Quasi-maximum likelihood estimation and penalized estimation under non-standard conditions

Junichiro Yoshida · Nakahiro Yoshida

Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan.

Supplementary material

6 Appendix: proof of (22)

Let

$$\mathcal{K}(\lambda, a, b) = \frac{b^\lambda}{2}\int_0^\infty t^{\lambda-1}\exp\left[-\frac{1}{2}\left(bt+\frac{a}{t}\right)\right]dt$$

Then $\mathcal{K}(\lambda, \delta^2, \gamma^2) = (\gamma\delta)^\lambda K_\lambda(\gamma\delta)$.

Simply denoted by $p(x;\lambda,\delta,\gamma)$, the density $p_{\mathrm{GIG}}(x;\lambda,\delta,\gamma)$ is expressed as

$$p(x;\lambda,\delta,\gamma) = \frac{\gamma^{2\lambda}}{2\mathcal{K}(\lambda,\delta^2,\gamma^2)}x^{\lambda-1}\exp\left[-\frac{1}{2}\left(\frac{\delta^2}{x}+\gamma^2 x\right)\right] \qquad (x>0)$$

This model is a curved exponential family:

$$p(x;\lambda,\delta,\gamma) = \exp\left[(\lambda-1)\log x - \frac{\delta^2}{2x} - \frac{\gamma^2}{2}x - \Psi(\lambda,\delta^2,\gamma^2)\right] \qquad (x>0)$$

with the potential $\Psi(\lambda,a,b) = -\log\frac{b^\lambda}{2\mathcal{K}(\lambda,a,b)}$.

We have

$$\begin{aligned}
&\Psi(\lambda,\delta^2,\gamma^2) - \Psi(\lambda^*,0,(\gamma^*)^2)\\
=\ &\log\int_0^\infty \exp\left[(\lambda-1)\log x - \frac{\delta^2}{2x} - \frac{\gamma^2}{2}x\right]dx\\
&\quad -\log\int_0^\infty \exp\left[(\lambda^*-1)\log x - \frac{0}{2x} - \frac{(\gamma^*)^2}{2}x\right]dx\\
=\ &D\big[\Delta(\lambda,\delta^2,\gamma^2)\big] + \int_0^1 (1-s)H(s,\lambda,\delta^2,\gamma^2)\big[(\Delta(\lambda,\delta^2,\gamma^2)^{\otimes 2}\big]ds.
\end{aligned}$$

Here

$$\Delta(\lambda,\delta^2,\gamma^2) = \begin{pmatrix}\lambda-\lambda^*\\ \delta^2\\ \gamma^2-(\gamma^*)^2\end{pmatrix},$$

$$D = \big(\partial_{(\lambda,a,b)}\Psi\big)\big(\lambda^*,0,(\gamma^*)^2\big) = \begin{pmatrix}E[\log\xi_0]\\ -2^{-1}E[\xi_0^{-1}]\\ -2^{-1}E[\xi_0]\end{pmatrix}$$

and

$$
\begin{aligned}
&H(s,\lambda,\delta^2,\gamma^2)\\
&=\big(\partial^2_{(\lambda,a,b)}\Psi\big)(s\lambda+(1-s)\lambda^*,s\delta^2,s\gamma^2+(1-s)(\gamma^*)^2)\\
&=\begin{pmatrix}
\mathrm{Var}\big[\log\xi_s\big] & -2^{-1}\mathrm{Cov}\big[\log\xi_s,\xi_s^{-1}\big] & -2^{-1}\mathrm{Cov}\big[\log\xi_s,\xi_s\big]\\
-2^{-1}\mathrm{Cov}\big[\log\xi_s,\xi_s^{-1}\big] & 4^{-1}\mathrm{Var}\big[\xi_s^{-1}\big] & 4^{-1}\mathrm{Cov}\big[\xi_s^{-1},\xi_s\big]\\
-2^{-1}\mathrm{Cov}\big[\log\xi_s,\xi_s\big] & 4^{-1}\mathrm{Cov}\big[\xi_s^{-1},\xi_s\big] & 4^{-1}\mathrm{Var}\big[\xi_s\big]
\end{pmatrix},
\end{aligned}
$$

with $\xi_s = \xi_s(\lambda,\delta,\gamma) = \xi(s\lambda+(1-s)\lambda^*,\sqrt{s}\delta,\sqrt{s\gamma^2+(1-s)(\gamma^*)^2})$, where $\xi(\lambda,\delta,\gamma)$ denotes a random variable such that $\xi(\lambda,\delta,\gamma)\sim\mathrm{GIG}(\lambda,\delta,\gamma)$.

