Kukush, Maschke:

The Efficiency of Adjusted Least Squares in the Linear Functional Relationship

Sonderforschungsbereich 386, Paper 208 (2000)

Online unter: http://epub.ub.uni-muenchen.de/

The Efficiency of Adjusted Least Squares in the Linear Functional Relationship

Discussion Paper 208

Alexander Kukush, SFB 386
Institut für Statistik, Universität München

Erich Otto Maschke
Institut für Statistik, Universität München

Abstract

A linear functional errors-in-variables model with unknown slope parameter and Gaussian errors is considered. The measurement error variance is supposed to be known, while the variance of errors in the equation is unknown. In this model a risk bound of asymptotic minimax type for arbitrary estimators is established. The bound lies above that one which was found previously in the case of both variances known. The bound is attained by an adjusted least square estimator.

Keywords: linear functional errors-in-variables model; Hájek bound; asymptotic efficiency; adjusted least squares estimator.

1 Introduction

Suppose that a linear functional errors-in-variables model is given, with known measurement error variance and **unknown variance of errors in the equation**. Then there exists a natural modification of the least squares estimator, see e. g. Cheng and Van Ness (1999), p. 85, which is consistent and asymptotically normal. We call it an adjusted least squares (ALS) estimator due to the paper Cheng and Schneeweiss (1998), where it was developed in a more general setting, namely for a polynomial regression.

A natural question arises about asymptotic efficiency of such an estimator. A result of the asymptotic minimax type for estimation in the linear functional error-in-variables model with **both variances known** has been obtained by Nussbaum (1984) and by Hasminskii and Ibragimov (1983). It was shown there that the bound of Hájek type is attained by the maximum likelihood estimator.

In the present paper we follow the line of Hasminskii and Ibragimov (1983) and establish such a bound for the linear functional model with uncorrelated errors and unknown variance of the errors of the response variable. The asymptotic bound is attained by the ALS estimator. Thus the ALS estimator delivers the smallest possible averaged losses, and it is asymptotically efficient in the sense of Ibragimov and Has'minskii (1981).

In the next section the linear errors-in-variables model is introduced and the ALS estimator is presented. It is shown that it is asymptotically normal uniformly with respect to designs from a certain class. In Section 3 the asymptotic minimax bound is given. In Section 4 it is shown that the bound is attained by the ALS estimator. The crucial calculations of the inverse Fisher information matrix in the corresponding linear structural model are given in the Appendix, as well as an auxiliary convergence result.

2 The ALS estimator in linear model

Consider a linear functional relationship with errors in the variables and without intercept term:

$$\begin{aligned} y_i &= \beta\xi_i + \epsilon_i, \\ x_i &= \xi_i + \delta_i, \end{aligned} \tag{1}$$

$i = 1, \ldots, n$, where (δ_i, ϵ_i) are i.i.d. random errors with Gaussian distribution. We suppose that (δ_i, ϵ_i) have the expectation 0 and covariance matrix

$$\Omega = \begin{pmatrix} \sigma_\delta^2 & 0 \\ 0 & v \end{pmatrix}$$

with unknown $v > 0$. Thus we assume that the errors δ_i and ϵ_i are independent, and the variance of δ_i is known, while the variance v of ϵ_i is unknown. The design points ξ_i, $i = 1, \ldots, n$, are unobservable nonstochastic variables.

In the model (1), the values $v, \xi_1, \ldots, \xi_n$ are nuisance parameters, the number of which grows with the sample size. We are interested only in the slope parameter β. The adjusted least squares (ALS) estimator of β is given by

$$\hat{\beta} = \frac{\overline{xy}}{\overline{x^2} - \sigma_\delta^2}, \tag{2}$$

where $\overline{xy} = \frac{1}{n}\sum_1^n x_i y_i$, $\overline{x^2} = \frac{1}{n}\sum_1^n x_i^2$, see Cheng and Van Ness (1999). Under normal distributions of errors, the denominator in (2) does not equal 0 a. s., therefore $\hat{\beta}$ is well defined by (2).

Hereafter we use the denotations for averaged values like $\overline{\xi\epsilon} = n^{-1}\sum_{i=1}^n \xi_i\epsilon_i$, and similar ones. We want to show that the suitably normalized ALS estimators converge in distribution to the normal law uniformly with respect to β, v and ξ_i's (see the definition and properties of uniform convergence in distribution in Ibragimov and Has'minskii (1981)).

