Supplementary material to the paper entitled
Partial ML Estimation for Spatial Autoregressive Nonlinear Probit Models with Autoregressive Disturbances

Anna Gloria Billé[a,*], Samantha Leorato[b]

[a]*Faculty of Economics and Management, Free University of Bozen–Bolzano, Bolzano, Italy*
[b]*Department of Economics, Management, and Quantitative Methods, University of Milan, Italy*

1. List of Assumptions

Assumption 1. *(a) All diagonal elements of* $\mathbf{W}_n$ *and* $\mathbf{M}_n$ *are zero. (b)* $\rho \in (-1/\overline{\tau}_{\mathbf{W}_n}, 1/\overline{\tau}_{\mathbf{W}_n})$ *and* $\lambda \in (-1/\overline{\tau}_{\mathbf{M}_n}, 1/\overline{\tau}_{\mathbf{M}_n})$.

Assumption 2. *Matrices* $\mathbf{W}_n$ *and* $\mathbf{M}_n$ *and* $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}$ *and* $(\mathbf{I}_n - \lambda\mathbf{M}_n)^{-1}$ *are uniformly bounded in both row and column sum norms.*

Assumption 3. *Elements of* $\mathbf{X}_n$ *are uniformly bounded constants,* $\mathbf{X}_n$ *has full column rank, and* $lim_{n\to\infty}(\mathbf{X}_n'\mathbf{X}_n)/n$ *exists and is nonsingular.*

Assumption 4. $\ell = \lim_n \mathbb{E}\ell_n$ *exists.* ℓ *attains a unique maximum over the compact set* Θ *at the interior point* $\boldsymbol{\theta}_0$.

Assumption 5. *(a) Every subset of the sampling area of size* c_n *contains at most* m_n *units, where* $\lim_n m_n/c_n < C < \infty$. *(b) Moreover,*

$$\sup_{1\leq g\leq G_n}\left|\sum_{d_1,d_2=0}^{1}\frac{1}{p_g(d_1,d_2)}\right| < \infty.$$

Assumption 6. $\sup_{n,g,h}|\mathbb{C}ov(y_{gi}, y_{hi})| \leq \alpha(d_{gh})$*, where* d_{gh} *is the distance between group* g *and* h *and* $\alpha(c) \to 0$ *as* $c \to \infty$

Assumption 7. *(a) There exists a sequence* $\{q_n\}$*, with* $\lim_{n\to\infty} q_n = \infty$*, such that the matrix* $\sum_{h=0}^{q_n}\rho^h\mathbf{W}_n^h$ *is nonsingular (and* $\sum_{h=0}^{q_n}\lambda^h\mathbf{M}_n^h$ *is nonsingular) for all* n *and for all* $\rho \in (-1/\overline{\tau}, 1/\overline{\tau})$ *(and* $\lambda \in (-1/\overline{\tau}, 1/\overline{\tau})$*). (b) There exists a* $\delta > 0$ *such that* $\lim_{n\to\infty} n^\delta/q_n < \infty$.

*Corresponding author. E–mail: annagloria.bille@unibz.it, web page: https://www.unibz.it/it/faculties/economics-management/academic-staff/person/38038-anna-gloria-bille

Preprint submitted to *October 15, 2019*

Assumption 8. *For all fixed $d > 0$,*

$$\lim_{k \to \infty} \frac{k^2 \alpha(kd)}{\alpha(d)} = 0.$$

Assumption 9. *The sampling area grows uniformly at a rate of $\sqrt{n}$ in two non–opposing directions.*

Assumption 10. *The matrices*

$$J(\boldsymbol{\theta}_0) = \lim_n G_n \mathbb{E}\left(\frac{\partial \ell_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) \frac{\partial \ell_n}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) \right)$$

and

$$\mathbf{H}(\boldsymbol{\theta}_0) = -\lim_{n \to \infty} \mathbb{E}\, H(\boldsymbol{\theta}_0) = -\lim_{n \to \infty} \mathbb{E}\left(\frac{\partial^2}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} \ell_n \right)$$

are positive definite.

2. Technical Lemmas

Lemmas 1–5 are used in proofs of Theorems 5.1 and 5.2. All quantities with a "tilde" superscript have to be interpreted as approximated versions of the same quantity, obtained by replacing the inverse matrices $\mathbf{A}_\rho$ or $\mathbf{B}_\lambda$ by the finite truncation (details on the truncation can be found in the supplemental material). For example, $\tilde{\ell}_n$ is the approximated partial maximum likelihood computed by finite sum approximations of the inverse matrices, $\tilde{\boldsymbol{\theta}}_n$ is the corresponding maximizer, and so on.

Lemma 1. *Under Assumptions 1–7,*

$$\frac{1}{G} \sum_{g=1}^{G} KL(f_g || \tilde{f}_g) \leq (1 + \|\mathbf{X}\boldsymbol{\beta}\|_2^2)\, O\left(|\tau\rho|^{2(q+1)} \right).$$

Lemma 2. *Under Assumptions 1–6, $\frac{\partial \ell_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = o_p(1)$.*

Lemma 3. *Under Assumptions 1–6, fon any $\gamma \in \mathbb{R}^k$, $\gamma \neq 0$,*

$$\sup_{g \leq G} \max_{i=1,2} \|\mathbf{X}_{\rho,g} \mathbf{X}_{\rho,g}'\|_2 \phi(\mathbf{X}_{\rho,g_i} \gamma) < \infty$$

Lemma 4. *Under Assumptions 1–6 and 8–10, for all $\boldsymbol{\theta} \in \Theta$, and for all $g = 1, \ldots, G$ and $d_1, d_2 \in \{0,1\}^2$,*

$$\left\| \frac{\partial^2 p_g(d_1, d_2)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|_2 < \infty$$

If further Assumption 5 holds, then

$$\left\| \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{\partial^2 \tilde{\ell}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|_2 = o(1)$$

Lemma 5. *Under Assumptions 1–6*

$$\frac{\partial \ell_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \tilde{\ell}_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} + o_p(1) = o_p(1)$$

3. Proofs of technical Lemmas

Proof of Lemma 1

For any g and any vector parameter $\theta = (\boldsymbol{\beta}', \rho)$, $\bar{f}_g \sim \mathcal{N}(\mu_g, \boldsymbol{\Sigma}_g)$ and $f_g \sim \mathcal{N}(\tilde{\mu}_g, \tilde{\boldsymbol{\Sigma}}_g)$, where $\mu_g = \mathbf{X}_{\rho,g}\boldsymbol{\beta} = \left(\mathbf{A}_\rho^{-1}\right)_g \mathbf{X}\boldsymbol{\beta} = \left(\sum_{k=0}^{\infty} \rho^k \mathbf{W}^k\right)_g \mathbf{X}\boldsymbol{\beta}$ and $\mu_g = \left(\sum_{k=0}^{q} \rho^k \mathbf{W}^k\right)_g \mathbf{X}\boldsymbol{\beta}$ while $\boldsymbol{\Sigma}_g$ is the $g-$th diagonal block of $\boldsymbol{\Sigma}$ and $\tilde{\boldsymbol{\Sigma}}_g$ is the $g-$th diagonal block of the approximating matrix $\tilde{\boldsymbol{\Sigma}} = \sum_{h=0}^{q}\sum_{k=0}^{q} \rho^{h+k}\mathbf{W}^k(\mathbf{W}')^h$.

Thus,

$$\begin{aligned}
KL(\bar{f}_g||f_g) &= \frac{1}{2}\left[\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g^{-1}\boldsymbol{\Sigma}_g - \mathbf{I}_2) + (\mu_g - \tilde{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mu_g - \tilde{\mu}_g) + \log\frac{|\boldsymbol{\Sigma}_g|}{|\tilde{\boldsymbol{\Sigma}}_g|}\right] \\
&= \frac{1}{2}\left[\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g^{-1}\boldsymbol{\Sigma}_g - \mathbf{I}_2) - \log\left(|\mathbf{I}_2 + \left(\tilde{\boldsymbol{\Sigma}}_g^{-1}\boldsymbol{\Sigma}_g - \mathbf{I}_2\right)|\right) + (\mu_g - \tilde{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mu_g - \tilde{\mu}_g)\right] \\
&\le \frac{1}{4}\mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g^{-1}\boldsymbol{\Sigma}_g - \mathbf{I}_2\right)\left(\tilde{\boldsymbol{\Sigma}}_g^{-1}\boldsymbol{\Sigma}_g - \mathbf{I}_2\right)'(1+o(1)) + (\mu_g - \tilde{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mu_g - \tilde{\mu}_g) \\
&\le \frac{1}{4}\mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g^{-1}\left(\boldsymbol{\Sigma}_g - \tilde{\boldsymbol{\Sigma}}_g\right)\left(\boldsymbol{\Sigma}_g - \tilde{\boldsymbol{\Sigma}}_g\right)'\tilde{\boldsymbol{\Sigma}}_g^{-1}\right)(1+o(1)) + (\mu_g - \tilde{\mu}_g)'\boldsymbol{\Sigma}_g^{-1}(\mu_g - \tilde{\mu}_g)
\end{aligned} \tag{1}$$

where $\|\cdot\|_2$ is the induced 2-norm while $\|A\|_\infty = \max_i \sum_j a_{ij}$ and where we have used inequality:

$$\mathrm{tr}(\mathbf{A}) - \log|\mathbf{I}+\mathbf{A}| = \sum_i \lambda_i - \sum_i \log(1+\lambda_i) = \sum_i\left(\lambda_i - \lambda_i + \frac{\lambda_i^2}{2} - \frac{\lambda^3}{3} + \cdots\right) \le \sum_i \frac{\lambda_i^2}{2}(1+o(1))$$

which is true if $\sup|\lambda_i| = o(1)$. Now, from $\mathrm{tr}(\mathbf{AB})^2 \le \mathrm{tr}(\mathbf{A})^2\mathrm{tr}(\mathbf{B})^2$, for two positive semidefinite matrices $\mathbf{A}, \mathbf{B}$, we get

$$\mathrm{tr}\left(\tilde{\boldsymbol{\Sigma}}_g^{-1}\left(\boldsymbol{\Sigma}_g - \tilde{\boldsymbol{\Sigma}}_g\right)\left(\boldsymbol{\Sigma}_g - \tilde{\boldsymbol{\Sigma}}_g\right)'\tilde{\boldsymbol{\Sigma}}_g^{-1}\right) \le \mathrm{tr}\left[\left(\tilde{\boldsymbol{\Sigma}}_g^{-1}\right)^2\right]\mathrm{tr}\left[\left(\boldsymbol{\Sigma}_g - \tilde{\boldsymbol{\Sigma}}_g\right)^2\right]$$

We then use the Theorem 1 in Bai and Golub (1997), to get

$$\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g^{-1}) \le [\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g), 2]\begin{pmatrix} \|\tilde{\boldsymbol{\Sigma}}_g\|_F^2 & \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g) \\ \underline{\lambda}^2(\tilde{\boldsymbol{\Sigma}}_g) & \underline{\lambda}(\tilde{\boldsymbol{\Sigma}}_g)\end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g)\imath'\tilde{\boldsymbol{\Sigma}}_g\imath + 4\underline{\lambda}^2(\tilde{\boldsymbol{\Sigma}}_g) + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g)^2 + 2\underline{\lambda}(\tilde{\boldsymbol{\Sigma}}_g),$$

where $\underline{\lambda}(\tilde{\boldsymbol{\Sigma}}_g)$ is the minimum eigenvalue of $\tilde{\boldsymbol{\Sigma}}_g$. Then, because of $\underline{\lambda}(\mathbf{A}) \le \mathrm{tr}(\mathbf{A})/k$, for any pd $k-$dimensional matrix $\mathbf{A}$,

$$\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g^{-1}) \le 2\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g)\|\tilde{\boldsymbol{\Sigma}}_g\|_F^2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g)^2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g)^2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g).$$

Moreover, since $\tilde{\boldsymbol{\Sigma}}$ is symmetric and pd, $\|\tilde{\boldsymbol{\Sigma}}_g\|_F^2 = \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g^2) \le \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g)^2$, and

$$\begin{aligned}
\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_g) &= \sum_{h=0}^{q}\sum_{k=0}^{q}\rho^{h+k}\mathrm{tr}\left((\mathbf{W}^k)_g(\mathbf{W}_g^h)'\right) \le \sum_{h=0}^{q}|\rho|^h\sum_{k=0}^{h}\mathrm{tr}\left((\mathbf{W}^k)_g(\mathbf{W}_g^{h-k})'\right) \\
&\le \sum_{h=0}^{\infty}|\rho|^h\sum_{k=0}^{h}\mathrm{tr}\left((\mathbf{W}^k)_g(\mathbf{W}_g^{h-k})'\right) \\
&= \|[(\mathbf{I} - |\rho|\mathbf{W})^{-1}]_{g,\cdot}\|_F^2 \le 2\|[(\mathbf{I} - |\rho|\mathbf{W})^{-1}]_{g,\cdot}\|_\infty^2 < 2c_{|\rho|}
\end{aligned}$$

where $[(\mathbf{I}-\rho\mathbf{W})^{-1}]_{g,\cdot}$ indicates the G blocks of $[(\mathbf{I}-\rho\mathbf{W})^{-1}]$ corresponding to the $g-$th rows. The last inequality follows from the assumption that $\boldsymbol{\Sigma}$ in row sum norm assumption (assumption 3 Kelejan and Prucha).

