2})^{>} V^{-1/2} \geq 0.$$

Moreover we have the spectral representation

$$\rho V^{-1} - V^{1/2} G V^{1/2} = \sum_{i=1}^{p} \mu_i f_i f_i^t, \quad \mu_1 \geq \mu_2 \geq ... \geq \mu_l > 0 \geq \mu_{l+1} \geq ... \geq \mu_p$$

for orthonormal eigenvectors f_i and the corresponding eigenvalues μ_i. Consequently we have

$$A_\rho = \rho V^{-2} - G - \sum_{i=l+1}^{p} \mu_i V^{-1/2} f_i f_i^t V^{-1/2}.$$

From this follows

$$\frac{1}{\rho}(G + A_\rho) \geq V^{-2}.$$

Therefore, a positive semidefinite matrix B exists with

$$\frac{1}{\sqrt{\rho}}(G + A_\rho)^{1/2} = V^{-1} + B. \tag{4.23}$$

For

$$C := \frac{1}{\sqrt{\rho}}(G + A_\rho)^{1/2} V A_\rho$$

we get $C = A_\rho$, because

$$C^t C = \frac{1}{\rho} A_\rho V (G + A_\rho) V A_\rho = A_\rho^2$$

and for any $c \in \mathbb{R}^p$ we have

$$c^t C^t C c = c^t A_\rho^2 c$$

and so

$$|(I + BV) A_\rho c|^2 = |A_\rho c|^2. \tag{4.24}$$

Here we have $B \geq 0, V > 0$ and hence $I + BV$ only has real eigenvalues ≥ 1. From (4.24) it follows $BVA_\rho = 0$. From (4.23) we get

$$C = \frac{1}{\sqrt{\rho}}(G + A_\rho)^{1/2} V A_\rho = A_\rho. \tag{4.25}$$

We see that A_ρ depends continuously and monotone increasing on ρ. We have $A_\rho = 0$ for small ρ and for $\rho \geq \lambda_{\max}(G)\lambda_{\max}(V)^2$ we have

$$A_\rho = \rho V^{-2} - G$$

and then

$$\frac{1}{\sqrt{\rho}}(G + A_\rho)^{1/2} = V^{-1}.$$

Hence, there exists a ρ_0 from (2.7), i.e.

$$\mathrm{tr}\,[(G + A_{\rho_0})^{1/2}] = \sqrt{\rho_0}(1 + \mathrm{tr}\,(V^{-1})).$$

For such a ρ_0 we have from (4.23)

$$\frac{1}{\sqrt{\rho_0}}(G + A_{\rho_0})^{1/2} \geq V^{-1}$$

and it follows from (4.25)

$$\frac{1}{\sqrt{\rho_0}}(G + A_{\rho_0})^{1/2} V A_{\rho_0} = A_{\rho_0}.$$

In Läuter (1975) it was shown that with such a A_{ρ_0} the matrix T^* in (2.8) yields a MILE for $\gamma = CX\beta + c$.

□

Proof of Lemma 1: The matrices $\tilde{G}, \tilde{V}$ are given by (2.4) and (2.3). For h_i, γ_i from section (2.1) we define

$$g_i := X(X^tX)^{-1}H^{1/2}h_i$$

and it follows from this

$$MC^tZCg_i = \gamma_i g_i$$

and consequently

$$M^{1/2}C^tZCM^{1/2}M^{1/2+}g_i = \gamma_i M^{1/2+}g_i.$$

This means that $\{M^{1/2+}g_i, i = 1...p\}$ is a system of orthogonal eigenvectors and the γ_i are eigenvalues of $M^{1/2}C^tZCM^{1/2}$. From (2.11) follows

