SUPPLEMENT TO "ROBUST PRIORS IN NONLINEAR PANEL DATA MODELS": SUPPLEMENTARY APPENDIX

By Manuel Arellano and Stéphane Bonhomme

This supplementary appendix contains proofs of some results contained in the paper. Specifically, section S1 provides proofs of Theorem 4 and its corollary, concerning the asymptotic distribution of flexible random effects estimators. Section S1 also proves Theorem 5, its corollary, and Theorem 6 concerning the bias and the asymptotic distribution of estimated marginal effects. Section S2 proves results stated in the paper for the autoregressive and logit models that we use as illustrations. It also contains results for a Poisson counts model as a further example. We keep the same notation as in the paper.

S1. PROOFS ON FLEXIBLE RANDOM EFFECTS AND POLICY PARAMETERS

An intermediate lemma to show Theorem 4. The following lemma gives the first terms of the asymptotic expansion of the score of the concentrated random effects likelihood when N and T go to infinity.

Lemma S1 *Let us assume that*

$$
\begin{array}{ll}
i) & \lim_{\alpha\to\pm\infty} \rho_i(\theta_0,\alpha)\pi_0(\alpha) = 0 \\
ii) & \lim_{\alpha\to\pm\infty} \rho_i(\theta_0,\alpha)\pi_0(\alpha)\left(\ln\widetilde{\pi}_0(\alpha)-\ln\pi_0(\alpha)\right) = 0 \\
iii) & \mathbb{E}_{\pi_0}\left[\left(\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i(\theta_0,\alpha_i)+\frac{\partial\ln\pi_0(\alpha_{i0})}{\partial\alpha_i}\rho_i(\theta_0,\alpha_{i0})\right)'\left(\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i(\theta_0,\alpha_i)+\frac{\partial\ln\pi_0(\alpha_{i0})}{\partial\alpha_i}\rho_i(\theta_0,\alpha_{i0})\right)\right]<\infty.
\end{array}
$$

Then:

$$
\frac{1}{N}\sum_{i=1}^N \frac{\partial}{\partial\theta}\Big|_{\theta_0}\ell_i^{RE}\left(\theta,\widehat{\xi}\left(\theta\right)\right)=\frac{1}{N}\sum_{i=1}^N\frac{\partial}{\partial\theta}\Big|_{\theta_0}\overline{\ell}_i\left(\theta\right)+O\left(\frac{\mathcal{K}(\pi_0,\widetilde{\pi}_0)}{T}\right)+o_p\left(\frac{1}{T}\right).
$$

$i)$ and $ii)$ are limit conditions that are satisfied if the tails of π_0 are thin enough. Condition $iii)$ requires the existence of some moments. As a particular case, the conditions are satisfied if π_0 has compact support. Lemma S1 will allow us to derive the asymptotic properties of the REML estimator of θ when N and T tend to infinity at the same rate.

Proof of Lemma S1: We have:

$$
\frac{1}{N}\sum_{i=1}^N \frac{\partial}{\partial\theta}\Big|_{\theta_0}\ell_i^{RE}\left(\theta,\widehat{\xi}\left(\theta_0\right)\right) \quad = \quad \frac{1}{N}\sum_{i=1}^N\frac{\partial}{\partial\theta}\Big|_{\theta_0}\overline{\ell}_i\left(\theta\right)+\frac{b}{T}+o_p\left(\frac{1}{T}\right)
$$

where

$$\begin{aligned}
b &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta}\Big|_{\theta_0}\ln\pi\left(\overline{\alpha}_i(\theta);\widehat{\xi}\left(\theta_0\right)\right)+\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i\left(\theta_0,\alpha_i\right)\right)\\
&= \underbrace{\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta}\Big|_{\theta_0}\ln\pi_0\left(\overline{\alpha}_i(\theta)\right)+\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i\left(\theta_0,\alpha_i\right)\right)}_{A}\\
&\quad \underbrace{+\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta}\Big|_{\theta_0}\left(\ln\pi\left(\overline{\alpha}_i(\theta);\widehat{\xi}\left(\theta_0\right)\right)-\ln\pi_0\left(\overline{\alpha}_i(\theta)\right)\right)\right)}_{B}.
\end{aligned}$$

We have:

$$\begin{aligned}
A &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta}\Big|_{\theta_0}\ln\pi_0\left(\overline{\alpha}_i(\theta)\right)+\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i\left(\theta_0,\alpha_i\right)\right)\\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\pi_0}\left[\frac{\partial}{\partial\theta}\Big|_{\theta_0}\ln\pi_0\left(\overline{\alpha}_i(\theta)\right)+\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i\left(\theta_0,\alpha_i\right)\right].
\end{aligned}$$

Using that

$$\frac{\partial}{\partial\theta}\Big|_{\theta_0}\overline{\alpha}_i(\theta)=\rho_i(\theta_0,\alpha_{i0})$$

we find that:

$$\begin{aligned}
\mathbb{E}_{\pi_0}\left[\frac{\partial}{\partial\theta}\Big|_{\theta_0}\ln\pi_0\left(\overline{\alpha}_i(\theta)\right)+\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i\left(\theta_0,\alpha_i\right)\right] &= \mathbb{E}_{\pi_0}\left[\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\ln\pi_0\left(\alpha_i\right)+\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i\left(\theta_0,\alpha_i\right)\right]\\
&= \int\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\left(\rho_i\left(\theta_0,\alpha_i\right)\pi_0\left(\alpha_i\right)\right)d\alpha_{i0}.
\end{aligned}$$

This term is zero as $\rho_i(\theta_0,\alpha)\pi_0(\alpha)\underset{\alpha\to\pm\infty}{\to}0$. So:

$$A=\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta}\Big|_{\theta_0}\ln\pi_0\left(\overline{\alpha}_i(\theta)\right)+\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\rho_i\left(\theta_0,\alpha_i\right)\right) = 0.$$

Now:

$$\begin{aligned}
B &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial}{\partial\theta}\Big|_{\theta_0}\left(\ln\pi\left(\overline{\alpha}_i(\theta);\widehat{\xi}\left(\theta_0\right)\right)-\ln\pi_0\left(\overline{\alpha}_i(\theta)\right)\right)\right)\\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\left(\ln\pi\left(\alpha_i;\widehat{\xi}\left(\theta_0\right)\right)-\ln\pi_0\left(\alpha_i\right)\right)\\
&= \underbrace{\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\left(\ln\pi\left(\alpha_i;\widehat{\xi}\left(\theta_0\right)\right)-\ln\widetilde{\pi}_0\left(\alpha_i\right)\right)}_{C}\\
&\quad \underbrace{+\operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial\alpha_i}\Big|_{\alpha_{i0}}\left(\ln\widetilde{\pi}_0\left(\alpha_i\right)-\ln\pi_0\left(\alpha_i\right)\right)}_{D}.
\end{aligned}$$

Now the proof of Lemma 2 shows that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln \pi\left(\alpha_{i0};\widehat{\xi}\left(\theta_0\right)\right)}{\partial \xi}=O_p\left(\frac{1}{T}\right).$$

So:

$$\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln \pi\left(\alpha_{i0};\widehat{\xi}_0\right)}{\partial \xi} \\
&\approx \frac{1}{N}\sum_{i=1}^{N}\frac{\partial \ln \pi\left(\alpha_{i0};\widehat{\xi}\left(\theta_0\right)\right)}{\partial \xi}+\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 \ln \pi\left(\alpha_{i0};\widehat{\xi}\left(\theta_0\right)\right)}{\partial \xi\partial \xi'}\left(\widehat{\xi}_0-\widehat{\xi}\left(\theta_0\right)\right).
\end{aligned}$$

