Supplemental material to "Generalized Spatial Matrix Specifications"

Samantha Leorato[a], Andrea Martinelli[b]
[a]Department of Economics, Management and Quantitative Methods, University of Milan
[b]Department of Science and High Technology, Insubria University

November 19, 2024

S.1 Defining the spatial specification M from the marginal effects

Given the coefficients s_k, the coefficients of M are obtained by applying Theorem 17 in Birmajer et al. (2012), with $a = 0$ and $b = 1$: if for all k, $y_k = \sum_{j=1}^{n} j! B_{k,j}(x_1, \ldots, x_{k-j+1})$, then

$$x_k = \sum_{j=1}^{k} (-1)^{j-1} j! B_{k,j}(y_1, \ldots, y_{k-j+1}),$$

where $B_{k,j}(x_1, \ldots, x_{k-j+1})$ is the exponential Bell polynomial, that is

$$B_{k,j}(x_1, \ldots, x_{k-j+1}) = \frac{k!}{j!} \hat{B}_{k,j}\left(\frac{x_1}{1!}, \ldots, \frac{x_{k-j+1}}{(k-j+1)!}\right).$$

In our framework, setting $n!(1-a_0)\, s_n = y_n$ and noticing that in this case $x_k = k!\, a_k/(1-a_0)$, we obtain

$$a_k = (1-a_0) \sum_{j=1}^{k} (-1)^{j-1} \frac{j!}{k!} B_{k,j}\left(1!\,(1-a_0)\, s_1, \ldots, (k-j+1)!\,(1-a_0)\, s_{k-j+1}\right). \tag{S.1}$$

In particular, we can construct a function M with inverse that is a polynomial of finite order: given $M^{-1}(\rho) = \sum_{j=0}^{K} s_j (\rho W)^j$ where $K < +\infty$, applying (S.1) we see that the coefficients of M are

$$a_k = \sum_{j=1}^{k} \left(-(1-a_0)\right)^{j+1} \hat{B}_{k,j}\left(s_1, s_2, \ldots, s_{k-j+1}\right)$$

for $k \leq K$ and

$$a_k = \left\{ \sum_{j=1}^{k-K} \left(-\left(1-a_0\right)\right)^{j+1} \hat{B}_{k,j}\left(s_1, s_2, \ldots, s_K, 0, \ldots, 0\right) + \sum_{j=k-K+1}^{k} \left(-(1-a_0)\right)^{j+1} \hat{B}_{k,j}\left(s_1, s_2, \ldots, s_{k-j+1}\right) \right\}$$
$$= \left\{ \sum_{j=1}^{k-K} \left(-\left(1-a_0\right)\right)^{j+1} \hat{B}_{K+j-1,j}\left(s_1, s_2, \ldots, s_K\right) + \sum_{j=k-K+1}^{k} \left(-(1-a_0)\right)^{j+1} \hat{B}_{k,j}\left(s_1, s_2, \ldots, s_{k-j+1}\right) \right\}$$

for $k > K$.

S.2 Power representations of GB distributions

S.2.1 Power representation of the beta family

To simplify the illustration, we consider the case $a = 1$.

We have in this case that our target function is related to the CDF of the beta family:

$$G\left(x; b, c, p, q\right) = \frac{B\left(\frac{x/b}{1+c\,x/b}; p, q\right)}{B(p,q)} = \frac{1}{B(p,q)} \int_0^{\frac{x/b}{1+c\,x/b}} u^{p-1}(1-u)^{q-1} du, \quad 0 < x < b/(1-c). \tag{S.2}$$

We point out once again that, by setting $b = p = q = 1$ and $c = 0$, one gets the SAR model.

To define M and derive the coefficients a_k and s_k, we exploit the series expansion of the beta function:

$$G\left(x; b, c, p, q\right) = \frac{1}{B(p,q)} \left(\frac{x/b}{1+c\,x/b}\right)^p \sum_{k=0}^{\infty} \frac{(1-q)_k}{k!(p+k)} \left(\frac{x/b}{1+c\,x/b}\right)^k, \tag{S.3}$$

where $(u)_k = \Gamma(u+k)/\Gamma(u)$ is the Pochhammer symbol, see (8.391) in Gradshtein et al. (1965).

In order to have a real and monotone function, we have to impose some constraints on the parameters space: by (S.2) and the integral representation (9.111) in Gradshtein et al. (1965) we have

$$G\left(x; b, c, p, q\right) = \frac{1}{B\left(q, 1-q\right)} \int_0^1 t^{-q}(1-t)^{p+q-1} \left(\frac{1+c\,x/b}{x/b} - t\right)^{-p} dt. \tag{S.4}$$

From (S.3) it is easy to see that the analytic continuation of the incomplete beta function to $(-b/(1+c), 0]$ is one–valued and real for $p \in \mathbb{N}_0$ and furthermore, its density is

$$g(x) = B(p,q)^{-1} B'\left(\frac{x/b}{1+c\,x/b}; b, c, p, q\right) \cdot b^{-1} \left(1 + \frac{cx}{b}\right)^2.$$

By deriving the expression in (S.4) w.r. to $u = \frac{x/b}{1+c\,x/b}$, it is easily proved that the function $B(u; b, c, p, q)$ is monotone increasing only if p is odd, otherwise it is monotone decreasing for any value of q.

Under these conditions and from this (S.3), we can derive the coefficients a_k in (5):

$$\begin{aligned} G\left(x;b,c,p,q\right) &= \frac{1}{B(p,q)}\sum_{k=p}^{\infty}\frac{(1-q)_{k-p}}{(k-p)!k}\left(\frac{x/b}{1+c\,x/b}\right)^{k} \\ &= \frac{1}{B(p,q)}\sum_{k=p}^{\infty}\frac{x^k}{b^k}\frac{(1-q)_{k-p}}{(k-p)!k}\sum_{h=0}^{\infty}\binom{h+k-1}{k-1}\left(-c\frac{x}{b}\right)^{h} \\ &= \sum_{m=p}^{\infty}x^m\frac{1}{b^mB(p,q)}\sum_{h=0}^{m-p}\binom{m-1}{h}\frac{(-c)^h(1-q)_{m-p-h}}{(m-p-h)!(m-h)} \end{aligned}$$

that implies $a_k = \frac{1}{b^kB(p,q)}\sum_{h=0}^{k-p}\binom{k-1}{h}\frac{(-c)^h(1-q)_{k-p-h}}{(k-p-h)!(k-h)}$ for $k \geq p$ and zero otherwise. In particular, if $c = 0$, the coefficients simplify to

$$a_k = \frac{1}{b^kB(p,q)}\frac{(1-q)_{k-p}}{(k-p)!\,k}$$

for $k \geq p$.

S.2.2 Power representation of the generalized gamma specification

In this section, we study the power series representation of the GG family, prove the representation in equation (10) and study the set of admissible values for the parameters.

Using the identities (8.354) of Gradshtein et al. (1965) we get the series representation:

$$G\left(x;a,d,p\right) = \sum_{n\geq 0}\frac{(-1)^n\left(\frac{x}{d}\right)^{a(p+n)}}{\Gamma\left(p\right)n!\left(p+n\right)}.$$

We assume, to define the coefficients a_k, that $ap + an \in \mathbb{N}$ for all $n \geq 0$, which requires both $r \in \mathbb{N}$ and $a \in \mathbb{N}$. In this case, it is easily seen that

$$a_k = \begin{cases} \frac{(-1)^{(k-ap)/a}d^{-k}}{\Gamma(p)(k/a-p)!}\frac{a}{k} & k = a(p+n), \quad n = 0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}.$$

When $a = 1$ (gamma distribution), the structural equation for y includes only h^{th}–order neighbors if $h \geq r$, while if for example $a > 1$ only h^{th}–order neighbors that are multiple of a and larger than or equal to r are included.

Now, exploiting this expression we can extend the function to the negative axis, but we have to choose the parameter region for which the function will be real and, if possible, monotone.

If $a \in \mathbb{N}$ is even, the function is naturally extended to the whole real axis because of:

$$G(-x;a,d,p) = \frac{\gamma\left(p,(-x/d)^a\right)}{\Gamma\left(p\right)} = G(x;a,d,p).$$

If instead a is odd, and by assuming $d=1$, without loss of generality, we have, for all $x>0$

$$
\begin{aligned}
G(-x;a,1,p) &= \sum_{n\geq 0} \frac{(-1)^n\left((-x)^a\right)^{p+n}}{\Gamma(p)\,n!\,(p+n)} = \frac{1}{\Gamma(p)}\sum_{n\geq 0}\frac{(-1)^{ap+(a+1)n}\left(x^a\right)^{p+n}}{n!\,(p+n)} \\
&= \frac{(-1)^{ap}}{\Gamma(p)}\sum_{n\geq 0}\frac{1}{n!}\int_0^{x^a} y^{n+p-1}dy = \frac{(-1)^{ap}}{\Gamma(p)}\int_0^{x^a} y^{p-1}\sum_{n\geq 0}\frac{y^n}{n!}dy \\
&= \frac{(-1)^{ap}}{\Gamma(p)}\int_0^{x^a} y^{p-1}e^{y}dy = -\frac{a}{\Gamma(p)}\int_0^{x}(-z)^{ap-1}e^{-(-z)^a}dz,
\end{aligned}
$$

where the last identity is obtained by the change of variable $y=z^a$ and the integrand in the right-hand-side is the GG density with $d=1$ evaluated at $-z$.

We point out that the assumption $ap\in\mathbb{N}$ guarantees $(-1)^{ap}\in\mathbb{R}$ and thus the polynomial coefficients are real and one valued. Further, we want $G(x;a,d,p)$ to be monotone non-decreasing, which is satisfied if the derivative of G is nonnegative also for negative values of x. By differentiating G, it is easy to see that the function is monotone nondecreasing if $ap-1$ is even. Thus, the parameters of the GG family must satisfy the restriction to odd values of ap.

S.3 Some useful Lemmas

The following Lemmas 1–5 may be found in Debarsy et al. (2015b) (Lemmas B.1–B.5) and are presented here for the convenience of the reader. Lemmas 6–8 are slightly modified versions of Lemmas B.6–B.8 in Debarsy et al. (2015b). We recall that a matrix is said to be RCB if it is bounded in both rows and columns sum norms.

In this Section, we use the following matrix functions

$$
\begin{aligned}
H_n(\tau) &:= I_n - V_nX_n\left(X_n'V_n'V_nX_n\right)^{-1}X_n'V_n' \\
P_{2n}(\tau) &:= X_n\left(X_n'V_n'V_nX_n\right)^{-1}X_n' \\
T_n(\rho,\tau) &:= M_n'(\rho)V_n'(\tau)H_n(\tau)V_n(\tau)M_n(\rho)
\end{aligned}
$$

and their derivatives are collected in Table S.1.

Derivatives	
$\frac{\partial}{\partial\tau}H_n(\rho,\tau)$	$-\left(\frac{\partial}{\partial\tau}P_{2n}(\tau)V_n'(\tau)\right)^s + V_n(\tau)P_{2n}(\tau)\left(\frac{\partial}{\partial\tau}V_n'(\tau)V_n(\tau)\right)^s P_{2n}(\tau)V_n'(\tau)$
$\frac{\partial}{\partial\rho}T_n(\rho,\tau)$	$\left(\frac{\partial}{\partial\rho}M_n'(\rho)V_n'(\tau)H_n(\tau)V_n(\tau)M_n(\rho)\right)^s$
$\frac{\partial}{\partial\tau}T_n(\rho,\tau)$	$M_n'(\rho)\left[\left(\frac{\partial}{\partial\tau}V_n'(\tau)H_n(\tau)V_n\right)^s + V_n'(\tau)\frac{\partial}{\partial\tau}H_n(\tau)V_n(\tau)\right]M_n(\rho)$

Table S.1: Table of most important derivatives

Lemma S.1. Given sequences of matrices $A_n \in \mathbb{R}^{n\times n}$ and $X_n \in \mathbb{R}^{n\times k}$ if

(i) A_n is RCB[1]

(ii) X_n is uniformly bounded, i.e. $\sup_n \max_{i,j} X_{n;i,j} < \infty$ and

$$\lim_n \frac{1}{n} X_n' X_n = L$$

where L is a nonsingular matrix.

Then $M_n := I_n - X_n \left(X_n' X_n\right)^{-1} X_n'$ is RCB and $tr\left(M_n A_n\right) = \operatorname{tr}\left(A_n\right) + O(1)$.