Introduce the class F_n of admissible design points $\xi^{(n)} = (\xi_1, \dots, \xi_n)$. Fix $H > 0$ and a sequence $\{\alpha_n\}$, s. t. $\frac{1}{n} \le \alpha_n \le 1$, $n = 1, 2, \dots$, and $\alpha_n \to 0$, $n \to \infty$. We set

$$F_n = \left\{ \xi^{(n)} \middle| \frac{1}{n} \sum_1^n \xi_k^2 \ge H \text{ and } \frac{\max_{1 \le k \le n} \ \xi_k^2}{\sum_1^n \xi_k^2} \le \alpha_n \right\}. \tag{3}$$

Sometimes we need additionally that

$$\lim_{n \to \infty} \inf \frac{\alpha_n \cdot n}{\ln n} > 2. \tag{4}$$

Under (4), for the r. v. $\tilde{\xi}_k = \tilde{H}^{1/2} \gamma_k$, where γ_k are i. i. $N(0,1)$, and $\tilde{H} > H$, we have a. s. for all $n \ge n_0(\omega)$

$$\frac{1}{n} \sum_1^n \tilde{\xi}_k^2 > H,$$

and

$$\frac{\max_{1 \le k \le n} \ \tilde{\xi}_k^2}{\sum_1^n \tilde{\xi}_k^2} \cdot \frac{n}{\ln n} \xrightarrow{P} 2 \tag{5}$$

(see Appendix). Therefore in this case $\tilde{\xi}^{(n)} = (\tilde{\xi}_1, \dots, \tilde{\xi}_n) \in F_n$ with probability tending to 1 as $n \to \infty$.

Note also that under (4) for a sequence $\eta_n = n^c$, $n = 1, 2, \dots$, $c > 0$, we have $\eta^{(n)} = (\eta_1, \dots, \eta_n) \in F_n$, for all sufficiently large n.

Lemma 1. Fix $K > 0$ and an interval $[v_1, v_2] \subset (0, +\infty)$. Then

$$\frac{\overline{\xi^2}}{\sqrt{2\beta^2 \sigma_\delta^4 + v\sigma_\delta^2 + \overline{\xi^2}(v + \beta\sigma_\delta^2)}} \sqrt{n}(\hat{\beta} - \beta) \to N(0,1)$$

in distribution, uniformly with respect to $|\beta| \le K$, $v \in [v_1, v_2]$ and $\xi^{(n)} \in F_n$, $n \ge 1$, where F_n is given in (3).

Proof. Substituting (1) in (2), we have that

$$\overline{\xi^2} \sqrt{n}(\hat{\beta} - \beta) = \frac{\sqrt{n}[-\beta(\overline{\xi\delta} + \overline{\delta^2} - \sigma_\delta^2) + \overline{\xi\epsilon} + \overline{\delta\epsilon}]}{1 + \frac{2\overline{\xi\delta}}{\overline{\xi^2}} + \frac{\overline{\delta^2} - \sigma_\delta^2}{\overline{\xi^2}}}. \tag{6}$$

Consider firstly the denominator.

$$\mathrm{E} \left(\frac{\overline{\xi\delta}}{\overline{\xi^2}} \right)^2 = \frac{\sigma_\delta^2}{n\overline{\xi^2}},$$

but $\frac{1}{\overline{\xi^2}} \leq \frac{1}{H}$ for $\xi^{(n)} \in F_n$, therefore $\overline{\xi\delta}/\overline{\xi^2}$ converges to 0 in probability uniformly for $\xi^{(n)} \in F_n$, $n \geq 1$. And

$$\mathrm{E}\left(\frac{\overline{\delta^2} - \sigma_\delta^2}{\overline{\xi^2}}\right)^2 \to 0, \quad n \to \infty$$

uniformly for $\xi^{(n)} \in F_n$, therefore $(\overline{\delta^2} - \sigma_\delta^2)/\overline{\xi^2}$ also converges to 0 in probability uniformly for $\xi^{(n)} \in F_n$. Thus the denominator of (6) converges to 1 in probability uniformly for $\xi^{(n)} \in F_n$. To prove Lemma 1, it is enough to show that

$$\frac{\frac{1}{\sqrt{n}}\sum_1^n[-\beta(\xi_i\delta_i + \delta_i^2 - \sigma_\delta^2) + \xi_i\epsilon_i + \delta_i\epsilon_i]}{\sqrt{2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + \overline{\xi^2}(v + \beta^2\sigma_\delta^2)}} \to N(0,1) \tag{7}$$

uniformly for $|\beta| \leq K$, $v \in [v_1, v_2]$, $\xi^{(n)} \in F_n$. Denote

$$\varphi_i = -\beta(\xi_i\delta_i + \delta_i^2 - \sigma_\delta^2) + \xi_i\epsilon_i + \delta_i\epsilon_i, \quad i \geq 1.$$