It follows that:

$$\widehat{\xi}\left(\theta_0\right)=\widehat{\xi}_0+O_p\left(\frac{1}{T}\right)$$

and

$$\begin{aligned}
C &= \operatorname*{plim}_{N\rightarrow\infty}\ \frac{1}{N}\sum_{i=1}^{N}\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\left(\ln \pi\left(\alpha_i;\widehat{\xi}\left(\theta_0\right)\right)-\ln \widetilde{\pi}_0\left(\alpha_i\right)\right) \\
&= \operatorname*{plim}_{N\rightarrow\infty}\ \frac{1}{N}\sum_{i=1}^{N}\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\left(\ln \pi\left(\alpha_i;\widehat{\xi}\left(\theta_0\right)\right)-\ln \pi\left(\alpha_i;\overline{\xi}_0\right)\right) \\
&= \operatorname*{plim}_{N\rightarrow\infty}\ \frac{1}{N}\sum_{i=1}^{N}\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\left(\ln \pi\left(\alpha_i;\widehat{\xi}\left(\theta_0\right)\right)-\ln \pi\left(\alpha_i;\widehat{\xi}_0\right)\right)=O\left(\frac{1}{T}\right).
\end{aligned}$$

We also have

$$\begin{aligned}
D &= \operatorname*{plim}_{N\rightarrow\infty}\ \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\pi_0}\left[\rho_i(\theta_0,\alpha_{i0})\frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\left(\ln \widetilde{\pi}_0\left(\alpha_i\right)-\ln \pi_0\left(\alpha_i\right)\right)\right] \\
&= \operatorname*{plim}_{N\rightarrow\infty}\ \frac{1}{N}\sum_{i=1}^{N}\int \rho_i(\theta_0,\alpha_{i0})\pi_0(\alpha_{i0})\frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\left(\ln \widetilde{\pi}_0\left(\alpha_i\right)-\ln \pi_0\left(\alpha_i\right)\right)d\alpha_{i0} \\
&= \operatorname*{plim}_{N\rightarrow\infty}\ \frac{1}{N}\sum_{i=1}^{N}\int \frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\Big(\rho_i(\theta_0,\alpha_i)\pi_0(\alpha_i)\left(\ln \widetilde{\pi}_0\left(\alpha_i\right)-\ln \pi_0\left(\alpha_i\right)\right)\Big)d\alpha_{i0} \\
&\quad -\operatorname*{plim}_{N\rightarrow\infty}\ \frac{1}{N}\sum_{i=1}^{N}\int \left(\frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\rho_i(\theta_0,\alpha_i)\pi_0(\alpha_i)\right)\left(\ln \widetilde{\pi}_0\left(\alpha_{i0}\right)-\ln \pi_0\left(\alpha_{i0}\right)\right)d\alpha_{i0}
\end{aligned}$$

The first term is zero as $\rho_i(\theta_0,\alpha)\pi_0(\alpha)\left(\ln \widetilde{\pi}_0(\alpha)-\ln \pi_0(\alpha)\right)\underset{\alpha\rightarrow\pm\infty}{\rightarrow}0$. As for the second term remark that, using the Cauchy-Schwarz inequality:

$$\left|\int\left(\pi_0^{-\frac{1}{2}}\frac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\rho_i(\theta_0,\alpha_i)\pi_0(\alpha_i)\right)\Big(\pi_0^{\frac{1}{2}}\left(\ln \widetilde{\pi}_0\left(\alpha_{i0}\right)-\ln \pi_0\left(\alpha_{i0}\right)\right)\Big)d\alpha_{i0}\right| \ \leq\ C^{st}\mathcal{K}\left(\pi_0,\widetilde{\pi}_0\right)$$

where C^{st} is

$$\left\{\mathbb{E}_{\pi_0}\left[\left(\tfrac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\rho_i(\theta_0,\alpha_i)+\tfrac{\partial \ln \pi_0(\alpha_{i0})}{\partial \alpha_i}\rho_i(\theta_0,\alpha_{i0})\right)'\left(\tfrac{\partial}{\partial \alpha_i}\Big|_{\alpha_{i0}}\rho_i(\theta_0,\alpha_i)+\tfrac{\partial \ln \pi_0(\alpha_{i0})}{\partial \alpha_i}\rho_i(\theta_0,\alpha_{i0})\right)\right]\right\}^{1/2}<\infty.$$

So

$$B = C + D = O\left(\mathcal{K}\left(\pi_0, \widetilde{\pi}_0\right)\right) + O\left(\frac{1}{T}\right).$$

The lemma follows. *Q.E.D.*

PROOF OF THEOREM 4: An expansion of the score around the truth yields:

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\Big|_{\widehat{\theta}} \ell_i^{RE}\left(\theta, \widehat{\xi}(\theta)\right) \\ &\approx \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\Big|_{\theta_0} \ell_i^{RE}\left(\theta, \widehat{\xi}(\theta)\right) + \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\Big|_{\theta_0} \ell_i^{RE}\left(\theta, \widehat{\xi}(\theta)\right)\left(\widehat{\theta} - \theta_0\right), \end{aligned}$$

so:

$$\widehat{\theta} - \theta_0 \approx \left[-\frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\Big|_{\theta_0} \ell_i^{RE}\left(\theta, \widehat{\xi}(\theta)\right)\right]^{-1} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\Big|_{\theta_0} \ell_i^{RE}\left(\theta, \widehat{\xi}(\theta)\right).$$

Next, remark that using a Laplace approximation argument we immediately obtain, when N and T tend to infinity:

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\Big|_{\theta_0} \ell_i^{RE}\left(\theta, \widehat{\xi}(\theta)\right) = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\Big|_{\theta_0} \overline{\ell}_i(\theta) + O_p\left(\frac{1}{T}\right).$$

Using this result together with Lemma S1 yields:

$$\widehat{\theta} - \theta_0 = \left[-\frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\Big|_{\theta_0} \overline{\ell}_i(\theta)\right]^{-1} \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\Big|_{\theta_0} \overline{\ell}_i(\theta) + O\left(\frac{\mathcal{K}(\pi_0, \widetilde{\pi}_0)}{T}\right) + o_p\left(\frac{1}{T}\right).$$

So, as:

$$\sqrt{NT}\left(\overline{\theta} - \theta_0\right) = \left[-\frac{1}{N}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta\partial\theta'}\Big|_{\theta_0} \overline{\ell}_i(\theta)\right]^{-1} \sqrt{NT}\frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial\theta}\Big|_{\theta_0} \overline{\ell}_i(\theta) + o_p(1)$$

we obtain:

$$\sqrt{NT}\left(\widehat{\theta} - \theta_0\right) = \sqrt{NT}\left(\overline{\theta} - \theta_0\right) + O\left(\sqrt{\frac{N}{T}}\mathcal{K}(\pi_0, \widetilde{\pi}_0)\right) + o_p\left(\sqrt{\frac{N}{T}}\right).$$

The theorem then follows, as $N/T \to C^{st}$. *Q.E.D.*

PROOF OF COROLLARY 3: Theorem 4.2. in Ghosal and Van der Vaart (2001) shows that if $K \geq C \log N$ for C large enough, then the convergence rate of the discrete sieve MLE is $(\log N)^{\kappa} N^{-1/2}$ for some $\kappa > 0$, where convergence is defined in terms of the Hellinger distance:

$$H(f, g) = \left(\int \left(f^{1/2}(\alpha) - g^{1/2}(\alpha)\right)^2 d\alpha\right)^{1/2}.$$

The result then comes from Theorem 5 in Wong and Shen (1995), that bounds the L^2 Kullback-Leibler loss $\mathcal{K}(\pi_0, \widetilde{\pi}_0)$ by the Hellinger distance $H(\pi_0, \widetilde{\pi}_0)$, under condition (25). *Q.E.D.*

Let us define

$$\widehat{M}_{BFE} = \int \frac{1}{N}\sum_{i=1}^{N} \underbrace{\left[\frac{\int m_i(\theta, \alpha_i) f_i(\theta, \alpha_i)\pi_i(\alpha_i|\theta)d\alpha_i}{\int f_i(\theta, \alpha_i)\pi_i(\alpha_i|\theta)d\alpha_i}\right]}_{\widehat{M}_i(\theta)} L^I(\theta)d\theta \Big/ \int L^I(\theta)d\theta.$$

where

$$L^I(\theta) = \prod_{i=1}^{N} \exp\left[T\ell_i^I(\theta)\right]$$

is the integrated likelihood function.

