MAXIMUM LQ-LIKELIHOOD ESTIMATION IN FUNCTIONAL MEASUREMENT ERROR MODELS

Patricia Giménez[1], Lucas Guarracino[1] and Manuel Galea[2]

[1] *Universidad Nacional de Mar del Plata, Argentina and*

[2] *Pontificia Universidad Católica de Chile, Chile*

Supplementary Material

This supplementary material contains: (i) the regularity conditions used to prove consistency and asymptotic normality, (ii) proofs of Theorem 1 and Propositions 1 and 2 and (iii) derivation of some results in the simple linear functional model. Equations that are specific to this supplement are labeled as (S*). All other referenced equations correspond to those of the main paper.

S1 Regularity conditions

We state regularity conditions similar in spirit to the corresponding conditions required in the case of maximum likelihood estimation in the presence of incidental parameters given by Mak (1982) (see also, Giménez and Bolfarine (1997)), although in our case, expectations are taken under the true distributions $G_j,\ j = 1, 2, \ldots$.

C0. $\frac{1}{n}\sum_{j=1}^{n} E_{G_j}[h_j(\boldsymbol{Z}_j;.)]$ converges uniformly to a function $\bar{h}(.)$ in a neighborhood of $\boldsymbol{\theta}^{\dagger}$, where $\boldsymbol{\theta}^{\dagger} \in \Theta^{\circ}$ is a local maximum of $\bar{h}$.

C1. Jacobian matrix of $\bar{h}$ evaluated in $\boldsymbol{\theta}^{\dagger}$ is nonsingular.

C2. For each $\boldsymbol{\theta} \in \Theta^{\circ}$, there exists $\delta > 0$, and functions d_j and d_{jkl}, such that

$$|h_j(\boldsymbol{Z}_j;\boldsymbol{\theta}^{'})| < d_j(\boldsymbol{Z}_j) \quad \text{and} \quad |I^{\dagger}_{jkl}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{'})| < d_{jkl}(\boldsymbol{Z}_j) \quad \text{a.e.}$$

for all $\boldsymbol{\theta}^{'} \in B(\boldsymbol{\theta},\delta) = \{\boldsymbol{\theta}' : \ \|\boldsymbol{\theta}' - \boldsymbol{\theta}\| < \delta\} \subset \Theta^{\circ}$. Moreover $\limsup \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\left[d_j(\boldsymbol{Z}_j)^2\right] < \infty$ and $\limsup \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\left[d_{jkl}(\boldsymbol{Z}_j)^2\right] < \infty$, $k,l = 1,\ldots,p$.

C3.
$$\begin{aligned} 0 &< \liminf \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\left[U^{\dagger}_{jk}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})U^{\dagger}_{jl}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})\right] \\ &\le \limsup \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\left[U^{\dagger}_{jk}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})U^{\dagger}_{jl}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})\right] < \infty, \quad k,l = 1,\ldots,p. \end{aligned}$$

C4. $0 < \liminf \det\left(\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^{\dagger}\right)$ and $0 < \liminf \left\|\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^{\dagger})^{-1}\right\| \le \limsup \left\|\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^{\dagger})^{-1}\right\|$ $< \infty$, where $\det(A)$ is the determinant of A and $\|.\|$ represents the euclidean matrix norm.

C5. Given $\epsilon > 0$, there exists $\delta > 0$, such that

$$\limsup \left| \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\left[\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^{\dagger},\delta)} I^{\dagger}_{jkl}(\boldsymbol{Z}_j;\boldsymbol{\theta}) - I^{\dagger}_{jkl}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})\right]\right| < \epsilon,$$

$k,l = 1,\ldots,p$, and the same is true when $\limsup$ is replaced by $\liminf$.

C6. There exists $\gamma > 0$ such that

$$\lim_{n\to\infty} \frac{1}{n^{1+\frac{\gamma}{2}}} \sum_{j=1}^{n} E_{G_j}\left[|U_{jk}^{\dagger}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})|^{2+\gamma}\right] = 0, \quad k = 1,\ldots,p.$$

Remark 1. Conditions C0 to C2 imply that (2.5) has a maximum $\hat{\boldsymbol{\theta}}_n^*$ which converges in probability to $\boldsymbol{\theta}^{\dagger}$. Conditions C3 to C5 are related to the asymptotic behavior of the matrices $\overline{\boldsymbol{\Lambda}}_n$ and $\overline{\boldsymbol{\Gamma}}_n$. It is not required that these matrices converge to any limit. If conditions C0 to C4 are verified, for the uniqueness of the sequence of roots of (2.6), which converges to $\boldsymbol{\theta}^{\dagger}$, it follows that with probability tending to one, this sequence is equal to the sequence of maximizers of H_n in (2.5). Condition C6 is stated for application of Liapunov's central limit theorem.