Then $\mathrm{E}\varphi_i = 0, \quad B_n^2 = \sum_1^n D\varphi_i = [2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + \overline{\xi^2}(v + \beta^2\sigma_\delta^2)]n.$
We bound Liapunov's ratio

$$\frac{1}{(B_n^2)^{3/2}}\sum_1^n \mathrm{E}|\varphi_i|^3. \tag{8}$$

For $|\beta| \leq K$, $v \in [v_1, v_2]$, $\xi^{(n)} \in F_n$ consider for instance the moments of the first summand of φ_i.

$$\begin{aligned}(B_n^{-2})^{3/2}\sum_1^n \mathrm{E}|\beta\xi_i\epsilon_i|^3 &\leq \frac{const}{(\sum_1^n \xi_i^2)^{3/2}}\sum_1^n |\xi_i|^3 \\ &\leq const\left(\frac{\max_{1\leq i\leq n}\xi_i^2}{\sum_1^n \xi_i^2}\right)^{1/2} \leq const\ \alpha_n^{1/2},\end{aligned}$$

with α_n given in (3). Similar calculations for other summands of φ_i show that Liapunov's ratio (8) converges to 0 uniformly for $|\beta| \leq K$, $v \in [v_1, v_2]$, $\xi^{(n)} \in F_n$. Now, by Theorem 15 from Ibragimov and Has'minskii (1981), p. 369, uniform convergence (7) holds. This implies (5).

Corollary. Let l be a bounded Borel measurable function which is continuous a. e. with respect to Lebesgue measure. Then

$$\mathrm{E}_{\beta v\xi^{(n)}}\left\{l\left(\frac{\overline{\xi^2}\sqrt{n}(\hat{\beta} - \beta)}{\sqrt{2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + \overline{\xi^2}(v + \beta^2\sigma_\delta^2)}}\right)\right\} \to \mathrm{E}\ l(\gamma_1) \tag{9}$$

uniformly with respect to $|\beta| \leq K$, $v \in [v_1, v_2]$, $\xi^{(n)} \in F_n$, $n \geq 1$, where $\gamma_1 \sim N(0,1)$.

Hereafter $\mathrm{E}_{\beta v\xi^{(n)}}$ denotes the expectation under the condition that β, v and $\xi^{(n)} = (\xi_1, \ldots, \xi_n)$ are the true values of unknown parameters in the regression model

(1). The uniform convergence (9) is an immediate consequence of Lemma 1 and of an evident modification of Theorem 8 from Ibragimov and Has'minskii (1981), p. 366, where the limit distribution $\mathcal{P}_\theta$ does not depend upon θ, and in this case the parametric set Θ may be arbitrary, not necessarily compact. In the situation of Corollary, $\mathcal{P}_\theta = N(0,1)$ and $\Theta = \{(\beta, v, \xi) : |\beta| \leq K, v \in [v_1, v_2], \xi^{(n)} \in F_n \text{ for all } n \geq 1\}$, where $\xi = (\xi_1, \ldots, \xi_n, \ldots)$.

3 Asymptotic minimax bound

Here we follow the line of Hasminskii and Ibraginov (1983). In that paper it was assumed that $v = 1$. But now we consider the model (1) with Gaussian errors and **unknown** $v = D\epsilon_1$. Denote by Λ the class of functions $l(x)$ on $I\!R$, such that $l(x) = -l(x) \geq 0$, $x \in I\!R$, and $l(x)$ is nondecreasing for $x > 0$.

Theorem 1. Fix $l \in \Lambda$, $\beta \in I\!R$, $v > 0$ and the set F_n of designs given in (3), (4). Then for every estimator β_n which is based on observations coming from model (1),

$$\lim_{\delta \to 0} \quad \liminf_{n \to \infty} \quad \sup_{|b-\beta|<\delta, |w-v|<\delta, \xi^{(n)} \in F_n} \quad \mathrm{E}_{\beta w \xi^{(n)}} \left\{ l \left(\frac{H\sqrt{n}(\beta_n - b)}{\sqrt{2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + H(v + \beta^2\sigma_\delta^2)}} \right) \right\}$$
$$\geq \quad \mathrm{E}\ l(\gamma_1), \tag{10}$$

where $\gamma_1 \sim N(0,1)$.