In preparation for the proof of Theorem 5, we need the following lemma, that gives the first-order expansion of $\widehat{M}_i(\theta_0)$ when N and T go to infinity.

Lemma S2 *When T tends to infinity:*

$$\begin{aligned}\widehat{M}_i(\theta_0) &= m_i(\theta_0,\alpha_{i0}) + m_i^{\alpha_i}(\theta_0,\alpha_{i0})\left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha_{i0})\right)\right]^{-1} v_i(\theta_0,\alpha_{i0}) \\ &\quad + \frac{1}{T\pi_i(\alpha_{i0}|\theta_0)}\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}} \left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}\pi_i(\alpha|\theta_0) m_i^{\alpha_i}(\theta_0,\alpha) + o_p\left(\frac{1}{T}\right).\end{aligned}$$

Proof of Lemma S2: Using a second-order Laplace expansion (e.g., Tierney *et al.*, 1989, eq. 2.6.) we obtain

$$\begin{aligned}\widehat{M}_i(\theta_0) &= m_i(\theta_0,\widehat{\alpha}_i(\theta_0)) + \frac{H^{-1}}{T}\frac{\partial \ln \pi_i(\alpha_{i0}|\theta_0)}{\partial\alpha_{i0}} m_i^{\alpha_i}(\theta_0,\alpha_{i0}) \\ &\quad + \frac{H^{-1}}{2T} m_i^{\alpha_i\alpha_i}(\theta_0,\alpha_{i0}) - \frac{H^{-2}H_2}{2T} m_i^{\alpha_i}(\theta_0,\alpha_{i0}) + O_p\left(\frac{1}{T^2}\right).\end{aligned}$$

where $H(\alpha) = \mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)$, $H = H(\alpha_{i0})$, and $H_2 = \mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i\alpha_i}(\theta_0,\alpha_{i0})\right)$.

Now, expanding $m_i(\theta_0,\widehat{\alpha}_i(\theta_0))$ and the score identity $v_i\left(\theta_0,\widehat{\alpha}_i(\theta_0)\right) = 0$ around α_{i0} yields

$$\begin{aligned}m_i(\theta_0,\widehat{\alpha}_i(\theta_0)) &= m_i(\theta_0,\alpha_{i0}) + m_i^{\alpha_i}(\theta_0,\alpha_{i0})\left[\widehat{\alpha}_i(\theta_0) - \alpha_{i0}\right] \\ &\quad + \frac{1}{2} m_i^{\alpha_i\alpha_i}(\theta_0,\alpha_{i0})\left[\widehat{\alpha}_i(\theta_0) - \alpha_{i0}\right]^2 + o_p\left(\frac{1}{T}\right) \\ &= m_i(\theta_0,\alpha_{i0}) + m_i^{\alpha_i}(\theta_0,\alpha_{i0}) H^{-1} v_i(\theta_0,\alpha_{i0}) - \frac{H^{-3}H_2}{2} m_i^{\alpha_i}(\theta_0,\alpha_{i0}) v_i^2(\theta_0,\alpha_{i0}) \\ &\quad + \frac{1}{2} m_i^{\alpha_i\alpha_i}(\theta_0,\alpha_{i0}) H^{-2} v_i^2(\theta_0,\alpha_{i0}) + o_p\left(\frac{1}{T}\right).\end{aligned}$$

Next, information equality at the truth yields:

$$T v_i^2(\theta_0,\alpha_{i0}) = H + o_p(1).$$

Also, remark that

$$\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}} H(\alpha) = H_2.$$

So:

$$\begin{aligned}\widehat{M}_i(\theta_0) &= m_i(\theta_0,\alpha_{i0}) + m_i^{\alpha_i}(\theta_0,\alpha_{i0}) H^{-1} v_i(\theta_0,\alpha_{i0}) - \frac{H^{-2}}{T}\left(\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}} H(\alpha)\right) m_i^{\alpha_i}(\theta_0,\alpha_{i0}) \\ &\quad + \frac{H^{-1}}{T}\frac{\partial \ln \pi_i(\alpha_{i0}|\theta_0)}{\partial\alpha_{i0}} m_i^{\alpha_i}(\theta_0,\alpha_{i0}) + \frac{H^{-1}}{T} m_i^{\alpha_i\alpha_i}(\theta_0,\alpha_{i0}) + o_p\left(\frac{1}{T}\right) \\ &= m_i(\theta_0,\alpha_{i0}) + m_i^{\alpha_i}(\theta_0,\alpha_{i0}) H^{-1} v_i(\theta_0,\alpha_{i0}) + \frac{1}{T}\left(\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}} H(\alpha)^{-1}\right) m_i^{\alpha_i}(\theta_0,\alpha_{i0}) \\ &\quad + \frac{H^{-1}}{T\pi_i(\alpha_{i0}|\theta_0)}\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}} \pi_i(\alpha|\theta_0) m_i^{\alpha_i}(\theta_0,\alpha) + o_p\left(\frac{1}{T}\right) \\ &= m_i(\theta_0,\alpha_{i0}) + m_i^{\alpha_i}(\theta_0,\alpha_{i0}) H^{-1} v_i(\theta_0,\alpha_{i0}) \\ &\quad + \frac{1}{T\pi_i(\alpha_{i0}|\theta_0)}\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}} H(\alpha)^{-1}\pi_i(\alpha|\theta_0) m_i^{\alpha_i}(\theta_0,\alpha) + o_p\left(\frac{1}{T}\right).\end{aligned}$$

Q.E.D.

Proof of Theorem 5:

Given Lemma S2, using a large-NT approximation we obtain:

$$\begin{aligned}\widehat{M}_{BFE} &= \int \left\{ \frac{1}{N}\sum_{i=1}^{N} \widehat{M}_i(\theta) \right\} L^I(\theta)d\theta \Big/ \int L^I(\theta)d\theta \\ &= \frac{1}{N}\sum_{i=1}^{N} \widehat{M}_i(\widehat{\theta}) + O_p\left(\frac{1}{NT}\right),\end{aligned}$$

where $\widehat{\theta}$ is the mode of the integrated likelihood: $\widehat{\theta} = \text{argmax}_\theta\, L^I(\theta)$. Note that the approximation comes from:

$$L^I(\theta) = \prod_{i=1}^{N} \exp\left[T\ell_i^I(\theta)\right] = \exp\left(NT\frac{1}{N}\sum_{i=1}^{N}\ell_i^I(\theta)\right),$$

with: $\ell_i^I(\theta) = \frac{1}{T}\ln\int \exp\left[T\ell_i(\theta,\alpha_i)\right]\pi_i(\alpha_i|\theta)d\alpha_i$, so: $\frac{1}{N}\sum_{i=1}^{N}\ell_i^I(\theta) = O_p(1)$.