Lemma S.2. Given two sequences of matrices (A_n) and (B_n) in $\mathbb{R}^{n\times n}$ and a sequence of random vectors $\boldsymbol{\epsilon} \in \mathbb{R}^n$ with independent elements s.t.

$$\mathbb{E}\left[\boldsymbol{\epsilon}_n\right] = \mathbf{0} \text{ and } cov\left(\boldsymbol{\epsilon}_n\right) := \Sigma_n = diag\left(\sigma^2_{n,11}, \sigma^2_{n,22}, \ldots, \sigma^2_{n,nn}\right),$$

then

$$\mathbb{E}\left[\boldsymbol{\epsilon}_n \boldsymbol{\epsilon}_n' A_n \boldsymbol{\epsilon}_n\right] = \left(a_{n,11}\,\mathbb{E}\left(\epsilon^3_{n1}\right), a_{n,22}\,\mathbb{E}\left(\epsilon^3_{n2}\right), \ldots, a_{n,nn}\,\mathbb{E}\left(\epsilon^3_{nn}\right)\right)$$

and

$$\mathbb{E}\left[\boldsymbol{\epsilon}' A_n \boldsymbol{\epsilon}_n \boldsymbol{\epsilon}' B_n \boldsymbol{\epsilon}_n\right] = \sum_{i=1}^{n} a_{n,ii} b_{n,ii}\left[\mathbb{E}\left(\boldsymbol{\epsilon}^4_{ni}\right) - 3\sigma^4_{n,i}\right] + \operatorname{tr}\left(\Sigma_n A_n\right) tr\left(\Sigma_n B_n\right) + \operatorname{tr}\left[\Sigma_n A_n \Sigma_n\left(B_n + B_n'\right)\right].$$

Lemma S.3. Given two sequences of matrices $(A_n) \in \mathbb{R}^{n\times n}$ and $(C_n) \in \mathbb{R}^{n\times k}$ and a sequence $\boldsymbol{\epsilon} \in \mathbb{R}^n$ of random vectors with independent elements s.t.

$$\mathbb{E}\left[\boldsymbol{\epsilon}_n\right] = \mathbf{0}, \sup_{i,n} \mathbb{E}\left[\left|\epsilon_{ni}\right|^4\right] < \infty \text{ and } cov\left(\boldsymbol{\epsilon}_n\right) := \Sigma_n = diag\left(\sigma^2_{n,11}, \sigma^2_{n,22}, \ldots, \sigma^2_{n,nn}\right).$$

If: (i) A_n is RCB; (ii) C_n is uniformly bounded; (iii) the sequence $\mathbb{E}\left(\epsilon^4_{ni}\right)$ is bounded, then

1. $\boldsymbol{\epsilon}_n' A_n \boldsymbol{\epsilon}_n = O_P(n)$ and $\mathbb{E}\left[\boldsymbol{\epsilon}_n' A_n \boldsymbol{\epsilon}_n\right] = O(n)$
2. $\left[\boldsymbol{\epsilon}_n' A_n \boldsymbol{\epsilon}_n - \mathbb{E}\left(\boldsymbol{\epsilon}_n' A_n \boldsymbol{\epsilon}_n\right)\right] = o_P(1)$ and
3. $n^{-1/2} C_n' A_n \boldsymbol{\epsilon}_n = O_P(1)$.

Lemma S.4. Given a sequence (A_n) in $\mathbb{R}^{n\times n}$, a sequence $\boldsymbol{b}_n \in \mathbb{R}^n$ and a sequence $\boldsymbol{\epsilon} \in \mathbb{R}^n$ of random vectors with independent elements and s.t. $\mathbb{E}\left[\boldsymbol{\epsilon}_n\right] = \mathbf{0}$ and $cov\left(\boldsymbol{\epsilon}_n\right) := \Sigma_n = diag\left(\sigma^2_{n,11}, \sigma^2_{n,22}, \ldots, \sigma^2_{n,nn}\right)$. If

(i) A_n is RCB

[1] Given $A \in M_n$ we define $\|A\|_\infty := \max_{i\le n} \|A_{i,\cdot}\|_1$ and $\|A\|_1 := \max_{j\le n} \|A_{\cdot,j}\|_1$; a sequence of matrices A_n is Uniformly bounded in both row and columns sum norms (RCB) if $\sup_n \|A_n\|_\infty < \infty$ and $\sup_n \|A_n\|_1 < \infty$

(ii) there exists a constant η_1 such that

$$\sup_n \frac{1}{n^{2+\eta_1}} \, \|\boldsymbol{b}_n\|_{2+\eta_1} < \infty$$

(iii) there exists a constant $\eta_2 > 0$ such that

$$\sup_{i,n} \mathbb{E}\left[|\epsilon_{ni}|^{4+\eta_2}\right] < \infty$$

(iv) letting $c_n := \boldsymbol{\epsilon}_n' A_n \boldsymbol{\epsilon}_n + \boldsymbol{b}_n' \boldsymbol{\epsilon}_n - \operatorname{tr}(A_n \Sigma_n)$ and $\sigma_{c_n}^2 := Var(c_n)$, the sequence $n^{-1}\sigma_{c_n}^2$ is bounded away from 0

then

$$\frac{c_n}{\sigma_{c_n}} \xrightarrow{d} Gauss\,(0,1)$$

Lemma S.5. Let $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n^*$ be the minimizers of $Q_n(\boldsymbol{\theta})$ and $Q_n^*(\boldsymbol{\theta})$ in $\boldsymbol{\Theta}$, respectively. If $|Q_n(\boldsymbol{\theta}) - \bar{Q}_n(\boldsymbol{\theta})| = o_p(1)$ uniformly in a convex set $\boldsymbol{\Theta}$ and $\bar{Q}_n(\boldsymbol{\theta})$ is uniquely identified at $\boldsymbol{\theta}_0$ and if $|Q_n(\boldsymbol{\theta}) Q_n^*(\boldsymbol{\theta})| = o_p(1)$ uniformly in $\boldsymbol{\Theta}$, then $\hat{\boldsymbol{\theta}}_n$ and $\hat{\boldsymbol{\theta}}_n^*$ converge in probability at $\boldsymbol{\theta}_0$.

Furthermore, if $\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ converges uniformly to a well defined, nonsingular matrix at $\boldsymbol{\theta}_0$; $\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial} = O_p(1)$; $\left|\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2 Q_n^*(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right| = o_p(1)$ uniformly in $\boldsymbol{\theta}$; $\left|\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial} - \sqrt{n}\frac{\partial Q_n^*(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial}\right| = o_p(1)$, then

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \sqrt{n}(\hat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_0) + o_p(1).$$

Lemma S.6. Let $V_n := V_n(\tau)$ a sequence of matrices that depend continuously on τ. If

(i) V_n is RCB

(ii) there exist $\delta > 0$ such that $\sigma(V_n(\tau)'V_n(\tau))$, (where $\sigma(A)$ is the spectrum of A) is uniformly bounded away from 0 on $[-\delta,\delta]$, this means that $\inf_{\tau\in[-\delta,\delta]} \min|\sigma(V_n(\tau)'V_n(\tau))| > 0$.

(iii) X_n are uniformly bounded.

(iv) for any $\tau \in [-\delta,\delta]$ we have

$$\lim_n \frac{1}{n} X_n' V_n' V_n X_n = L_\tau$$

where L_τ is a nonsingular matrix.

then

1. $X_n \left(X_n' V_n' V_n X_n\right)^{-1} X_n'$
2. $H_n(\tau) := I_n - V_n X_n \left(X_n' V_n' V_n X_n\right)^{-1} X_n' V_n'$

are RCB uniformly in $\tau \in [-\delta,\delta]$.

Proof. By assumption (ii) the symmetric matrix $W_n(\tau) := V_n'(\tau)V_n(\tau)$ is diagonalizable for all $\tau \in [-\delta, \delta]$ as $W_n(\tau) = U'(\tau)\Lambda_\tau U(\tau)$, where $U(\tau)$ is the matrix of normalized eigenvectors of $W_n(\tau)$ and Λ_τ is the diagonal matrix with its eigenvalues. From this, there exist a $\kappa > 0$ such that $W_n(\tau) - \kappa I_n$ is positive definite, then it follows that

$$X_n'\left(W_n(\tau) - \kappa I_n\right)X_n \geq 0 \Rightarrow X_n'W_n(\tau)X_n - \kappa X_n'X_n \geq 0 \Rightarrow \left(\frac{1}{n}X_n'W_n(\tau)X_n\right)^{-1} \leq \left(\kappa\frac{1}{n}X_n'X_n\right)^{-1}$$

where the inequalities are in matrix definiteness sense. From this inequality and assumption (ii) we get 1. and 2. follow from assumption (i). □

Lemma S.7. Given

(i) $M_n(\rho), V_n(\tau), A_n$ and B_n sequences of matrices in $\mathbb{R}^{n\times n}$ that are RCB

(ii) $\frac{\partial}{\partial\rho}M_n$ is continuous and RCB in $[-\delta, \delta]$ and $\frac{\partial}{\partial\tau}V_n$ is continuous and RCB in $[-\delta, \delta]$

(iii) $\boldsymbol{b}_n \in \mathbb{R}^n$ elementwise uniformly bounded

(iv) X_n a sequence of matrices in $\mathbb{R}^{n\times k}$ such that for any $\tau \in [-\delta, \delta]$ we have

$$\lim_n \frac{1}{n}X_n'V_n'(\tau)V_n(\tau)X_n = L_\tau$$

where L_τ is a nonsingular matrix.

(v) a $\boldsymbol{\epsilon} \in \mathbb{R}^n$ a sequence of random vectors with independent elements and s.t. $\mathbb{E}\left[\boldsymbol{\epsilon}_n\right] = \mathbf{0}$ and $cov\left(\boldsymbol{\epsilon}_n\right) := \Sigma_n = diag\left(\sigma^2_{n,11}, \sigma^2_{n,22}, \ldots, \sigma^2_{n,nn}\right)$ and

$$\sup_{n,i}\mathbb{E}\left(\epsilon^4_{ni}\right) < \infty$$

then

1. $\frac{1}{n}\boldsymbol{b}_n'M_n'(\rho)V_n'(\tau)H_n(\tau)V_n(\tau)M_n(\rho)A_n\boldsymbol{\epsilon}_n = o_P(1)$ uniformly on $[-\delta, \delta]^2$
2. $\frac{1}{n}\boldsymbol{b}_n'M_n'(\rho)V_n'(\tau)B_nV_n(\tau)M_n(\rho)A_n\boldsymbol{\epsilon}_n = o_P(1)$ uniformly on $[-\delta, \delta]^2$
3. $$\frac{1}{n}\left[\boldsymbol{\epsilon}_n'A_n'M_n'(\rho)V_n'(\tau)H_n(\tau)V_n(\tau)M_n(\rho)A_n\boldsymbol{\epsilon}_n - \mathrm{tr}\left(A_n'M_n'(\rho)V_n'(\tau)H_n(\tau)V_n(\tau)M_n(\rho)A_n\boldsymbol{\Sigma}_n\right)\right] = o_P(1)$$
   uniformly on $[-\delta, \delta]^2$
4. $$\frac{1}{n}\left[\boldsymbol{\epsilon}_n'A_n'M_n'(\rho)V_n'(\tau)B_nV_n(\tau)M_n(\rho)A_n\boldsymbol{\epsilon}_n - \mathrm{tr}\left(A_n'M_n'(\rho)V_n'(\tau)B_nV_n(\tau)M_n(\rho)A_n\boldsymbol{\Sigma}_n\right)\right] = o_P(1)$$
   uniformly on $[-\delta, \delta]^2$

5.
$$\frac{1}{n}\,\mathrm{tr}\left(A_n' M_n'(\rho) V_n'(\tau)\left(I_n - H_n(\tau)\right) V_n(\tau) M_n(\rho) A_n \mathbf{\Sigma}_n\right) = o_P(1)$$
uniformly on $[-\delta, \delta]^2$

Proof. To prove the uniform convergence in probability we prove the pointwise convergence in probability and a "Lipschitz–type" condition on the growth of the function.

We begin with 1., calling $T_n(\rho, \tau) := M_n'(\rho) V_n'(\tau) H_n(\tau) V_n(\tau) M_n(\rho)$ we can write the expression in 1. as $\boldsymbol{b}_n' T_n(\rho, \tau) A_n \epsilon_n / n$ and by Lemma S.3 we have

$$\frac{1}{n}\,\boldsymbol{b}_n' T_n(\rho, \tau) A_n \epsilon_n = o_P(1).$$

Then we have to prove the growth condition.