Proof. Let H be a lower bound from (3). Consider the sequence $\{\tilde{\xi}_i\}$ of i.i.d. $(0, \tilde{H})$-normal random variables, independent on $\{\delta_i, \epsilon_i; i \geq 1\}$. The variance $\tilde{H}$ is unknown, we know only that $\tilde{H} > H$. In Section 2 it was mentioned that $\tilde{\xi}^{(n)} = (\tilde{\xi}_1, \ldots, \tilde{\xi}_n) \in F_n$, $n \geq n_0(w)$, a. s.. Now, consider the problem of estimation of the parameters $\tilde{H}$, β, v on the basis of independent observations

$$x_i = \tilde{\xi}_i + \delta_i, \quad y_i = \beta\tilde{\xi}_i + \epsilon_i, \quad i = 1, \ldots, n. \tag{11}$$

Denote

$$\Delta = \tilde{H}v + \tilde{H}\beta^2\sigma_\delta^2 + v\sigma_\delta^2. \tag{12}$$

The observations (11) are Gaussian with density function

$$p(x, y; \beta, \tilde{H}, v)$$
$$= \frac{1}{2\pi\sqrt{\Delta}} \exp\left\{ -\frac{1}{2\Delta} \left[x^2 \left(\beta^2\tilde{H} + v \right) - 2xy\beta\tilde{H} + y^2 \left(\tilde{H} + \sigma_\delta^2 \right) \right] \right\}.$$

The Fisher information matrix I of the density has the form (see Lemma 3 in Appendix)

$$2I \quad = \quad \frac{1}{\Delta^2} A,$$

$$A = \begin{pmatrix} 2\tilde H^2(\Delta + 2\beta^2\sigma_\delta^4) & 2\tilde H\beta\sigma_\delta^2(v+\beta^2\sigma_\delta^2) & 2\tilde H\sigma_\delta^2(\tilde H+\sigma_\delta^2)\beta \\ 2\tilde H\beta\sigma_\delta^2(v+\beta^2\sigma_\delta^2) & (v+\beta^2\sigma_\delta^2)^2 & \beta^2\sigma_\delta^4 \\ 2\tilde H\sigma_\delta^2(\tilde H+\sigma_\delta^2)\beta & \beta^2\sigma_\delta^4 & (\tilde H+\sigma_\delta^2)^2 \end{pmatrix}.$$

As $\det A > 0$ (see Lemma 4 in Appendix), the observations (11) satisfy Le Cam's LAN conditions with the norming factors $n^{-1/2}I^{-1/2}$.

Denote $z = (1,\ 0,\ 0)'$. We are interested in

$$\begin{aligned} &(I^{-1/2})^2_{11} + (I^{-1/2})^2_{12} + (I^{-1/2})^2_{13} \\ = \ &(I^{-1/2}z,\ I^{-1/2}z) = (I^{-1}z,\ z) = (I^{-1})_{11} = \frac{\Delta + 2\beta^2\sigma_\delta^4}{\tilde H^2}, \end{aligned} \tag{13}$$

see Lemma 4 in Appendix. Introduce the class of bounded loss functions

$$\Lambda_b = \{l \in \Lambda : l \text{ is bounded}\},$$

and let

$$B = \frac{(\Delta + 2\beta^2\sigma_\delta^4)I}{H^2},$$

with Δ given in (12). We apply Theorem 2.12.1 from Ibragimov and Has'minskii (1981) to the model (11) with the loss function

$$w(x) = l((B^{-1/2}x)_1), \quad l \in \Lambda_b,$$

where $(x)_1$ denotes the first component of the vector $x \in I\!\!R^3$. Then by (13) we have

$$\sum_1^3 (B^{-1/2})^2_{1i} = (B^{-1})_{11} = \left(\frac{H}{\tilde H}\right)^2. \tag{14}$$

Taking into account (14) we find that for every estimators β_n,

$$\begin{aligned} \lim_{\delta\to 0}\ \liminf_{n\to\infty}\ \sup_{(|b-\beta|<\delta,|w-v|<\delta,|h-\tilde H|<\delta)} \mathrm{E}_{bwh}\left\{ l\left(\frac{H\cdot\sqrt{n}(\beta_n - b)}{\sqrt{\tilde H(v+\beta^2\sigma_\delta^2) + v\sigma_\delta^2 + 2\beta^2\sigma_\delta^4}}\right)\right\} \\ \ge \mathrm{E}\{l((B^{-1/2}\gamma)_1)\} = E\ l\left(\frac{H}{\tilde H}\gamma_1\right). \end{aligned} \tag{15}$$

Here γ is a standard normal random vector in $I\!\!R^3$, and E_{bwh} denotes the expectation under the condition that in the model (11) $\beta = b$, $D\epsilon_1 = w$, $D\tilde\xi_1 = h$.