So:

$$\widehat{M}_{BFE} = \frac{1}{N}\sum_{i=1}^{N}\widehat{M}_i(\theta_0) + \left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Big|_{\theta_0}\widehat{M}_i(\theta)\right]\left(\widehat{\theta}-\theta_0\right) + o_p\left(\frac{1}{T}\right).$$

Then:

$$\operatorname*{plim}_{N\to\infty}\ \widehat{\theta} = \theta_0 + \frac{B}{T} + o\left(\frac{1}{T}\right)$$

and

$$\operatorname*{plim}_{N\to\infty}\ \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Big|_{\theta_0}\widehat{M}_i(\theta) = \operatorname*{plim}_{N\to\infty}\ \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Big|_{\theta_0} m_i(\theta,\alpha_{i0}) + o(1).$$

So:

$$\operatorname*{plim}_{N\to\infty}\ \widehat{M}_{BFE} = \operatorname*{plim}_{N\to\infty}\ \frac{1}{N}\sum_{i=1}^{N}\widehat{M}_i(\theta_0) + \left[\operatorname*{plim}_{N\to\infty}\ \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Big|_{\theta_0} m_i(\theta,\alpha_{i0})\right]\frac{B}{T} + o\left(\frac{1}{T}\right).$$

Lastly, Lemma S2 implies,using that $\mathbb{E}_{\theta_0,\alpha_{i0}}\left(v_i(\theta_0,\alpha_{i0})\right) = 0$:

$$\begin{aligned}\operatorname*{plim}_{N\to\infty}\ \frac{1}{N}\sum_{i=1}^{N}\widehat{M}_i(\theta_0) &= \operatorname*{plim}_{N\to\infty}\ \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta_0,\alpha_{i0}}\left(\widehat{M}_i(\theta_0)\right) \\ &= \operatorname*{plim}_{N\to\infty}\ M + \frac{\widetilde{B}_M}{T} + o\left(\frac{1}{T}\right),\end{aligned}$$

with

$$\widetilde{B}_M = \operatorname*{plim}_{N\to\infty}\ \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\pi_i(\alpha_{i0}|\theta_0)}\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}\pi_i(\alpha|\theta_0)m_i^{\alpha_i}(\theta_0,\alpha).$$

Q.E.D.

PROOF OF COROLLARY 4: Let $\pi_i(.;\xi)$ be a random effects specification. We assume

$$
\begin{array}{ll}
i) & \lim_{\alpha\to\pm\infty}\left\{\mathbb{E}_{\theta_0,\alpha_{i0}}\left[-v_i^{\alpha_i}(\theta_0,\alpha)\right]\right\}^{-1} m_i^{\alpha_i}(\theta_0,\alpha)\pi_0(\alpha)=0,\\
ii) & \lim_{\alpha\to\pm\infty}\left\{\mathbb{E}_{\theta_0,\alpha_{i0}}\left[-v_i^{\alpha_i}(\theta_0,\alpha)\right]\right\}^{-1} m_i^{\alpha_i}(\theta_0,\alpha)\pi_0(\alpha)\left(\ln\widetilde{\pi}_0(\alpha)-\ln\pi_0(\alpha)\right)=0,\\
iii) & \mathbb{E}_{\pi_0}\left(\left(\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\left\{\mathbb{E}_{\theta_0,\alpha}\left[-v_i^{\alpha_i}(\theta_0,\alpha)\right]\right\}^{-1} m_i^{\alpha_i}(\theta_0,\alpha)+\frac{\partial\ln\pi_0(\alpha_{i0})}{\partial\alpha_i}\left\{\mathbb{E}_{\theta_0,\alpha}\left(-v_i^{\alpha_i}\right)\right\}^{-1}m_i^{\alpha_i}\right)^2\right)<\infty.
\end{array}
$$

These conditions are very similar to the ones of Lemma S1, and impose restrictions on the tails of π_0 and $\widetilde{\pi}_0$. As before, they are clearly satisfied if π_0 is compactly supported. Note that in condition *iii)* we have left the dependence on true parameter values implicit, to simplify the notation.

It follows from Theorem 4 that $B=O\left(\mathcal{K}\left(\pi_0,\widetilde{\pi}_0\right)\right)$. Moreover:

$$
\begin{aligned}
\widetilde{B}_M &\equiv \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\frac{1}{\pi_i\left(\alpha_{i0};\widehat{\xi}(\widehat{\theta})\right)}\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}\pi_i\left(\alpha;\widehat{\xi}(\widehat{\theta})\right)m_i^{\alpha_i}(\theta_0,\alpha)\\
&= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\mathbb{E}_{\pi_0}\left(\frac{1}{\widetilde{\pi}_0(\alpha_{i0})}\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}\widetilde{\pi}_0(\alpha)m_i^{\alpha_i}(\theta_0,\alpha)\right)+o(1)\\
&= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\int\frac{\pi_0(\alpha_{i0})}{\widetilde{\pi}_0(\alpha_{i0})}\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}\widetilde{\pi}_0(\alpha)m_i^{\alpha_i}(\theta_0,\alpha)d\alpha_{i0}+o(1)\\
&= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\int\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}\pi_0(\alpha)m_i^{\alpha_i}(\theta_0,\alpha)d\alpha_{i0}+o(1)\\
&\quad+\operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\int\pi_0(\alpha_{i0})\left[\mathbb{E}_{\theta_0,\alpha_{i0}}\left(-v_i^{\alpha_i}(\theta_0,\alpha_{i0})\right)\right]^{-1}m_i^{\alpha_i}(\theta_0,\alpha_{i0})\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\left(\ln\frac{\widetilde{\pi}_0(\alpha)}{\pi_0(\alpha)}\right)d\alpha_{i0}.
\end{aligned}
$$

Conditions *i)* and *ii)* of the corollary imply, as in the proof of Theorem 4, that

$$
\widetilde{B}_M = \operatorname*{plim}_{N\to\infty}-\frac{1}{N}\sum_{i=1}^N\int\left(\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\pi_0(\alpha)\left[\mathbb{E}_{\theta_0,\alpha}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}m_i^{\alpha_i}(\theta_0,\alpha)\right)\ln\frac{\widetilde{\pi}_0(\alpha_{i0})}{\pi_0(\alpha_{i0})}d\alpha_{i0}+o(1).
$$

Lastly, the Cauchy-Schwarz inequality implies

$$
\frac{1}{N}\sum_{i=1}^N\int\left(\frac{\partial}{\partial\alpha}\Big|_{\alpha_{i0}}\pi_0(\alpha)\left[\mathbb{E}_{\theta_0,\alpha}\left(-v_i^{\alpha_i}(\theta_0,\alpha)\right)\right]^{-1}m_i^{\alpha_i}(\theta_0,\alpha)\right)\ln\frac{\widetilde{\pi}_0(\alpha_{i0})}{\pi_0(\alpha_{i0})}d\alpha_{i0} = O\left(\mathcal{K}\left(\pi_0,\widetilde{\pi}_0\right)\right),
$$

provided that condition *iii)* above holds. *Q.E.D.*

PROOF OF THEOREM 6: We have:

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty}\widehat{M}_{RE} &= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\int m_i(\widehat{\theta},\alpha_i)\pi_i\left(\alpha_i;\widehat{\xi}\left(\widehat{\theta}\right)\right)d\alpha_i\\
&= \operatorname*{plim}_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\int m_i(\theta_0,\alpha_i)\pi_i\left(\alpha_i;\widehat{\xi}\left(\theta_0\right)\right)d\alpha_i+O\left(\frac{1}{T}\right),
\end{aligned}
$$

as $\widehat{\theta}$ is large-T consistent even if $\pi_i(.;\xi)$ is misspecified. Next, using that

$$
\widehat{\xi}\left(\theta_0\right)=\widehat{\xi}_0+O_p\left(\frac{1}{T}\right)
$$

we obtain

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \widehat{M}_{RE} &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\int m_i(\theta_0,\alpha_i)\pi_i\left(\alpha_i;\widehat{\xi}_0\right)d\alpha_i + O\left(\frac{1}{T}\right) \\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\int m_i(\theta_0,\alpha_i)\pi_i\left(\alpha_i;\overline{\xi}_0\right)d\alpha_i + O\left(\frac{1}{T}\right) \\
&= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N}\int m_i(\theta_0,\alpha_i)\widetilde{\pi}_0\left(\alpha_i\right)d\alpha_i + O\left(\frac{1}{T}\right).
\end{aligned}$$

Q.E.D.