As usual we estimate the growth of the function with the mean value theorem: given $(\rho_0 \tau_0)$ and (ρ_1, τ_1), we have

$$\begin{aligned}\frac{1}{n}\,\boldsymbol{b}_n'\left[T_n(\rho_1, \tau_1) - T_n(\rho_0, \tau_0)\right] A_n \epsilon_n &= \frac{1}{n}\,\boldsymbol{b}_n' \frac{\partial}{\partial \rho} T_n\left(\tilde{\rho}, \tilde{\tau}\right) A_n \epsilon_n (\rho_1 - \rho_0) \\ &\quad + \frac{1}{n}\,\boldsymbol{b}_n' \frac{\partial}{\partial \tau} T_n\left(\tilde{\rho}, \tilde{\tau}\right) A_n \epsilon_n \left(\tau_1 - \tau_0\right)\end{aligned}$$

where, setting $P_{2n}(\tau) := X_n \left(X_n' V_n' V_n X_n\right)^{-1} X_n'$,

$$\begin{aligned}\frac{\partial}{\partial \rho} T_n\left(\rho, \tau\right) &= \left(\frac{\partial}{\partial \rho} M_n'(\rho) V_n'(\tau) H_n(\tau) V_n(\tau) M_n(\rho)\right)^s \\ \frac{\partial}{\partial \tau} T_n\left(\rho, \tau\right) &= M_n'(\rho)\left[\left(\frac{\partial}{\partial \tau} V_n'(\tau) H_n(\tau) V_n\right)^s + V_n'(\tau) \frac{\partial}{\partial \tau} H_n(\tau) V_n(\tau)\right] M_n(\rho)\end{aligned}$$

and

$$\frac{\partial}{\partial \tau} H_n(\tau) = -\left(\frac{\partial}{\partial \tau} P_{2n}(\tau) V_n'(\tau)\right)^s + V_n(\tau) P_{2n}(\tau)\left(\frac{\partial}{\partial \tau} V_n'(\tau) V_n(\tau)\right)^s P_{2n}(\tau) V_n'(\tau)$$

then all the matrices in partial derivatives are RCB and by Lemma S.6 the partial derivatives too are RCB, i.e. there is a constant c such that

$$\left|\frac{1}{n}\,\boldsymbol{b}_n'\left[T_n(\rho_1, \tau_1) - T_n(\rho_0, \tau_0)\right] A_n \epsilon_n\right| \le \frac{c}{n}\,\left|\boldsymbol{b}_n' A_n \epsilon_n\right|$$

and from this we get the stochastic equicontinuity.

For point 3. the function is

$$\frac{1}{n}\left[\boldsymbol{\epsilon}' A_n' T_n\left(\rho, \tau\right) A_n \boldsymbol{\epsilon}_n - \mathrm{tr}\left(A_n' T_n\left(\rho, \tau\right) A_n \Sigma_n\right)\right]$$

this is easily proved that converge to 0 in probability and for stochastic equicontinuity, with the above notation, and applying mean value theorem we get

$$
\begin{aligned}
\frac{1}{n}&\left|\boldsymbol{\epsilon}' A_n' \left[T_n\left(\rho_1,\tau_1\right) - T_n\left(\rho_0,\tau_0\right)\right] A_n \boldsymbol{\epsilon}_n - \operatorname{tr}\left(A_n'\left[T_n\left(\rho_0,\tau_0\right) - T_n\left(\rho_1,\tau_1\right)\right] A_n \Sigma_n\right)\right| = \\
&= \frac{1}{n}\left|\boldsymbol{\epsilon}' A_n' \left[\frac{\partial}{\partial\rho} T_n\left(\tilde{\rho},\tilde{\tau}\right)\left(\rho_1-\rho_0\right) + \frac{\partial}{\partial\tau} T_n\left(\tilde{\rho},\tilde{\tau}\right)\left(\tau_1-\tau_0\right)\right] A_n \boldsymbol{\epsilon}_n \right. \\
&\qquad \left. - \operatorname{tr}\left[A_n'\left(\frac{\partial}{\partial\rho} T_n\left(\tilde{\rho},\tilde{\tau}\right)\left(\rho_1-\rho_0\right) + \frac{\partial}{\partial\tau} T_n\left(\tilde{\rho},\tilde{\tau}\right)\left(\tau_1-\tau_0\right)\right) A_n \Sigma_n\right]\right| \\
&\le \left|\frac{1}{2n}\boldsymbol{\epsilon}' A_n' \left[\nabla T_n\left(\tilde{\rho},\tilde{\tau}\right)\nabla T_n'\left(\tilde{\rho},\tilde{\tau}\right)\right] A_n \boldsymbol{\epsilon}_n\right| \|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_0\| + \\
&\qquad + \frac{1}{n}\left|\operatorname{tr}\left(A_n' \nabla T_n\left(\tilde{\rho},\tilde{\tau}\right) A_n \Sigma_n\right)\right| \|\boldsymbol{\varphi}_1 - \boldsymbol{\varphi}_0\|
\end{aligned}
$$

then the first term in the sum can be diagonalized because is symmetric (see the proof of Lemma B.7 in Debarsy et al. (2015b)) and can be majorized by RCB hypothesis on M_n, V_n and their derivatives and the second one by the same hypothesis but directly. □

Lemma S.8. If $(V_n)_{n\ge1}$ is a family of equicontinuous matrix valued functions with respect to the norm $\|\cdot\|$, then for all sequences τ_n such that $\tau_n \to \tau_0$

$$\|V_n(\tau_n) - V_n(\tau_0)\| = o_P(1).$$

Proof. With the notation above, for all $\varepsilon > 0$ there exist n_0 such that $|\tau_n - \tau_0| < \delta$ for all $n \ge n_0$ and by equicontinuity

$$\|V_n(\tau_n) - V_n(\tau_0)\| < \varepsilon$$

for all $n \ge n_0$. □

S.4 Proofs

S.4.1 Proof of Lemma 1

We recall that an ordinary Bell's polynomial and an exponential Bell polynomial are given by

$$\hat{B}_{n,k}(x_1, x_2, \ldots, x_{n-k+1}) = \sum_{(j_1,\ldots,j_{n-k+1})\in S_{n,k}} \frac{k!}{j_1! j_2! \ldots j_{n-k+1}!} x_1^{j_1} x_2^{j_2} \ldots x_{n-k+1}^{j_{n-k+1}}$$

and

$$B_{n,k}(x_1, x_2, \ldots, x_{n-k+1}) = \sum_{(j_1,\ldots,j_{n-k+1})\in S_{n,k}} \frac{n!}{j_1! j_2! \ldots j_{n-k+1}!} x_1^{j_1} \left(\frac{x_2}{2}\right)^{j_2} \cdots \left(\frac{x_{n-k+1}}{(n-k+1)!}\right)^{j_{n-k+1}}$$

where $S_{n,k} := \left\{(j_1,\ldots,j_{n-k+1}) : \sum_{h=1}^{n-k+1} j_h = k \text{ and } \sum_{h=1}^{n-k+1} hj_h = n\right\}$ and a complete exponential Bell polynomial is

$$B_n\left(x_1,\ldots,x_n\right) = \sum_{k=1}^{n} B_{n,k}\left(x_1,x_2,\ldots,x_{n-k+1}\right)$$

see, for example, Wang and Wang (2009).

First of all we remark that

$$\left(1-a_0\right) M_n^{-1}\left(\rho\right) = \sum_{n\geq 0}\left[G_{0,n}\left(\rho W\right)\right]^n = \left(h\circ f\right)\left(\rho W\right)$$

where

$$f(X) = \sum_{k\geq 1}\underbrace{a_k k!/(1-a_0)}_{p_k}\frac{X^k}{k!} \text{ and } h(X) = \sum_{n\geq 0}\underbrace{n!}_{b_n}\frac{X^n}{n!}$$

We recall that Fàa di Bruno's formula for formal power series guarantees that, given two functions $f(x) = \sum_{n\geq 1} p_n x^n/n!$ and $h(x) = \sum_{k\geq 0} b_k x^k/k!$, then

$$h\left(f(x)\right) = \sum_{n\geq 0} b_n\left[f\left(x\right)\right]^n = b_0 + \sum_{n\geq 1}\left[\sum_{k=1}^{n}\frac{b_k}{n!}B_{n,k}\left(p_1,p_2,\ldots,p_{n-k+1}\right)\right]x^n$$

where $B_{k,j}$ are exponential Bell polynomial. Applying this formula to our case we get

$$\left(1-a_0\right) M_n^{-1}\left(\rho\right) = 1 + \sum_{n\geq 1} c_n(\rho W)^n$$

where the coefficients c_n are given by

$$\begin{aligned} c_n &= \sum_{k=1}^{n}\frac{k!}{n!}B_{n,k}\left(p_1,\ldots,p_{n-k+1}\right) = \sum_{k=1}^{n}\frac{k!}{n!}\left\{\sum_{\boldsymbol{j}\in S_{n,k}}\frac{n!}{j_1!\ldots j_{n-k+1}!}\prod_{i=1}^{n-k+1}\left(\frac{a_i}{1-a_0}\right)^{j_i}\right\} \\ &= \sum_{k=1}^{n}\hat{B}_{n,k}\left(\frac{a_1}{1-a_0},\frac{a_2}{1-a_0},\ldots,\frac{a_{n-k+1}}{1-a_0}\right) \end{aligned}$$

and from this we have the result.

S.4.2 Proof of Proposition 5

We can write

$$\begin{aligned} \left|\ell_n^{K_n}(\rho) - \ell_n(\rho)\right| &= \left|\log|det(M_n^{K_n}(\rho))| - \log|det(M_n(\rho))| - \frac{n}{2}\left(\log\|\hat{\boldsymbol{e}}_{K_n}(\rho)\|^2 - \log\|\hat{\boldsymbol{e}}(\rho)\|^2\right)\right| \\ &= \left|\log|det(M_n^{K_n}(\rho))| - \log|det(M_n(\rho))| - \frac{n}{2}\log\left(1 - \frac{\hat{\boldsymbol{e}}_{K_n}(\rho)\|^2 - \|\hat{\boldsymbol{e}}(\rho)\|^2}{\|\hat{\boldsymbol{e}}(\rho)\|^2}\right)\right| \\ &\leq \left|\log|det(M_n^{K_n}(\rho))| - \log|det(M_n(\rho))|\right| + \frac{n}{2}\left|\frac{\|\hat{\boldsymbol{e}}_{K_n}(\rho)\|^2 - \|\hat{\boldsymbol{e}}(\rho)\|^2}{\|\hat{\boldsymbol{e}}(\rho)\|^2}\right| \end{aligned}$$

By expanding the function $\log(1-u)$, we obtain:

$$
\begin{aligned}
\log|det(M_n(\rho))| &= \sum_{i=1}^{n} \log\left|1-\sum_k a_k\rho^k\lambda_i^k\right| \\
&= \sum_{i=1}^{n} \log\left(1-\sum_k a_k\rho^k\lambda_i^k\right) \text{ (because } \textstyle\sum_k a_k\rho^k\lambda_i^k \leq 1-\eta) \\
&= -\sum_{i=1}^{n}\left[\sum_{h=1}^{\infty}\left(\sum_k a_k\rho^k\lambda_i^k\right)^h \frac{1}{h}\right] \\
&= -\sum_{i=1}^{n}\sum_{k=1}^{\infty} a_k\rho^k\lambda_i^k \left[\sum_{h=1}^{\infty}\left(\sum_k a_k\rho^k\lambda_i^k\right)^{h-1}\frac{1}{h}\right].
\end{aligned}
$$

We set $g_i = \sum_{k=1}^{\infty} a_k\rho^k\lambda_i^k$, and note that $g_i < 1$ because $|\rho^k\lambda_i^k < r_G|$. Thus, we easily see that

$$
\begin{aligned}
\log|det(M_n(\rho))| &= -\sum_{i=1}^{n} g_i \sum_{h=1}^{\infty} g_i^{h-1}\frac{1}{h} = -\sum_{i=1}^{n} g_i \sum_{h=1}^{\infty}\frac{\partial}{\partial g_i} g_i^h \\
&= -\sum_{i=1}^{n} g_i \frac{\partial}{\partial g_i}\left(\frac{1}{1-g_i}-1\right) = -\sum_{i=1}^{n}\frac{g_i}{(1-g_i)^2}.
\end{aligned}
$$

Thus,

$$
\log\left|det(M_n^{K_n}(\rho))\right| = -\sum_{i=1}^{n}\frac{g_i^{K_n}}{(1-g_i^{K_n})^2}.
$$

By denoting by $g_i^{K_n} = \sum_{k\leq K_n} a_k\rho^k\lambda_i^k$, we also get

$$
\log|M_n^{K_n}(\rho)| = -\sum_{i=1}^{n}\frac{g_i^{K_n}}{(1-g_i^{K_n})^2}.
$$

Then, since $0<\eta\leq \min_i \min\{g_i^{K_n}, g_i\} \leq \max_i \max\{g_i^{K_n}, g_i\} \leq 1-\eta$:

$$
\begin{aligned}
\left|\log|det(M_n^{K_n}(\rho))| - \log|det(M_n(\rho))|\right| &= \left|\sum_{i=1}^{n}\frac{\left(g_i-g_i^{K_n}\right)(1-g_i^{K_n}g_i)}{(1-g_i^{K_n})^2(1-g_i)^2}\right| \\
&\leq \sum_{i=1}^{n}\left|\frac{\left(g_i-g_i^{K_n}\right)(1-1+\eta)}{\eta^2}\right| \\
&= \eta^{-1}\sum_{i=1}^{n}\left|\sum_{k=K_n+1}^{\infty} a_k\rho^k\lambda_i^k\right|.
\end{aligned}
$$

If the series $\sum_k a_k u^k$ is absolutely convergent for all $|u| < r_G$, then the series $\sum_k |a_k u^k|$ is uniformly convergent for $|u| \leq \gamma < r_G$. Therefore, if $|\rho| \, \|W\| < \gamma$

$$\eta^{-1} \sum_{i=1}^{n} \left| \sum_{k=K_n+1}^{\infty} a_k \rho^k \lambda_i^k \right| \leq \eta^{-1} \sum_{i=1}^{n} \left| \frac{\rho \lambda_i}{\gamma} \right|^{K_n} \sum_{k=K_n+1}^{\infty} \left| a_k \gamma^k \right| \leq \eta^{-1} n \left| \frac{\rho \|W\|}{\gamma} \right|^{K_n} |R_{K_n}| \qquad \text{(S.5)}$$

where $|R_{K_n}| = o(1)$ because the series is convergent.