For $h > H$, $P_{\tilde H}\{\tilde\xi^{(n)} \in F_n\} \to 1$, $n\to\infty$ uniformly for $\tilde H \in (h-\delta_0, h+\delta_0)$, $\delta_0 = \frac{h-H}{2}$. Because of the boundedness of l we have now for small δ,

$$\begin{aligned} &\sup_{(|b-\beta|<\delta,|w-v|<\delta,\xi^{(n)}\in F_n)} \mathrm{E}_{bw\xi^{(n)}}\left\{ l\left(\frac{H\cdot\sqrt{n}(\beta_n-b)}{\sqrt{2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + H(v+\beta^2\sigma_\delta^2)}}\right)\right\} \\ \ge\ &\sup_{(|b-\beta|<\delta,|w-v|<\delta,|h-\tilde H|<\delta)} \mathrm{E}_{bwh}\left\{ l\left(\frac{H\cdot\sqrt{n}(\beta_n-b)}{\sqrt{2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + H(v+\beta^2\sigma_\delta^2)}}\right)\right\} \\ &-o(1), \quad n\to\infty \end{aligned} \tag{16}$$

Then the following chain of inequalities holds, see (16) and (15)

$$\begin{aligned}
&\lim_{\delta\to 0}\quad \liminf_{n\to\infty}\quad \sup_{(|b-\beta|<\delta,|w-v|<\delta,\xi^{(n)}\in F_n)} \mathrm{E}_{bw\xi^{(n)}}\left\{ l\left(\frac{H\cdot\sqrt{n}(\beta_n-b)}{\sqrt{2\beta^2\sigma_\delta^4+v\sigma_\delta^2+H(v+\beta^2\sigma_\delta^2)}}\right)\right\}\\
\geq\ &\lim_{\delta\to 0}\quad \liminf_{n\to\infty}\quad \sup_{(|b-\beta|<\delta,|w-v|<\delta,|h-\tilde{H}|<\delta)} \mathrm{E}_{bwh}\left\{ l\left(\frac{H\cdot\sqrt{n}(\beta_n-b)}{\sqrt{2\beta^2\sigma_\delta^4+v\sigma_\delta^2+H(v+\beta^2\sigma_\delta^2)}}\right)\right\}\\
\geq\ &\lim_{\delta\to 0}\quad \liminf_{n\to\infty}\quad \sup_{(|b-\beta|<\delta,|w-v|<\delta,|h-\tilde{H}|<\delta)} \mathrm{E}_{bwh}\left\{ l\left(\frac{H\cdot\sqrt{n}(\beta_n-b)}{\sqrt{2\beta^2\sigma_\delta^4+v\sigma_\delta^2+\tilde{H}(v+\beta^2\sigma_\delta^2)}}\right)\right\}\\
\geq\ &\mathrm{E}\ l(\frac{H}{\tilde{H}}\gamma_1)\to \mathrm{E}\ l(\gamma_1),\quad \tilde{H}\to H+\,.
\end{aligned}$$

We proved (10) for all bounded $l\in\Lambda$. At last consider an unbounded loss function $f\in\Lambda$. Denote by $\phi(l)$, $l\in\Lambda$, the left hand side of the inequality (10) and by f_c the truncated function $f_c(t)=\min(c,f(t))$, $t\in I\!R$; $c>0$. The function f_c belongs to the class Λ_b, therefore

$$\phi(f)\geq\phi(f_c)\geq \mathrm{E}\ f_c(\gamma_1)\to \mathrm{E}\ f(\gamma_1),\quad c\to+\infty.$$

This proves Theorem 1.

4 Asymptotic efficiency of the ALS estimator

Suppose for a moment that in the model (1) the variance v is known. Then the corresponding minimax bound can be obtained by a modification of Theorem 1 if the summand $2\beta^2\sigma_\delta^4$ under the root is canceled, see Hasminskii and Ibragimov (1983). The additional summand $2\beta^2\sigma_\delta^4$ in (10) reflects the lack of information in the model (1) with unknown variance.