S2. EXAMPLES

This section proves results stated in the paper for the autoregressive and logit models that we use as examples. As an additional example, the section also contains results for a Poisson counts model. For notational simplicity we drop the indices of the expectation terms when they are evaluated at true parameter values.

S2.1. *Dynamic AR(p)*

Let $y_i^0 = (y_{i,1-p}, ..., y_{i0})'$ be the vector of initial conditions, that we assume observed. In matrix form, we have:

$$y_i = X_i\mu_0 + \alpha_{i0}\iota + \varepsilon_i,$$

where the tth row of X_i is $x_{it}' = (y_{i,t-p}, ..., y_{i,t-1})$, $\mu_0 = \left(\mu_{10} ... \mu_{p0}\right)'$, and ι is a $T\times 1$ vector of ones. The scaled individual log-likelihood is given by:

$$\ell_i(\mu,\sigma^2,\alpha_i) = \frac{1}{T}\ln f(y_i|y_i^0,\alpha_i;\mu,\sigma^2) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2T}\sum_{t=1}^{T}\frac{(y_{it}-x_{it}'\mu-\alpha_i)^2}{\sigma^2}.$$

We thus have:

$$v_i(\mu,\sigma^2,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}\frac{(y_{it}-x_{it}'\mu-\alpha_i)}{\sigma^2},$$

and hence:

$$\mathbb{E}\left[-v_i^{\alpha_i}(\mu,\sigma^2,\alpha_i)\right] = \frac{1}{\sigma^2}.$$

Moreover:

$$\begin{aligned}
\mathbb{E}\left[v_i^2(\mu,\sigma^2,\alpha_i)\right] &= \frac{1}{T^2\sigma^4}\iota'\mathbb{E}\left(\left(y_i - X_i\mu - \alpha_i\iota\right)\left(y_i - X_i\mu - \alpha_i\iota\right)'\right)\iota, \\
&= \frac{1}{T^2\sigma^4}\iota'\mathbb{E}\left(\left(X_i(\mu_0-\mu)+(\alpha_{i0}-\alpha_i)\iota+\varepsilon_i\right)\left(X_i(\mu_0-\mu)+(\alpha_{i0}-\alpha_i)\iota+\varepsilon_i\right)'\right)\iota.
\end{aligned}$$

Note that this expectation depends on the true values of the parameters. Note also that the expectation is taken for i fixed. The same will be true of the variances and covariances that we will consider in this section of the appendix.

Computation of $\mathbb{E}\left[v_i^2(\mu,\sigma^2,\alpha_i)\right]$. One has:

$$\mathrm{Var}\left(\varepsilon_i + X_i(\mu_0-\mu)\right) = \mathrm{Var}\left(\varepsilon_i + \left[(\mu_0-\mu)'\otimes I_T\right]\mathrm{vec}\, X_i\right).$$

Let $B(\mu_0,\mu) = (\mu_0-\mu)'\otimes I_T$. Then:

$$\begin{aligned}\mathrm{Var}\left(\varepsilon_i + X_i(\mu_0-\mu)\right) &= \sigma^2 I_T + \mathbb{E}\left(\varepsilon_i\left(\mathrm{vec}\, X_i\right)'\right)B(\mu_0,\mu)' + B(\mu_0,\mu)\mathbb{E}\left(\varepsilon_i\left(\mathrm{vec}\, X_i\right)'\right)' \\ &\quad + B(\mu_0,\mu)\,\mathrm{Var}\left(\mathrm{vec}\, X_i\right)B(\mu_0,\mu)'.\end{aligned}$$

To compute these expressions, we shall write the model as (see Alvarez and Arellano, 2004, appendix A.3):

$$\begin{pmatrix} I_p & 0 \\ B_{Tp} & B_T \end{pmatrix}\begin{pmatrix} y_i^0 \\ y_i \end{pmatrix} = \begin{pmatrix} y_i^0 \\ \alpha_i\iota + \varepsilon_i \end{pmatrix},$$

where

$$\begin{pmatrix} B_{Tp} & B_T \end{pmatrix} = \begin{pmatrix} -\mu_{p0} & -\mu_{(p-1)0} & \cdots & -\mu_{10} & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -\mu_{p0} & \cdots & -\mu_{20} & -\mu_{10} & 1 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 & \ldots & -\mu_{10} & 1 \end{pmatrix}.$$

Inverting the system yields:

$$y_i = \overline{C}_{Tp}y_i^0 + \alpha_i\overline{C}_T\iota + \overline{C}_T\varepsilon_i,$$

where $\overline{C}_T = B_T^{-1}$ and $\overline{C}_{Tp} = -B_T^{-1}B_{Tp}$.

At this stage, it is convenient to introduce the $(T+p)\times(Tp)$ selection matrix such that

$$\mathrm{vec}(X_i) = P'\begin{pmatrix} y_i^0 \\ y_i \end{pmatrix}.$$

Moreover, the matrix $B(\mu_0,\mu)P'$ reads:

$$\begin{pmatrix} \mu_{10}-\mu_1 & \mu_{20}-\mu_2 & \ldots & \mu_{p0}-\mu_p & 0 & 0 & \ldots & 0 & 0 \\ 0 & \mu_{10}-\mu_1 & \mu_{20}-\mu_2 & \ldots & \mu_{p0}-\mu_p & 0 & \ldots & 0 & 0 \\ 0 & 0 & \mu_{10}-\mu_1 & \mu_{20}-\mu_2 & \ldots & \mu_{p0}-\mu_p & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & \ldots & \mu_{p0}-\mu_p & 0 \end{pmatrix}.$$

We shall write:

$$B(\mu_0,\mu)P' = \begin{pmatrix} \overline{A}(\mu_0,\mu) & \overline{B}(\mu_0,\mu) \end{pmatrix},$$

where $\overline{A}(\mu_0,\mu)$ is $T\times p$ and $\overline{B}(\mu_0,\mu)$ is $T\times T$. Now:

$$\mathrm{vec}(X_i) = P'\begin{pmatrix} y_i^0 \\ y_i \end{pmatrix} = P'\begin{pmatrix} I_p \\ \overline{C}_{Tp} \end{pmatrix}y_i^0 + \alpha_i P'\begin{pmatrix} 0 \\ \overline{C}_T\iota \end{pmatrix} + P'\begin{pmatrix} 0 \\ \overline{C}_T\varepsilon_i \end{pmatrix}. \qquad \text{(S1)}$$

It thus follows that

$$\begin{aligned}\mathbb{E}\left[\varepsilon_i\left(\mathrm{vec}\, X_i\right)'\right]B(\mu_0,\mu)' &= \sigma_0^2\begin{pmatrix} 0_p & \overline{C}_T' \end{pmatrix}PB(\mu_0,\mu)' \\ &= \sigma_0^2\overline{C}_T'\overline{B}(\mu_0,\mu)'.\end{aligned}$$