Moreover, we clearly have

$$\frac{1}{G}\sum_{g=1}^{G}\mathrm{tr}\left(\mathbf{\Sigma}_g-\tilde{\mathbf{\Sigma}}_g\right)^2=\frac{1}{G}\sum_{g=1}^{G}\|\mathbf{\Sigma}_g-\tilde{\mathbf{\Sigma}}_g\|_F^2\le\frac{1}{G}\|\mathbf{\Sigma}-\tilde{\mathbf{\Sigma}}\|_F^2\le\frac{n}{G}\|\mathbf{\Sigma}-\tilde{\mathbf{\Sigma}}\|_2^2=O(|\rho\tau|^{2(q+1)})$$

where $\tilde{\mathbf{\Sigma}}=\sum_{h=0}^{q}\rho^h\sum_{k=0}^{k}\mathbf{W}^k(\mathbf{W}')^{h-k}$, is the finite order approximation of the whole covariance matrix $\mathbf{\Sigma}$.

Finally,

$$\begin{aligned}\frac{1}{G}\sum_{g=1}^{G}(\mu_g-\tilde{\mu}_g)'\mathbf{\Sigma}_g^{-1}(\mu_g-\tilde{\mu}_g)&\le\frac{1}{G}\sum_{g=1}^{G}\|\mu_g-\tilde{\mu}_g\|_2^2\|\mathbf{\Sigma}_g^{-1}\|_2\\&\le\|\mathbf{I}-\rho\mathbf{W}\|_2^2\|\mathbf{X}\boldsymbol{\beta}\|_2\|\mathbf{A}_\rho^{-1}-\tilde{\mathbf{A}}_\rho^{-1}\|_F^2\le O(|\rho\tau|^{2(q+1)})\end{aligned}$$

Putting all these together in (1) and summing with respect to g:

$$\frac{1}{G}\sum_{g=1}^{G}KL(f_g||\tilde{f}_g)\le\ O\left(|\tau\rho|^{2(q+1)}\right)\left(1+\|\mathbf{X}\boldsymbol{\beta}\|_2^2\right).$$

Proof of Lemma 2

For the numerator, it is easy to see, from equations in the Appendix F of the main paper, that all terms $\partial p_g(d_1,d_2)/\partial\boldsymbol{\beta}$ are bounded, provided that $\mathbf{X}$ is bounded (Assumption 3) and each $\mathbf{\Sigma}_g$ is nonsingular (which implies all variances $\sigma_{g_i}>0$), because both the standard Gaussian CDF and the density of a Gaussian distribution with finite positive variance are bounded over the whole real line.

We now focus on $\partial p_g(1,1)/\partial\rho$. From Eq. (F.6),

$$\begin{aligned}\frac{\partial}{\partial\rho}p_g(1,1)=&\,\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}-\frac{\sigma_{g_1,g_2}}{\sigma_{g,1}^2}\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sqrt{\sigma_{g,2}^2-\sigma_{g_1,g_2}/\sigma_{g,1}^2}}\right)\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta}\\&+\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\frac{u^2\,\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^4}\phi\left(\frac{u}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(u)\right)du\\&+\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right)\frac{\partial}{\partial\rho}\varphi_{2,g}(u)du\end{aligned}\tag{2}$$

where $\dot{\mathbf{x}}_{\rho,g_1}$ (or $\dot{\mathbf{x}}_{\rho,g_2}$) follows from:

$$\dot{\mathbf{X}}_\rho=\mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{A}_\rho^{-1}\mathbf{X}=\sum_{h=0}^{\infty}(h+1)\rho^h\mathbf{W}^h\mathbf{W}\mathbf{X}.$$

Then using Assumption 2, Lemma 3 and arguments similar to those in the proof of Lemma 4, we have, for g_1 (or equivalently, g_2),

$$\|\dot{\mathbf{x}}_{\rho,g_1}\|_2\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\le C\,\|\mathbf{X}_\rho\mathbf{X}_\rho'\|_2\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)=O(1).$$

Thus, the first term of (2) is bounded above because Φ is bounded. For the second term, we have:

$$\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\frac{u^2\,\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^4}\phi\left(\frac{u}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(u)\right)du\le\int_{-\infty}^{\infty}\frac{u^2\,\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^4}\phi\left(\frac{u}{\sigma_{g_1}}\right)du\le\frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^3}$$

Further, boundedness of $(\mathbf{I} - \rho\mathbf{W})^{-1}$ implies boundedness of $\dot{\boldsymbol{\Sigma}}$ and thus of $\dot{\sigma}^2_{g_1}$.

Finally, the last term of (2): $\partial\varphi_{2,g}(u)/\partial\rho$ is an affine function of u with bounded coefficients, and thus the integral can be bounded above by a linear transformation of the expectation $\int_{\mathbb{R}} u\phi(u/\sigma_{g_1})du$ with bounded coefficients.

We can apply similar arguments to the other terms $\partial p_g(1,0)/\partial\rho$, $\partial p_g(0,1)/\partial\rho$ and $\partial p_g(0,0)/\partial\rho$.

Proof of Lemma 3

We first prove that under the Assumptions, we have that the minimum eigenvalue of $\mathbf{X}'\mathbf{X}$, say $\underline{\lambda}(n)$, is equal to $\underline{\lambda}(n) == n\underline{\lambda}_n$ where $\underline{\lambda}_n = \underline{\lambda}(n)/n$ is the minimum eigenvalue of $\frac{\mathbf{X}'\mathbf{X}}{n}$ and is continuous, in the sense that $\underline{\lambda}_n - \underline{\lambda}_{n-1} = O(n^{-1})$. Then, because of Assumption 3 $\lim \underline{\lambda}_n = \underline{\lambda} > 0$, we have that $\underline{\lambda}(n) = O(n)$. Moreover, since $\mathbf{X}'\mathbf{X}/n$ converges to a positive definite matrix, also $\overline{\lambda}_n = O(n)$.[1]

This also implies that the minimum and maximum eigenvalues of $\mathbf{X}_\rho'\mathbf{X}_\rho = \mathbf{X}'(\mathbf{A}_\rho^{-1})'\mathbf{A}_\rho^{-1}\mathbf{X}$ are of the same order, in light of uniform boundedness of $\mathbf{A}_\rho^{-1}$. Moreover, since $\mathbf{x}_{\rho,i} = (\mathbf{A}_\rho^{-1})_i\mathbf{X}$, also $\|\mathbf{x}_{\rho,i}'\mathbf{x}_{\rho,i}\| = O(n)$ and

$$\frac{\gamma'\mathbf{x}_{\rho,i}'\mathbf{x}_{\rho,i}\gamma}{\|\gamma\|} \geq \inf_{\|\mathbf{z}\|=1} \mathbf{z}'\mathbf{x}_{\rho,i}'\mathbf{x}_{\rho,i}\mathbf{z} = \inf_{\mathbf{z}} \mathbf{z}'\mathbf{X}'(\mathbf{A}_\rho^{-1})_i'(\mathbf{A}_\rho^{-1})_i\mathbf{X}\mathbf{z} \geq \underline{\lambda}(n) \inf_{\mathbf{z}} \mathbf{z}'(\mathbf{A}_\rho^{-1})_i'(\mathbf{A}_\rho^{-1})_i\mathbf{z} = O(n).$$

Now we show that this implies

$$\sup_{i\leq n} \|\mathbf{X}_\rho\mathbf{X}_\rho'\|_2\phi(\mathbf{X}_{\rho,i}\gamma) < \infty$$

and thus the claim. Note first that the above display is always bounded if n is finite, because $\mathbf{X}$ is assumed to take finite values and $\mathbf{A}_\rho^{-1}$ is uniformly bounded. Then, we need to prove that it doesn't explode as $n \to \infty$.

Note that, for any $0 \leq i \leq n$,

$$\begin{aligned}\|\mathbf{X}_\rho\mathbf{X}_\rho'\|_2\phi(\mathbf{X}_{\rho,i}\gamma) &= (2\pi)^{-1/2} \sup_{\|z\|=\|\gamma\|} \mathbf{z}'\mathbf{X}_\rho'\mathbf{X}_\rho\mathbf{z}\exp\left\{-\frac{1}{2}\gamma'\mathbf{x}_{\rho,i}'\mathbf{x}_{\rho,i}\gamma\right\}\\ &\leq (2\pi)^{-1/2}O(n)\overline{\lambda}\exp\left\{-\frac{\|\gamma\|_2}{2}O(n)\right\} \to_n 0.\end{aligned}$$

Finally, the inequality

$$\sup_{g\leq G}\max_{i=1,2} \|\mathbf{X}_{\rho,g}'\mathbf{X}_{\rho,g}\|_2\phi(\mathbf{X}_{\rho,g_i}\gamma) < \infty$$

is obtained by applying iteratively the following argument. Given a block matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

$$\|\mathbf{A}\|_2 = \sup_{\|\mathbf{z}\|=1} \mathbf{z}'\mathbf{A}\mathbf{z} \geq \sup_{\|\mathbf{z}\|=\|(\mathbf{z}_1,\mathbf{z}_2)\|=1,\ \mathbf{z}_1=0} \mathbf{z}_2'\mathbf{A}_{22}\mathbf{z}_2 = \|\mathbf{A}_{22}\|_2.$$

clearly implies $\sup_{g\leq G}\|\mathbf{X}_{\rho,g}\mathbf{X}_{\rho,g}'\|_2 \leq \|\mathbf{X}_\rho\mathbf{X}_\rho'\|_2 = \|\mathbf{X}_\rho'\mathbf{X}_\rho\|_2$.

[1] In fact, $\overline{\lambda}_n \leq n\mathrm{tr}(\mathbf{X}'\mathbf{X})/n = O(n)$.

Proof of Lemma 4

We limit ourselves to the case $d_1 = d_2 = 1$. All other cases follow with similar steps.

In order to prove the claim, it is sufficient to show that each of the components

$$\begin{pmatrix} \frac{\partial^2 p_g(1,1)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \frac{\partial^2 p_g(1,1)}{\partial \boldsymbol{\beta} \partial \rho} \\ \frac{\partial^2 p_g(1,1)}{\partial \rho \partial \boldsymbol{\beta}'} & \frac{\partial^2 p_g(1,1)}{\partial \rho^2} \end{pmatrix}$$

has bounded norm.