In the case of known v, the maximum likelihood estimator of β attains the corresponding bound, i. e. it is asymptotically efficient in that case. We show now that in the case of unknown v, the ALS estimator attains the bound, but we prove it for bounded loss functions only.

Theorem 2. Fix $l\in\Lambda_b$, $\beta\in I\!R$, $v>0$ and the set F_n of designs given in (3), (4). Then for the ALS estimator $\hat{\beta}$ defined in (2), equality in (10) holds.

Proof. Following the line of Ibragimov and Has'minskii (1981), p. 177, we induce from Corollary of Lemma 1 that for $l\in\Lambda_b$,

$$\begin{aligned}
\lim_{\delta\to 0}\quad \liminf_{n\to\infty}\quad \sup\nolimits_{(|b-\beta|<\delta,|w-v|<\delta,\xi^{(n)}\in F_n)}\quad &\mathrm{E}_{bw\xi^{(n)}}\left\{ l\left(\frac{\overline{\xi^2}\sqrt{n}(\hat{\beta}_n-\beta)}{\sqrt{2\beta^2\sigma_\delta^4+v\sigma_\delta^2+\overline{\xi^2}(v+\beta^2\sigma_\delta^2)}}\right)\right\}\\
=\qquad &\mathrm{E}\ l(\gamma_1). \qquad (17)
\end{aligned}$$

The function $\varphi(u) = \frac{u}{\sqrt{A^2+u^2}}$, $u \geq 0$, is increasing, and $\overline{\xi^2} \geq H$, for $\xi^{(n)} \in F_n$. Therefore from (17) we get

$$\lim_{\delta\to 0} \quad \liminf_{n\to\infty} \quad \sup_{(|b-\beta|<\delta, |w-v|<\delta, \xi^{(n)}\in F_n)} \quad \mathrm{E}_{bw\xi^{(n)}} \left\{ l\left(\frac{H\sqrt{n}(\hat\beta_n - \beta)}{\sqrt{2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + H(v+\beta^2\sigma_\delta^2)}} \right) \right\}$$
$$\leq \qquad \mathrm{E}\, l(\gamma_1). \tag{18}$$

But it follows from (10) that actually in (18) equality holds, and the theorem is proved.

Thus we showed for the model (1) that the ALS estimator $\hat\beta$ is asymptotically efficient in the sense of Hájek bound, i. e. $\hat\beta$ attains the minimax bound (10). This means that, under suitable normalzation, $\hat\beta$ has the least possible averaged losses from imprecise estimation of β.

5 Appendix

5.1 Auxiliary matrix calculations

Consider a random vector $X \sim N(0, \Sigma)$ with probability density p.

Lemma 2. Suppose that the entries of Σ are C^2-smooth functions of χ, and χ belongs to an open set $G \subset I\!R^d$, and for each $\chi \in G$ the covariance matrix Σ is nonsingular. Then for each $i, k = 1, 2, \ldots, d$,

$$-2\mathrm{E}\frac{\partial^2 \ln p}{\partial\chi_k\partial\chi_i} = \mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\right).$$

Proof. Let X be distributed in $I\!R^m$. Denote

$$l(x,\chi) = \ln p = -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln\det\Sigma - \frac{1}{2}x'\Sigma^{-1}x.$$

We have

$$\frac{\partial l}{\partial\chi_i} = -\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\right) + \frac{1}{2}x'\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\Sigma^{-1}x,$$

and

$$\begin{aligned} 2\frac{\partial^2 l}{\partial\chi_k\partial\chi_i} &= \mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\right) - \mathrm{tr}\left(\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\chi_k\partial\chi_i}\right) \\ &\quad -x'\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\Sigma^{-1}x + x'\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\chi_k\partial\chi_i}\Sigma^{-1}x \\ &\quad -x'\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\Sigma^{-1}x. \end{aligned}$$

Now,

$$
\begin{aligned}
-2\mathrm{E}\frac{\partial^2 l}{\partial\chi_k\partial\chi_i} &= \mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\right) - \mathrm{tr}\left(\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\chi_k\partial\chi_i}\right) \\
&\quad -\mathrm{tr}\ \Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\Sigma^{-1}\mathrm{E}xx' \\
&\quad +\mathrm{tr}\ \Sigma^{-1}\frac{\partial^2\Sigma}{\partial\chi_k\partial\chi_i}\Sigma^{-1}\mathrm{E}xx' - \mathrm{tr}\ \Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\Sigma^{-1}\mathrm{E}xx' \\
&= -\mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\right).
\end{aligned}
$$

This proves Lemma 2.