S2 Proofs

S2.1 Proof of Theorem 1

(i) First, we can see that condition C0 to C2 implies that (2.5) has a maximum $\hat{\boldsymbol{\theta}}_n^*$, which converges in probability to $\boldsymbol{\theta}^{\dagger}$. The proof of this result follows along the line of the proof of Theorem 2.1A in Mak (1982) and the details are thus omitted. On the other hand, C4 implies that exist $K > 0$ and $n_0 \in \mathbb{N}$, such that $\left\|\overline{\boldsymbol{\Lambda}}_n^{-1}(\boldsymbol{\theta}^{\dagger})\right\| < K$, $\forall n \geq n_0$. As $\theta^{\dagger} \in \Theta^{\circ}$, from C2 and Chebyshev's inequality, it follows that exists $\delta_0 > 0$ such that

$\forall\, \boldsymbol{\theta} \in B(\boldsymbol{\theta}^{\dagger}, \delta_0)$

$$\overline{\boldsymbol{I}}_n^{\dagger}(\boldsymbol{\theta}) - \overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}) \xrightarrow{P} 0. \qquad \text{(S2.1)}$$

Applying Lemma 2.2 in Mak (1982), we have with probability going to 1 as $n \to \infty$, that $\overline{\boldsymbol{I}}_n^{\dagger}(\boldsymbol{\theta}^{\dagger})$ is invertible and

$$\left\| \overline{\boldsymbol{I}}_n^{\dagger\,-1}(\boldsymbol{\theta}^{\dagger}) \right\| < 2K. \qquad \text{(S2.2)}$$

Setting $\lambda = \dfrac{1}{4K}$ and $\lambda_n = 1/4\left\| \overline{\boldsymbol{I}}_n^{\dagger\,-1}(\boldsymbol{\theta}^{\dagger}) \right\|$, by (S2.2) we have, with probability going to 1 as $n \to \infty$, that $\lambda < 2\lambda_n$. On the other hand, C5 implies that exists $\delta_1 > 0$, such that

$$\left\| \overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}) - \overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^{\dagger}) \right\| < \frac{\lambda}{2}, \qquad \text{(S2.3)}$$

$\forall\, \boldsymbol{\theta} \in B(\boldsymbol{\theta}^{\dagger}, \delta_1)$.

Then, from (S2.1) and (S2.3) we have, with probability going to 1 as $n \to \infty$, that

$$\begin{aligned} \left\| \overline{\boldsymbol{I}}_n^{\dagger}(\boldsymbol{\theta}) - \overline{\boldsymbol{I}}_n^{\dagger}(\boldsymbol{\theta}^{\dagger}) \right\| \leq \left\| \overline{\boldsymbol{I}}_n^{\dagger}(\boldsymbol{\theta}) - \overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}) \right\| + \left\| \overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}) - \overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^{\dagger}) \right\| \\ + \left\| \overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^{\dagger}) - \overline{\boldsymbol{I}}_n^{\dagger}(\boldsymbol{\theta}^{\dagger}) \right\| < \lambda < 2\lambda_n, \end{aligned}$$

$\forall\, \boldsymbol{\theta} \in B(\boldsymbol{\theta}^{\dagger}; \delta)$, with $\delta = \min\{\delta_0, \delta_1\}$.

Applying the inverse function theorem (Rudin (1964)) to $\overline{\boldsymbol{U}}_n^{\dagger}$, we have, with probability going to 1 as $n \to \infty$, that

(1) Exists $\delta_2 > 0$ such that $\overline{\boldsymbol{U}}_n^{\dagger} \colon B(\boldsymbol{\theta}^{\dagger}, \delta_2) \longrightarrow \overline{\boldsymbol{U}}_n^{\dagger}\left(B(\boldsymbol{\theta}^{\dagger}, \delta_2)\right)$ is invertible.

(2) $B\big(\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger), \frac{\lambda\delta_2}{2}\big) \subset B\big(\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger), \lambda_n\delta_2\big) \subset \overline{\boldsymbol{U}}_n^\dagger\left(B(\boldsymbol{\theta}^\dagger, \delta_2)\right)$.

Since $E_{G_j}\big[\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)\big] = \boldsymbol{0}\ \forall\, j \in \mathbb{N}$, from Chebyshev's inequality and C3, it follows that $\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger) \xrightarrow{P} \boldsymbol{0}$. Then $\big\|\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger) - \boldsymbol{0}\big\| < \frac{\lambda\delta_2}{2}$, with probability going to 1, i.e. $\boldsymbol{0} \in B\big(\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger), \frac{\lambda\delta_2}{2}\big)$. From (2) it follows, with probability going to 1 as $n \to \infty$, that $\boldsymbol{0} \in \overline{\boldsymbol{U}}_n^\dagger\left(B(\boldsymbol{\theta}^\dagger, \delta_2)\right)$.

From (1), we can consider ${\overline{\boldsymbol{U}}_n^\dagger}^{-1} : \overline{\boldsymbol{U}}_n^\dagger\left(B(\boldsymbol{\theta}^\dagger, \delta_2)\right) \longrightarrow B(\boldsymbol{\theta}^\dagger, \delta_2)$. Since $\boldsymbol{0} \in \overline{\boldsymbol{U}}_n^\dagger\left(B(\boldsymbol{\theta}^\dagger, \delta_2)\right)$ with probability going to 1 as $n \to \infty$, we conclude that $\lim_{n\to\infty} P\Big(\big\|{\overline{\boldsymbol{U}}_n^\dagger}^{-1}(\boldsymbol{0}) - \boldsymbol{\theta}^\dagger\big\| < \delta_2\Big) = 1$. Since δ_2 can be taken arbitrarily small, it follows that ${\overline{\boldsymbol{U}}_n^\dagger}^{-1}(\boldsymbol{0}) \xrightarrow{P} \boldsymbol{\theta}^\dagger$ and we can take $\hat{\boldsymbol{\theta}}_n^* = {\overline{\boldsymbol{U}}_n^\dagger}^{-1}(\boldsymbol{0})$, $\forall\, n \in \mathbb{N}$ as the consistent sequence.