We start from: $\frac{\partial^2 p_g(1,1)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$:

$$\begin{aligned} \frac{\partial^2 p_g(1,1)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= \frac{\partial}{\partial \boldsymbol{\beta}'} \phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) \Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) \mathbf{x}'_{\rho,g_1} \\ &\quad + \frac{\partial}{\partial \boldsymbol{\beta}'} \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \phi\left(\frac{u}{\sigma_{g_1}}\right) \phi\left(\varphi_{2,g}(u)\right) \frac{\mathbf{x}_{\rho,g_2}}{\sqrt{\sigma^2_{g_2} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}} du \end{aligned}$$

where

$$\varphi_{2,g}(u) = \frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} + \frac{\sigma_{g_1,g_2}}{\sigma^2_{g,1}} u}{\sqrt{\sigma^2_{g_2} - \sigma^2_{g_1,g_2}/\sigma^2_{g_1}}}.$$

Thus,

$$\begin{aligned} \frac{\partial^2 p_g(1,1)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= -\sigma_{g_1}^{-1} \phi'\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) \Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) \mathbf{x}'_{\rho,g_1}\mathbf{x}_{\rho,g_1} \\ &\quad + \phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) \phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) \mathbf{x}'_{\rho,g_1} \frac{\partial \varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \\ &\quad + \phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) \phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) \frac{\mathbf{x}'_{\rho,g_2}\mathbf{x}_{\rho,g_2}}{\sqrt{\sigma^2_{g_2} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}} du \\ &\quad + \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \phi\left(\frac{u}{\sigma_{g_1}}\right) \phi'\left(\varphi_{2,g}(u)\right) \frac{\mathbf{x}'_{\rho,g_2}}{\sqrt{\sigma^2_{g_2} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}} \frac{\partial \varphi_{2,g}(u)}{\partial \boldsymbol{\beta}'} du \\ &= \frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}} \phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) \Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) \mathbf{x}'_{\rho,g_1}\mathbf{x}_{\rho,g_1} \\ &\quad + 2\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) \phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) \frac{\mathbf{x}'_{\rho,g_2}\mathbf{x}_{\rho,g_2}}{\sqrt{\sigma^2_{g_2} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}} du \\ &\quad - \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \varphi_{2,g}(u) \phi\left(\frac{u}{\sigma_{g_1}}\right) \phi\left(\varphi_{2,g}(u)\right) \frac{\mathbf{x}'_{\rho,g_2}\mathbf{x}_{\rho,g_2}}{\sigma^2_{g_2} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}} du \\ &\leq \frac{\|\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}\|}{\sigma_{g_1}} \phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) \|\mathbf{x}'_{\rho,g_1}\mathbf{x}_{\rho,g_1}\|_2 + 2\phi(0)\phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} - \frac{\sigma_{g_1,g_2}}{\sigma^2_{g,1}}\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sqrt{\sigma^2_{g_2} - \sigma^2_{g_1,g_2}/\sigma^2_{g_1}}}\right) \frac{\|\mathbf{x}'_{\rho,g_2}\mathbf{x}_{\rho,g_2}\|_2}{\sqrt{\sigma^2_{g_2} - \sigma^2_{g_1,g_2}/\sigma^2_{g_1}}} \\ &\quad + \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} |\varphi_{2,g}(u)| \phi\left(\frac{u}{\sigma_{g_1}}\right) \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g,1}} u}{\sqrt{\sigma^2_{g_2} - \sigma^2_{g_1,g_2}/\sigma^2_{g_1}}}\right) \frac{\|\mathbf{x}'_{\rho,g_2}\mathbf{x}_{\rho,g_2}\|_2}{\sqrt{\sigma^2_{g_2} - \sigma^2_{g_1,g_2}/\sigma^2_{g_1}}} du \end{aligned}$$

The result now easily follows because of Lemma 3, noting that $\|\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}\|_2^2 \leq \|\boldsymbol{\beta}\|_2^2\|\mathbf{x}'_{\rho,g_1}\mathbf{x}_{\rho,g_1}\|_2 \leq \|\boldsymbol{\beta}\|_2^2\|\mathbf{X}'_\rho\mathbf{X}_\rho\|_2$:

$$\left\|\frac{\partial^2 p_g(1,1)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right\| \leq \sigma_{g_1}^{-1}\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\|\mathbf{x}'_{\rho,g_1}\mathbf{x}_{\rho,g_1}\|_2^{3/2} + 2\phi(0)\phi\left(\mathbf{X}_{\rho,g}\gamma_1\right)\frac{\|\mathbf{X}'_{\rho,g}\mathbf{X}_{\rho,g}\|_2}{\sqrt{\sigma_{g_2}^2 - \sigma_{g_1,g_2}^2/\sigma_{g_1}^2}}$$
$$+ \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}|\varphi_{2,g}(u)|\,\phi\left(\frac{u}{\sigma_{g_1}}\right)du\phi\left(\mathbf{X}_{\rho,g}\gamma_2\right)\frac{\|\mathbf{x}'_{\rho,g_2}\mathbf{x}_{\rho,g_2}\|_2}{\sqrt{\sigma_{g_2}^2 - \sigma_{g_1,g_2}^2/\sigma_{g_1}^2}} < \infty$$

We now consider the partial derivative $\frac{\partial^2}{\partial\rho\partial\boldsymbol{\beta}'}p_g(1,1)$ where $\frac{\partial}{\rho}p_g(1,1)$ is computed in Appendix F:

$$\frac{\partial^2}{\partial\rho\partial\boldsymbol{\beta}'}p_g(1,1) = \frac{\partial}{\partial\boldsymbol{\beta}'}\left[\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(x_{\rho,g_1}\boldsymbol{\beta})\right)\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta} + \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\frac{u^2\,\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^4}\phi\left(\frac{u}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(u)\right)du\right.$$
$$\left.+ \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right)\frac{\partial}{\partial\rho}\varphi_{2,g}(u)du\right]$$
$$= \left(\phi'\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\frac{-\mathbf{x}_{\rho,g_1}}{\sigma_{g_1}}\Phi\left(\varphi_{2,g}(x_{\rho,g_1}\boldsymbol{\beta})\right) + \phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(x_{\rho,g_1}\boldsymbol{\beta})\right)\frac{\partial\varphi_{2,g}(x_{\rho,g_1}\boldsymbol{\beta}}{\partial\boldsymbol{\beta}'}\right)\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta}$$
$$+ \phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(x_{\rho,g_1}\boldsymbol{\beta})\right)\dot{\mathbf{x}}_{g_1} + \mathbf{x}_{\rho,g_1}\left[\frac{u^2\,\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^4}\phi\left(\frac{u}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(u)\right)\right]_{|_{u=-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}}$$
$$+ \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\frac{u^2\,\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^4}\phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right)\frac{\partial\varphi_{2,g}(u)}{\partial\boldsymbol{\beta}'}du$$
$$+ \mathbf{x}_{\rho,g_1}\left[\phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right)\frac{\partial}{\partial\rho}\varphi_{2,g}(u)\right]_{|_{u=-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}}$$
$$+ \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u}{\sigma_{g_1}}\right)\left(\phi'\left(\varphi_{2,g}(u)\right)\frac{\partial\varphi_{2,g}(u)}{\partial\rho}\frac{\partial\varphi_{2,g}(u)}{\partial\boldsymbol{\beta}'} + \phi\left(\varphi_{2,g}(u)\right)\frac{\partial^2\varphi_{2,g}(u)}{\partial\rho\partial\boldsymbol{\beta}'}\right)du.$$

Note that, from the equations in Appendix F, $\partial\varphi_{2,g}(u)/\partial\rho$ depends on combinations of u (linear in u) and coefficients of $\dot{\boldsymbol{\Sigma}}_g$, $\boldsymbol{\Sigma}_g$ and $\mathbf{X}_g\boldsymbol{\beta}$ and $\dot{\mathbf{X}}_g\boldsymbol{\beta}$. Therefore, all terms of the form $\phi(\varphi)\frac{\partial\varphi}{\partial\rho}$ are bounded, provided $\|\dot{\boldsymbol{\Sigma}}_g\| < \infty$ and $\|\mathbf{X}_g\| < \infty$. Moreover, for the same reason, also the terms that can be bounded by $C\cdot\phi(\mathbf{X}_{\rho,g}\gamma)\|\mathbf{X}_{\rho,g}\|_2^2$ or $C\cdot\phi(\mathbf{X}_{\rho,g}\gamma)\|\dot{\mathbf{X}}_{\rho,g}\|_2^2$ are bounded. It then remains to show that $\|\dot{\boldsymbol{\Sigma}}_g\| < \infty$ and $\|\dot{\mathbf{X}}_{\rho,g}\| < \infty$. Recall that,

$$\dot{\mathbf{X}}_\rho = \frac{\partial\mathbf{X}_\rho}{\partial\rho} = \mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{A}_\rho^{-1}\mathbf{X} \quad \text{and} \quad \dot{\boldsymbol{\Sigma}} = \mathbf{A}_\rho^{-1}\mathbf{W}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{W}'(\mathbf{A}'_\rho)^{-1}.$$

Thus,

$$\|\dot{\boldsymbol{\Sigma}}\|_2 \leq 2\|\boldsymbol{\Sigma}\|\,\|\mathbf{A}_\rho^{-1}\|\,\|\mathbf{W}\|_2 < \infty$$

implies boundedness of the terms $\dot{\sigma}$ and

$$\|\dot{\mathbf{X}}_\rho\dot{\mathbf{X}}'_\rho\|_2 = \|\dot{\mathbf{X}}'_\rho\dot{\mathbf{X}}_\rho\|_2 \leq \|\mathbf{X}'_\rho\mathbf{X}_\rho\|_2\|\mathbf{W}'\mathbf{W}\|_2\|\boldsymbol{\Sigma}\|_2$$

and then because of Lemma 3:

$$\max_{i=1,2}\|\dot{\mathbf{X}}_{\rho,g_i}\dot{\mathbf{X}}'_{\rho,g_i}\|_2\phi(\mathbf{X}_{\rho,g_i}\gamma) < \infty.$$

Finally,

$$\frac{\partial^2}{\partial\rho^2}p_g(1,1) = \frac{\partial}{\partial\rho}\left[\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(x_{\rho,g_1}\boldsymbol{\beta})\right)\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta} + \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\frac{u^2\,\dot{\sigma}^2_{g_1}}{2\sigma^4_{g_1}}\phi\left(\frac{u}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(u)\right)du\right.$$
$$\left. + \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right)\frac{\partial}{\partial\rho}\varphi_{2,g}(u)du\right]$$

is again bounded in norm. With respect to the other derivatives, we only need to check that $\phi\left(\mathbf{x}_{\rho,i}\gamma\right)\frac{\partial^2}{\partial\rho^2}\varphi_{2,g}(u)$ has bounded coefficients and is integrable in u and that all the terms of $\ddot{\boldsymbol{\Sigma}} = \frac{\partial^2\boldsymbol{\Sigma}}{\partial\rho^2}$ are bounded and that $\ddot{\mathbf{X}}_\rho = \frac{\partial^2\mathbf{X}_\rho}{\partial\rho^2}$ satisfies $\sup_i\|\ddot{\mathbf{x}}_{\rho,i}\ddot{\mathbf{x}}'_{\rho,i}\|_2\phi(\ddot{\mathbf{x}}_{\rho,i}\gamma) < \infty$ for every $\gamma \neq 0$.

Using matrix calculus, we have:

$$\ddot{\boldsymbol{\Sigma}} = \frac{\partial}{\partial\rho}(\mathbf{A}_\rho^{-1}\mathbf{W}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{W}'(\mathbf{A}'_\rho)^{-1}) = 2\left(\mathbf{A}_\rho^{-1}\mathbf{W}\right)^2\boldsymbol{\Sigma} + 2\mathbf{A}_\rho^{-1}\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}'(\mathbf{A}'_\rho)^{-1} + 2\boldsymbol{\Sigma}\left(\mathbf{W}'(\mathbf{A}'_\rho)^{-1}\right)^2$$

that clearly has bounded norm if $\|\boldsymbol{\Sigma}\|_2 < \infty$, $\|\mathbf{A}_\rho^{-1}\|_2 < \infty$ and $\|\mathbf{W}\|_2 < \infty$, which is implied by Assumption 2.