Now, consider random variables, corresponding to the model (11), with slightly different denotations. Let

$$\delta \sim N(0,\sigma_\delta^2), \quad \epsilon \sim N(0,v), \quad \xi \sim N(0,H)$$

be independent r. v. with positive variances, and

$$x = \xi + \delta, \quad y = \beta\xi + \epsilon.$$

Here β is a fixed real parameter. Then

$$(x,y)' \sim N(0,\Sigma), \quad \Sigma = \begin{pmatrix} H+\sigma_\delta^2 & H\beta \\ H\beta & H\beta^2+v \end{pmatrix}, \tag{19}$$

and

$$
\begin{aligned}
\det\Sigma &= \Delta = H\beta^2\sigma_\delta^2 + (H+\sigma_\delta^2)v, \quad \Delta > 0, \\
\Sigma^{-1} &= \frac{1}{\Delta}\begin{pmatrix} H\beta^2+v & -H\beta \\ -H\beta & H+\sigma_\delta^2 \end{pmatrix}.
\end{aligned} \tag{20}
$$

The probability density p of random vector (x,y) depends upon the parameter $\chi = (\beta, H, v) \in I\!R \times (0,+\infty) \times (0,+\infty)$.

Lemma 3. The Fisher information matrix I of the density $p(x,y;\chi)$ has the form

$$2I = \frac{1}{\Delta^2}A,$$

with

$$A = \begin{pmatrix} 2H^2(\Delta+2\beta^2\sigma_\delta^4) & 2H\beta\sigma_\delta^2(v+\beta^2\sigma_\delta^2) & 2H\sigma_\delta^2(H+\sigma_\delta^2)\beta \\ 2H\beta\sigma_\delta^2(v+\beta^2\sigma_\delta^2) & (v+\beta^2\sigma_\delta^2)^2 & \beta^2\sigma_\delta^4 \\ 2H\sigma_\delta^2(H+\sigma_\delta^2)\beta & \beta^2\sigma_\delta^4 & (H+\sigma_\delta^2)^2 \end{pmatrix}.$$

Proof. By Lemma 2

$$(2I)_{ik} = -2\mathrm{E}\frac{\partial^2 ln\ p}{\partial\chi_k\partial\chi_i} \quad = \quad \mathrm{tr}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\chi_k}\right). \tag{21}$$

Using (19) and (20) we have consequently

$$\begin{aligned}
\Sigma^{-1}\frac{\partial\Sigma}{\partial\beta} &= \frac{H}{\Delta}\begin{pmatrix} -H\beta & v - H\beta^2 \\ H+\sigma_\delta^2 & \beta(H+2\sigma_\delta^2) \end{pmatrix}, \\
\Sigma^{-1}\frac{\partial\Sigma}{\partial H} &= \frac{1}{\Delta}\begin{pmatrix} v & v\beta \\ \beta\sigma_\delta^2 & \beta^2\sigma_\delta^2 \end{pmatrix}, \\
\Sigma^{-1}\frac{\partial\Sigma}{\partial v} &= \frac{1}{\Delta}\begin{pmatrix} 0 & -H\beta \\ 0 & H+\sigma_\delta^2 \end{pmatrix}.
\end{aligned}$$

Now, a direct calculation of the entries of $2I$ using (21) accomplishes the proof.

Lemma 4. The Fisher information matrix I of the density $p(x, y; \chi)$ is nonsingular, and

$$(I^{-1})_{11} = \frac{\Delta + 2\beta^2\sigma_\delta^4}{H^2} = \frac{2\beta^2\sigma_\delta^4 + v\sigma_\delta^2 + H(v+\beta^2\sigma_\delta^2)}{H^2}.$$

Proof. Due to Lemma 3 we have to show nonsingularity of A and to compute $(A^{-1})_{11}$. Find the algebraic complements in A for the entries of the right row.

$$\begin{aligned}
A_{11} &= \Delta(\Delta + 2\beta^2\sigma_\delta^4), \quad A_{12} = -2H\sigma_\delta^2(H+\sigma_\delta^2)\beta\Delta, \\
A_{13} &= -2H\beta\sigma_\delta^2(v+\beta^2\sigma_\delta^2)\Delta.
\end{aligned}$$