Furthermore, since $\overline{\boldsymbol{U}}_n^\dagger$ is one to one in a neighborhood of $\boldsymbol{\theta}^\dagger$, if we have $(\tilde{\boldsymbol{\theta}}_n)_n$ such that $\lim_{n\to\infty} P\big(\overline{\boldsymbol{U}}_n^\dagger(\tilde{\boldsymbol{\theta}}_n) = \boldsymbol{0}\big) = 1$, then $\lim_{n\to\infty} P\big(\hat{\boldsymbol{\theta}}_n^* = \tilde{\boldsymbol{\theta}}_n\big) = \lim_{n\to\infty} P\big({\overline{\boldsymbol{U}}_n^\dagger}^{-1}(\boldsymbol{0}) = \tilde{\boldsymbol{\theta}}_n\big) = \lim_{n\to\infty} P\big(\overline{\boldsymbol{U}}_n^\dagger(\tilde{\boldsymbol{\theta}}_n) = \boldsymbol{0}\big) = 1$.

(ii) For proving asymptotic normality, let $\boldsymbol{h} \in \mathbb{R}^p$ and consider for each $n \in \mathbb{N}$, the function given by $\boldsymbol{\theta} \to \sqrt{n}\ \boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta})$. From the mean value theorem and because $\overline{\boldsymbol{U}}_n^\dagger(\hat{\boldsymbol{\theta}}_n^*) = \boldsymbol{0}$, it exists $\tilde{\boldsymbol{\theta}}_n \in \Theta$, such that

$$\sqrt{n}\ \boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger) = -\sqrt{n}\ \boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{I}}_n^\dagger(\tilde{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}^\dagger), \tag{S2.4}$$

where $\big\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\dagger\big\| < \big\|\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}^\dagger\big\|$. Notice that

$$\sqrt{n}\ \boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger) = \sum_{j=1}^n \frac{1}{\sqrt{n}}\ \boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger) = \sum_{j=1}^n W_j,$$

where $W_j = \frac{1}{\sqrt{n}} \; \boldsymbol{h}^T \overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)$, for $j = 1, \ldots, n$ are independent random variables with zero mean and variance given by $v_n^2 = \sum_{j=1}^n \mathrm{Var}(W_j) = \boldsymbol{h}^T\boldsymbol{h}$, because $E_{G_j}\left[\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)\right] = \boldsymbol{0}$, for $j = 1, \ldots, n$. Moreover, for $\gamma > 0$,

$$
\begin{aligned}
\frac{1}{v_n^{2+\gamma}} \sum_{j=1}^n E_{G_j}\left[\, |W_j|^{2+\gamma}\,\right] &= \frac{1}{(\boldsymbol{h}^T\boldsymbol{h})^{1+\frac{\gamma}{2}}} \sum_{j=1}^n E_{G_j}\left[\left|\frac{1}{\sqrt{n}}\boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)\right|^{2+\gamma}\right] \\
&= \frac{1}{(\boldsymbol{h}^T\boldsymbol{h})^{1+\frac{\gamma}{2}} n^{1+\frac{\gamma}{2}}} \sum_{i=1}^n E_{G_j}\left[\left|\boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)\right|^{2+\gamma}\right] \\
&\leq \frac{\left\|\boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\right\|^{2+\gamma}}{(\boldsymbol{h}^T\boldsymbol{h})^{1+\frac{\gamma}{2}} n^{1+\frac{\gamma}{2}}} \sum_{j=1}^n E_{G_j}\left[\left\|\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)\right\|^{2+\gamma}\right] \\
&\leq \frac{\mathrm{cte}}{n^{1+\frac{\gamma}{2}}} \sum_{j=1}^n \sum_{k=1}^p E_{G_j}\left[\left|U_{jk}^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)\right|^{2+\gamma}\right] \longrightarrow 0,
\end{aligned}
$$

as $n \to \infty$. The last inequality follows from C3 and Rao (1973, 8a, p.149). Thus, by Liapunov's central limit theorem,

$$
\frac{\sum_{i=1}^n W_i}{v_n} \xrightarrow{\mathcal{D}} N(0,1),
$$

or equivalently,

$$
\sqrt{n}\; \boldsymbol{h}^T\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{U}}_n^\dagger(\boldsymbol{\theta}^\dagger) \xrightarrow{\mathcal{D}} N(0, \boldsymbol{h}^T\boldsymbol{h}).
$$

From (S2.4), it follows that

$$
\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{I}}_n^\dagger(\tilde{\boldsymbol{\theta}}_n)\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}^\dagger) \xrightarrow{\mathcal{D}} N_p(\boldsymbol{0}, \boldsymbol{I}_p). \qquad \text{(S2.5)}
$$

Now, given $\epsilon > 0$ and δ as in C5, with ϵ replaced by $\epsilon/2$, it follows from

Chebyshev's inequality, that $\forall\, k, l = 1, \ldots, p$

$$P\Big(\Big|\frac{1}{n}\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^\dagger,\delta)}\sum_{j=1}^{n} I^\dagger_{jkl}(\boldsymbol{\theta}) - \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\Big[\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^\dagger,\delta)} I^\dagger_{jkl}(\boldsymbol{\theta})\Big]\Big| < \frac{\epsilon}{2}\Big) \longrightarrow 1, \tag{S2.6}$$

as $n \to \infty$. A similar result holds when sup is replaced by inf. Thus, from (S2.6) and the fact that $\hat{\boldsymbol{\theta}}^*_n \xrightarrow{P} \boldsymbol{\theta}^\dagger$, it follows that