Then:

$$\begin{aligned} B(\mu_0,\mu)\operatorname{Var}(\operatorname{vec} X_i)\, B(\mu_0,\mu)' &= \sigma_0^2 B(\mu_0,\mu)P' \begin{pmatrix} 0 & 0 \\ 0 & \overline{C}_T\overline{C}_T' \end{pmatrix} PB(\mu_0,\mu)' \\ &= \sigma_0^2 \overline{B}(\mu_0,\mu)\overline{C}_T\overline{C}_T'\overline{B}(\mu_0,\mu)'. \end{aligned}$$

Hence:

$$\begin{aligned} \operatorname{Var}(\varepsilon_i + X_i(\mu_0-\mu)) &= \sigma_0^2 I_T + \sigma_0^2\overline{C}_T'\overline{B}(\mu_0,\mu)' + \sigma_0^2\overline{B}(\mu_0,\mu)\overline{C}_T \\ &\quad + \sigma_0^2\overline{B}(\mu_0,\mu)\overline{C}_T\overline{C}_T'\overline{B}(\mu_0,\mu)'. \end{aligned}$$

Now:

$$\begin{aligned} &\mathbb{E}\left((X_i(\mu_0-\mu) + (\alpha_{i0}-\alpha_i)\iota + \varepsilon_i)\,(X_i(\mu_0-\mu) + (\alpha_{i0}-\alpha_i)\iota + \varepsilon_i)'\right) \\ &= \operatorname{Var}(\varepsilon_i + X_i(\mu_0-\mu)) + \mathbb{E}\,(X_i(\mu_0-\mu) + (\alpha_{i0}-\alpha_i)\iota + \varepsilon_i)\,\mathbb{E}\,(X_i(\mu_0-\mu) + (\alpha_{i0}-\alpha_i)\iota + \varepsilon_i)'. \end{aligned}$$

Since:

$$\operatorname{vec}(X_i) = P'\begin{pmatrix} I_p \\ \overline{C}_{Tp} \end{pmatrix} y_i^0 + \alpha_{i0}P'\begin{pmatrix} 0 \\ \overline{C}_T\iota \end{pmatrix} + P'\begin{pmatrix} 0 \\ \overline{C}_T\varepsilon_i \end{pmatrix},$$

it follows that

$$\begin{aligned} \mathbb{E}\,[X_i(\mu_0-\mu)] &= B(\mu_0,\mu)\mathbb{E}\,[\operatorname{vec}(X_i)] \\ &= \left(\overline{A}(\mu_0,\mu) + \overline{B}(\mu_0,\mu)\overline{C}_{Tp}\right) y_i^0 + \alpha_{i0}\overline{B}(\mu_0,\mu)\overline{C}_T\iota. \end{aligned}$$

The previous results yield:

$$\begin{aligned} \mathbb{E}\left[v_i^2(\mu,\sigma^2,\alpha_i)\right] &= \frac{1}{T^2\sigma^4}\iota'\Big\{\sigma_0^2 I_T + \sigma_0^2\overline{C}_T'\overline{B}(\mu_0,\mu)' + \sigma_0^2\overline{B}(\mu_0,\mu)\overline{C}_T \\ &\quad + \sigma_0^2\overline{B}(\mu_0,\mu)\overline{C}_T\overline{C}_T'\overline{B}(\mu_0,\mu)' \\ &\quad + \left[\left(\overline{A}(\mu_0,\mu) + \overline{B}(\mu_0,\mu)\overline{C}_{Tp}\right) y_i^0 + \alpha_{i0}\overline{B}(\mu_0,\mu)\overline{C}_T\iota + (\alpha_{i0}-\alpha_i)\iota\right] \times \\ &\quad \left[\left(\overline{A}(\mu_0,\mu) + \overline{B}(\mu_0,\mu)\overline{C}_{Tp}\right) y_i^0 + \alpha_{i0}\overline{B}(\mu_0,\mu)\overline{C}_T\iota + (\alpha_{i0}-\alpha_i)\iota\right]'\Big\}\iota. \end{aligned}$$

The infeasible robust prior is thus given by:

$$\begin{aligned} \pi_i^{IR}\left(\alpha_i|\mu,\sigma^2\right) &\propto \Big(\iota'\Big\{\sigma_0^2 I_T + \sigma_0^2\overline{C}_T'\overline{B}(\mu_0,\mu)' + \sigma_0^2\overline{B}(\mu_0,\mu)\overline{C}_T \\ &\quad + \sigma_0^2\overline{B}(\mu_0,\mu)\overline{C}_T\overline{C}_T'\overline{B}(\mu_0,\mu)' \\ &\quad + \left[\left(\overline{A}(\mu_0,\mu) + \overline{B}(\mu_0,\mu)\overline{C}_{Tp}\right) y_i^0 + \alpha_{i0}\overline{B}(\mu_0,\mu)\overline{C}_T\iota + (\alpha_{i0}-\alpha_i)\iota\right] \times \\ &\quad \left[\left(\overline{A}(\mu_0,\mu) + \overline{B}(\mu_0,\mu)\overline{C}_{Tp}\right) y_i^0 + \alpha_{i0}\overline{B}(\mu_0,\mu)\overline{C}_T\iota + (\alpha_{i0}-\alpha_i)\iota\right]'\Big\}\iota\Big)^{-1/2}, \\ &\propto \Big(1 + a(\mu-\mu_0) + b(\mu-\mu_0,\alpha_i-\alpha_{i0})\Big)^{-1/2}, \end{aligned}$$

where

$$a(\mu-\mu_0) = \frac{1}{T}\iota'\Big\{\overline{C}_T'\overline{B}(\mu_0,\mu)' + \overline{B}(\mu_0,\mu)\overline{C}_T\Big\}\iota \tag{S2}$$

is a linear function of $\mu-\mu_0$, and

$$\begin{aligned} b(\mu-\mu_0,\alpha_i-\alpha_{i0}) &= \frac{1}{T\sigma_0^2}\iota'\Big\{\sigma_0^2\overline{B}(\mu_0,\mu)\overline{C}_T\overline{C}_T'\overline{B}(\mu_0,\mu)' \\ &\quad + \left[\left(\overline{A}(\mu_0,\mu) + \overline{B}(\mu_0,\mu)\overline{C}_{Tp}\right) y_i^0 + \alpha_{i0}\overline{B}(\mu_0,\mu)\overline{C}_T\iota + (\alpha_{i0}-\alpha_i)\iota\right] \times \\ &\quad \left[\left(\overline{A}(\mu_0,\mu) + \overline{B}(\mu_0,\mu)\overline{C}_{Tp}\right) y_i^0 + \alpha_{i0}\overline{B}(\mu_0,\mu)\overline{C}_T\iota + (\alpha_{i0}-\alpha_i)\iota\right]'\Big\}\iota \end{aligned} \tag{S3}$$

is a quadratic function of $\mu-\mu_0$ and $\alpha_i-\alpha_{i0}$.