Now, let us consider

$$\left| \frac{\|\hat{\boldsymbol{e}}(\rho)\|^2 - \|\hat{\boldsymbol{e}}_{K_n}(\rho)\|^2}{\|\hat{\boldsymbol{e}}(\rho)\|^2} \right| \leq \frac{\|\hat{\boldsymbol{e}}(\rho) - \hat{\boldsymbol{e}}_{K_n}(\rho)\|^2}{\|\hat{\boldsymbol{e}}(\rho)\|^2}.$$

By using Cauchy-Schwartz inequality to $\boldsymbol{y}^\top (W^k)^\top H_n W^h \boldsymbol{y}$,

$$\begin{aligned}
\|\hat{\boldsymbol{e}}(\rho) - \hat{\boldsymbol{e}}_{K_n}(\rho)\|^2 &= \left\| \sum_{k=K_n+1}^{\infty} a_k \rho^k \boldsymbol{e}_W(k) \right\|^2 = \sum_{h,k=K_n+1}^{\infty} a_k a_h \rho^{h+k} \boldsymbol{y}^\top (W^k)^\top H_n W^h \boldsymbol{y} \\
&\leq \sum_{h,k=K_n+1}^{\infty} \left| a_k a_h \rho^{h+k} \sqrt{\boldsymbol{y}^\top (W^k)^\top H_n W^k \boldsymbol{y}} \sqrt{\boldsymbol{y}^\top (W^h)^\top H_n W^h \boldsymbol{y}} \right| \\
&\leq \|\boldsymbol{y}\|^2 \sum_{h,k=K_n+1}^{\infty} \left| a_k a_h \rho^{h+k} \right| \sqrt{\|H_n W^k\|} \sqrt{\|H_n W^h\|} \\
&\leq \|\boldsymbol{y}\|^2 \sum_{h,k=K_n+1}^{\infty} \left| a_k a_h \rho^{h+k} \|W\|^{h+k} \right| \leq \left| \frac{|\rho| \|W\|}{\gamma} \right|^{2(K_n+1)} \left(\sum_{h=K_n+1}^{\infty} |a_h| \gamma^k \right)^2 \\
&\leq \|\boldsymbol{y}\|^2 \sup_{|\rho| \|W\|| < \gamma} \left| \frac{|\rho| \|W\|}{\gamma} \right|^{2(K_n+1)} |R_{K_n}|^2.
\end{aligned}$$

We also note that $\|\hat{\boldsymbol{e}}(\rho)\|^2 = n \hat{\sigma}_n^2$ (see equation (13)), and thus $\hat{\sigma}_n^2 = \sigma_0^2 + o_p(1)$, so that

$$\frac{\|\hat{\boldsymbol{e}}(\rho) - \hat{\boldsymbol{e}}_{K_n}(\rho)\|^2}{\|\hat{\boldsymbol{e}}(\rho)\|^2} \leq \frac{1}{n(\sigma_0^2 + o_p(1))} \|\boldsymbol{y}\|^2 \gamma^{2(K_n+1)} |R_{K_n}|^2 \qquad \text{(S.6)}$$

Putting (S.5) and (S.6) together, we find:

$$\left| \ell_n^{K_n}(\rho) - \ell_n(\rho) \right| \leq \eta^{-1} n \left| \frac{\rho \|W\|_2}{\gamma} \right|^{K_n} |R_{K_n}| + \frac{n}{2\sigma_0^2 + o_p(1)} \frac{\|\boldsymbol{y}\|^2}{n} \left(\frac{|\rho| \|W\|_2}{\gamma} \right)^{2(K_n+1)} |R_{K_n}|^2$$

Both the terms in the right-hand-side have $|R_{K_n}| = o(1)$ and $\frac{\|\boldsymbol{y}\|^2}{n} = O_p(1)$ as a consequence of the assumptions. Thus, for any $\delta > 0$ and $|\rho| \leq \gamma / \|W_n\| - \delta$,

$$\begin{aligned}
\left| \ell_n^{K_n}(\rho) - \ell_n(\rho) \right| &\leq \eta^{-1} n \, |1 - \delta/\gamma|^{K_n} \, |R_{K_n}| + \frac{n}{2\sigma_0^2 + o_p(1)} \frac{\|\boldsymbol{y}\|^2}{n} (1 - \delta/\gamma)^{2(K_n+1)} \, |R_{K_n}|^2 \\
&\leq n(1 - \delta/\gamma)^{K_n} o(1) + n(1 - \delta/\gamma)^{2(K_n+1)} o_p(1).
\end{aligned}$$

The sequence of sets $\Phi_n = \{\rho : |\rho| \leq \gamma/\|W_n\| - \delta, \ \forall \delta > 0\} = \{\rho : |\rho| < \gamma/\|W_n\|\}$ has nonempty limsup:

$$\Phi = \bigcap_n \bigcup_{m \geq n} \Phi_m = \{|\rho| < \gamma/\tilde{\lambda}\}$$

if $\|W_n\| \to \tilde{\lambda} < \infty$, that is, if the spectral norm of W_n converges to a finite limit as n increases. Under this condition, we can write:

$$\sup_{\rho \in \Phi} \left| \ell_n^{K_n}(\rho) - \ell_n(\rho) \right| \leq n(\rho\tilde{\lambda}/\gamma)^{K_n} o(1) + n(\rho\tilde{\lambda}/\gamma)^{2(K_n+1)} o_p(1) = o_p(1).$$

because of $n(\rho\tilde{\lambda}/\gamma)^{K_n} = \exp\left\{ K_n \left(\frac{\log n}{K_n} + \log(\rho\tilde{\lambda}/\gamma) \right) \right\} \to 0$

S.4.3 Proof of Theorem 1

The proof is analogous to the one of Proposition 1 in Debarsy et al. (2015a) (Supplement). It consists of two steps: first, we prove that there exists a sequence $\bar{Q}_n$ such that $\left[Q_n - \bar{Q}_n\right]/n$ converges to zero uniformly, then we apply Theorem 3.4 of White (1994), by checking that the unique maximizer of $\bar{Q}_n$ tends to $\boldsymbol{\theta}_0$ (identification uniqueness condition).

Let $\bar{Q}_n(\rho, \tau) = \min_{\boldsymbol{\beta}, \sigma^2} \mathbb{E}_0 \left[Q_n(\rho, \tau, \boldsymbol{\beta}, \sigma^2) \right]$, where $\mathbb{E}_0$ denotes the expected value with respect to the true parameters, i.e. ρ_0, τ_0, β_0 and σ_0^2. Then, we can write the expected value as:

$$\begin{aligned}
\mathbb{E}_0 \left[Q_n(\rho, \tau, \boldsymbol{\beta}, \sigma^2) \right] &= n \log 2\pi\sigma^2 - \log |M_n' M_n| - \log |V_n' V_n| \\
&+ \frac{1}{\sigma^2} (M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0 - X\boldsymbol{\beta})' V_n' V_n (M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0 - X\boldsymbol{\beta}) \\
&+ \frac{\sigma_0^2}{\sigma^2} \mathrm{tr} \left((V_{n,0}^{-1})' (M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1} \right)
\end{aligned}$$

and it attains the minimum at

$$\begin{aligned}
\boldsymbol{\beta}_n^* &= (X' V_n' V_n X)^{-1} X' V_n' V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0 \\
\sigma_n^{2*} &= \frac{1}{n} \boldsymbol{\beta}_0' X' (M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0 + \frac{\sigma_0^2}{n} \mathrm{tr} \left((V_{n,0}^{-1})' (M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1} \right).
\end{aligned} \tag{S.7}$$

Thus, we can easily compute $\bar{Q}_n(\rho, \tau)$:

$$\bar{Q}_n(\rho, \tau) = n \log 2\pi - \log |M_n' M_n| - \log |V_n' V_n| + n \log(\sigma_n^{2*}) + n. \tag{S.8}$$

From (14) and (S.8), we get:

$$\frac{1}{n} \left(\tilde{Q}_n - \bar{Q}_n \right) = \log(\hat{\sigma}_n^2) - \log(\sigma_n^{2*}). \tag{S.9}$$

Note that Assumption 7 readily implies that $\sigma_n^{2*} > 0$ for all n uniformly in the parameter space Φ and that the difference in (S.9) does not depend on $\log|M_n' M_n|$ nor $\log|V_n' V_n|$.

Uniform Convergence. We have (equation (S.9)):

$$
\begin{aligned}
\frac{1}{n}\left(\tilde{Q}_n - \bar{Q}_n\right) &= \log(\hat{\sigma}_n^2) - \log(\sigma_n^{2*}) = \log\left(\frac{\boldsymbol{y}' M_n' V_n' H_n V_n M_n \boldsymbol{y} - n\sigma_n^{2*}}{n\sigma_n^{2*}} + 1\right) \\
&\leq \left|\frac{\boldsymbol{y}' M_n' V_n' H_n V_n M_n \boldsymbol{y} - n\sigma_n^{2*}}{n\sigma_n^{2*}}\right| \\
&= \frac{1}{n\sigma_n^{2*}}\left|(\boldsymbol{u} + X\boldsymbol{\beta}_0)'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1}(X\boldsymbol{\beta}_0 + \boldsymbol{u}) - n\sigma_n^{2*}\right| \\
&= \frac{1}{n\sigma_n^{2*}}\boldsymbol{\epsilon}'(V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\boldsymbol{\epsilon} + \\
&\quad + \frac{2}{n\sigma_n^{2*}}\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\boldsymbol{\epsilon} \\
&\quad - \frac{1}{n\sigma_n^{2*}}\sigma_0^2 \mathrm{tr}\left(V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right) \\
&\leq \frac{1}{n\sigma_n^{2*}}(A + B + C)
\end{aligned}
\tag{S.10}
$$

where

$$
\begin{aligned}
A &= \left|\boldsymbol{\epsilon}'(V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\boldsymbol{\epsilon} - \sigma_0^2 \mathrm{tr}\left(V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right)\right| \\
B &= \sigma_0^2 \mathrm{tr}\left(V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n'(I_n - H_n) V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right) \\
C &= 2\left|\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\boldsymbol{\epsilon})\right|
\end{aligned}
$$

It is therefore enough to show that A, B, C are bounded uniformly over $\boldsymbol{\Theta}$ and $\sigma_n^{2*} > \eta > 0$ for all n to have uniform convergence of $n^{-1}|\tilde{Q}_n - \bar{Q}_n|$. The first follows from the application of Lemma S.7, which is a generalization of Lemma A.7 in Debarsy et al. (2015a), while $\sigma_n^{2*} > \eta > 0$ for all n if either for all n

$$
n^{-1}\boldsymbol{\beta}_0'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X\boldsymbol{\beta}_0 > 0,
$$

or if

$$
\sigma_0^2 \frac{\mathrm{tr}\left((V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right)}{n} > 0
$$

for all n.