$$\begin{aligned}
&P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n} E_{G_j}\Big[I^\dagger_{jkl}(\boldsymbol{\theta}^\dagger)\Big] - \frac{1}{n}\sum_{j=1}^{n} I^\dagger_{jkl}(\tilde{\boldsymbol{\theta}}_n)\Big| < \epsilon\Big) \\
&\geq P\Big(\Big|\frac{1}{n}\sum_{j=1}^{n}\inf_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^\dagger,\delta)} I^\dagger_{jkl}(\boldsymbol{\theta}) - \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\Big[\inf_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^\dagger,\delta)} I^\dagger_{jkl}(\boldsymbol{\theta})\Big]\Big| < \frac{\epsilon}{2}, \\
&\Big|\frac{1}{n}\sum_{j=1}^{n}\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^\dagger,\delta)} I^\dagger_{jkl}(\boldsymbol{\theta}) - \frac{1}{n}\sum_{j=1}^{n} E_{G_j}\Big[\sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}^\dagger,\delta)} I^\dagger_{jkl}(\boldsymbol{\theta})\Big]\Big| < \frac{\epsilon}{2}, \|\hat{\boldsymbol{\theta}}^*_n - \boldsymbol{\theta}^\dagger\| < \delta\Big) \longrightarrow 1,
\end{aligned}$$

as $n \to \infty$. Then,

$$\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^\dagger) - \overline{\boldsymbol{I}}^\dagger_n(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{P} \mathbf{0}, \tag{S2.7}$$

as $n \to \infty$. C4, Lemma 2.2. in Mak (1982), and (S2.7) imply that $\overline{\boldsymbol{I}}^{\dagger\,-1}_n(\tilde{\boldsymbol{\theta}}_n)$ exists and

$$\big[\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^\dagger) - \overline{\boldsymbol{I}}^\dagger_n(\tilde{\boldsymbol{\theta}}_n)\big]\overline{\boldsymbol{I}}^{\dagger\,-1}_n(\tilde{\boldsymbol{\theta}}_n) \xrightarrow{P} \mathbf{0}. \tag{S2.8}$$

Then, from C3 and (S2.8), it follows that

$$\begin{aligned}
&\overline{\boldsymbol{\Gamma}}^{-\frac{1}{2}}_n(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{I}}^{\dagger\,-1}_n(\tilde{\boldsymbol{\theta}}_n)\overline{\boldsymbol{\Gamma}}^{\frac{1}{2}}_n(\boldsymbol{\theta}^\dagger) \\
&\qquad = \boldsymbol{I}_p + \overline{\boldsymbol{\Gamma}}^{-\frac{1}{2}}_n(\boldsymbol{\theta}^\dagger)\big[\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^\dagger) - \overline{\boldsymbol{I}}^\dagger_n(\tilde{\boldsymbol{\theta}}_n)\big]\overline{\boldsymbol{I}}^{\dagger\,-1}_n(\tilde{\boldsymbol{\theta}}_n)\overline{\boldsymbol{\Gamma}}^{\frac{1}{2}}_n(\boldsymbol{\theta}^\dagger) \xrightarrow{P} \boldsymbol{I}_p.
\end{aligned} \tag{S2.9}$$

Finally, to conclude the proof, (S2.5) and (S2.9) imply that

$$\begin{aligned}
&\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^\dagger)\sqrt{n}(\hat{\boldsymbol{\theta}}_n^\dagger - \boldsymbol{\theta}^\dagger) \\
&\quad = \left\{\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{I}}_n^{\dagger\,-1}(\tilde{\boldsymbol{\theta}}_n)\overline{\boldsymbol{\Gamma}}_n^{\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\right\}\left\{\overline{\boldsymbol{\Gamma}}_n^{-\frac{1}{2}}(\boldsymbol{\theta}^\dagger)\overline{\boldsymbol{I}}_n^\dagger(\tilde{\boldsymbol{\theta}}_n)\sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}^\dagger)\right\} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \xrightarrow{\mathcal{D}} N_p(\mathbf{0}, \boldsymbol{I}_p).
\end{aligned}$$

Remark 2. A simple calculation shows that the expectations in the matrices $\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta})$ and $\overline{\boldsymbol{\Gamma}}_n(\boldsymbol{\theta})$ can be expressed as $E_{G_j}[\boldsymbol{I}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta})] = (1-q)A_j^{(1)} + B_j$ and $E_{G_j}[\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta})\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta})^T] = A_j^{(2)}$, respectively, where $A_j^{(r)} = E_{G_j}\left[\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^T\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{r(1-q)}\right]$, for $r = 1, 2$; $B_j = E_{G_j}\left[\nabla\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^T\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q}\right]$, and $\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})$ and $\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})$ are given in (2.7).

Remark 3. As a consistent estimator of the asymptotic covariance matrix, we can consider the sandwich estimator given by $\overline{\boldsymbol{V}}_n^{-1}(\hat{\boldsymbol{\theta}}_n^*)\overline{\boldsymbol{S}}_n(\hat{\boldsymbol{\theta}}_n^*)\overline{\boldsymbol{V}}_n^{-1}(\hat{\boldsymbol{\theta}}_n^*)^T$, proposed by Giménez and Bolfarine (1997) in a similar context, where $\overline{\boldsymbol{V}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{j=1}^n \left\{(1-q)\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^T + \nabla\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^T\right\}\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q}$ and $\overline{\boldsymbol{S}}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{j=1}^n \tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^T\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{2(1-q)}$ are the sample counterparts of expectations in matrices $\overline{\boldsymbol{\Lambda}}_n(\boldsymbol{\theta})$ and $\overline{\boldsymbol{\Gamma}}_n(\boldsymbol{\theta})$, respectively.