$$k^2H^{-1/2}X^tXH^{-1/2}D_{\rho_0}H^{-1/2}X^te_i = \lambda_i H^{-1/2}X^te_i.$$

From the commutability of $\tilde{V}$ and D_{ρ_0} we obtain now that

$$H^{-1/2}X^te_i = \alpha_i h_i$$

for some constants α_i. This means now $e_i = \alpha_i g_i$ and so

$$MC^tZCQ = \sum_{i=1}^{p} e_i e_i^t = Q^{1/2}LQ^{1/2}.$$

From this representation we get directly

$$M^{1/2+}QC^tZCM^{1/2} = \sum_{i=1}^{p} \frac{\gamma_i\mu_i}{1 + \hat{\lambda}_i} M^{1/2+}e_ie_i^tM^{1/2+} \tag{4.26}$$

and the orthogonality of $M^{1/2+}e_i, i = 1, ..., p$ was found already. From (4.26) follows (3.18).

□

Proof of Lemma 2:
With $\hat\kappa^2$ in (3.15) and $\hat\lambda_i$ in (3.16) and

$$Q \;=\; Q(y^tMy) \;=\; \sum_{i=1}^{p} \frac{\mu_i}{1+\hat\lambda_i}$$

we have

$$R(\vartheta,\hat T^*) \;=\; \mathsf{E}\Big(CY - CX\beta - CQY\Big)^t Z\Big(CY - CX\beta - CQY\Big) =$$
$$= \sigma^2 \;\mathrm{tr}\, ZCC^t - 2\mathsf{E}Y^tQC^tZC(Y-X\beta) + \mathsf{E}Y^tQC^tZCQY. \quad (4.27)$$

Using $\tilde Q = \tilde Q(x) := x\,Q(x)$ we have that $\tilde Q$ is bounded, continuous and piecewise continuously differentiable. Here $\frac{\mathrm{d}}{\mathrm{d}x}\tilde Q =: \tilde Q'$ is positive semidefinite where it exists. This $\tilde Q'$ has the same eigenvectors as $\tilde Q$, only the eigenvalues change. By construction there exist disjoint subsets $A_1,...,A_{p+1}$ of $\mathbb{R}^n$ with

$$\mathbb{R}^n = \bigcup A_j$$

and with $\tilde Q$ is differentiable in the interior of A_j. With Stein's identity based on integration by parts we obtain

$$\mathsf{E}Y^tQC^tZC(Y-X\beta) = \;\mathrm{tr}\,\mathsf{E}\frac{\tilde QC^tZC}{Y^tMY} - 2\,\mathrm{tr}\,\mathsf{E}\frac{MYY^t\tilde QC^tZC}{(Y^tMY)^2} +$$
$$+\,2\sum_j \int_{A_j} \frac{\mathrm{tr}\, M\xi\xi^t\tilde Q'C^tZC}{\xi^tM\xi} p(\xi|X\beta)\,\mathrm{d}\xi \quad (4.28)$$

The eigenvalues of $\tilde Q'$ are nonnegative. Hence the last term in (4.28) is nonnegative. This leads to

$$R(\vartheta,\hat T^*) \le \sigma^2\,\mathrm{tr}\,ZCC^t - 2\,\mathrm{tr}\,\mathsf{E}\frac{\tilde QC^tZC}{Y^tMY} + 4\,\mathrm{tr}\,\mathsf{E}\frac{MYY^t\tilde QC^tZC}{(Y^tMY)^2} +$$
$$+\,\mathrm{tr}\,\mathsf{E}\frac{YY^t\tilde QC^tZC\tilde Q}{(Y^tMY)^2}$$
$$= \sigma^2\,\mathrm{tr}\,ZCC^t - 2\,\mathrm{tr}\,\mathsf{E}QC^tZC + 4\mathsf{E}\,\frac{Y^tQC^tZCMY}{Y^tMY} +$$
$$+\mathsf{E}\,Y^tQC^tZCQY.$$

From (3.17) the desired relation (3.19) follows.