Uniform equicontinuity of $n^{-1}\bar{Q}_n$. Since for any arbitrary $(\rho_i, \tau_i) \in [-\delta, \delta]^2$,

$$\begin{aligned} &n^{-1}\left|\bar{Q}_n(\rho_2,\tau_2) - \bar{Q}_n(\rho_1,\tau_1)\right| \\ &\quad\le \frac{1}{n}\left|\log\left|V_n(\tau_2)'V_n(\tau_2)\right| - \log\left|V_n(\tau_1)'V_n(\tau_1)\right|\right| + \frac{1}{n}\left|\log\left|M_n(\rho_2)'M_n(\rho_2)\right| - \log\left|M_n(\rho_1)'M_n(\rho_1)\right|\right| \\ &\qquad + \left|\log(\sigma_n^{2*}(\rho_2,\tau_2)) - \log(\sigma_n^{2*}(\rho_1,\tau_1))\right| \end{aligned} \tag{S.11}$$

Let us assume (w.l.o.g.) $|V_n(\tau_2)'V_n(\tau_2)| \ge |V_n(\tau_1)'V_n(\tau_1)|$

We can therefor apply Klein's trace inequality (Petz (1994)) to the operator convex function $f(A) = -\log(A)$:

$$-\log\left|V_n(\tau_2)'V_n(\tau_2)\right| + \log\left|V_n(\tau_1)'V_n(\tau_1)\right| \ge -\mathrm{tr}\left(\left(V_n(\tau_2)'V_n(\tau_2) - V_n(\tau_1)'V_n(\tau_1)\right)\left(V_n(\tau_1)'V_n(\tau_1)\right)^{-1}\right)$$

which can be rewritten as

$$\begin{aligned} &\left|\log\left|V_n(\tau_2)'V_n(\tau_2)\right| - \log\left|V_n(\tau_1)'V_n(\tau_1)\right|\right| \le \mathrm{tr}\left[\left(V_n(\tau_2)'V_n(\tau_2) - V_n(\tau_1)'V_n(\tau_1)\right)\left(V_n(\tau_1)'V_n(\tau_1)\right)^{-1}\right] \\ &\quad\le \mathrm{tr}\left|\left(V_n(\tau_2)'V_n(\tau_2) - V_n(\tau_1)'V_n(\tau_1)\right)\left(V_n(\tau_1)'V_n(\tau_1)\right)^{-1}\right| \\ &\quad\le \left\|\left(V_n(\tau_2)'V_n(\tau_2) - V_n(\tau_1)'V_n(\tau_1)\right)\right\|_F \left\|\left(V_n(\tau_1)'V_n(\tau_1)\right)^{-1}\right\|_F \end{aligned}$$

because of the following version of Cauchy-Schwartz inequality for symmetric matrices A and B, $\mathrm{tr}(AB) \le \sqrt{\mathrm{tr}(A^2)\mathrm{tr}(B^2)} = \|A\|_F\|B\|_F$, and where $\|A\|_F$ is the Frobenius norm of A. Then, we apply the norm inequality $\|A\|_F \le \sqrt{n}\|A\|$ (with $\|A\|$ either the row or column sum norm). Finally, pointing out that $(V_n(\tau_1)'V_n(\tau_1))^{-1}$ is RCB and because of equicontinuity of V_n, we can say that for $|\tau_2 - \tau_1| < \delta$ there exists a $\varepsilon > 0$ such that

$$\frac{1}{n}\Big|\log|V_n(\tau_2)'V_n(\tau_2)| - \log|V_n(\tau_1)'V_n(\tau_1)|\Big| \le n^{-1}\left\|(V_n(\tau_1)'V_n(\tau_1))^{-1}\right\| \|(V_n(\tau_2)'V_n(\tau_2) - V_n(\tau_1)'V_n(\tau_1))\| \le K\varepsilon.$$

The same arguments can be repeated for M_n to get $\frac{1}{n}\Big|\log|M_n(\rho_2)'M_n(\rho_2)| - \log|M_n(\rho_1)'M_n(\rho_1)|\Big| \le K\varepsilon$.

Similarly, since σ_n^{2*} is a quadratic (bounded) function of the equicontinuous matrices V_n and M_n, we can also obtain a similar bound for the last term from:

$$\begin{aligned} &\left|\log(\sigma_n^{2*}(\rho_2,\tau_2)) - \log(\sigma_n^{2*}(\rho_1,\tau_1))\right| \le \left|\frac{\sigma_n^{2*}(\rho_2,\tau_2) - \sigma_n^{2*}(\rho_1,\tau_1)}{\sigma_n^{2*}(\rho_1,\tau_1)}\right| \\ &\le \left\|M_{n,0}^{-1}X\boldsymbol{\beta}_0\right\|_F^2 \left\|M_{n,2}'V_{n,2}'H_{n,2}V_{n,2}M_{n,2} - M_{n,1}'V_{n,1}'H_{n,1}V_{n,1}M_{n,1}\right\|_F \\ &\quad + \frac{\sigma_0^2}{n\sigma_n^{2*}}\left\|(V_{n,0}^{-1})'(M_{n,0}^{-1})'M_{n,0}^{-1}V_{n,0}^{-1}\right\|_F \left\|M_{n,2}'V_{n,2}'V_{n,2}M_{n,2} - M_{n,1}'V_{n,1}'V_{n,1}M_{n,1}\right\|_F \end{aligned}$$

where $M_{n,i} = M_n(\rho_i)$, $i = 1, 2$ and $V_{n,1}, H_{n,i}$ are defined accordingly.

Identification uniqueness. Finally, we need to show that $(\rho_n^*, \tau_n^*) = \arg\max_{\rho,\tau} \bar{Q}_n$ is unique (and coincides with (ρ_0, τ_0)). In particular, we show that $\frac{1}{n}\bar{Q}_n(\rho_0, \tau_0,) > \frac{1}{n}\bar{Q}_n(\rho, \tau)$ for all $(\rho, \tau) \neq (\rho_0, \tau_0)$ and $n \to \infty$.

Note in fact that minimizing $n^{-1}\bar{Q}_n$ is equivalent to minimizing

$$\begin{aligned}
&\frac{1}{n}\left(\bar{Q}_n - \log|\mathbf{\Sigma}_0|\right) - \log|\sigma_0^2| - \log(2\pi) - 1 \\
&\quad = \log\left(\frac{1}{n\sigma_0^2}\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0 + \frac{1}{n}\mathrm{tr}\left((V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right)\right) \\
&\qquad\qquad + \frac{1}{n}\log|\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}| \\
&\quad = \log\left(\frac{1}{n}\frac{\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0}{\sigma_0^2|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} + \frac{1}{n}\frac{\mathrm{tr}\left((V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right)}{|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}}\right),
\end{aligned}$$

where $\mathbf{\Sigma}_0 = M_{n,0}^{-1} V_{n,0}^{-1} (V_{n,0}^{-1})'(M_{n,0}^{-1})'$.

The term

$$\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0 \geq 0$$

(in particular, $\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0 = 0$ only if $\rho = \rho_0$ (in which case $H_n V_n M_n M_{n,0}^{-1} X = H_n V_n X = \mathbf{0}$), because we assumed that $\boldsymbol{\beta}_0 \neq \mathbf{0}$ and that both V_n and $M_n V_n$ are full rank for all $\tau, \rho \in [-\delta, \delta]^2$). Further,

$$\frac{1}{n}\frac{\mathrm{tr}\left((V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right)}{|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} \geq 1,$$

follows from the arithmetic-geometric mean inequality.

Then, the minimum of $\frac{1}{n}\left(\bar{Q}_n - \log|\mathbf{\Sigma}_0| - \log(2\pi) - 1\right)$ is zero and clearly is achieved at (ρ_0, τ_0).

The identification uniqueness is therefore satisfied if, for any $(\rho, \tau) \neq (\rho_0, \tau_0)$, the limit of $\frac{1}{n}\left(\bar{Q}_n - \log|\mathbf{\Sigma}_0| - \log(2\pi) - 1\right)$ is strictly positive. This means we have to ensure that

$$\lim_{n\to\infty}\frac{1}{n}\frac{\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0}{\sigma_0^2|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} + \frac{1}{n}\frac{\mathrm{tr}\left((V_{n,0}^{-1})'(M_{n,0}^{-1})' M_n' V_n' V_n M_n M_{n,0}^{-1} V_{n,0}^{-1}\right)}{|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} > 1.$$

This follows if either

$$\lim_n \frac{\boldsymbol{\beta}_0' X'(M_{n,0}^{-1})' M_n' V_n' H_n V_n M_n M_{n,0}^{-1} X \boldsymbol{\beta}_0}{n|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} > 0,$$

(which can't happen if $\rho = \rho_0$), or if

$$\lim_n \frac{1}{n} \frac{\operatorname{tr}\left((V_{n,0}^{-1})'(M_{n,0}^{-1})'M_n'V_n'V_nM_nM_{n,0}^{-1}V_{n,0}^{-1}\right)}{|(V_{n,0}^{-1})'(M_{n,0}^{-1})'M_n'V_n'V_nM_nM_{n,0}^{-1}V_{n,0}^{-1}|^{1/n}} = \lim_n \frac{n^{-1}\operatorname{tr}(\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1})}{|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} := \lim_{n\to\infty} \frac{A_n}{G_n} > 1.$$

From assumption 3, easily follows that $|\mathbf{\Sigma}|^{1/n}$ and $|\mathbf{\Sigma}^{-1}|^{1/n}$ are bounded, for all (ρ, τ). Then, we can find $c > 0$ and $C < \infty$ such that $c < |\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n} < C$, for all n. This means that

$$\lim_{n\to\infty} \frac{1}{n} \frac{\boldsymbol{\beta}_0'X'(M_{n,0}^{-1})'M_n'V_n'H_nV_nM_nM_{n,0}^{-1}X\boldsymbol{\beta}_0}{\sigma_0^2|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} > 0$$

if and only if

$$\lim_n \frac{1}{n}\boldsymbol{\beta}_0'X'(M_{n,0}^{-1})'M_n'V_n'H_nV_nM_nM_{n,0}^{-1}X\boldsymbol{\beta}_0 > 0,$$

thus yielding Assumption 7.1.

On the other hand, we can write:

$$\lim_n \frac{n^{-1}\operatorname{tr}(\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1})}{|\mathbf{\Sigma}_0\mathbf{\Sigma}^{-1}|^{1/n}} := \lim_{n\to\infty} \frac{A_n}{G_n} > 1,$$

where $A_n = n^{-1}\operatorname{tr}(\boldsymbol{S}_n)$ and $G_n = \sqrt[n]{|\boldsymbol{S}_n|}$ are the arithmetic and geometric means of the eigenvalues of $\boldsymbol{S}_n = (V_{n,0}^{-1})'(M_{n,0}^{-1})'M_n'V_n'V_nM_nM_{n,0}^{-1}V_{n,0}^{-1}$, respectively, that we denote by $0 < \lambda_{n:1}^S \leq \cdots \leq \lambda_{n:n}^S < \infty$.

We note that, for any $n \geq 2$, $\frac{A_n}{G_n} = 1$ if and only if all terms of the two means are the same, and this occurs if and only if $\boldsymbol{S}_n$ has one eigenvalue with multiplicity n. This happens of course if $\rho = \rho_0$ and $\tau = \tau_0$, because $\boldsymbol{S}_n = I_n$. If $\rho \neq \rho_0$ (and/or $\tau \neq \tau_0$), then $\mathbf{\Sigma}_n \neq \mathbf{\Sigma}_0$ for finite n, and we can bound the difference between arithmetic and geometric means using a result in Tung (1975):

$$A_n - G_n \geq \frac{1}{n}\left(\sqrt{\lambda_{n:n}^S} - \sqrt{\lambda_{n:1}^S}\right)^2.$$

Therefore:

$$\frac{A_n}{G_n} = \frac{A_n - G_n}{G_n} + 1 \geq 1 + \frac{1}{n}\frac{\left(\sqrt{\lambda_{n:n}^S} - \sqrt{\lambda_{n:1}^S}\right)^2}{G_n}$$

and a sufficient condition is that the difference between the largest and smallest eigenvalues of $\boldsymbol{S}_n$ increases with n at a rate at least equal to $O(n\,G_n) = O(n)$, because from Assumption 3 $G_n = O(1)$.

S.4.4 Computation of the Information matrix

Before the proof of Theorem 2, we need the following preliminary Lemma, whose proof is in the supplemental material.