Further,

$$\ddot{\mathbf{X}}_\rho = 2\mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{A}_\rho^{-1}\mathbf{X}$$

implies $\|\ddot{\mathbf{X}}_\rho\|_2 \leq 2\|\mathbf{X}_\rho\|_2\|\mathbf{W}\|_2^2\|\boldsymbol{\Sigma}\|_2$, therefore $\sup_i\|\ddot{\mathbf{x}}_{\rho,i}\ddot{\mathbf{x}}'_{\rho,i}\|_2\phi(\ddot{\mathbf{x}}_{\rho,i}\gamma) < \infty$, in view of Lemma 3.

Now, an inspection of $\partial\varphi/\partial\rho$ easily shows that $\phi\left(\mathbf{x}_{\rho,i}\gamma\right)\partial^2\varphi/\partial\rho^2$ is bounded because of boundedness of $\boldsymbol{\Sigma}$, $\dot{\boldsymbol{\Sigma}}$, $\ddot{\boldsymbol{\Sigma}}$ and of $\sup_i\|\mathbf{z}_i\mathbf{z}'_i\|\phi(\mathbf{x}_{\rho,i}\gamma)$ for $\mathbf{z}_i$ equal to any of the vectors $\mathbf{x}_{\rho,i}, \dot{\mathbf{x}}_{\rho,i}, \ddot{\mathbf{x}}_{\rho,i}$ and because of nonsingularity of each $\boldsymbol{\Sigma}_g$.

Finally, if Assumption 7 holds, then all elements of $\boldsymbol{\Sigma}_\rho, \mathbf{A}_\rho$ $\mathbf{X}_\rho$ and their derivatives ($\dot{\boldsymbol{\Sigma}}_\rho$, etc...) can be replaced by their truncated sum approximation, with (focusing on $\boldsymbol{\Sigma}$ only, w.l.o.g.), $\|\boldsymbol{\Sigma}_\rho - \tilde{\boldsymbol{\Sigma}}_\rho\| \to 0$ because of Lemma 1.

Proof of Lemma 5

The claim $\frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{\partial\tilde{\ell_n}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} + o_p(n^{-1/2}) = o_p(n^{-1/2})$, follows if we prove that

$$\sup_{\boldsymbol{\theta}}\left\|\frac{\partial\ell_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} - \frac{\partial\tilde{\ell_n}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| = o_p(n^{-1/2}).$$

The above result follows if we can show that, for all g, and for any $d_1, d_2 = 0, 1$:

$$\left\|\frac{\partial p_g(d_1,d_2;\boldsymbol{\theta})/\partial\boldsymbol{\theta}}{p_g(d_1,d_2;\boldsymbol{\theta})} - \frac{\partial\tilde{p}_g(d_1,d_2;\boldsymbol{\theta})/\partial\boldsymbol{\theta}}{\tilde{p}_g(d_1,d_2;\boldsymbol{\theta})}\right\| = o_p(n^{-1/2}) \tag{3}$$

where $p_g(d_1,d_2;\boldsymbol{\theta}) = p_g(d_1,d_2)$ and $\tilde{p}_g(d_1,d_2;\boldsymbol{\theta}) = \tilde{p}_g(d_1,d_2)$ follow equations:

$$p_g\left(d_1,d_2\right) = \int_{\{s_{g_1}u > -s_{g_1}\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}\}}\frac{1}{\sigma_{g_1}}\phi\left(\frac{u}{\sigma_{g_1}}\right)\Phi\left(s_{g_2}\varphi_{2,g}(u)\right)du$$

computed using the exact and approximated values of $\mathbf{X}_\rho$ and σ_{g_1,g_2}, $\sigma^2_{g_1}$, $\sigma^2_{g_2}$ respectively.

We will prove (3) for $d_1 = d_2 = 1$. The other cases follow by the same arguments. First note that

$$\left\|\frac{\partial p_g(1,1)/\partial\boldsymbol{\theta}}{p_g(1,1)} - \frac{\partial \tilde{p}_g(1,1)/\partial\boldsymbol{\theta}}{\tilde{p}_g(1,1)}\right\| \leq \left\|\frac{\partial(p_g(1,1)-\tilde{p}_g(1,1))/\partial\boldsymbol{\theta}}{p_g(1,1)}\right\| + \left\|\frac{\partial\tilde{p}_g(1,1)}{\partial\boldsymbol{\theta}}\left(\frac{1}{\tilde{p}_g(1,1)} - \frac{1}{p_g(1,1)}\right)\right\|$$
$$\leq C\left\|\frac{\partial(p_g(1,1)-\tilde{p}_g(1,1))}{\partial\boldsymbol{\theta}}\right\| + \left\|\frac{\partial\tilde{p}_g(1,1)}{\partial\boldsymbol{\theta}}\right\|\left|\left(\frac{1}{\tilde{p}_g(1,1)} - \frac{1}{p_g(1,1)}\right)\right|$$

because for assumption 5. Moreover, $\left\|\frac{\partial\tilde{p}_g(1,1)}{\partial\boldsymbol{\theta}}\right\| = O(1)$ if the first term is $o(1)$, because we already proved (see the proof of Theorem 5.1) that $\|\frac{\partial p_g(1,1)}{\partial\boldsymbol{\theta}}\| < \infty$. So, we need to prove $\left\|\frac{\partial(p_g(1,1)-\tilde{p}_g(1,1))}{\partial\boldsymbol{\theta}}\right\| = o(n^{-1/2})$ and $\left|\left(\frac{1}{\tilde{p}_g(1,1)} - \frac{1}{p_g(1,1)}\right)\right| = o(n^{-1/2})$. To prove the latter it is enough to show that $|p_g(1,1) - \tilde{p}_g(1,1))| = o(n^{-1/2})$. From the definition of p_g and $\tilde{p}_g$, we have in fact:

$$|p_g(1,1)-\tilde{p}_g(1,1))| \leq \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\left|\phi\left(\frac{u_{g_1}}{\sigma_{g_1}}\right) - \phi\left(\frac{u_{g_1}}{\tilde{\sigma}_{g_1}}\right)\right|\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} + \frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right)du_{g_1}$$
$$+\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u_{g_1}}{\tilde{\sigma}_{g_1}}\right)\left|\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} + \frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2} - \frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right) - \Phi\left(\frac{\tilde{\mathbf{x}}_{\rho,g_2}\boldsymbol{\beta} + \frac{\tilde{\sigma}_{g_1,g_2}}{\tilde{\sigma}^2_{g_1}}u_{g_1}}{\sqrt{\tilde{\sigma}^2_{g_2} - \frac{\tilde{\sigma}^2_{g_1,g_2}}{\tilde{\sigma}^2_{g_1}}}}\right)\right|du_{g_1}$$
$$+\int_{-\tilde{\mathbf{x}}_{\rho,g_1}\boldsymbol{\beta}}^{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}\phi\left(\frac{u_{g_1}}{\tilde{\sigma}_{g_1}}\right)\Phi\left(\frac{\tilde{\mathbf{x}}_{\rho,g_2}\boldsymbol{\beta} + \frac{\tilde{\sigma}_{g_1,g_2}}{\tilde{\sigma}^2_{g_1}}u_{g_1}}{\sqrt{\tilde{\sigma}^2_{g_2} - \frac{\tilde{\sigma}^2_{g_1,g_2}}{\tilde{\sigma}^2_{g_1}}}}\right)du_{g_1} = A + B + C$$

where we used the tilde to distinguish between exact and approximated terms and we have assumed (w.l.o.g.) that $-\tilde{\mathbf{x}}_{\rho,g_1}\boldsymbol{\beta} < -\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}$. Then, because of $|e^x - e^y| \leq e^y|y-x|\frac{e^{|y-x|}-1}{|y-x|} = e^y|y-x|\left(1 + \sum_{k\geq 1}(y-x)^k(k+1)!\right)$, we get

$$A \leq \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u_{g_1}}{\sigma_{g_1}}\right)\left|\frac{u_{g_1}}{\sigma_{g_1}} - \frac{u_{g_1}}{\tilde{\sigma}_{g_1}}\right|\left(1 + O\left(\left|\frac{u_{g_1}}{\sigma_{g_1}} - \frac{u_{g_1}}{\tilde{\sigma}_{g_1}}\right|\right)\right)du_{g_1} = O\left(|\tilde{\sigma}_{g_1} - \sigma_{g_1}|\right)$$

Similar bounds can be found for B and C, because of the continuity of the functions ϕ, Φ and of the integral:

$$B \leq O(|(\tilde{\mathbf{x}}_{\rho,i} - \mathbf{x}_{\rho,i})\boldsymbol{\beta}|) + O(|det(\tilde{\boldsymbol{\Sigma}}_g) - det(\boldsymbol{\Sigma}_g))|) + O(\|\tilde{\boldsymbol{\Sigma}}_g - \boldsymbol{\Sigma}_g\|_F) = O(|\rho\tau|^{2(q_n+1)})$$

and $C \leq O(|(\tilde{\mathbf{x}}_{\rho,i} - \mathbf{x}_{\rho,i})\boldsymbol{\beta}|)$. All these terms are negligible because of Assumption 7 and $|\rho\tau| < 1$.

The same steps and the continuity of all functions entering in $\partial p_g(d_1,d_2)/\partial\boldsymbol{\theta}$ also yield

$$\left\|\frac{\partial(p_g(1,1)-\tilde{p}_g(1,1))}{\partial\boldsymbol{\theta}}\right\| = o(n^{-1/2}).$$

4. Approximation of $\boldsymbol{\Sigma}_g$ and $\mathbf{X}_\rho$.

As already pointed out by Wang et al. (2013), the terms σ_{g_1}, σ_{g_2} and σ_{g_1,g_2}, that are essential for the computation of the probabilities p_g, can not be easily written in closed form as functions of ρ and the weight matrix. In our case, things are made even worse by the fact that the vectors $\mathbf{x}_{\rho,g_i}$ are complex transformations of the whole design matrix that also depends on ρ.

In this section we give details on the approximation of the terms $\mathbf{X}_{\rho,g}$ and $\sigma_{g,\cdot}$ based on finite series expansion for $(\mathbf{I}-\rho\mathbf{W})^{-1}$, mentioned in section 3 of the main paper.

The contribution of each pair g on the loglikelihood depends on the rows g_1, g_2 of the matrix $\mathbf{A}_\rho^{-1}$ and on the submatrix $\mathbf{\Sigma}_g$. Under Assumption 1,

$$\mathbf{A}_\rho^{-1} = \sum_{k=0}^{\infty} \rho^k \mathbf{W}^k,$$

implies,

$$\mathbf{X}_\rho = \sum_{k=0}^{\infty} \rho^k \mathbf{W}^k \mathbf{X}$$

and, for $i = 1, 2$,

$$\mathbf{x}_{\rho,g_i} = \sum_{j=1}^{n} \sum_{k=0}^{\infty} \rho^k w_{g_i,j}^{(k)} \mathbf{x}_j$$

where $w_{l,j}^{(k)}$ is the $(l,j)-$term of the matrix $\mathbf{W}^k$. Then $\mathbf{x}_{\rho,g_i}$ can be approximated by truncating the series expansion to the $q-$th term.

Similarly, since

$$\mathbf{\Sigma} = \mathbf{A}_\rho^{-1}(\mathbf{A}'_\rho)^{-1} = \sum_{k=0}^{\infty} \rho^k \mathbf{W}^k \sum_{h=0}^{\infty} \rho^h (\mathbf{W}')^h,$$

we could approximate,

$$\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{A}_\rho}^{-1}(\tilde{\mathbf{A}_\rho}')^{-1} = \sum_{k=0}^{q} \rho^k \mathbf{W}^k \sum_{h=0}^{q} \rho^h (\mathbf{W}')^h = \sum_{k=0}^{2q} \rho^k \sum_{h=0}^{\min(k,q)} \mathbf{W}^h (\mathbf{W}')^{k-h}.$$

This approach can be convenient in the case of large samples, to avoid inversion of large matrices and is especially useful in the dense matrix case.