□

Proof of Theorem 3: We calculate with (3.17) and (3.18)

$$\operatorname{tr} \mathsf{E} QC^t ZC = \operatorname{tr} \mathsf{E} \frac{\tilde{Q}C^t ZC}{Y^t MY} = \operatorname{tr} \mathsf{E} \frac{M^{1/2+}\tilde{Q}^{1/2} L \tilde{Q}^{1/2} M^{1/2+}}{Y^t MY}.$$

Moreover, from (3.17) it follows

$$\frac{Y^t QC^t ZCMY}{Y^t MY} = \frac{Y^t Q^{1/2} L Q^{1/2} Y}{Y^t MY} \le \frac{1}{Y^t MY} \lambda_{\max}\Big(M^{1/2+}\tilde{Q}^{1/2} L \tilde{Q}^{1/2} M^{1/2+}\Big)$$

and in the same way

$$Y^t QC^t ZCQY \le \frac{1}{Y^t MY} \lambda_{\max}\Big(M^{1/2+}\tilde{Q} L \tilde{Q} M^{1/2+}\Big).$$

From Lemma 2 the assertion follows.

□

Proof of Theorem 4: We put $\tilde{Y} = Y^t MY$. Then for $i = 1, ..., q$ easy calculations give

$$\frac{\mu_i \gamma_i \, a_i \tilde{Y}}{1 + (\eta \tilde{Y} \nu_i \sqrt{\gamma_i} - 1)_+} = \begin{cases} c\,\nu_i \sqrt{\gamma_i}\, \tilde{Y} & \text{for} \quad \eta\, \nu_i \sqrt{\gamma_i}\, \tilde{Y} < 1 \\ \dfrac{c}{\eta} & \text{otherwise ,} \end{cases}$$

$$\frac{\mu_i^2 \gamma_i \, a_i \tilde{Y}^2}{[1 + (\eta \tilde{Y} \nu_i \sqrt{\gamma_i} - 1)_+]^2} = \begin{cases} \dfrac{c^2\, \nu_i^2\, \tilde{Y}^2}{a_i} & \text{for} \quad \eta\, \nu_i \sqrt{\gamma_i}\, \tilde{Y} < 1 \\ \dfrac{c^2}{\eta^2 \gamma_i a_i} & \text{otherwise .} \end{cases}$$

Hence we have

$$\max_i \{ \frac{\mu_i \gamma_i \, a_i \tilde{Y}}{1 + (\eta \tilde{Y} \nu_i \sqrt{\gamma_i} - 1)_+} \} = \begin{cases} c\,\nu_1 \sqrt{\gamma_1}\, \tilde{Y} & \text{for} \quad \eta\, \nu_1 \sqrt{\gamma_1}\, \tilde{Y} < 1 \\ \dfrac{c}{\eta} & \text{otherwise ,} \end{cases}$$

$$\max_i \{ \frac{\mu_i^2 \gamma_i \, a_i \tilde{Y}^2}{[1 + (\eta \tilde{Y} \nu_i \sqrt{\gamma_i} - 1)_+]^2} \} \le \frac{c^2}{\eta^2} \max\{ \frac{1}{\gamma_i a_i}; i = 1, ..., q \}. \tag{4.29}$$

Because of

$$\operatorname{tr} M^{1/2+}\tilde{Q}^{1/2} L \tilde{Q}^{1/2} M^{1/2+} = \sum_{i=1}^{q} \frac{\mu_i \gamma_i \, a_i \tilde{Y}}{1 + (\eta \tilde{Y} \nu_i \sqrt{\gamma_i} - 1)_+}, \tag{4.30}$$

$$\lambda_{\max}\Big(M^{1/2+}\tilde{Q}^{1/2} L \tilde{Q}^{1/2} M^{1/2+}\Big) = \max_i \{ \frac{\mu_i \gamma_i \, a_i \tilde{Y}}{1 + (\eta \tilde{Y} \nu_i \sqrt{\gamma_i} - 1)_+} \}, \tag{4.31}$$

$$\lambda_{\max}\Big(M^{1/2+}\tilde{Q}L\tilde{Q}M^{1/2+}\Big) = \max_i \{ \frac{\mu_i^2\gamma_i\, a_i\tilde{Y}^2}{[1+(\eta\tilde{Y}\nu_i\sqrt{\gamma_i}-1)_+]^2} \}. \qquad (4.32)$$