Then

$$\begin{aligned}
\det A &= a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} \\
&= 2H^2(\Delta + 2\beta^2\sigma_\delta^4)^2\Delta - 4H^2\beta^2\sigma_\delta^4(v+\beta^2\sigma_\delta^2) \times (H+\sigma_\delta^2)\Delta \\
&\quad -4H^2\sigma_\delta^4(H+\sigma_\delta^2)\beta^2(v+\beta^2\sigma_\delta^2)\Delta \\
&= 2H^2\Delta[(\Delta + 2\beta^2\sigma_\delta^4)^2 - 4\beta^2\sigma_\delta^4(H+\sigma_\delta^2) \times (v+\beta^2\sigma_\delta^2)] \\
&= 2H^2\Delta[(\Delta + 2\beta^2\sigma_\delta^4)^2 - 4\beta^2\sigma_\delta^4(\Delta+\beta^2\sigma_\delta^4)] \\
&= 2H^2\Delta^3,
\end{aligned}$$

and $\det A > 0$. Therefore I is also nonsingular. At last

$$(A^{-1})_{11} = \frac{A_{11}}{\det A} = \frac{\Delta + 2\beta^2\sigma_\delta^4}{2H^2\Delta^2}$$

and by Lemma 3

$$(I^{-1})_{11} = 2\Delta^2(A^{-1})_{11} = \frac{\Delta + 2\beta^2\sigma_\delta^4}{H^2}.$$

5.2 Proof of convergence (5)

Lemma 5. Let γ_k, $k = 1, 2, \dots$ be i.i. $N(0, 1)$ distributed random values. Then

$$\frac{\max_{1\le k\le n}\gamma_k^2}{\sum_1^n \gamma_k^2}\frac{n}{\ln n} \xrightarrow{P} 2. \tag{22}$$

Proof. As $\sum_1^n \gamma_k^2/n \to 1$, a. s., (22) is equivalent to

$$\eta_n = \frac{\max_{1\le k\le n}\gamma_k^2}{\ln n} \xrightarrow{P} 2. \tag{23}$$

Find the d. f. of η_n. A r. v. γ_k^2 has a d. f. F with density

$$f(t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{t}{2}}, \quad t > 0.$$

Then $\max_{1\le k\le n}\gamma_k^2$ has a d. f. $F^n(t)$, and the d. f. F_n of η_n equals to

$$F_n(t) = F^n(t \ln n), \quad n \ge 2.$$

To prove (23) it is sufficient to show that for each $\epsilon > 0$,

$$F_n(2-\epsilon) \to 0, \text{ and } \quad F_n(2+\epsilon) \to 1, \quad n \to \infty. \tag{24}$$

For $t > 0$ we have

$$F_n(t) = [(1-(1-F(t\ln n))^{\frac{1}{1-F(t\ln n)}}]^{n(1-F(t\ln n))} = A_n^{B_n}.$$

Here $A_n \to e^{-1}$, $n \to \infty$, and

$$\begin{aligned}
\lim_{n\to\infty} B_n &= \lim_{z\to+\infty}\frac{1-F(t\ln z)}{1/z} = \lim_{z\to+\infty}\frac{-f(t\ln z)\frac{t}{z}}{-1/z^2} \\
&= \frac{\sqrt{t}}{\sqrt{2\pi}}\lim_{z\to+\infty}\frac{z}{\sqrt{\ln z}}e^{-\frac{t}{2}\ln z} \\
&= \sqrt{\frac{t}{2\pi}}\lim_{z\to+\infty}\frac{z^{1-\frac{t}{2}}}{\sqrt{\ln z}} = \begin{cases} +\infty, & \text{if} \quad t < 2 \\ 0, & \text{if} \quad t > 2. \end{cases}
\end{aligned}$$

Therefore $A_n^{B_n} \to 0$ if $t < 2$, and $A_n^{B_n} \to 1$ if $t > 2$. This proves (24), and (23) holds true. Lemma is proved.

Acknowledgement

We are grateful to Silvelyn Zwanzig for a good advice.

References

[1] Cheng, C.-L. and Schneeweiss, H. (1998). Polynomial regression with errors in the variables. J. R. Statist. Soc. B **60**, 189–199.

[2] Cheng, C.-L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error.* Arnold, London.

[3] Hasminskii, R. Z. and Ibragimov, I. A. (1983). On asymptotic efficiency in the presence of an infinitedimensional nuisance parameter. Springer Lecture Notes in Mathematics **1021**, 195–229.

[4] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, Berlin.

[5] Nussbaum, M. (1984) An asymptotic minimax risk bound for estimation of a linear functional relationship. J. Multivar. Anal., **14**, 3, 300–314.