S2.2 Proof of Proposition 1

We define the function $\mathbb{H}_n\colon \Theta \times \Theta \to \mathbb{R}$, by

$$\mathbb{H}_n(\boldsymbol{u}, \boldsymbol{v}) = H_n(\boldsymbol{u}) + \mathbb{Q}_n(\boldsymbol{u}, \boldsymbol{v}).$$

Suppose that $M_n(\boldsymbol{v}) \neq \boldsymbol{v}$, then we have that

$$\mathbb{Q}_n(M_n(\boldsymbol{v}), \boldsymbol{v}) \leq \mathbb{Q}_n(\boldsymbol{v}, \boldsymbol{v}). \tag{S2.10}$$

Considering the function $h\colon \mathbb{R}_{>0} \times \mathbb{R}_{>0} \to \mathbb{R}$, defined by $h(x, y) = L_q x - y^{1-q} \log x$, it is simple to see that for all $x \neq y \in \mathbb{R}_{>0}$, we have that $h(x, x) < h(y, x)$. Particularly, for all $j = 1, \dots n$

$$h\left(\tilde{f}_j(\boldsymbol{z}_j, \boldsymbol{v}), \tilde{f}_j(\boldsymbol{z}_j, \boldsymbol{v})\right) < h\left(\tilde{f}_j(\boldsymbol{z}_j, M_n(\boldsymbol{v})), \tilde{f}_j(\boldsymbol{z}_j, \boldsymbol{v})\right)$$

and adding for $j = 1, \dots, n$, we have that

$$\mathbb{H}_n\left(\boldsymbol{v}, \boldsymbol{v}\right) < \mathbb{H}_n\left(M_n(\boldsymbol{v}), \boldsymbol{v}\right). \tag{S2.11}$$

From (S2.10) and (S2.11), it follows that

$$H_n(\boldsymbol{v}) = \mathbb{H}_n\left(\boldsymbol{v}, \boldsymbol{v}\right) - \mathbb{Q}_n(\boldsymbol{v}, \boldsymbol{v}) < \mathbb{H}_n\left(M_n(\boldsymbol{v}), \boldsymbol{v}\right) - \mathbb{Q}_n(M_n(\boldsymbol{v}), \boldsymbol{v}) = H_n(M_n(\boldsymbol{v})).$$

S2.3 Proof of Proposition 2

The global convergence theorem in Zangwill (1969, p. 91), can be applied considering the following facts:

(i) The solution set is the set of all the fixed points of M_n, or equivalently, the stationary points of H_n.

(ii) The objective function H_n is a continuous function and satisfies the monotonicity property by Proposition 1.

(iii) The mapping M_n, which generates the sequence $\boldsymbol{\theta}^{(s+1)} = M_n(\boldsymbol{\theta}^{(s)})$, can be treated as a point-to-set mapping and it can be easily seen that is closed.

S3 Derivation of some results in the simple linear functional model

In this section we derive some results related with the application of Section 6. We denote the generic elements by $\boldsymbol{\theta} = (\alpha, \beta, \phi)^T \in \Theta$ and $\xi_j \in \Xi$; and the true values by $\theta^0 = (\alpha^0, \beta^0, \phi^0)^T$ and ξ_j^0 . Also, $\boldsymbol{\mu}_j = \boldsymbol{a} + \boldsymbol{b}\xi_j,$ with $\boldsymbol{a} = (0, \alpha)^T$ and $\boldsymbol{b} = (1, \beta)^T$, and analogously, we have the values $\boldsymbol{\mu}_j^0,$ $\boldsymbol{a}^0,$ $\boldsymbol{b}^0$ and ξ_j^0.

S3.1 Computational aspects

The MLqE $\hat{\boldsymbol{\theta}}_n^* = (\hat{\alpha}_n^*, \hat{\beta}_n^*, \hat{\phi}_n^*)^T$ is obtained as the maximizer of (6.2), or equivalently, as the solution of the estimating equation

$$\sum_{j=1}^{n} \tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q} = \mathbf{0},$$

where $\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = (\tilde{U}_{j1}(\boldsymbol{Z}_j;\boldsymbol{\theta}), \tilde{U}_{j2}(\boldsymbol{Z}_j;\boldsymbol{\theta}), \tilde{U}_{j3}(\boldsymbol{Z}_j;\boldsymbol{\theta}))^T$, with

$$\begin{aligned} \tilde{U}_{j1}(\boldsymbol{Z}_j;\boldsymbol{\theta}) &= \frac{1}{c\phi}(Y_j - \alpha - \beta X_j) && \text{(S3.1)} \\ \tilde{U}_{j2}(\boldsymbol{Z}_j;\boldsymbol{\theta}) &= \frac{1}{\phi c^2}(Y_j - \alpha - \beta X_j)[X_j + \beta(Y_j - \alpha)] && \text{(S3.2)} \\ \tilde{U}_{j3}(\boldsymbol{Z}_j;\boldsymbol{\theta}) &= -\frac{1}{\phi} + \frac{1}{2c\phi^2}(Y_j - \alpha - \beta X_j)^2. && \text{(S3.3)} \end{aligned}$$