5. Other proofs

This Section includes some proofs that are omitted in the main paper.

5.1. Proof of Theorem 3.1

Repeating the steps in Wang et al. (2013), we can easily get

$$\begin{aligned}
p_g(1,1) &= P(y_{g_1}=1, y_{g_2}=1) = P(y_{g_1}=1)P(y_{g_2}=1 \mid y_{g_1}=1) \\
&= \Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\mathbb{E}_{u_{g_1}}\left(P(y_{g_2}=1 \mid u_{g_1}) \mid y_{g_1}=1\right) \\
&= \Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\mathbb{E}_{u_{g_1}\mid y_{g_1}=1}\left[\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}+\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2}-\frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right)\right] \\
&= \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\frac{1}{\sigma_{g_1}}\phi\left(\frac{u_{g_1}}{\sigma_{g_1}}\right)\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}+\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2}-\frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right)du_{g_1} \qquad (4)
\end{aligned}$$

from $u_{g_2} \mid u_{g_1} \sim \mathcal{N}\left(\tau_g\frac{\sigma_{g_2}}{\sigma_{g_1}}u_{g_1}, (1-\tau_g^2)\sigma^2_{g,2}\right)$ and noting that the conditional density of $u_{g_1} \mid y_{g_1}=1$ is:

$$p(u_{g_1} \mid y_{g_1}=1) = \mathbb{I}\{u_{g_1} \geq -\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}\}\frac{\frac{1}{\sigma_{g_1}}\phi\left(\frac{u_{g_1}}{\sigma_{g_1}}\right)}{\Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)}.$$

In a similar way:

$$\begin{aligned}
p_g(1,0) &= \Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\mathbb{E}_{u_{g_1}\mid \mathbf{X}, y_{g_1}=1}\left(1-\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}+\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2}-\frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right)\right) \\
&= \Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) - p_g(1,1), \qquad (5)
\end{aligned}$$

$$\begin{aligned}
p_g(0,1) &= \left(1-\Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\right)\mathbb{E}_{u_{g_1}\mid \mathbf{X}, y_{g_1}=0}\left(\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}+\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2}-\frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right)\right) \\
&= \int_{-\infty}^{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}\frac{1}{\sigma_{g_1}}\phi\left(\frac{u_{g_1}}{\sigma_{g_1}}\right)\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}+\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2}-\frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right)du_{g_1}, \qquad (6)
\end{aligned}$$

$$\begin{aligned}
p_g(0,0) &= \left(1-\Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\right) - \int_{-\infty}^{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}\frac{1}{\sigma_{g_1}}\phi\left(\frac{u_{g_1}}{\sigma_{g_1}}\right)\Phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}+\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}u_{g_1}}{\sqrt{\sigma^2_{g_2}-\frac{\sigma^2_{g_1,g_2}}{\sigma^2_{g_1}}}}\right)du_{g_1} \\
&= \left(1-\Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\right) - p_g(0,1). \qquad (7)
\end{aligned}$$

The identity of all the formulas with the two equivalent expressions in the statement of the Theorem is straightforward.

5.2. Proof of Theorem 5.1

Following the notation introduced in section 2 above, let

$$\tilde{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \tilde{\ell}_n(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X}),$$

where $\tilde{\ell}_n$ is the approximated partial maximum loglikelihood defined there. Moreover, to discriminate between the exact bivariate probabilities and those based on the q–th order finite sum approximation, we denote the last by $\tilde{p}_g(d_1, d_2)$, and the former by $p_g(d_1, d_2)$[2]. There is the following link between the $\tilde{\ell}_n$ and ℓ_n:

$$\begin{aligned}\mathbb{E}\tilde{\ell}_n(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X}) &= \sum_g \mathbb{E}\left(\sum_{d=(d_1,d_2)} \frac{1}{G}\mathbb{I}\{y_{g_1} = d_1, y_{g_2} = d_2\} \log \frac{\tilde{p}_{g,}(d_1, d_2)}{p_{g,}(d_1, d_2)} \mid \mathbf{X}\right) + \mathbb{E}\left(\ell_n(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X})\right) \\ &= \mathbb{E}\left(\ell_n(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X})\right) - \frac{1}{G}\sum_g KL(p_{g,}||\tilde{p}_{g,}).\end{aligned} \tag{8}$$

Thus, consistency and asymptotic normality of $\tilde{\boldsymbol{\theta}}$ come from the analogous properties of the PML estimator, and from negligibility of the term $\frac{1}{G}\sum_g KL(p_{g,}\|\tilde{p}_{g,})$.

We first prove consistency of the PML estimator $\hat{\boldsymbol{\theta}}_n = \arg\max \ell_n(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X})$. This can be proved by using the same arguments of Wang et al. (2013): in particular, given Assumption 4, we need to prove $\ell_n(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}) = o_p(1)$ and stochastic equicontinuity of $\ell_n(\boldsymbol{\theta})$. The first result follows by repeating exactly the same arguments as those of Lemma 1 in Wang et al. (2013).

In order to prove stochastic equicontinuity, following Wang et al. (2013), we need to show that

$$\sup_{\boldsymbol{\theta}} \left| \frac{1}{G} \sum_{g=1}^{G} y_{g_1} y_{g_2} \frac{\partial p_g(d_1, d_2)/\partial\boldsymbol{\theta}}{p_g(d_1, d_2)} \right| = O_p(1)$$

for all d_1, d_2.

The term at the denominator of the above equation is bounded away from zero because of Assumption 5. Moreover, all derivatives $\partial p_g(d_1, d_2)/\partial\boldsymbol{\theta}$ are $O_p(1)$ from Lemma 2.

Thus, stochastic equicontinuity of ℓ_n follows as in Lemma 3 of Wang et al. (2013) and this implies consistency of $\hat{\boldsymbol{\theta}}$.

The proof of the consistency of $\tilde{\boldsymbol{\theta}}_n$ follows from the asymptotic equivalence of $\tilde{\boldsymbol{\theta}}_n$ with the PML estimator $\hat{\boldsymbol{\theta}}_n = \arg\max \ell_n(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X})$, which is a consequence of Lemma 1. In fact, Assumption 3 implies, for all finite n/G, that $\|\mathbf{X}\boldsymbol{\beta}\|_2 = O(1)$ for all $\boldsymbol{\beta}$ in the interior of the compact parameter space Θ. Thus, by Assumption 7 (a), $\frac{1}{G}\sum_{g=1}^{G} KL(\tilde{f}_g||f_g) = o(1)$ which implies, because of (8) and Theorem 4.1–(i), that $\|\hat{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\theta}}_n\| = o_p(1)$.

[2] Note that the only difference between the two probabilities $\tilde{p}_g(d_1, d_2)$ and $p_g(d_1, d_2)$ is the computation of terms $\mathbf{X}_\rho$ and of the elements of $\boldsymbol{\Sigma}_g$.

5.3. Proof of Theorem 5.2

The result is proven in two steps. First, we decompose $\sqrt{G_n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \sqrt{G_n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \sqrt{G_n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, and show that the first term is negligible with respect to the second, where $\hat{\boldsymbol{\theta}}$ is the pairwise ML estimator based on the exact computation of ℓ_n.

Second, we prove that $\sqrt{G_n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ has the asymptotic Gaussian distribution $\sqrt{G_n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to_d \mathcal{N}\left(0, \mathbf{H}(\boldsymbol{\theta}_0)^{-1}\mathrm{J}(\boldsymbol{\theta}_0)\mathbf{H}(\boldsymbol{\theta}_0)^{-1}\right)$. This second part follows the lines of the proof of Theorem 2 in Wang et al. (2013) and those of Pinkse and Slade (1998).

We start by showing that $\sqrt{G_n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = o_p(1)$. Using the mean value theorem,

$$\frac{\partial \ell_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \ell_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = \frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$$

implies

$$\sqrt{G_n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) = \left(\frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right)^{-1} \sqrt{G_n}\, \frac{\partial \ell_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}.$$

Boundedness of $\left(\frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right)^{-1}$ follows from Lemma 4, then if

$$\frac{\partial \ell_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \tilde{\ell_n}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} + o_p(n^{-1/2}) = o_p(n^{-1/2}) \tag{9}$$

the negligibility of $\sqrt{G_n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$ follows. The proof of (9) is in Lemma 5.

In order to prove

$$\sqrt{G_n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to \mathcal{N}\left(0, \mathbf{H}(\boldsymbol{\theta}_0)^{-1}\mathrm{J}(\boldsymbol{\theta}_0)\mathbf{H}(\boldsymbol{\theta}_0)^{-1}\right),$$

we can repeat the same steps in Wang et al. (2013).

We sketch the main steps of the proof. For more details we refer to Wang et al. (2013).

Using the mean value theorem,

$$0 = \frac{\partial \ell_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Thus, for some $\boldsymbol{\theta}^*$ such that $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\| \leq \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|$,

$$\sqrt{G_n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left(\frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right)^{-1} \frac{\partial \ell_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}.$$

Then, we first need to prove that all terms composing $\frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ are bounded (therefore integrable), in order to conclude, by invoking consistency of $\hat{\boldsymbol{\theta}}$ and the law of large numbers, that

$$\lim_{n\to\infty} \frac{\partial^2 \ell_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \mathbf{H}(\boldsymbol{\theta}_0). \tag{10}$$

To prove this, the same exact arguments of Theorem 2 of Wang et al. (2013) apply. First, the bounds $\left\|\frac{\partial p_g(d_1,d_2)}{\partial \boldsymbol{\theta}}\right\| < \infty$ and $\left\|\frac{\partial^2 p_g(d_1,d_2)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right\| < \infty$ come from Lemma 2 and Lemma 4 respectively.

In order to have the weak limit of $\sqrt{G_n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, we finally need to show that

$$\mathrm{J}^{-1/2}(\boldsymbol{\theta}_0)\sqrt{G_n}\frac{\partial \ell_n(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} \to_d \mathcal{N}(0, \mathbf{I}).$$

Following Wang et al. (2013), and as in Theorem 1 of Pinkse and Slade (1998), we invoke Bernstein's blocking methods and the McLeish's central limit theorem for dependent processes (McLeish, 1974). This states that, if, for the triangular array $T_{nk_n} = \prod_{j=1}^{k_n}(1 + \imath\gamma D_{n,j})$, where $\imath^2 = -1$ and γ is a real constant, the following conditions are satisfied: (i) $\{T_{n,k_n}\}$ is uniformly integrable; (ii) $\mathbb{E}\, T_{n,k_n} \to_n 1$; (iii) $\sum_{j=1}^{k_n} D_{n,j}^2 \to_p 1$; (iv) $\max_{j\leq k_n} |D_{n,j}| \to 0$, then $\sum_{j=1}^{k_n} D_{n,j} \to_d N(0, 1)$.

Following the reasoning in Wang et al. (2013), the (sequence of) regions where the observations are located is split into a_n areas of size $\sqrt{b_n} \times \sqrt{b_n}$, with a_n growing at a faster rate than b_n and such that $a_n b_n = n$. Moreover, a_n and b_n are chosen so that $b_n < n^{1/2-\varepsilon}$ uniformly in n and $\alpha(\sqrt{b_n})a_n \to 0$. Let $\Lambda_{n,j}$ represents the set of indices of observations falling into the $j-$th area, and write $D_{n,j} = G^{-1/2}\sum_{g\in\Lambda_{n,j}} A_{n,g}$ where $A_{n,g}$ is implicitly defined by $\mathbf{z}'\sqrt{G}\mathrm{J}(\boldsymbol{\theta}_0)^{-1/2}\left(\frac{\partial \ell_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) = G^{-1/2}\sum_{g=1}^{G} A_{n,g}$, for an arbitrary vector s.t. $\|\mathbf{z}\| = 1$. It then remains to prove conditions (i)–(iv) to ascertain that the sum $\sum_{j=0}^{k_n} D_{n,j} = G^{-1/2}\sum_{g=1}^{G} A_{n,g}$ is asymptotically normal.