Lemma S.9. The matrix

$$\Omega_n(\boldsymbol{\theta}_0) = \frac{1}{n}\,\mathbb{E}\left[\frac{\partial Q_n\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\theta}}\,\frac{\partial Q_n\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\theta}'}\right] = \left[2C_n(\boldsymbol{\theta}_0) + \Delta_n(\boldsymbol{\theta}_0)\right],$$

where $C_n(\boldsymbol{\theta}_0)$ is given by

$$\begin{aligned}
C_n(\boldsymbol{\theta}_0) &= \frac{1}{n}\,\mathbb{E}\left(\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) \\
&= \begin{bmatrix}
\frac{1}{n}\,\frac{2}{\sigma_0^2}X'V_{n,0}'V_{n,0}X & \mathbf{0} & -\frac{2X'V_{n,0}'\bar{B}_0V_{n,0}X\boldsymbol{\beta}_0}{n\sigma_0^2} & \mathbf{0} \\
\cdot & \frac{1}{n}\mathrm{tr}\left((A_0^s)^2\right) & \frac{1}{n}\mathrm{tr}\left(\bar{B}_0^sA_0^s\right) & -\frac{1}{n\sigma_0^2}\mathrm{tr}\left(A_0^s\right) \\
\cdot & \cdot & \frac{1}{n}\left[\mathrm{tr}\left((\bar{B}_0^s)^2\right) + \frac{2m_0}{\sigma_0^2}\right] & -\frac{1}{n\sigma_0^2}\mathrm{tr}\left(B_0^s\right) \\
\cdot & \cdot & \cdot & \frac{1}{\sigma_0^4}
\end{bmatrix}, \qquad \text{(S.12)}\\
m_0 &= \boldsymbol{\beta}_0'X'B_0'V_n'V_nB_0X\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0'X'V_n'\bar{B}_0'\bar{B}_0V_nX\boldsymbol{\beta}_0
\end{aligned}$$

and $\Delta_n(\boldsymbol{\theta}_0)$ is defined by

$$\Delta_n(\boldsymbol{\theta}_0) = \tfrac{1}{n}\begin{bmatrix}
\mathbf{0} & -\frac{2\mu_3}{\sigma_0^4}X'V_{n,0}'\boldsymbol{a}_0 & -\frac{4\mu_3}{\sigma_0^4}X'V_{n,0}'\boldsymbol{b}_0 & \frac{2\mu_3}{\sigma_0^6}X'V_{n,0}'\imath_n \\
\cdot & \frac{\kappa_4}{\sigma_0^4}\boldsymbol{a}_0'\boldsymbol{a}_0 & \frac{2\kappa_4}{\sigma_0^2}\boldsymbol{b}_0'\boldsymbol{a}_0 + \frac{2\mu_3}{\sigma_0^4}\beta_0'X'V_{n,0}'\bar{B}_0'\boldsymbol{a}_0 & -\frac{\kappa_4}{\sigma_0^6}\mathrm{tr}(A_0^s) \\
\cdot & \cdot & \frac{4\kappa_4}{\sigma_0^4}\boldsymbol{b}_0'\boldsymbol{b}_0 + \frac{8\mu_3}{\sigma_0^4}\boldsymbol{b}_0'\bar{B}_0V_{n,0}X\boldsymbol{\beta}_0 & -\frac{1}{\sigma_0^6}\left(2\mu_3\imath_n'\bar{B}_0V_{n,0}X\boldsymbol{\beta}_0 + \kappa_4\mathrm{tr}(\bar{B}_0^s)\right) \\
\cdot & \cdot & \cdot & \frac{n\kappa_4}{\sigma_0^8}
\end{bmatrix} \qquad \text{(S.13)}$$

with $\mu_3 = \mathbb{E}_0\,\epsilon_i^3$ and $\kappa_4 = \mathbb{E}_0\,\epsilon_i^4 - 3\sigma_0^4$ the excess kurtosis.

Proof. First, we compute the derivatives

$$\begin{aligned}
\frac{\partial Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}'} &= -\frac{2}{\sigma^2}\left(M_n\boldsymbol{y} - X\boldsymbol{\beta}\right)'V_n'V_nX \\
\frac{\partial Q_n(\boldsymbol{\theta})}{\partial\tau} &= \frac{1}{\sigma^2}\left(M_n\boldsymbol{y} - X\boldsymbol{\beta}\right)'\left(\frac{\partial V_n'}{\partial\tau}V_n\right)^s\left(M_n\boldsymbol{y} - X\boldsymbol{\beta}\right) - 2\frac{\partial\log|V_n|}{\partial\tau} \\
\frac{\partial Q_n(\boldsymbol{\theta})}{\partial\rho} &= \frac{2}{\sigma^2}\left(M_n\boldsymbol{y} - X\boldsymbol{\beta}\right)'V_n'V_n\frac{\partial M_n}{\partial\rho}\boldsymbol{y} - 2\frac{\partial\log|M_n|}{\partial\rho} \\
\frac{\partial Q_n(\boldsymbol{\theta})}{\partial\sigma^2} &= -\frac{1}{\sigma^4}\left(M_n\boldsymbol{y} - X\boldsymbol{\beta}\right)'V_n'V_n\left(M_n\boldsymbol{y} - X\boldsymbol{\beta}\right) + \frac{n}{\sigma^2}.
\end{aligned} \qquad \text{(S.14)}$$

Then, we take all expectations in $\mathbb{E}_0\left[\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right]$, where $\mathbb{E}_0$ is the expectation with respect to the true data generating distribution. Thus, recalling that $\boldsymbol{b}_0$ and $\boldsymbol{a}_0$ denote the diagonal terms of $\bar{B}_0$ and A_0. , we find:

$$n\Omega_n(1,1) = \frac{4}{\sigma_0^4}\,\mathbb{E}_0\left[X'V_{n,0}'V_{n,0}(M_{n,0}\boldsymbol{y}_n - X\boldsymbol{\beta}_0)(M_{n,0}\boldsymbol{y}_n - X\boldsymbol{\beta}_0)'V_{n,0}'V_{n,0}X\right] = \frac{4}{\sigma_0^4}\,\mathbb{E}_0\left(X'V_{n,0}'\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'V_{n,0}X\right)$$

$$= \frac{4}{\sigma_0^2}\,X'V_{n,0}'V_{n,0}X = 2C_n(1,1)$$

$$n\Omega_n(2,1) = \mathbb{E}_0\left[-\frac{2}{\sigma_0^2}X'V_{n,0}'\boldsymbol{\epsilon}_n\left(\frac{1}{\sigma_0^2}\boldsymbol{u}_n'\left(\frac{\partial V_{n,0}}{\partial\tau}V_{n,0}\right)^s\boldsymbol{u}_n - 2\frac{\partial\log|V_{n,0}|}{\partial\tau}\right)\right]$$

$$= -\frac{2}{\sigma_0^4}\,\mathbb{E}_0\left[X'V_{n,0}'\boldsymbol{\epsilon}_n\boldsymbol{u}_n'V_{n,0}'\left(V_{n,0}^{-1}\frac{\partial V_{n,0}}{\partial\tau}\right)^s V_{n,0}\boldsymbol{u}_n\right] = -\frac{2}{\sigma_0^4}\,\mathbb{E}_0\,X'V_{n,0}'\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'A_0^s\boldsymbol{\epsilon}_n$$

$$= -\frac{2}{\sigma_0^4}\,\mathbb{E}_0\,X'V_{n,0}'\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'A_0^s\boldsymbol{\epsilon}_n = -\frac{2\mu_3}{\sigma_0^4}X'V_n'(\boldsymbol{a}_0)$$

$$n\Omega_n(3,1) = \mathbb{E}_0\left[-\frac{2}{\sigma_0^2}X'V_{n,0}'\boldsymbol{\epsilon}_n\left(\frac{2}{\sigma_0^2}\boldsymbol{\epsilon}_n'V_{n,0}\frac{\partial M_{n,0}}{\partial\rho}\boldsymbol{y} - 2\frac{\partial\log|M_{n,0}|}{\partial\rho}\right)\right]$$

$$= -\frac{4}{\sigma_0^4}X'V_{n,0}'\,\mathbb{E}_0\left(\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'V_nB_0(X\boldsymbol{\beta}_0+\boldsymbol{u}_n)\right) - \frac{4}{\sigma_0^2}X'V_{n,0}'V_{n,0}B_0X\boldsymbol{\beta}_0 - \frac{4}{\sigma_0^4}X'V_{n,0}'\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'\bar{B}_0\boldsymbol{\epsilon}_n$$

Using Lemma S.2

$$= -\frac{4}{\sigma_0^2}X'V_{n,0}'V_{n,0}B_0X\boldsymbol{\beta}_0 - \frac{4\mu_3}{\sigma_0^4}X'V_{n,0}'\,\boldsymbol{b}_0$$

$$n\Omega_n(4,1) = \mathbb{E}_0\left[-\frac{2}{\sigma_0^2}X'V_{n,0}'\boldsymbol{\epsilon}_n\left(\frac{n}{\sigma_0^2} - \frac{1}{\sigma_0^4}\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}\right)\right] = \frac{2}{\sigma_0^6}\,\mathbb{E}_0\,X'V_{n,0}'\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n = \frac{2\mu_3}{\sigma_0^6}X'V_{n,0}'\imath_n$$

$$n\Omega_n(2,2) = \mathbb{E}_0\left[\frac{1}{\sigma_0^2}\boldsymbol{\epsilon}_n'A_0^s\boldsymbol{\epsilon}_n - 2\frac{\partial\log|V_{n,0}|}{\partial\tau}\right]^2 = \frac{\mathbb{E}_0\left[\boldsymbol{\epsilon}_n'A_0^s\boldsymbol{\epsilon}_n\right]^2}{\sigma_0^4} + 4\left(\mathrm{tr}\left(V_n^{-1}\frac{\partial V_{n,0}}{\partial\tau}\right)\right)^2 - 4\mathrm{tr}\left(V_n^{-1}\frac{\partial V_{n,0}}{\partial\tau}\right)\mathrm{tr}(A_0^s)$$

$$= \frac{1}{\sigma_0^4}\boldsymbol{a}_0'\boldsymbol{a}_0\kappa_4 + \sigma_0^4\left[\mathrm{tr}(A_0^s)\right]^2 + 2\sigma_0^4\mathrm{tr}[(A_0^s)^2] - 4\mathrm{tr}(A_0^s)\mathrm{tr}(A_0) + 4\mathrm{tr}(A_0)\mathrm{tr}(A_0)$$

$$= 2\mathrm{tr}[(A_0^s)^2] + \frac{1}{\sigma_0^4}\boldsymbol{a}_0'\boldsymbol{a}_0\kappa_4$$

$$n\Omega_n(2,3) = \mathbb{E}_0\left[\frac{2}{\sigma_0^4}\boldsymbol{\epsilon}_n'V_{n,0}'\frac{\partial M_{n,0}}{\partial\rho}\boldsymbol{y}\boldsymbol{u}_n'(\frac{\partial V_{n,0}}{\partial\tau}V_{n,0})^s\boldsymbol{u}_n - \frac{2}{\sigma_0^2}\frac{\partial\log|M_{n,0}|}{\partial\rho}\boldsymbol{\epsilon}_n'A_0^s\boldsymbol{\epsilon}_n - \frac{4}{\sigma_0^2}\frac{\partial\log|V_{n,0}|}{\partial\tau}\boldsymbol{\epsilon}_n'V_n'\frac{\partial M_{n,0}}{\partial\rho}\boldsymbol{y}\right.$$

$$\left. + 4\frac{\partial\log|M_{n,0}|}{\partial\rho}\frac{\partial\log|V_{n,0}|}{\partial\tau}\right]$$

$$= \mathbb{E}_0\left[\frac{2}{\sigma_0^4}\boldsymbol{\epsilon}_n'V_n'B_0(X\boldsymbol{\beta}_0+\boldsymbol{u}_n)\boldsymbol{\epsilon}_n'A_0^s\boldsymbol{\epsilon}_n\right] - 2\frac{\partial\log|M_{n,0}|}{\partial\rho}\mathrm{tr}(A_0^s)$$

$$- \frac{4}{\sigma_0^2}\frac{\partial log|V_{n,0}|}{\partial\tau}\,\mathbb{E}_0\left[\boldsymbol{\epsilon}_n'V_nB_0(X\boldsymbol{\beta}_0+\boldsymbol{u}_n)\right] + 4\mathrm{tr}(\bar{B}_0)\mathrm{tr}(A_0)$$

$$= \frac{2}{\sigma_0^4}\left[\mathbb{E}_0\left(\boldsymbol{\epsilon}_n'A_0^s\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n\right)V_nB_0X\boldsymbol{\beta}_0 + \boldsymbol{b}_0'\boldsymbol{a}_0\kappa_4 + \sigma_0^4\mathrm{tr}(A_0^s)\mathrm{tr}(\bar{B}_0) + \sigma_0^4\mathrm{tr}(A_0\bar{B}_0^s)\right]$$

$$- 2\mathrm{tr}(A_0^s)\mathrm{tr}(\bar{B}_0) - \frac{4}{\sigma_0^2}\mathrm{tr}(A_0)\sigma_0^2\mathrm{tr}(\bar{B}_0) + 4\mathrm{tr}(A_0)\mathrm{tr}(\bar{B}_0)$$

$$= 2\mathrm{tr}(A_0\bar{B}_0^s) + \frac{2\mu_3}{\sigma_0^2}(\boldsymbol{a}_0)'V_nB_0X\boldsymbol{\beta}_0 + \frac{2\kappa_4}{\sigma_0^4}\boldsymbol{b}_0'\boldsymbol{a}_0$$

$$
\begin{aligned}
n\Omega_n(2,4) &= -\mathbb{E}_0\left[\frac{1}{\sigma_0^2}\boldsymbol{\epsilon}_n' A_0^s \boldsymbol{\epsilon}_n - \frac{\partial \log|V_{n,0}|}{\partial \tau}\right]\left(\frac{1}{\sigma_0^4}\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n - \frac{n}{\sigma_0^2}\right) \\
&= -\frac{1}{(\sigma_0^2)^3}\,\mathbb{E}_0\left(\boldsymbol{\epsilon}_n' A_0^s \boldsymbol{\epsilon}_n \boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n\right) + \frac{2}{\sigma_0^4}\frac{\partial \log|V_{n,0}|}{\partial \tau}\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n + \frac{n}{(\sigma_0^2)^2}\,\mathbb{E}_0(\boldsymbol{\epsilon}_n' A_0^s \boldsymbol{\epsilon}_n) - \frac{2n}{\sigma_0^2}\,\mathbb{E}_0\,\frac{\partial \log|V_{n,0}|}{\partial \tau} \\
&= -\frac{1}{(\sigma_0^2)^3}\left[\imath'\boldsymbol{a}_0\kappa_4 + n(\sigma_0^2)^2\mathrm{tr}(A_0^s) + 2(\sigma_0^2)^2\mathrm{tr}(A_0^s)\right] \\
&\quad + \frac{2}{(\sigma_0^2)^2}\mathrm{tr}(A_0)n\sigma_0^2 + \frac{n}{\sigma_0^2}\mathrm{tr}(A_0^s) - \frac{2n}{\sigma_0^2}\mathrm{tr}(A_0) \\
&= -\frac{2}{\sigma_0^2}\mathrm{tr}(A_0^s) - \frac{\kappa_4}{(\sigma_0^2)^3}\boldsymbol{a}_0'\imath\kappa_4
\end{aligned}
$$