Let $a_n=\mathrm{diag}\big[n^{-1/2},n^{-1/4},n^{-1/2}\big]$, and define U_n and U as (5) and (7), respectively, for $\Theta:=[\underline{\lambda},\overline{\lambda}]\times[0,\overline{\delta}]\times[\underline{\gamma},\overline{\gamma}]$ and $\theta^*:=(\lambda^*,0,\gamma^*)$. For $\mathbb{H}_n(\theta)=\sum_{j=1}^n\log p(X_j;\theta)$ ($\theta=(\lambda,\delta,\gamma)\in\Theta$) and $u=(u_1,u_2,u_3)\in U_n$, we obtain

$$
\begin{aligned}
\log\mathbb{Z}_n(u) &= \mathbb{H}_n(\theta^*+a_nu)-\mathbb{H}_n(\theta^*)\\
&= u_1n^{-1/2}\sum_{j=1}^n\widetilde{\log X_j}-u_2^2n^{-1/2}\sum_{j=1}^n2^{-1}\widetilde{X_j^{-1}}-u_3\gamma^*n^{-1/2}\sum_{j=1}^n\widetilde{X_j}\\
&\quad -n\int_0^1(1-s)H(s,\lambda^*+n^{-1/2}u_1,n^{-1/2}u_2^2,(\gamma^*+n^{-1/2}u_3)^2))\\
&\qquad\times\left[\left(n^{-1/2}u_1,n^{-1/2}u_2^2,(\gamma^*+n^{-1/2}u_3)^2-(\gamma)^*)^2\right)^{\otimes 2}\right]ds\\
&\quad -u_3^22^{-1}n^{-1}\sum_{j=1}^n\widetilde{X}_j,
\end{aligned}
$$

where we are writing $\widetilde{F(X_j)}=F(X_j)-E[F(X_j)]$ for a function $F(X_j)$ of a random variable X_j satisfying $X_j\sim GIG(\lambda^*,0,\gamma^*)=\Gamma(\lambda^*,(\gamma^*)^2/2)$. Then it is possible to write it as

$$
\begin{aligned}
\log\mathbb{Z}_n(u) &= u_1n^{-1/2}\sum_{j=1}^n\widetilde{\log X_j}-u_2^2n^{-1/2}\sum_{j=1}^n2^{-1}\widetilde{X_j^{-1}}-u_3\gamma^*n^{-1/2}\sum_{j=1}^n\widetilde{X_j}\\
&\quad -\frac{1}{2}C\big[(u_1,u_2^2,u_3)^{\otimes}\big]+r_n(u)
\end{aligned}
$$

with the positive-definite covariance matrix

$$
C=\begin{pmatrix}
\mathrm{Var}\big[\log\xi_0\big] & -2^{-1}\mathrm{Cov}\big[\log\xi_0,\xi_0^{-1}\big] & -\gamma^*\mathrm{Cov}\big[\log\xi_0,\xi_0\big]\\
-2^{-1}\mathrm{Cov}\big[\log\xi_0,\xi_0^{-1}\big] & 4^{-1}\mathrm{Var}\big[\xi_0^{-1}\big] & 2^{-1}\gamma^*\mathrm{Cov}\big[\xi_0^{-1},\xi_0\big]\\
-\gamma^*\mathrm{Cov}\big[\log\xi_0,\xi_0\big] & 2^{-1}\gamma^*\mathrm{Cov}\big[\xi_0^{-1},\xi_0\big] & (\gamma^*)^2\mathrm{Var}\big[\xi_0\big]
\end{pmatrix}, \tag{85}
$$

and the term $r_n(u)$ satisfying

$$
\sup_{u\in U_n}\frac{|r_n(u)|}{1+|u_1|^2+|u_2|^4+|u_3|^2}=O_p(n^{-1/2}).
$$

Then, Condition [**A1**] is verified by the estimate

$$\lim_{R\to\infty}\limsup_{n\to\infty} P\left[M_n \geq \left(2^{-1}\lambda_{\min}[C] + O_p(n^{-1/2})\right)R\right] = 0$$

for $M_n = \left|n^{-1/2}\sum_{j=1}^n \widetilde{\log X_j}\right| + \left|n^{-1/2}\sum_{j=1}^n 2^{-1}\widetilde{X_j^{-1}}\right| + \left|n^{-1/2}\gamma^* \sum_{j=1}^n \widetilde{X_j}\right|$.

Condition [**A2**] is satisfied with

$$\begin{aligned}\mathbb{V}_n(u) = \exp\Bigg(&u_1 n^{-1/2}\sum_{j=1}^n \widetilde{\log X_j} - u_2^2 n^{-1/2}\sum_{j=1}^n 2^{-1}\widetilde{X_j^{-1}} \\ &-u_3\gamma^* n^{-1/2}\sum_{j=1}^n \widetilde{X_j} - \frac{1}{2}C\left[(u_1, u_2^2, u_3)^{\otimes 2}\right]\Bigg],\\ \mathbb{Z}(u) = \exp\Bigg(&\Delta\cdot(u_1, u_2^2, u_3) - \frac{1}{2}C\left[(u_1, u_2^2, u_3)^{\otimes 2}\right]\Bigg)\end{aligned}$$

with a three-dimensional random vector $\Delta \sim N_3(0, C)$. Now from Example 1, [**A3**] holds, and we obtain $U = \mathbb{R}\times[0,\infty)\times\mathbb{R}$. Condition [**A4**] obviously holds. Therefore, Theorem 1 concludes that the MLE $\left(\widehat{\lambda}_n, \widehat{\delta}_n, \widehat{\gamma}_n\right)$ admits (22). □