For the proofs of conditions (iv) and (i), we just follow Wang et al. (2013). Conditions (ii)-(iii) follow from Lemmas 4–7 in Wang et al. (2013).

5.4. Proof of the inequality (14): $\mathbb{C}ov(y_1, y_2) \leq C\mathbb{C}ov(y_1^*, y_2^*)$ for some $C < \infty$

Proof. From the definition of y_1 and y_2 we clearly have that

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(y_1, y_2) &= \mathbb{E}(y_1 y_2) - \Pr(y_1 = 1)\Pr(y_2 = 1) \\
&= \int_{-x_{\rho,1}\boldsymbol{\beta}}^{\infty} \frac{1}{\sigma_1}\phi\left(\frac{u}{\sigma_1}\right)\left[\Pr\left(y_2 = 1\,|y_1^* \in du\right) - \Pr\left(y_2 = 1\right)\right] du \\
&= \int_{-x'_{\rho,1}\boldsymbol{\beta}}^{\infty} \frac{1}{\sigma_1}\phi\left(\frac{u}{\sigma_1}\right)\left[\Phi\left(\varphi_2(u)\right) - \Phi\left(\frac{x'_{\rho,2}\boldsymbol{\beta}}{\sigma_2}\right)\right] du \\
&= \int_{-x_{\rho,1}\boldsymbol{\beta}}^{\infty} \tfrac{1}{\sigma_1}\phi\left(\tfrac{u}{\sigma_1}\right)\left[\phi\left(\tfrac{x'_{\rho,2}\boldsymbol{\beta}}{\sigma_2}\right)\left(\varphi_2(u) - \tfrac{x'_{\rho,2}\boldsymbol{\beta}}{\sigma_2}\right) + \phi'\left(\tfrac{x'_{\rho,2}\boldsymbol{\beta}}{\sigma_2}\right)\tfrac{1}{2}\left(\varphi_2(u) - \tfrac{x'_{\rho,2}\boldsymbol{\beta}}{\sigma_2}\right)^2 + ...\right] du \\
&\leq \phi\left(\frac{x'_{\rho,2}\boldsymbol{\beta}}{\sigma_2}\right)\int_{-x_{\rho,1}\boldsymbol{\beta}}^{\infty} \tfrac{1}{\sigma_1}\phi\left(\tfrac{u}{\sigma_1}\right)\left|\tfrac{u}{\sigma_1}\tfrac{\rho_{12}}{\sqrt{1-\rho_{12}^2}} - He_2\left(\tfrac{u}{\sigma_1}\tfrac{\rho_{12}}{\sqrt{1-\rho_{12}^2}}\right) + He_3\left(\tfrac{u}{\sigma_1}\tfrac{\rho_{12}}{\sqrt{1-\rho_{12}^2}}\right) + \cdots\right| \\
&\leq \rho_{12}\phi(0)C
\end{aligned}$$

where He_k are the probabilistic Hermite polynomials, $C < \infty$ because the Gaussian distribution has finite moments of all orders and where ρ_{12} is the correlation of the latent variables y_1^* and y_2^*. □

6. Score vector

In this section we are going to derive the score vectors of the SAR(1) probit and SARAR(1,1) probit models. These formulas will be used to easily study the behavior of the score vector, and to perform a more efficient computation of the partial maximum likelihood estimator.

6.1. SAR(1) probit

In order to compute the score vector for the optimization of the quasi pairwise loglikelihood, we need to compute the derivatives of $p_g(d_1, d_2)$ with respect to $\boldsymbol{\beta}$ and ρ. We recall some notation used in the main paper:

$$\varphi_{1,g}(u) = \frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta} + u\frac{\sigma_{g_1,g_2}}{\sigma_{g_2}^2}}{\sqrt{\sigma_{g_1}^2 - \sigma_{g_1,g_2}^2/\sigma_{g_2}^2}} \quad \text{and} \quad \varphi_{2,g}(u) = \frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} + u\frac{\sigma_{g_1,g_2}}{\sigma_{g_1}^2}}{\sqrt{\sigma_{g_2}^2 - \sigma_{g_1,g_2}^2/\sigma_{g_1}^2}}, \tag{11}$$

where $s_{g_i} = 2(d_i - 1/2)$. We first consider differencing with respect to $\boldsymbol{\beta}$:

$$\begin{aligned}
\frac{\partial p_g(1,1)}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \frac{1}{\sigma_{g_1}} \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \phi\left(\frac{u}{\sigma_{g_1}}\right) \Phi\left(\varphi_{2,g}(u)\right) du \\
&= \frac{1}{\sigma_{g_1}}\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\mathbf{x}'_{\rho,g_1} + \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \frac{1}{\sigma_{g_1}}\phi\left(\frac{u}{\sigma_{g_1}}\right)\frac{\partial}{\partial \boldsymbol{\beta}}\Phi\left(\varphi_{2,g}(u)\right) du \\
&= \frac{1}{\sigma_{g_1}}\phi\left(\frac{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\mathbf{x}'_{\rho,g_1} + \frac{1}{\sigma_{g_1}}\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right)\frac{\mathbf{x}'_{\rho,g_2}}{\sqrt{\sigma_{g_2}^2 - \frac{\sigma_{g_1,g_2}^2}{\sigma_{g_1}^2}}} du
\end{aligned}$$

After some algebra, we obtain:

$$\begin{aligned}
\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right) du &= \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\left(u\sigma_{g_2} + \mathbf{x}_{\rho,g_2}\boldsymbol{\beta}\sigma_{g_1,g_2}/\sigma_{g_2}\right)^2}{2\left(\sigma_{g_1}^2\sigma_{g_2}^2 - \sigma_{g_1,g_2}^2\right)}\right\} du \\
&= \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\sqrt{\sigma_{g_1}^2 - \sigma_{g_1,g_2}^2/\sigma_{g_2}^2},
\end{aligned}$$

Thus,

$$\frac{\partial p_g(1,1)}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma_{g_1}}\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\mathbf{x}'_{\rho,g_1} + \frac{1}{\sigma_{g_2}}\phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\mathbf{x}'_{\rho,g_2}. \tag{12}$$

Similarly,

$$\begin{aligned}
\frac{\partial p_g(1,0)}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}}\Phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right) - \frac{\partial p_g(1,1)}{\partial \boldsymbol{\beta}} = \phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\frac{\mathbf{x}'_{\rho,g_1}}{\sigma_{g_1}} - \frac{\partial p_g(1,1)}{\partial \boldsymbol{\beta}} \\
&= \tfrac{1}{\sigma_{g_1}}\phi\left(\tfrac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\left(1 - \Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\right)\mathbf{x}'_{\rho,g_1} - \tfrac{1}{\sigma_{g_2}}\phi\left(\tfrac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\mathbf{x}'_{\rho,g_2},
\end{aligned} \tag{13}$$

and, by repeating the same steps,

$$\frac{\partial p_g(0,1)}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma_{g_1}}\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\mathbf{x}'_{\rho,g_1} + \frac{1}{\sigma_{g_2}}\phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\left(1 - \Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\right)\mathbf{x}'_{\rho,g_2}, \tag{14}$$

$$\tfrac{\partial p_g(0,0)}{\partial \boldsymbol{\beta}} = -\tfrac{1}{\sigma_{g_1}}\phi\left(\tfrac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\left(1 - \Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\right)\mathbf{x}'_{\rho,g_1} - \tfrac{1}{\sigma_{g_2}}\phi\left(\tfrac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\left(1 - \Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\right)\mathbf{x}'_{\rho,g_2}. \tag{15}$$

In order to compute the derivatives with respect to ρ, we need to define:

$$\dot{\mathbf{X}}_\rho := \frac{\partial \mathbf{X}_\rho}{\partial \rho} = \frac{\partial \mathbf{A}_\rho^{-1}\mathbf{X}}{\partial \rho} = -\mathbf{A}_\rho^{-1}\frac{\partial \mathbf{A}_\rho}{\partial \rho}\mathbf{A}_\rho^{-1}\mathbf{X} = \mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{A}_\rho^{-1}\mathbf{X},$$

$$\begin{aligned}\dot{\mathbf{\Sigma}} := \frac{\partial \mathbf{\Sigma}}{\partial \rho} &= \frac{\partial}{\partial \rho}\left(\mathbf{I}-\rho \mathbf{W}\right)^{-1}\left(\mathbf{I}-\rho \mathbf{W}'\right)^{-1} = \left(\frac{\partial}{\partial \rho}\mathbf{A}_{\rho}^{-1}\right)\left(\mathbf{I}-\rho \mathbf{W}'\right)^{-1} + \left(\mathbf{I}-\rho \mathbf{W}\right)^{-1}\left(\frac{\partial}{\partial \rho}\left(\mathbf{A}_{\rho}'\right)^{-1}\right)\\ &= -\mathbf{A}_{\rho}^{-1}\frac{\partial \mathbf{A}_{\rho}}{\partial \rho}\mathbf{A}_{\rho}^{-1}\left(\mathbf{A}_{\rho}'\right)^{-1} - \mathbf{A}_{\rho}^{-1}\left(\mathbf{A}_{\rho}'\right)^{-1}\frac{\partial \mathbf{A}_{\rho}'}{\partial \rho}\left(\mathbf{A}_{\rho}'\right)^{-1} \\ &= \mathbf{A}_{\rho}^{-1}\mathbf{W}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{W}'\left(\mathbf{A}_{\rho}'\right)^{-1}.\end{aligned} \tag{16}$$

Now, we denote by $\dot{\mathbf{X}}_g = (\dot{\mathbf{x}}_{g_1}', \dot{\mathbf{x}}_{g_2}')'$ and

$$\dot{\mathbf{\Sigma}}_g = \begin{pmatrix} \dot{\sigma}_{g_1}^2 & \dot{\sigma}_{g_1,g_2} \\ \dot{\sigma}_{g_1,g_2} & \dot{\sigma}_{g_2}^2 \end{pmatrix}$$

the submatrix corresponding to rows g_1, g_2 and the g–th diagonal block matrix, respectively of $\dot{\mathbf{X}} = \dot{\mathbf{X}}_\rho$ and $\dot{\mathbf{\Sigma}} = \frac{\partial \mathbf{\Sigma}}{\partial \rho}$. We further note that, from

$$\dot{\sigma}_{g_1}^2 = \frac{\partial \sigma_{g_1}^2}{\partial \rho} = \frac{\partial \sigma_{g_1}^2}{\partial \sigma_{g_1}}\frac{\partial \sigma_{g_1}}{\partial \rho} = 2\sigma_{g_1}\frac{\partial \sigma_{g_1}}{\partial \rho}$$

we have $\frac{\partial \sigma_{g_1}}{\partial \rho} = \frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}}$ and thus $\frac{\partial}{\partial \rho}\frac{1}{\sigma_{g_1}} = -\frac{1}{\sigma_{g_1}^2}\frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}}$, and

$$\frac{\partial}{\partial \rho}\frac{1}{\sigma_{g_1}}\phi\left(\frac{u}{\sigma_{g_1}}\right) = \frac{u}{\sigma_{g_1}^2}\phi\left(\frac{u}{\sigma_{g_1}}\right)\frac{u\,\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^3} - \phi\left(\frac{u}{\sigma_{g_1}}\right)\frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^3} = \phi\left(\frac{u}{\sigma_{g_1}}\right)\left(\frac{u^2}{\sigma_{g_1}^2}-1\right)\frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^3}.$$

Then, we can write down the derivatives with respect to ρ,

$$\begin{aligned}\frac{\partial}{\partial \rho}p_g(1,1) &= \frac{1}{\sigma_{g_1}}\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta} \\ &\quad + \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u}{\sigma_{g_1}}\right)\left(\frac{u^2}{\sigma_{g_1}^2}-1\right)\frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^3}\Phi\left(\varphi_{2,g}(u)\right)du \\ &\quad + \frac{1}{\sigma_{g_1}}\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}\phi\left(\frac{u}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(u)\right)\frac{\partial}{\partial \rho}\varphi_{2,g}(u)du \\ &= \frac{1}{\sigma_{g_1}}\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta} + B + C.\end{aligned}$$