$$
\begin{aligned}
n\Omega_n(3,3) &= \mathbb{E}_0\left[\frac{2}{\sigma_0^2}\boldsymbol{\epsilon}_n' V_{n,0}' B_0(X\boldsymbol{\beta}_0 + \boldsymbol{u}_n) - 2\frac{\partial \log|M_{n,0}|}{\partial \rho}\right]^2 \\
&= \mathbb{E}_0\left[\frac{4}{\sigma_0^4}\left(\boldsymbol{\epsilon}_n' V_{n,0}' B_0(X\boldsymbol{\beta}_0 + \boldsymbol{u}_n)\right)^2 + 4\left(\frac{\partial \log|M_{n,0}|}{\partial \rho}\right)^2 - 4\frac{\partial \log|M_{n,0}|}{\partial \rho}\frac{2}{\sigma_0^2}\boldsymbol{\epsilon}_n' V_{n,0}' B_0(X\boldsymbol{\beta}_0 + \boldsymbol{u}_n)\right] \\
&= \frac{4}{\sigma_0^2}\mathrm{tr}\left(V_{n,0}B_0X\boldsymbol{\beta}_0\boldsymbol{\beta}_0'X'B_0'V_{n,0}'\right) + \frac{4}{\sigma_0^4}\left[\boldsymbol{b}_0'\boldsymbol{b}_0(\mathbb{E}\,\epsilon_{ni} - 3\sigma_0^4) + \sigma_0^4\mathrm{tr}(\bar{B}_0)^2 + \sigma_0^4\mathrm{tr}(\bar{B}_0\bar{B}_0^s)\right] \\
&\quad + \frac{8}{\sigma_0^2}\,\mathbb{E}_0(\boldsymbol{\epsilon}_n'\bar{B}_0\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n')V_{n,0}B_0X\boldsymbol{\beta}_0 + 4\mathrm{tr}\left(M_{n,0}^{-1}\frac{\partial M_{n,0}}{\partial \rho}\right)^2 - 8\mathrm{tr}\left(M_{n,0}^{-1}\frac{\partial M_{n,0}}{\partial \rho}\right)\mathrm{tr}\left(\bar{B}_0\right) \\
&= \frac{4}{\sigma_0^2}\boldsymbol{\beta}_0'X'B_0'V_{n,0}'V_{n,0}B_0X\boldsymbol{\beta}_0 + 2\mathrm{tr}((\bar{B}_0^s)^2) + \frac{4\kappa_4}{\sigma_0^4}\boldsymbol{b}_0'\boldsymbol{b}_0 + \frac{8}{\sigma_0^4}\mu_3\boldsymbol{b}_0'V_nB_0X\boldsymbol{\beta}_0
\end{aligned}
$$

$$
\begin{aligned}
n\Omega_n(3,4) &= \mathbb{E}_0\left[\frac{2}{\sigma_0^2}\boldsymbol{\epsilon}_n' V_{n,0}' B_0(X\boldsymbol{\beta}_0 + \boldsymbol{u}_n) - 2\frac{\partial \log|M_{n,0}|}{\partial \rho}\right]\left(\frac{n}{\sigma_0^2} - \frac{\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n}{(\sigma_0^2)^2}\right) \\
&= \frac{2n}{(\sigma_0^2)^2}\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n' V_{n,0}' B_0(X\boldsymbol{\beta}_0 + \boldsymbol{u}_n) - \frac{2\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n' V_{n,0}' B_0(X\boldsymbol{\beta}_0 + \boldsymbol{u}_n)\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n}{(\sigma_0^2)^3} + 2\frac{\partial \log|M_{n,0}|}{\partial \rho}\left(\mathbb{E}_0\,\frac{\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n}{(\sigma_0^2)^2} - \frac{n}{\sigma_0^2}\right) \\
&= \frac{2n}{(\sigma_0^2)^2}\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n'\bar{B}_0\boldsymbol{\epsilon}_n - \frac{2n}{\sigma_0^2}\mathrm{tr}(\bar{B}_0) - \frac{2}{(\sigma_0^2)^3}\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'V_nB_0X\boldsymbol{\beta}_0 - \frac{2}{(\sigma_0^2)^3}\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n'\bar{B}_0\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n + \frac{2n}{\sigma_0^2}\mathrm{tr}(\bar{B}_0) \\
&= -\frac{4}{\sigma_0^2}\mathrm{tr}(\bar{B}_0) - \frac{2\mu_3}{\sigma_0^6}\imath_n'V_n'B_0X\boldsymbol{\beta}_0 - \frac{2\kappa_4}{\sigma_0^6}\mathrm{tr}(B_0)
\end{aligned}
$$

$$
\begin{aligned}
n\Omega_n(4,4) &= \mathbb{E}_0\left(\frac{n}{\sigma_0^2} - \frac{\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n}{(\sigma_0^2)^2}\right)^2 = \frac{1}{(\sigma_0^4)^2}\,\mathbb{E}_0\,\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n - 2\frac{n}{(\sigma_0^2)^3}\,\mathbb{E}_0(\boldsymbol{\epsilon}_n'\boldsymbol{\epsilon}_n) + \frac{n^2}{(\sigma_0^2)^2} \\
&= \frac{2n}{(\sigma_0^2)^2} + \frac{n\kappa_4}{(\sigma_0^2)^4}
\end{aligned}
$$

□

S.4.5 Proof of Theorem 2

Proof. **1.** We apply Taylor expansion of $\frac{\partial Q_n}{\partial \boldsymbol{\theta}}\left(\hat{\boldsymbol{\theta}}_n\right)$ around $\boldsymbol{\theta}_0$, then we get

$$
\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = -\left[\frac{1}{n}\frac{\partial^2 Q_n\left(\tilde{\boldsymbol{\theta}}_n\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right]^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\theta}}
$$

where $\tilde{\boldsymbol{\theta}}_n$ is a "point" between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_0$.

2. We show that

$$\frac{1}{n}\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = C_n(\boldsymbol{\theta}_0) + o_P(1) \tag{S.15}$$

where $C_n(\boldsymbol{\theta}_0) = \frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$. To prove (S.15) we need to write the second derivative as

$$\frac{1}{n}\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] + C_n(\boldsymbol{\theta}_0)$$

and show that

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = o_P(1) \tag{S.16}$$

for all components.

The proof of equation (S.16) is given in Lemma S.10.

3. Next, we note that

$$\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + o_p(1).$$

This in fact follows from Lemma S.3, because all second derivatives can be written as linear-quadratic functions of ϵ_n with RCB matrices.

By taking the expectations in the right-hand-side of the above equation, and denoting by

$$m_0 = \boldsymbol{\beta}_0' X' B_0' V_n' V_n B_0 X \boldsymbol{\beta}_0,$$

with $B_\rho = \frac{\partial M_n}{\partial \rho} M_n^{-1}$ and $B_0 = B_{\rho_0}$, we get:

$$C_n(\boldsymbol{\theta}_0) = \begin{bmatrix} \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}'\partial\tau} & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}'\partial\rho} & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\sigma^2} \\ \cdot & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\tau^2} & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\tau\partial\rho} & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\tau\partial\sigma^2} \\ \cdot & \cdot & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\rho^2} & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\rho\partial\sigma^2} \\ \cdot & \cdot & \cdot & \mathbb{E}\,\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{(\partial\sigma^2)^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n}\frac{2}{\sigma_0^2}X'V_n'VX & \mathbf{0}' & -\boldsymbol{\beta}_0'\frac{2X'V_n'V_n\frac{\partial M_n}{\partial\rho}M_n^{-1}X}{n\sigma_0^2} & \mathbf{0}' \\ \cdot & \frac{2}{n}\mathrm{tr}\left(A_{\tau_0}A_{\tau_0}^s\right) & \frac{2}{n}\mathrm{tr}\left(V_nB_{\rho_0}V_n^{-1}A_{\tau_0}^s\right) & -\frac{1}{n\sigma_0^2}\mathrm{tr}\left(A_{\tau_0}^s\right) \\ \cdot & \cdot & \frac{1}{n}\left[\mathrm{tr}\left(\bar{B}_0^s\bar{B}_0^s\right) + \frac{2m_0}{\sigma_0^2}\right] & -\frac{2}{n\sigma_0^2}\mathrm{tr}\left(B_{\rho_0}\right) \\ \cdot & \cdot & \cdot & \frac{1}{\sigma_0^4} \end{bmatrix}$$

where $A_\tau = \frac{\partial V_n}{\partial\tau}V_n^{-1}$ and where $\bar{B}_0 = V_nB_{\rho_0}V_n^{-1}$, while for any matrix M, $M^s = M + M'$, as above.

4. It now remains to prove invertibility of $\lim C_n(\boldsymbol{\theta}_0)$. Following Debarsy et al. (2015a), we prove that the equation

$$\lim_n C_n(\boldsymbol{\theta}_0)\boldsymbol{\eta} = 0$$

has unique solution $\boldsymbol{\eta} = \mathbf{0}$. Consider the linear system in 4 blocks, each one corresponding of one of the block rows of $C_n(\boldsymbol{\theta}_0)$. Then from the first block of equations, we get the identity:

$$\boldsymbol{\eta}_1 = (X'V_n'V_nX)^{-1}X'V_n'V_nB_{\rho 0}X\boldsymbol{\beta}_0\eta_3.$$

By replacing $\boldsymbol{\eta}_1$ in the equation of the third block, we get

$$\lim_n \frac{1}{n}\left[\frac{1}{\sigma_0^2}\boldsymbol{\beta}_0'X'B_{\rho 0}'V_n'H_nV_nB_{\rho 0}X\boldsymbol{\beta}_0 + \mathrm{tr}(\bar{B}_0^s\bar{B}_0^s)\right]\eta_3 + \frac{2}{n}\mathrm{tr}(\bar{B}_0A_{\tau 0}^s)\eta_2 - \frac{2}{n\sigma_0^2}\mathrm{tr}(\boldsymbol{B}_{\rho 0})\eta_4 = 0$$

where $H_n = I_n - V_nX\left(X'V_n'V_nX\right)^{-1}X'V_n'$.

Then, from the fourth block,

$$\eta_4 = \frac{2\sigma_0^2}{n}\left(\mathrm{tr}(B_{\rho 0})\eta_3 + \mathrm{tr}(A_{\tau 0})\eta_2\right) = \frac{\sigma_0^2}{n}\left(\mathrm{tr}(\bar{B}_0^s)\eta_3 + \mathrm{tr}(A_{\tau 0}^s)\eta_2\right). \tag{S.17}$$

We replace this to η_4 in the second block equation:

$$\begin{aligned}&\frac{2\eta_2}{n}\mathrm{tr}(A_{\tau 0}A_{\tau 0}^s) + \frac{2\eta_3}{n}\mathrm{tr}(\bar{B}_0A_{\tau 0}^s) - \frac{\eta_4}{n\sigma_0^2}\mathrm{tr}(A_{\tau 0}^s)\\ &= \eta_2\left[\frac{\mathrm{tr}(A_{\tau 0}^sA_{\tau 0}^s)}{n} - \left(\frac{\mathrm{tr}(A_{\tau 0}^s)}{n}\right)^2\right] + \eta_3\left[\frac{\mathrm{tr}(\bar{B}_0^sA_{\tau 0}^s)}{n} - \frac{\mathrm{tr}(\bar{B}_0^s)}{n}\frac{\mathrm{tr}(A_{\tau 0}^s)}{n}\right]\end{aligned}$$

Let us denote by $\mathcal{V}_A = \frac{\mathrm{tr}(A_{\tau 0}^sA_{\tau 0}^s)}{n} - \left(\frac{\mathrm{tr}(A_{\tau 0}^s)}{n}\right)^2$ the variance of the eigenvalues of $A_{\tau 0}$ and by $\mathcal{C}$ the *pseudo-covariance*[2] $\mathcal{C} = \frac{\mathrm{tr}(\bar{B}_0^sA_{\tau 0}^s)}{n} - \frac{\mathrm{tr}(\bar{B}_0^s)}{n}\frac{\mathrm{tr}(A_{\tau 0}^s)}{n}$.