The AR(1) case. Let us assume that $p=1$. Then:

$$\overline{C}_T = \begin{pmatrix} 1 & 0 & ... & 0 \\ \mu_{10} & 1 & ... & 0 \\ ... & ... & ... & ... \\ \mu_{10}^{T-1} & \mu_{10}^{T-2} & ... & 1 \end{pmatrix},$$

so that:

$$\overline{C}_T \iota = \frac{1}{1-\mu_{10}} \begin{pmatrix} 1-\mu_{10} \\ 1-\mu_{10}^2 \\ ... \\ 1-\mu_{10}^T \end{pmatrix}.$$

Moreover:

$$\overline{C}_{Tp} = \begin{pmatrix} \mu_{10} \\ \mu_{10}^2 \\ ... \\ \mu_{10}^T \end{pmatrix},$$

and

$$\overline{A}(\mu_0,\mu) = \begin{pmatrix} \mu_{10}-\mu_1 \\ 0 \\ ... \\ 0 \end{pmatrix}, \qquad \overline{B}(\mu_0,\mu) = \begin{pmatrix} 0 & 0 & ... & 0 & 0 \\ \mu_{10}-\mu_1 & 0 & ... & 0 & 0 \\ 0 & \mu_{10}-\mu_1 & ... & 0 & 0 \\ 0 & 0 & ... & \mu_{10}-\mu_1 & 0 \end{pmatrix}.$$

Hence $\pi_i^{IR}\left(\mu_i|\mu,\sigma^2\right)$ is proportional to

$$\Big\{\sigma_0^2 T + 2\sigma_0^2 \frac{\mu_{10}-\mu_1}{1-\mu_{10}} \cdot \sum_{t=1}^{T-1}\left(1-\mu_{10}^t\right) + \sigma_0^2 \left(\frac{\mu_{10}-\mu_1}{1-\mu_{10}}\right)^2 \cdot \sum_{t=1}^{T-1}\left(1-\mu_{10}^t\right)^2 +$$
$$\left[\left((\mu_{10}-\mu_1)\frac{1-\mu_{10}^T}{1-\mu_{10}}\right) y_i^0 + \alpha_{i0}\frac{\mu_{10}-\mu_1}{1-\mu_{10}} \cdot \sum_{t=1}^{T-1}\left(1-\mu_{10}^t\right) + (\alpha_{i0}-\alpha_i)T\right] \times$$
$$\left[\left((\mu_{10}-\mu_1)\frac{1-\mu_{10}^T}{1-\mu_{10}}\right) y_i^0 + \alpha_{i0}\frac{\mu_{10}-\mu_1}{1-\mu_{10}} \cdot \sum_{t=1}^{T-1}\left(1-\mu_{10}^t\right) + (\alpha_{i0}-\alpha_i)T\right]' \Big\}^{-1/2}.$$

We thus obtain:

$$\pi_i^{IR}\left(\overline{\alpha}_i(\mu,\sigma^2)|\mu,\sigma^2\right) \propto \Big\{T + 2\frac{\mu_{10}-\mu_1}{1-\mu_{10}} \cdot \sum_{t=1}^{T-1}\left(1-\mu_{10}^t\right)$$
$$+ \left(\frac{\mu_{10}-\mu_1}{1-\mu_{10}}\right)^2 \cdot \sum_{t=1}^{T-1}\left(1-\mu_{10}^t\right)^2 \Big\}^{-1/2}.$$

Hence, for π to reduce bias we need that:

$$\frac{\partial \ln \pi\left(\overline{\alpha}_i(\mu,\sigma^2)|\mu,\sigma^2\right)}{\partial \mu}\Big|_{\mu_{10},\sigma_0^2,\alpha_{i0}} = \frac{1}{T(1-\mu_{10})} \cdot \sum_{t=1}^{T-1}\left(1-\mu_{10}^t\right) = \frac{1}{T}\sum_{t=1}^{T-1}(T-t)\mu_{10}^{t-1}.$$

Gaussian REML. We have:

$$v_i(\mu, \sigma^2, \alpha_i) = \frac{1}{T}\sum_{t=1}^{T} \frac{(y_{it} - x_{it}'\mu - \alpha_i)}{\sigma^2},$$

and hence:

$$\mathbb{E}\left(-v_i^{\alpha_i}(\mu, \sigma^2, \alpha_i)\right) = \frac{1}{\sigma^2}; \qquad \mathbb{E}\left(-v_i^{\mu}(\mu, \sigma^2, \alpha_i)\right) = -\frac{1}{T\sigma^2}\sum_{t=1}^{T} x_{it}; \qquad \mathbb{E}\left(-v_i^{\sigma^2}(\mu, \sigma^2, \alpha_i)\right) = 0.$$

Dropping for simplicity the derivative with respect to σ^2 we obtain:

$$\rho_i(\mu, \alpha_i) = -\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(x_{it}).$$

Let us define the following $p \times (T+p)$ matrix:

$$Q = \left(\begin{array}{ccc} I_p & ... & I_p \end{array} \right) P'.$$

Then as

$$\sum_{t=1}^{T} x_{it} = \left(\begin{array}{ccc} I_p & ... & I_p \end{array} \right) \mathrm{vec}(X_i),$$

we obtain, using (S1):

$$\rho_i(\mu, \alpha_i) = -\frac{1}{T}\left(Q \left(\begin{array}{c} I_p \\ \overline{C}_{Tp} \end{array} \right) y_i^0 + \alpha_i Q \left(\begin{array}{c} 0 \\ \overline{C}_{T}\iota \end{array} \right) \right),$$

where $\overline{C}_{Tp}$ and $\overline{C}_{T}$ are functions of μ. Moreover, for a stationary process, the coefficient of α_i is $O(1)$ while the coefficient of y_{i0} is $O(1/T)$. For example, for a stationary AR(1) process the coefficient of y_{i0} is: $-(1 + \mu_{10} + \mu_{10}^2 + ... + \mu_{10}^{T-1})/T = O(1/T)$.

S2.2. *Linear model with one endogenous regressor*

The individual log-likelihood is given by (see, e.g., Hahn, 2000):

$$\ell_i(\theta, \alpha_i) = -\frac{1}{2}\ln|\Omega| - \frac{1}{2T}\omega_{11}\sum_{t=1}^{T}(y_{it} - \theta\alpha_i)^2 - \frac{1}{T}\omega_{12}\sum_{t=1}^{T}(y_{it} - \theta\alpha_i)(x_{it} - \alpha_i) - \frac{1}{2T}\omega_{22}\sum_{t=1}^{T}(x_{it} - \alpha_i)^2.$$

We thus have:

$$v_i(\theta, \alpha_i) = \frac{1}{T}\omega_{11}\theta\sum_{t=1}^{T}(y_{it} - \theta\alpha_i) + \frac{1}{T}\omega_{12}\sum_{t=1}^{T}(y_{it} - 2\theta\alpha_i + \theta x_{it}) + \frac{1}{T}\omega_{22}\sum_{t=1}^{T}(x_{it} - \alpha_i).$$

Then:

$$\mathbb{E}\left(-v_i^{\alpha_i}(\theta, \alpha_i)\right) = \omega_{11}\theta^2 + 2\omega_{12}\theta + \omega_{22},$$

and:

$$v_i^{\theta}(\theta, \alpha_i) = \frac{1}{T}\omega_{11}\sum_{t=1}^{T}(y_{it} - 2\theta\alpha_i) + \frac{1}{T}\omega_{12}\sum_{t=1}^{T}(-2\alpha_i + x_{it}).$$

Hence, at true values:

$$\mathbb{E}_{\theta_0, \alpha_{i0}}\left(v_i^{\theta}(\theta_0, \alpha_{i0})\right) = -\omega_{11}\theta_0\alpha_{i0} - \omega_{12}\alpha_{i0}.$$

We obtain that:

$$\rho_i(\theta, \alpha_i) = \alpha_i \frac{-\omega_{11}\theta - \omega_{12}}{\omega_{11}\theta^2 + 2\omega_{12}\theta + \omega_{22}}.$$

S.2.3. *Poisson counts*

Let the data consist of T Poisson counts y_{it} with individual means:

$$\mathbb{E}_{\theta_0,\alpha_{i0}}(y_{it}) = \alpha_{i0} \exp(x_{it}'\theta_0), \quad i = 1...N, \quad t = 1...T,$$

where x_{it} are known covariates. The individual log-likelihood is given by:

$$\ell_i(\theta, \alpha_i) \propto -\alpha_i \frac{1}{T} \sum_{t=1}^{T} \exp(x_{it}'\theta) + \frac{1}{T} \sum_{t=1}^{T} y_{it} \ln(\alpha_i) + \frac{1}{T} \sum_{t=1}^{T} y_{it} x_{it}'\theta.$$