The integral in B can be computed by parts and we obtain, after some computations:

$$\begin{aligned}B = \frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^2}\Bigg[&-\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) - \frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right) \\ &+\frac{\sigma_{g_1,g_2}|\mathbf{\Sigma}_g|^{1/2}}{\sigma_{g_1}^2\sigma_{g_2}^2}\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\Bigg]\end{aligned}$$

where $|\mathbf{\Sigma}_g| = \sigma_{g_1}^2\sigma_{g_2}^2 - \sigma_{g_1,g_2}^2$. Note moreover that

$$\frac{\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)}{|\mathbf{\Sigma}_g|^{1/2}} = \frac{\phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)}{|\mathbf{\Sigma}_g|^{1/2}} = f(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}, \mathbf{x}_{\rho,g_2}\boldsymbol{\beta}),$$

is the bivariate density of $(u_{g_1}, u_{g_2}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_g)$ at $(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}, \mathbf{x}_{\rho,g_2}\boldsymbol{\beta})$:

$$f(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}, \mathbf{x}_{\rho,g_2}\boldsymbol{\beta}) = \frac{1}{|\mathbf{\Sigma}_g|^{1/2}2\pi}\exp\left\{-\frac{1}{2}(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}, \mathbf{x}_{\rho,g_2}\boldsymbol{\beta})'\mathbf{\Sigma}_g^{-1}(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}, \mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right\}.$$

Thus,

$$B = \frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^2}\left[-\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right) - \frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right) + \frac{\sigma_{g_1,g_2}|\boldsymbol{\Sigma}_g|}{\sigma_{g_1}^2\sigma_{g_2}^2}f(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta},\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right].$$

To compute the integral in C, we first note that

$$\begin{aligned}\frac{\partial\varphi_{2,g}(u)}{\partial\rho} &= \frac{\dot{\mathbf{x}}_{g_2}\boldsymbol{\beta} + \left(\dot{\sigma}_{g_1,g_2}\sigma_{g_1}^2 - \sigma_{g_1,g_2}\dot{\sigma}_{g_1}^2\right)u/(\sigma_{g_1}^2)^2}{(\sigma_{g_2}^2 - \sigma_{g_1,g_2}^2/\sigma_{g_1}^2)^{1/2}} - \frac{1}{2}\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} + \frac{\sigma_{g_1,g_2}}{\sigma_{g_1}^2}u}{(\sigma_{g_2}^2 - \sigma_{g_1,g_2}^2/\sigma_{g_1}^2)^{3/2}}\left(\dot{\sigma}_{g_2}^2 - \frac{2\sigma_{g_1,g_2}\dot{\sigma}_{g_1,g_2}\sigma_{g_1}^2 - \sigma_{g_1,g_2}^2\dot{\sigma}_{g_1}^2}{(\sigma_{g_1}^2)^2}\right)\\ &= a + bu\end{aligned}$$

where

$$\begin{aligned} a &= \frac{\sigma_{g_1}}{|\boldsymbol{\Sigma}_g|^{1/2}}\dot{\mathbf{x}}_{g_2}\boldsymbol{\beta} + \frac{2\sigma_{g_1,g_2}\dot{\sigma}_{g_1,g_2}\sigma_{g_1}^2 - \sigma_{g_1,g_2}^2\dot{\sigma}_{g_1}^2 - \sigma_{g_1}^4\dot{\sigma}_{g_2}^2}{2\sigma_{g_1}|\boldsymbol{\Sigma}_g|^{3/2}}\\ b &= \frac{1}{2|\boldsymbol{\Sigma}_g|^{3/2}}\left[2\frac{\sigma_{g_1}^2\sigma_{g_2}^2}{\sigma_{g_1}}\dot{\sigma}_{g_1,g_2} - \dot{\sigma}_{g_1}^2\left(2\frac{\sigma_{g_2}^2\sigma_{g_1,g_2}}{\sigma_{g_1}} - \frac{\sigma_{g_1,g_2}^3}{\sigma_{g_1}^3}\right) - \frac{\sigma_{g_1,g_2}\sigma_{g_1}^2}{\sigma_{g_1}}\dot{\sigma}_{g_2}^2\right].\end{aligned}$$

Then, by noting that (performing a change of variable)

$$\begin{aligned} C &= \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}/\sigma_{g_1}}^{\infty}\phi(u)\,\phi\left(\varphi_{2,g}(u\sigma_{g_1})\right)\,(a + b\sigma_{g_1}u)du\\ &= \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty}(a + b\sigma_{g_1}u)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(u\sigma_{g_2} + \mathbf{x}_{\rho,g_2}\boldsymbol{\beta}\sigma_{g_1,g_2}/\sigma_{g_1}\sigma_{g_2})}{|\boldsymbol{\Sigma}_g|/\sigma_{g_1}^2}\right\}du\\ &= \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\int_{-\varphi_{1,g}(\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})}^{\infty}\frac{|\boldsymbol{\Sigma}_g|^{1/2}}{\sigma_{g_1}\sigma_{g_2}}\left[a + b\sigma_{g_1}\left(\frac{|\boldsymbol{\Sigma}_g|^{1/2}}{\sigma_{g_1}\sigma_{g_2}}v - \frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\frac{\sigma_{g_1,g_2}}{\sigma_{g_1}\sigma_{g_2}}\right)\right]\phi(v)dv\end{aligned}$$

we get

$$C = \frac{|\boldsymbol{\Sigma}|^{3/2}b\sigma_{g_1}}{\sigma_{g_1}^2\sigma_{g_2}^2}f(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta},\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}) + \frac{|\boldsymbol{\Sigma}|^{1/2}}{\sigma_{g_1}\sigma_{g_2}}\left(a - b\sigma_{g_1}\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\frac{\sigma_{g_1,g_2}}{\sigma_{g_1}\sigma_{g_2}}\right)\phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right).$$

Finally, by putting all the terms together and after some tedious calculations, we get

$$\begin{aligned}\frac{\partial p_g(1,1)}{\partial\rho} = &\frac{f(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta},\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})}{2}\left(2\dot{\sigma}_{g_1,g_2} - \dot{\sigma}_{g_1}^2\frac{\sigma_{g_1,g_2}}{\sigma_{g_1}^2} - \dot{\sigma}_{g_2}^2\frac{\sigma_{g_1,g_2}}{\sigma_{g_2}^2}\right)\\ &+ \phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\left(\frac{\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta}}{\sigma_{g_1}} - \frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^2}\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\\ &+ \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\left(\frac{\dot{\mathbf{x}}_{g_2}\boldsymbol{\beta}}{\sigma_{g_2}} - \frac{\dot{\sigma}_{g_2}^2}{2\sigma_{g_2}^2}\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right).\end{aligned} \tag{17}$$

Similar steps lead to,

$$\begin{aligned}\frac{\partial p_g(d_1,d_2)}{\partial\rho} = s_1s_2&\frac{s_1s_2f(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta},\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})}{2}\left(2\dot{\sigma}_{g_1,g_2} - \dot{\sigma}_{g_1}^2\frac{\sigma_{g_1,g_2}}{\sigma_{g_1}^2} - \dot{\sigma}_{g_2}^2\frac{\sigma_{g_1,g_2}}{\sigma_{g_2}^2}\right)\\ &- \phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\left(\frac{\dot{\mathbf{x}}_{g_1}\boldsymbol{\beta}}{\sigma_{g_1}} - \frac{\dot{\sigma}_{g_1}^2}{2\sigma_{g_1}^2}\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\\ &+ \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\left(1 - \Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\right)\left(\frac{\dot{\mathbf{x}}_{g_2}\boldsymbol{\beta}}{\sigma_{g_2}} - \frac{\dot{\sigma}_{g_2}^2}{2\sigma_{g_2}^2}\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right),\end{aligned} \tag{18}$$

for all $d_1, d_2 \in \{0,1\}^2$.

6.2. SARAR(1,1) probit

In the SARAR(1,1) specification, the probabilities $p_g(d_1, d_2)$ follow the same equation used for the SAR(1) case. However, when defining the quantities in $p_g(d_1, d_2)$, one has to bear in mind that the components $\sigma^2_{g_1}, \sigma^2_{g_2}, \sigma_{g_1,g_2}$ now depend on both ρ and λ through the variance covariance matrix:

$$\mathbf{\Sigma}_{\nu(\rho,\lambda)} = \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}\mathbf{B}_\lambda^{-1\prime}\mathbf{A}_\rho^{-1\prime} = (\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{I}-\lambda\mathbf{M})^{-1}(\mathbf{I}-\lambda\mathbf{M})^{-1\prime}(\mathbf{I}-\rho\mathbf{W})^{-1\prime}. \tag{19}$$

Thus, also in computing the derivatives of each $p_g(d_1, d_2)$ the simultaneous dependence of $\mathbf{\Sigma} = \mathbf{\Sigma}_{\nu(\rho,\lambda)}$ on ρ and λ has to be considered. This, however, does not in general alter the structure of the derivatives with respect to $\boldsymbol{\beta}$ and ρ. It is in fact easy to see that $\partial p_g(d_1, d_2)/\partial\boldsymbol{\beta}$ follows equations (12), (13), (14) and (15). Similarly, $\partial p_g(d_1, d_2)/\partial\rho$ follows equations (18). Note moreover that, by writing

$$\begin{aligned}
\frac{\partial\mathbf{\Sigma}}{\partial\rho} &= \frac{\partial}{\partial\rho}\left(\mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}(\mathbf{B}_\lambda^{-1})'(\mathbf{A}_\rho^{-1})'\right)\\
&= \mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}(\mathbf{B}_\lambda^{-1})'(\mathbf{A}_\rho^{-1})' + \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}(\mathbf{B}_\lambda^{-1})'(\mathbf{A}_\rho^{-1})'\mathbf{W}'(\mathbf{A}_\rho^{-1})'\\
&= \mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{W}'(\mathbf{A}_\rho^{-1})'
\end{aligned}$$

we can use the same equation as in the right–had–side of (16) to compute all the components of $\dot{\mathbf{\Sigma}}_\rho$. We now focus on the derivative $\dot{\mathbf{\Sigma}}_\lambda = \partial\mathbf{\Sigma}/\partial\lambda$:

$$\begin{aligned}
\dot{\mathbf{\Sigma}}_\lambda &= \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}\mathbf{M}\mathbf{B}_\lambda^{-1}(\mathbf{B}_\lambda^{-1})'(\mathbf{A}_\rho^{-1})' + \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}(\mathbf{B}_\lambda^{-1})'\mathbf{M}'(\mathbf{B}_\lambda^{-1})'(\mathbf{A}_\rho^{-1})'\\
&= \mathbf{A}_\rho^{-1}\mathbf{B}_\lambda^{-1}\mathbf{M}\mathbf{A}_\rho\mathbf{\Sigma} + \mathbf{\Sigma}\mathbf{A}_\rho'\mathbf{M}'(\mathbf{B}_\lambda^{-1})'(\mathbf{A}_\rho^{-1})'.
\end{aligned} \tag{20}$$