A simple reweighting algorithm for computing the estimators, is derived as follows. If $\boldsymbol{\theta}^{(s)} = (\alpha^{(s)}, \beta^{(s)}, \phi^{(s)})^T$ denotes the estimator in step s, then the estimator in step $s+1$, $\boldsymbol{\theta}^{(s+1)} = (\alpha^{(s+1)}, \beta^{(s+1)}, \phi^{(s+1)})^T$, is a solution of the equations $\sum_{j=1}^n \omega_j^{(s)} \tilde{U}_{ji}(\boldsymbol{Z}_j;\boldsymbol{\theta}) = 0$, for $i = 1, 2, 3$, where $\tilde{U}_{ji}(\boldsymbol{Z}_j;\boldsymbol{\theta})$, for $i = 1, 2, 3$, are given in (S3.1), (S3.2), and (S3.3), and the weights $\omega_j^{(s)} = \omega_j(\boldsymbol{Z}_j;\boldsymbol{\theta}^{(s)})$, computed using (3.2), are updated at each step. These equations are a weighted version of maximum likelihood equations (Kimura (1992); Gleser (1981)). Let $\bar{\boldsymbol{Z}}_\omega^{(s)} = \sum_{j=1}^n \omega_j^{(s)} \boldsymbol{Z}_j = (\bar{X}_\omega^{(s)}, \bar{Y}_\omega^{(s)})^T$ and

$$\boldsymbol{S}_\omega^{(s)} = \sum_{j=1}^{n} \omega_j^{(s)} (\boldsymbol{Z}_j - \bar{\boldsymbol{Z}}_\omega^{(s)})(\boldsymbol{Z}_j - \bar{\boldsymbol{Z}}_\omega^{(s)})^T = \begin{pmatrix} S_{\omega,XX}^{(s)} & S_{\omega,XY}^{(s)} \\ S_{\omega,XY}^{(s)} & S_{\omega,YY}^{(s)} \end{pmatrix}$$

the vector of the weighted means and weighted covariance matrix, respectively. Thus, after some calculations, the MLqE in step $s+1$, are given by

$$
\begin{aligned}
\beta^{(s+1)} &= \frac{S^{(s)}_{\omega,YY} - S^{(s)}_{\omega,XX} + \sqrt{\left(S^{(s)}_{\omega,XX} - S^{(s)}_{\omega,YY}\right)^2 + 4\left(S^{(s)}_{\omega,XY}\right)^2}}{2S^{(s)}_{\omega,XY}},\\
\alpha^{(s+1)} &= \bar{Y}^{(s)}_{\omega} - \beta^{(s+1)}\bar{X}^{(s)}_{\omega},\\
\phi^{(s+1)} &= \frac{1}{2(1+(\beta^{(s+1)})^2)}\big(S^{(s)}_{\omega,YY} - 2\beta^{(s+1)}S^{(s)}_{\omega,XY} + (\beta^{(s+1)})^2 S^{(s)}_{\omega,XX}\big).
\end{aligned}
$$

The algorithm can be initialized by setting $\beta^{(0)}$, $\alpha^{(0)}$, and $\phi^{(0)}$ as the MLE estimates.

S3.2 Derivation of some results when the true density belongs to the model

Assuming that the true density belongs to the model, that is $\boldsymbol{Z}_j \sim N_2(\boldsymbol{\mu}^0_j, \phi^0\boldsymbol{I}_2)$, we have that

$$
f_j(\boldsymbol{z}_j;\boldsymbol{\theta}^0,\boldsymbol{\xi}^0_j) = \frac{1}{2\pi\phi^0}\exp\Big\{-\frac{1}{2\phi^0}(\boldsymbol{z}_j-\boldsymbol{\mu}^0_j)^T(\boldsymbol{z}_j-\boldsymbol{\mu}^0_j)\Big\}. \tag{S3.4}
$$

Also, we have that

$$
\begin{aligned}
\tilde{f}_j(\boldsymbol{z}_j;\boldsymbol{\theta}) = f_j(\boldsymbol{z}_j;\boldsymbol{\theta},\hat{\xi}_j) &= \frac{1}{2\pi\phi}\exp\Big\{-\frac{1}{2\phi c}(y_j-\alpha-\beta x_j)^2\Big\}\\
&= \frac{1}{2\pi\phi}\exp\Big\{-\frac{1}{2\phi}(\boldsymbol{z}_j-\boldsymbol{a})^T A(\boldsymbol{z}_j-\boldsymbol{a})\Big\}, \qquad \text{(S3.5)}
\end{aligned}
$$

where $A = \boldsymbol{I}_2 - \dfrac{1}{c}\boldsymbol{b}\boldsymbol{b}^T$, with $c = \boldsymbol{b}^T\boldsymbol{b} = 1+\beta^2$.

We denote by E_j the expectation with respect to the model distribution $N_2(\boldsymbol{\mu}_j^0, \phi^0 \boldsymbol{I}_2)$.

Lemma 1. *The maximum of* $\frac{1}{n}\sum_{j=1}^n E_j[h_j(\boldsymbol{Z}_j;\boldsymbol{\theta})]$ *for all large* n*, where* $h_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = L_q\{\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\} = \frac{1}{1-q}\big\{\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q} - 1\big\}$*, is attained for* $\boldsymbol{\theta}^\dagger = (\alpha^0, \beta^0, k\phi^0)^T$*, with* $k = q - \frac{1}{2} > 0$.