So:

$$v_i(\theta, \alpha_i) = \frac{1}{T\alpha_i} \sum_{t=1}^{T} \left(y_{it} - \alpha_i \exp(x_{it}'\theta) \right).$$

Note that it follows that:

$$\overline{\alpha}_i(\theta) = \alpha_{i0} \frac{\sum_{t=1}^{T} \exp(x_{it}'\theta_0)}{\sum_{t=1}^{T} \exp(x_{it}'\theta)}. \tag{S4}$$

Moreover:

$$\mathbb{E}\left(-v_i^{\alpha_i}(\theta, \alpha_i)\right) = \frac{1}{T\alpha_i^2} \sum_{t=1}^{T} \alpha_{i0} \exp(x_{it}'\theta_0),$$

and:

$$\begin{aligned}
\mathbb{E}\left(v_i^2(\theta, \alpha_i)\right) &= \frac{1}{T^2\alpha_i^2} \sum_{t=1}^{T} \mathbb{E}\left((y_{it} - \alpha_i \exp(x_{it}'\theta))^2\right), \\
&= \frac{1}{T^2\alpha_i^2} \sum_{t=1}^{T} \left(\mathbb{E}\left((y_{it} - \mathbb{E}(y_{it}))^2\right) + \left(\mathbb{E}(y_{it}) - \alpha_i \exp(x_{it}'\theta)\right)^2 \right), \\
&= \frac{1}{T^2\alpha_i^2} \sum_{t=1}^{T} \alpha_{i0} \exp(x_{it}'\theta_0) + \left(\alpha_{i0} \exp(x_{it}'\theta_0) - \alpha_i \exp(x_{it}'\theta)\right)^2,
\end{aligned}$$

where we have used that $\mathrm{Var}(y_{it}) = \mathbb{E}(y_{it}) = \alpha_{i0} \exp(x_{it}'\theta_0)$. Hence a consistent estimate of the following quantity is robust:

$$\pi_i^{IR}(\alpha_i|\theta) \propto \frac{1}{\alpha_i} \left(\sum_{t=1}^{T} \alpha_{i0} \exp(x_{it}'\theta_0) + \left[\alpha_{i0} \exp(x_{it}'\theta_0) - \alpha_i \exp(x_{it}'\theta)\right]^2 \right)^{-1/2}. \tag{S5}$$

Then, by Proposition 1 one can add a quadratic adjustment in $(\theta - \theta_0)$ and $(\alpha_i - \alpha_{i0})$ to the logarithm of π_i^{IR} without altering its bias properties. It follows that:

$$\widetilde{\pi}(\alpha_i|\theta) \propto \frac{1}{\alpha_i} \tag{S6}$$

is also bias reducing. Note that π_i^{IR} is proper, while $\widetilde{\pi}$ is not.

As in Lancaster (2002), let us consider the reparameterization: $\psi_i = \alpha_i \sum_{t=1}^{T} \exp(x_{it}'\theta)$. Then it is straightforward to show that: $\frac{\partial^2 \ell_i(\theta, \psi_i)}{\partial\theta\partial\psi_i} = 0$. In this reparameterized model, parameters are fully orthogonal, not just information orthogonal. In particular, the uniform prior is bias reducing. Therefore, in terms of the original reparameterization, the following prior reduces bias:

$$\pi_i\left(\alpha_i|\theta\right) \propto \left|\frac{\partial \psi_i(\alpha_i,\theta)}{\partial \alpha_i}\right| = \sum_{t=1}^{T} \exp(x_{it}'\theta).$$

Interestingly, the robust prior and Lancaster's prior are directly related, as:[28]

$$\pi_i{}^{IR}(\overline{\alpha}_i(\theta)|\theta) \propto \widetilde{\pi}(\overline{\alpha}_i(\theta)|\theta) = \sum_{t=1}^{T} \exp(x_{it}'\theta) = \pi_i(\alpha_i|\theta).$$

REML. For the Poisson counts model, we have:

$$\rho_i(\theta,\alpha_i) = -\alpha_i h(x_i,\theta),$$

where:

$$h(x_i,\theta) = \frac{\sum_{t=1}^{T} \exp(x_{it}'\theta) x_{it}}{\sum_{t=1}^{T} \exp(x_{it}'\theta)}.$$

It follows that uncorrelated Gaussian REML is not bias reducing in this model in general, unless α_{i0} is independent of x_i. Note that if we let α_{i0} and x_i be dependent, then correlated REML (as introduced in Corollary 2) is not robust either.

In addition, remark that the local approximation to the robust prior:

$$\widetilde{\pi}(\alpha_i|\theta) = \frac{1}{\alpha_i}$$

is a bias reducing prior that is independent of θ. However, $\widetilde{\pi}$ is an improper prior which does not correspond to a random effects specification.

Assume now that π belongs to the $\Gamma(p,r)$ family, for some $p>0$, $r>0$. We have:

$$\pi(\alpha_i;p,r) = \frac{p^r \alpha_i^{r-1} \exp(-p\alpha_i)}{\Gamma(r)}.$$

It is straightforward to check that the left-hand side in equation (21) is equal to:

$$\operatorname*{plim}_{N\to\infty} \quad -\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\pi_0}\left[\left(\overline{r}(\theta_0) - \overline{p}(\theta_0)\alpha_{i0}\right) h(x_i,\theta_0)\right]. \tag{S7}$$

So Gamma REML is not bias reducing in general in the Poisson model. Here also, it is bias reducing only under the assumption that α_{i0} and x_i are independent.

S.2.4. *Static logit*

We have:

$$v_i(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \Lambda(x_{it}'\theta + \alpha_i)\right).$$

[28] This result follows directly from the expression of $\overline{\alpha}_i(\theta)$.

It follows that:

$$\mathbb{E}\left[-v_i^{\alpha_i}(\theta,\alpha_i)\right] = \frac{1}{T}\sum_{t=1}^{T}\Lambda(x_{it}'\theta+\alpha_i)(1-\Lambda(x_{it}'\theta+\alpha_i)), \tag{S8}$$

and:

$$\begin{aligned}\mathbb{E}\left[v_i^2(\theta,\alpha_i)\right] &= \mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T}\left(y_{it}-\Lambda(x_{it}'\theta+\alpha_i)\right)\right)^2 \\ &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}\left(\left(y_{it}-\Lambda(x_{it}'\theta+\alpha_i)\right)^2\right),\end{aligned} \tag{S9}$$

where we have used the fact that observations are i.i.d. across T.

CEMFI; Casado del Alisal, 5, 28014 Madrid, Spain; arellano@cemfi.es
and
CEMFI; Casado del Alisal, 5, 28014 Madrid, Spain; bonhomme@cemfi.es.

References

[1] Alvarez, J. and M. Arellano (2004): "Robust Likelihood Estimation of Dynamic Panel Data Models", unpublished manuscript.

[2] Ghosal, S., and A. W. Van der Vaart (2001): "Rates of Convergence for Bayes and Maximum Likelihood Estimation for Mixture of Normal Densities", *Annals of Statistics*, 29, 1233–1263.

[3] Hahn, J. (2000): "Parameter orthogonalization and Bayesian inference", unpublished manuscript.

[4] Lancaster, T. (2002): "Orthogonal Parameters and Panel Data", *Review of Economic Studies*, 69, 647–666.

[5] Tierney, L., R.E. Kass and J.B. Kadane (1989): "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions", *J. Am. Stat. Ass.,* 84, 710–716.

[6] Wong, W.H., and X. Shen (1995): "Probability Inequalities for Likelihood Ratios and Convergence Rates of Sieve MLEs", *Annals of Statistics*, 23, 339–362.