Finally, using equation (20), we can compute the elements of $\dot{\mathbf{\Sigma}}_\lambda$, namely $\dot\sigma^2_{g_1}(\lambda), \dot\sigma^2_{g_2}(\lambda), \dot\sigma_{g_1,g_2}(\lambda)$ to be used in the following derivatives $\partial p_g(d_1, d_2)/\partial\lambda$:

$$\begin{aligned}
\frac{\partial p_g(1,1)}{\partial\lambda} &= \int_{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}^{\infty} \frac{1}{\sigma_{g_1}}\phi\left(\frac{u}{\sigma_{g_1}}\right)\left[\left(\frac{u^2}{\sigma^2_{g_1}}-1\right)\frac{\dot\sigma^2_{g_1}(\lambda)}{2\sigma^2_{g_1}}\Phi\left(\varphi_{2,g}(u)\right) + \phi\left(\varphi_{2,g}(u)\right)\frac{\partial\varphi_{2,g}(u)}{\partial\lambda}\right]du\\
\frac{\partial p_g(1,0)}{\partial\lambda} &= -\phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\frac{\dot\sigma^2_{g_1}(\lambda)\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{2\sigma^3_{g_1}} - \frac{\partial p_g(1,1)}{\partial\lambda}\\
\frac{\partial p_g(0,1)}{\partial\lambda} &= \int_{-\infty}^{-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}} \frac{1}{\sigma_{g_1}}\phi\left(\frac{u}{\sigma_{g_1}}\right)\left[\left(\frac{u^2}{\sigma^2_{g_1}}-1\right)\frac{\dot\sigma^2_{g_1}(\lambda)}{2\sigma^2_{g_1}}\Phi\left(\varphi_{2,g}(u)\right) + \phi\left(\varphi_{2,g}(u)\right)\frac{\partial\varphi_{2,g}(u)}{\partial\lambda}\right]du\\
\frac{\partial p_g(0,0)}{\partial\lambda} &= \phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\frac{\dot\sigma^2_{g_1}(\lambda)\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{2\sigma^3_{g_1}} - \frac{\partial p_g(0,1)}{\partial\lambda},
\end{aligned} \tag{21}$$

with

$$\frac{\partial\varphi_{2,g}(u)}{\partial\lambda} = \frac{\dot\sigma_{g_1,g_2}(\lambda)\sigma^2_{g_1} - \sigma_{g_1,g_2}\dot\sigma^2_{g_1}(\lambda)}{\sigma_{g_1}\sqrt{\sigma^2_{g_1}\sigma^2_{g_2} - \sigma^2_{g_1,g_2}}}u - \frac{1}{2}\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta} + u\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}}}{\left(\sigma^2_{g_2} - \sigma^2_{g_1,g_2}/\sigma^2_{g_1}\right)^{3/2}}\left(\dot\sigma^2_{g_2}(\lambda) - \frac{2\sigma_{g_1,g_2}\dot\sigma_{g_1,g_2}(\lambda)\sigma^2_{g_1} - \sigma^2_{g_1,g_2}\dot\sigma^2_{g_1}(\lambda)}{\sigma^4_{g_1}}\right).$$

Formulas in equation (21) can be simplified through integration, as for the other terms of the score. Some calculations lead to a formula very similar to equation (17):

$$\begin{aligned}\frac{\partial p_g(1,1)}{\partial \lambda} = \frac{f(\mathbf{x}_{\rho,g_1}\boldsymbol{\beta},\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})}{2}\left(2\dot{\sigma}_{g_1,g_2}(\lambda) - \dot{\sigma}^2_{g_1}(\lambda)\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_1}} - \dot{\sigma}^2_{g_2}(\lambda)\frac{\sigma_{g_1,g_2}}{\sigma^2_{g_2}}\right) \\ - \phi\left(\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}}\right)\Phi\left(\varphi_{2,g}(-\mathbf{x}_{\rho,g_1}\boldsymbol{\beta})\right)\frac{\dot{\sigma}^2_{g_1}(\lambda)}{2\sigma^2_{g_1}}\frac{\mathbf{x}_{\rho,g_1}\boldsymbol{\beta}}{\sigma_{g_1}} - \phi\left(\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}\right)\Phi\left(\varphi_{1,g}(-\mathbf{x}_{\rho,g_2}\boldsymbol{\beta})\right)\frac{\dot{\sigma}^2_{g_2}(\lambda)}{2\sigma^2_{g_2}}\frac{\mathbf{x}_{\rho,g_2}\boldsymbol{\beta}}{\sigma_{g_2}}.\end{aligned} \tag{22}$$

The other derivatives can be easily derived adjusting equations (18) in the same way.

7. Other Figures

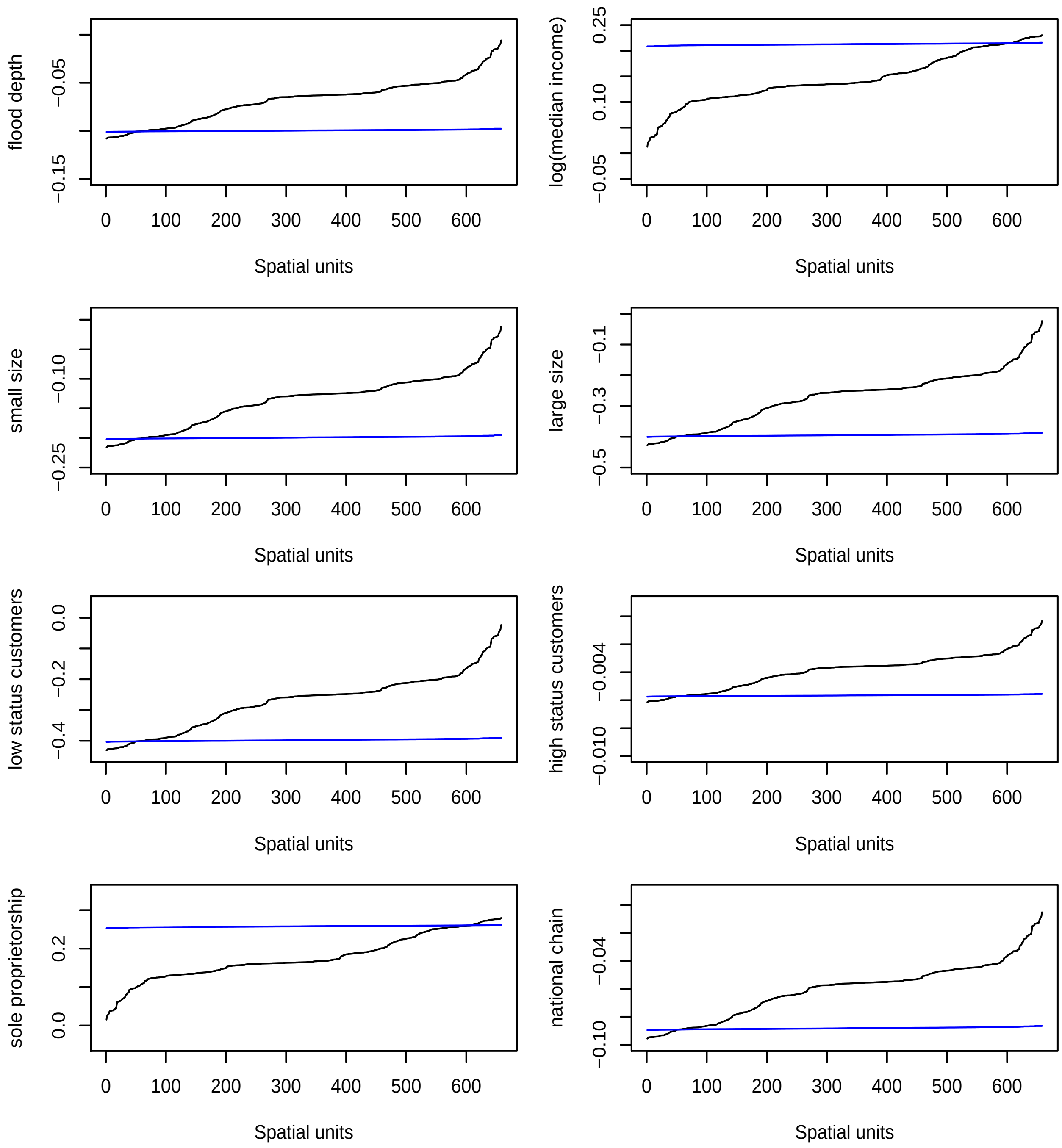

Figure 1: Spatial heterogeneity of the total marginal impacts for each regressor during the second time horizon. Blue lines represent marginal impacts relative to the mean value.

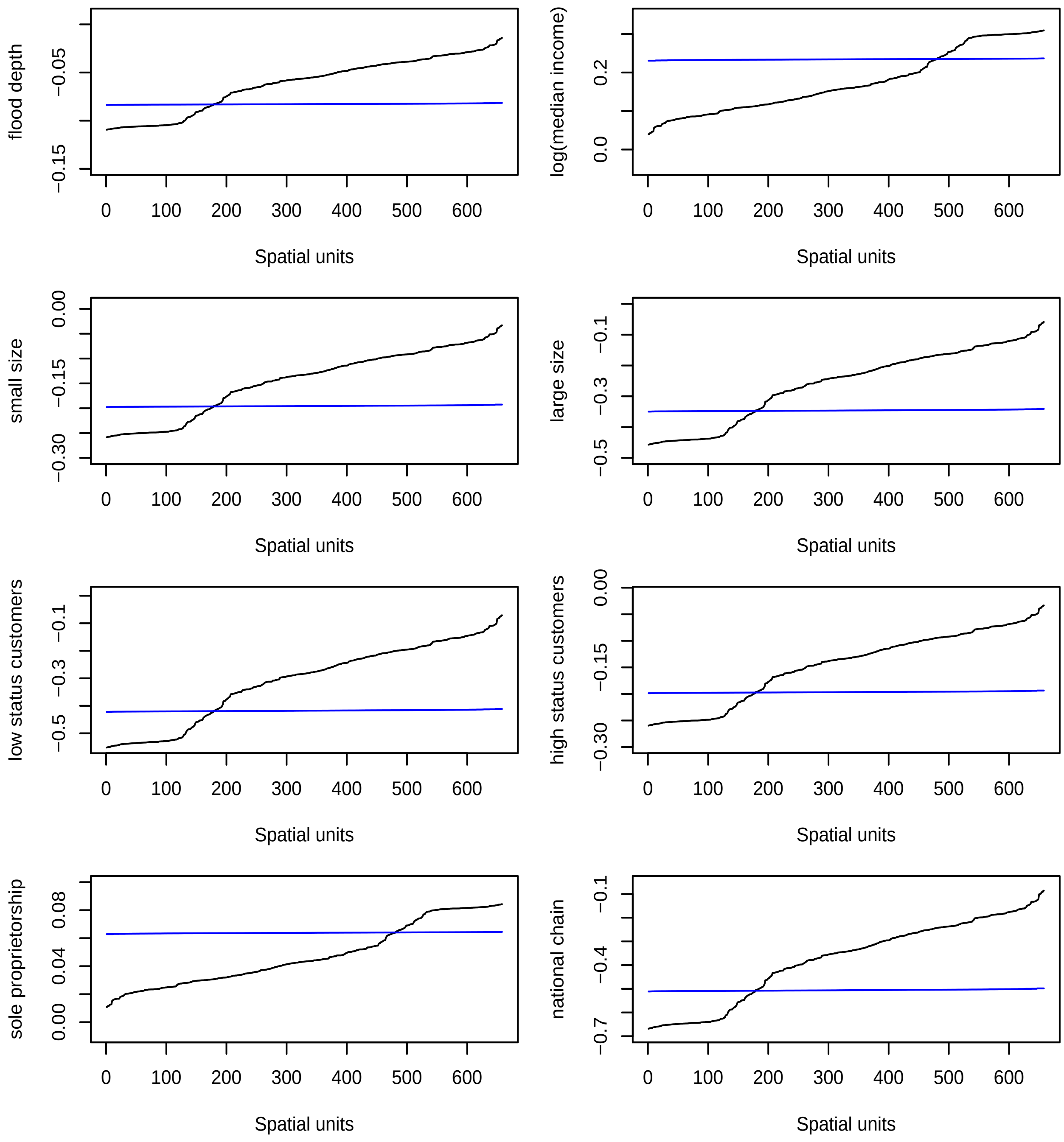

Figure 2: Spatial heterogeneity of the total marginal impacts for each regressor during the third time horizon. Blue lines represent marginal impacts relative to mean value.

References

Bai, Z. and Golub, G. H. (1997). Bounds for the Trace of the Inverse and the Determinant of Symmetric Positive Definite Matrices. *Annals of Numerical Mathematics*, 4:29–38.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.*, 2(4):620–628.

Pinkse, J. and Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, 85(1):125–154.

Wang, H., Iglesias, E. M., and Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, 172(1):77–89.