Proof. We need to find the maximum of $\frac{1}{n}\sum_{j=1}^n E_j[\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q}]$ for all large n. Using (S3.4) and (S3.5), we can see that

$$\begin{aligned}
E_j[\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q}] &= \int_{\mathbb{R}^2} \tilde{f}_j(\boldsymbol{z}_j;\boldsymbol{\theta})^{1-q} f_j(\boldsymbol{z}_j;\boldsymbol{\theta}^0,\boldsymbol{\xi}_j^0)\, d\boldsymbol{z}_j \\
&= \frac{1}{(2\pi\phi)^{1-q}}\frac{1}{2\pi\phi^0}\int_{\mathbb{R}^2} \exp\{-S(\boldsymbol{z}_j;\boldsymbol{\theta},\xi_j,\boldsymbol{\theta}^0,\xi_j^0,q)\}\, d\boldsymbol{z}_j,
\end{aligned} \tag{S3.6}$$

where

$$S(\boldsymbol{z}_j;\boldsymbol{\theta},\xi_j,\boldsymbol{\theta}^0,\xi_j^0,q) = \frac{(1-q)}{2\phi}(\boldsymbol{z}_j-\boldsymbol{a})^T A(\boldsymbol{z}_j-\boldsymbol{a}) + \frac{1}{2\phi^0}(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0)^T(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0).$$

Using that

$$(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0)^T(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0) = (\boldsymbol{z}_j-\boldsymbol{a}^0)^T(\boldsymbol{z}_j-\boldsymbol{a}^0) - 2(\boldsymbol{z}_j-\boldsymbol{a}^0)^T\boldsymbol{b}^0\xi_j^9 + \boldsymbol{b}^{0\,T}\boldsymbol{b}^0\xi_j^{0\,2},$$

we can write

$$S(\boldsymbol{z}_j;\boldsymbol{\theta},\xi_j,\boldsymbol{\theta}^0,\xi_j^0,q) = (\boldsymbol{z}_j-\boldsymbol{a}^0)^T B^*(\boldsymbol{z}_j-\boldsymbol{a}^0) + \boldsymbol{b}^{*\,T}(\boldsymbol{z}_j-\boldsymbol{a}^0) + b_0^*,$$

where

$$B^* = \frac{(1-q)}{2\phi}A + \frac{1}{2\phi^0}\boldsymbol{I}_2,$$
$$\boldsymbol{b}^* = \frac{(1-q)}{\phi}A(\boldsymbol{a}^0 - \boldsymbol{a}) - \frac{1}{\phi^0}\boldsymbol{b}^0\xi_j^0,$$
$$b_0^* = \frac{(1-q)}{2\phi}(\boldsymbol{a}_0 - \boldsymbol{a})^T A(\boldsymbol{a}_0 - \boldsymbol{a}) + \frac{1}{2\phi^0}\boldsymbol{b}^{0\,T}\boldsymbol{b}^0\xi_j^{0\,2}.$$

Using results in Harville (1997, p. 321), we have that

$$\int_{\mathbb{R}^2} \exp\{-S(\boldsymbol{z}_j;\boldsymbol{\theta},\xi_j,\boldsymbol{\theta}^0,\xi_j^0,q)\}\,d\boldsymbol{z}_j = \frac{1}{2}\pi|B^*|^{-\frac{1}{2}}\exp\{\frac{1}{4}\boldsymbol{b}^{*\,T}B^{*\,-1}\boldsymbol{b}^* - b_0^*\}.$$

Then, after some algebra, (S3.6) can be written as

$$\frac{1}{4(2\pi)^{1-q}}\,g_1(\phi,\phi^0,q)\exp\{-g_2(\phi,\phi^0,q)[(\boldsymbol{a}^0-\boldsymbol{a})^T(\boldsymbol{a}_0-\boldsymbol{a}) + (\boldsymbol{a}^0-\boldsymbol{a})A\boldsymbol{b}^{0\,T}\xi_j$$
$$+\,\boldsymbol{b}^{0\,T}A\boldsymbol{b}^0\xi_j^{0\,2}]\}, \tag{S3.7}$$

where

$$g_1(\phi,\phi^0,q) = \frac{2}{\phi^{(\frac{1}{2}-q)}[(1-q)\phi^0+\phi]^{\frac{1}{2}}} \text{ and } g_2(\phi,\phi^0,q) = \frac{1-q}{2[(1-q)\phi^0+\phi]}.$$

From expression (S3.7), we can easily seen that the maximum of (S3.6) is independent of ξ_j^0, and it is attained at $\boldsymbol{a} = \boldsymbol{a}^0$, $\boldsymbol{b} = \boldsymbol{b}^0$, and $\phi = k\phi^0$, with $k = q - \frac{1}{2}$. Then, we have $\boldsymbol{\theta}^\dagger = (\alpha^0,\beta^0,k\phi^0)^T$.

$\square$

Lemma 2. *Let $\tilde{f}_j(\boldsymbol{z}_j;\boldsymbol{\theta}^\dagger)$ given in (S3.5), with $\boldsymbol{\theta}^\dagger = (\alpha^0,\beta^0,k\phi^0)^T$, where $k = q - \frac{1}{2} > 0$ and $f_j(\boldsymbol{z}_j;\boldsymbol{\theta}^0,\boldsymbol{\xi}_j^0)$ in (S3.4). Then, for all function $V(\boldsymbol{Z}_j;\boldsymbol{\theta})$*

such that the expectation below exists, we have that

$$E_j[V(Z_j;\boldsymbol{\theta}^\dagger)\tilde{f}_j(\boldsymbol{z}_j;\boldsymbol{\theta}^\dagger)^{r(1-q)}] = c_{(r)}E_j^{(r)}[V(Z_j;\boldsymbol{\theta}^\dagger)], \quad r=1,2, \quad j=1,2,\ldots,$$

where $c_{(r)} = (2\pi k\phi^0)^{-\frac{r(1-q)(p+1)}{2}} a_{(r)}^{-\frac{p}{2}}$, $a_{(r)} = \dfrac{r(1-q)+k}{k}$, *and* $E_j^{(r)}$ *denotes expectation taken with respect to the distribution* $N_2(\boldsymbol{\mu}_j^0, \boldsymbol{\Sigma}_{(r)}^0)$, *with* $\boldsymbol{\Sigma}_{(r)}^0 = \dfrac{\phi^0}{a_{(r)}}\left(\boldsymbol{I}_2 + \dfrac{r(1-q)}{k\,c^0}\ \boldsymbol{b}^0\boldsymbol{b}^{0\,T}\right)$, $\quad c^0 = \boldsymbol{b}^{0\,T}\boldsymbol{b}^0 = 1+(\beta^0)^2$, *for* $r=1,2$.