Therefore under (3.22)

$$f(\tilde{Y}) := 2M^{1/2+}\tilde{Q}^{1/2}L\tilde{Q}^{1/2}M^{1/2+} - 4\lambda_{\max}\Big(M^{1/2+}\tilde{Q}^{1/2}L\tilde{Q}^{1/2}M^{1/2+}\Big) - \qquad (4.33)$$

is a strictly monotone increasing function in $\tilde{Y} \leq \frac{1}{\eta\nu_q\sqrt{\gamma_q}}$ and moreover in $\tilde{Y} \leq \frac{1}{\eta\nu_1\sqrt{\gamma_1}}$ it is linear in $\tilde{Y}$. With these representations we get by easy calculations that for any $\tilde{Y}$

$$2\sum_{i=1}^{q} \frac{\mu_i\gamma_i\, a_i\tilde{Y}}{1+(\eta\tilde{Y}\nu_i\sqrt{\gamma_i}-1)_+} - 4\max_i\{\frac{\mu_i\gamma_i\, a_i\tilde{Y}}{1+(\eta\tilde{Y}\nu_i\sqrt{\gamma_i}-1)_+}\} -$$

$$- \max_i \{ \frac{\mu_i^2\gamma_i\, a_i\tilde{Y}^2}{[1+(\eta\tilde{Y}\nu_i\sqrt{\gamma_i}-1)_+]^2} \} \geq 0. \qquad (4.34)$$

Consequently it follows

$$\mathsf{E}\frac{1}{\tilde{Y}}\Big[2\sum_{i=1}^{q} \frac{\mu_i\gamma_i\, a_i\tilde{Y}}{1+(\eta\tilde{Y}\nu_i\sqrt{\gamma_i}-1)_+} - 4\max_i\{\frac{\mu_i\gamma_i\, a_i\tilde{Y}}{1+(\eta\tilde{Y}\nu_i\sqrt{\gamma_i}-1)_+}\} -$$

$$- \max_i \{ \frac{\mu_i^2\gamma_i\, a_i\tilde{Y}^2}{[1+(\eta\tilde{Y}\nu_i\sqrt{\gamma_i}-1)_+]^2} \}\Big] \geq 0. \qquad (4.35)$$

With (4.30), (4.31), (4.32) and the assertion of Theorem 3 we get that the MILE (3.14) is a Stein's estimate.

□

References

Arnold, B.F. and Stahlecker, P. (2000). Another view of the Kuks-Olman estimator. J. Statist. Plann. Inference **89**, 169-174.

Drygas H.(1996). Spectral methods in linear minimax estimation,Acta Appl. Math. **43** 17-42.

Drygas H. and Läuter H. (1994). On the representation of the linear minimax estimator in the convex linear model. In: T. Caliński and R. Kala (eds). Proc. Internat. Conf. on Linear Statistical Inference LINSTAT '93, Kluwer Acad. Publ., Dordrecht, pp. 13 - 26.

Drygas, H. and Pilz, J. (1996). On the equivalence of spectral theory and bayesian analysis in minimax linear estimation. Acta Appl. Math. **43** 43-57.

Gaffke N. and Heiligers B.: 1989, Bayes, admissible and minimax linear estimators in linear models with restricted parameter space,Statistics **20** 478-508.

Girko V. L.(1996). Spectral theory of estimation, Acta Appl. Math. **43**, 59-69

Hoffmann, K. (1978). Characterization of minimax linear estimators in linear regression. Math. Oper. Statist. Ser. Statist. **10** 19-26.

Kuks J. and Olman V.(1972). Minimax linear estimation of regression coefficients. Izv. Akad. Nauk Est. SSR **21** 66-72 (inRussian)

Läuter, H. (1975). A linear minimax estimator for linear parameters under restrictions in form of inequalities. Math. Oper. Statist. Ser. Statist. **6** 689-695.

Pilz J.(1991). Bayesian Estimation and Experimental Design in Linear Regression Models, Wiley, Chichester, New York.