Proof.

$$E_j[V(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)^{\,r(1-q)}] = \int V(\boldsymbol{z}_j;\boldsymbol{\theta}^\dagger)\tilde{f}_j(\boldsymbol{z}_j;\boldsymbol{\theta}^\dagger)^{\,r(1-q)} f_j(\boldsymbol{z}_j;\boldsymbol{\theta}^0,\xi_j^0)\,d\boldsymbol{z}_j.$$

Moreover, from a direct calculation, we can show that

$$\begin{aligned}
&\tilde{f}_j(\boldsymbol{z}_j;\boldsymbol{\theta}^\dagger)^{\,r(1-q)} f_j(\boldsymbol{z}_j;\boldsymbol{\theta}^0,\xi_j^0)\\
&\quad = c_{(r)}\frac{1}{(2\pi)}\frac{a_{(r)}^{\frac{1}{2}}}{\phi^{0\,2}}\exp\left\{-\frac{a_{(r)}}{2\phi^0}(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0)^T\left(\boldsymbol{I}_2 - \frac{r(1-q)}{k\,c^0\,a_{(r)}}\boldsymbol{b}^0\boldsymbol{b}^{0\,T}\right)(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0)\right\}\\
&\quad = c_{(r)}\frac{1}{(2\pi)}\frac{1}{|\boldsymbol{\Sigma}_{(r)}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0)^T\boldsymbol{\Sigma}_{(r)}^{-1}(\boldsymbol{z}_j-\boldsymbol{\mu}_j^0)\right\}.
\end{aligned}$$

Then, the result follows inmediately. □

Lemma 3. *Let* $\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}) = \tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})\tilde{f}_j(\boldsymbol{Z}_j;\boldsymbol{\theta})^{1-q}$, *with* $\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}) = (\tilde{U}_{j1}(\boldsymbol{Z}_j;\boldsymbol{\theta}),\tilde{U}_{j2}(\boldsymbol{Z}_j;\boldsymbol{\theta}),\tilde{U}_{j3}(\boldsymbol{Z}_j;\boldsymbol{\theta}))^T$, *where* $\tilde{U}_{j1},\tilde{U}_{j2}$, *and* $\tilde{U}_{j3}$ *are given in (S3.1), (S3.2), and (S3.3), respectively. Then,* $E_j[\boldsymbol{U}_j^\dagger(\boldsymbol{Z}_j;\boldsymbol{\theta}^\dagger)] = 0$, *for* $j=1,2,\ldots$

Proof. From Lemma 2, we have that

$$E_j[\boldsymbol{U}_j^{\dagger}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})] = c_{(1)}E_j^{(1)}[\tilde{\boldsymbol{U}}_j(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})],$$

where $E_j^{(1)}$ denotes expectation with respect to the distribution $N_2(\boldsymbol{\mu}_j^0, \Sigma_{(1)})$. To calculate these expectations, it is convenient to rewrite expressions (S3.1), (S3.2), and (S3.3) in the following equivalent form

$$\tilde{U}_{j1}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger}) = \frac{1}{k\phi^0 c^0}\boldsymbol{d}^T(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0),$$

$$\tilde{U}_{j2}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger}) = \frac{1}{k\phi^0 c^0}\left\{\frac{1}{c^0}\boldsymbol{d}^T(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)^T\boldsymbol{b}^0 + \boldsymbol{d}^T(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)\xi_j^0\right\},$$

$$\tilde{U}_{j3}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger}) = -\frac{1}{k\phi^0} + \frac{1}{2(k\phi^0)^2}(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)^T A(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0),$$

with $\boldsymbol{d} = (-\beta^0, 1)^T$ and $A = \boldsymbol{I}_2 - \frac{1}{c^0}\boldsymbol{b}^0\boldsymbol{b}^{0\,T}$. Using that $E_j^{(1)}[\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0] = \boldsymbol{0}$, $E_j^{(1)}[(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)^T] = \Sigma_{(1)}$, and $E_j^{(1)}[(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)^T A(\boldsymbol{Z}_j - \boldsymbol{\mu}_j^0)] = \mathrm{tr}(A\Sigma_{(1)})$, after some simple algebra, we have that $E_j^{(1)}[\tilde{U}_{ji}(\boldsymbol{Z}_j;\boldsymbol{\theta}^{\dagger})] = 0$, for $i = 1, 2, 3$. □

References

Giménez, P. and Bolfarine, H. (1997). Corrected score functions in classical error-in-variables and incidental parameter models. *Australian Journal of Statistics* **39**, 325–344.

Gleser, L.J. (1981). Estimation in a multivariate "error in variables" regression model: Large sample results. *The Annals of Statistics* **9**, 24–44.

Harville, D.A. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer, N.Y.

Kimura, D.K. (1992). Functional comparative calibration using an EM algorithm. *Biometrics* **48**, 1263–1271.

Mak, T.K. (1982). Estimation in the presence of incidental parameters. *The Canadian Journal of Statistics* **10**, 121–132.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. 2nd ed. Wiley, N.Y.

Rudin, W. (1964). *Principles of Mathematical Analysis*. McGraw Hill, N.Y.

Zangwill, W. (1969). *Nonlinear Programming - A Unified Approach*. Prentice-Hall, Englewood Cliffs, N.Y.