Thus, we find

$$\eta_2 = -\frac{\mathcal{C}}{\mathcal{V}_A}\eta_3$$

We then get, from (S.17),

$$\eta_4 = \sigma_0^2\left[\frac{\mathrm{tr}\left(\bar{B}_0^s\right)}{n} - \frac{\mathcal{C}}{\mathcal{V}_A}\frac{\mathrm{tr}\left(A_{\tau 0}^s\right)}{n}\right]\eta_3.$$

Note that in analogy with the OLS estimates in a simple linear model, if we (improperly) consider $\mathcal{C}$ as a real covariance between eigenvalues, we might write $\eta_2 = -b\eta_3$ and $\eta_4 = \sigma_0^2a\eta_3$,

[2]This is a proper covariance, between the vectors of the eigenvalues of $A_{\tau 0}$ and of $\bar{B}_0$, (only) if these tow matrices commute. In general, because of Von Neumann's inequality $|\mathrm{tr}(AB)| \leq \sum_i \lambda_i(A)\lambda_i(B)$, we have that $\mathcal{C} \leq n^{-1}\sum_i \lambda_i(A_{\tau 0}^s)\lambda_i(\bar{B}_0^s) - n^{-2}\sum_i \lambda_i(A_{\tau 0}^s)\sum_i \lambda_i(\bar{B}_0^s)$

where b and a are analog to the OLS coefficients in a regression of the eigenvalues of $\bar{B}_0^s$ on the eigenvalues of A_{τ_0}.

Now, going back to equation 3 of the system, and replacing η_2 and η_4,

$$\frac{1}{n}\left[\frac{\boldsymbol{\beta}_0' X' B_{\rho_0}' V_n' H_n V_n X B_{\rho_0} \boldsymbol{\beta}_0}{\sigma_0^2} + \operatorname{tr}\left(\bar{B}_0^s \bar{B}_0^s\right) - \operatorname{tr}(\bar{B}_0^s A_{\tau_0}^s) b - \operatorname{tr}(\bar{B}_0^s) a\right] \eta_3$$
$$= \left[\frac{1}{n}\boldsymbol{\beta}_0' X' B_{\rho_0}' V_n' H_n V_n B_{\rho_0} X \boldsymbol{\beta}_0 + \frac{1}{\mathcal{V}_A}\left(\mathcal{V}_A \mathcal{V}_{\bar{B}} - \mathcal{C}^2\right)\right] \eta_3 = 0$$

where, in analogy to $\mathcal{V}_A$, we define $\mathcal{V}_{\bar{B}} = \frac{\operatorname{tr}(\bar{B}_0^s \bar{B}_0^s)}{n} - \left(\frac{\operatorname{tr}(\bar{B}_0^s)}{n}\right)^2$.

Because of $\mathcal{C}^2 \leq \operatorname{cov}\left(\lambda(A_{\tau_0}^s), \lambda(\bar{B}_0^s)\right)$, we readily find that the second term is ≥ 0 by Cauchy-Schwartz inequality. Further, $\boldsymbol{\beta}_0' X' B_{\rho_0}' V_n' H_n V_n X B_{\rho_0} \boldsymbol{\beta}_0 \geq 0$. Then, the equation

$$\lim_n \left[\frac{1}{n}\boldsymbol{\beta}_0' X' B_{\rho_0}' V_n' H_n V_n X B_{\rho_0} \boldsymbol{\beta}_0 + \frac{1}{\mathcal{V}_A}\left(\mathcal{V}_A \mathcal{V}_{\bar{B}} - \mathcal{C}^2\right)\right] \eta_3 = 0$$

has solution $\eta_3 = 0$ if

$$\lim_n \left[\frac{1}{n}\boldsymbol{\beta}_0' X' B_{\rho_0}' V_n' H_n V_n X B_{\rho_0} \boldsymbol{\beta}_0 + \frac{1}{\mathcal{V}_A}\left(\mathcal{V}_A \mathcal{V}_{\bar{B}} - \mathcal{C}^2\right)\right] > 0.$$

A sufficient condition is clearly that either $\lim_n \frac{1}{n}\boldsymbol{\beta}_0' B_{\rho_0}' V_n' H_{VX} V_n B_{\rho_0} \boldsymbol{\beta}_0 > 0$, or both $\lim_n \mathcal{V}_A > 0$ and $\left(\mathcal{V}_A \mathcal{V}_{\bar{B}} - \mathcal{C}^2\right) > 0$.

If $V_n = I$, the row and column associated to τ in $C_n(\boldsymbol{\theta}_0)$ disappear and by repeating the above argument one gets the simplified version of Assumption 8.2: $\mathcal{V}_{\bar{B}} > 0$. □

S.4.6 Proof of Proposition 1

The proof is an immediate consequence of the continuity of $C_n(\boldsymbol{\theta})$ and $\Delta_n(\boldsymbol{\theta})$) in a neighborhood of $\boldsymbol{\theta}_0$. The existence of the inverse of $\lim_n C_n(\boldsymbol{\theta}_0) = C_0$ also implies that $C_n(\hat{\boldsymbol{\theta}}_n)$ is nonsingular for n large enough. Let in fact $l = \min_{i \leq n} \sigma(C_0) > 0$. We then have, for all $\varepsilon > 0$ and $n \geq n_\varepsilon$,

$$l + \varepsilon > \min_{i \leq n} \sigma(C_n(\boldsymbol{\theta}_0)) > l - \varepsilon.$$

Suppose that there is a subsequence $\{n_j\}$ such that $\min_{i \leq n} \sigma(C_{n_j}(\boldsymbol{\theta}_{n_j})) = 0$ for all j. Then, even for $n_j > n_\varepsilon$,

$$l + \varepsilon > \min_{i \leq n} \sigma(C_{n_j}(\boldsymbol{\theta}_0)) - \min_{i \leq n} \sigma(C_{n_j}(\boldsymbol{\theta}_{n_j})) > l - \varepsilon$$

implying

$$\lim_n \| \min_{i \leq n} \sigma(C_{n_j}(\boldsymbol{\theta}_0)) - \min_{i \leq n} \sigma(C_{n_j}(\boldsymbol{\theta}_{n_j})) \| = l > 0$$

with probability one, which contradicts $|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| = o_p(1)$.

S.4.7 Proof of (S.16)

Lemma S.10. Under Assumptions 1–6, then

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]=o_P(1).$$

Proof. By computing the second derivatives of the loglikelihood, we get

$$\begin{aligned}
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}&=\frac{2}{\sigma^2}X'V_n'V_nX\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\rho}&=-\frac{2}{\sigma^2}X'V_n'V_n\frac{\partial M_n}{\partial\rho}\boldsymbol{y}\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\tau}&=-\frac{2}{\sigma^2}X'\left(\frac{\partial V_n'}{\partial\tau}V_n\right)^s(M_n\boldsymbol{y}-X\boldsymbol{\beta})\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\sigma^2}&=\frac{2}{\sigma^4}X'V_n'V_n(M_n\boldsymbol{y}-X\boldsymbol{\beta})\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\rho^2}&=\frac{2}{\sigma^2}\left[\boldsymbol{y}'\frac{\partial M_n'}{\partial\rho}V_n'V_n\frac{\partial M_n}{\partial\rho}\boldsymbol{y}+(M_n\boldsymbol{y}-X\boldsymbol{\beta})'V_n'V_n\frac{\partial^2 M_n}{\partial\rho^2}\boldsymbol{y}\right]-2\frac{\partial^2}{\partial\rho^2}\log|M_n|\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\rho\partial\tau}&=\frac{2}{\sigma^2}\boldsymbol{y}'\frac{\partial M_n'}{\partial\rho}\left(\frac{\partial V_n'}{\partial\tau}V_n\right)^s(M_n\boldsymbol{y}-X\boldsymbol{\beta})\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\rho\partial\sigma^2}&=-\frac{2}{\sigma^2}\boldsymbol{y}'\frac{\partial M_n'}{\partial\rho}V_n'V_n(M_n\boldsymbol{y}-X\boldsymbol{\beta})\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\tau^2}&=\frac{2}{\sigma^2}(M_n\boldsymbol{y}-X\boldsymbol{\beta})'\left[\frac{\partial^2 V_n'}{\partial\tau^2}V_n+\frac{\partial V_n'}{\partial\tau}\frac{\partial V_n}{\partial\tau}\right](M_n\boldsymbol{y}-X\boldsymbol{\beta})-2\frac{\partial^2}{\partial\tau^2}\log|V_n|\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial\tau\partial\sigma^2}&=-\frac{1}{(\sigma^2)^2}(M_n\boldsymbol{y}-X\boldsymbol{\beta})'\left(\frac{\partial V_n'}{\partial\tau}V_n\right)^s(M_n\boldsymbol{y}-X\boldsymbol{\beta})\\
\frac{\partial^2 Q_n(\boldsymbol{\theta})}{(\partial\sigma^2)^2}&=\frac{2}{(\sigma^2)^3}(M_n\boldsymbol{y}-X\boldsymbol{\beta})'V_n'V_n(M_n\boldsymbol{y}-X\boldsymbol{\beta})-\frac{n}{(\sigma^2)^2}
\end{aligned}$$

Now, we have to show that all these terms are $o_P(1)$. The proof is similar for all the components having derivatives of the same order, then we tackle the problem only for terms with different derivatives.

We begin with the second derivatives with respect to $\boldsymbol{\beta}$, we have

$$\frac{1}{n}\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}-\frac{1}{n}\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}=\frac{1}{n}\left[A_{11}+A_{12}\right]$$

where

$$\begin{aligned}
A_{11}&=\frac{2}{\tilde{\sigma}^2}X'\left(V_n(\tilde{\tau})'V_n(\tilde{\tau})-V_n(\tau_0)'V_n(\tau_0)\right)X\\
A_{12}&=2\left(\frac{1}{\tilde{\sigma}^2}-\frac{1}{\sigma_0^2}\right)X'V_n(\tau_0)'V_n(\tau_0)X
\end{aligned}$$

by Assumptions 3 and 4, we can apply Lemma S.8 and Lemma S.3 that the above difference is an $o_P(1)$.

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\tau}-\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\tau}\right]=A_{21}+A_{22}+A_{23}$$

where, setting $\boldsymbol{u}_0=M_n(\rho_0)\boldsymbol{y}-X\boldsymbol{\beta}_0$ and $\tilde{\boldsymbol{u}}=M_n(\tilde{\rho})\boldsymbol{y}-X\tilde{\boldsymbol{\beta}}$, we have

$$\begin{aligned}
A_{21}&=-\frac{2}{\tilde{\sigma}^2}X'\left(\frac{\partial V_n'(\tilde{\tau})}{\partial\tau}V_n(\tilde{\tau})\right)^s(\tilde{\boldsymbol{u}}-\boldsymbol{u}_0)\\
A_{22}&=-\frac{2}{\tilde{\sigma}^2}X'\left[\left(\frac{\partial V_n'(\tilde{\tau})}{\partial\tau}V_n(\tilde{\tau})\right)^s-\left(\frac{\partial V_n'(\tau_0)}{\partial\tau}V_n(\tau_0)\right)^s\right]\boldsymbol{u}_0\\
A_{23}&=2X'\left(\frac{\partial V_n'(\tau_0)}{\partial\tau}V_n(\tau_0)\right)^s\boldsymbol{u}_0\left(\frac{1}{\sigma_0^2}-\frac{1}{\tilde{\sigma}^2}\right)
\end{aligned}$$

For the second derivatives we have

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\rho}-\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\rho}\right]=\frac{1}{n}(A_{31}+A_{32}+A_{33})$$

where

$$\begin{aligned}
A_{31}&=-\frac{2}{\tilde{\sigma}^2}X'V_n'(\tilde{\tau})V_n(\tilde{\tau})\left[\frac{\partial M_n(\tilde{\rho})}{\partial\rho}-\frac{\partial M_n(\rho_0)}{\partial\rho}\right]\boldsymbol{y}\\
A_{32}&=-\frac{2}{\tilde{\sigma}^2}X'\left[V_n'(\tilde{\tau})V_n(\tilde{\tau})-V_n'(\tau_0)V_n(\tau_0)\right]\frac{\partial M_n(\rho_0)}{\partial\rho}\boldsymbol{y}\\
A_{33}&=-X'V_n'(\tau_0)V_n(\tau_0)\frac{\partial M_n(\rho_0)}{\partial\rho}\boldsymbol{y}\left(\frac{1}{\tilde{\sigma}^2}-\frac{1}{\sigma_0^2}\right)
\end{aligned}$$

The next one

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\beta}\partial\sigma^2}-\frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}\partial\sigma^2}\right]=A_{41}+A_{42}+A_{43}$$

where

$$\begin{aligned}
A_{41}&=-\frac{2}{(\tilde{\sigma}^2)^2}X'V_n'(\tilde{\tau})V_n(\tilde{\tau})(\tilde{\boldsymbol{u}}-\boldsymbol{u}_0)\\
A_{42}&=-\frac{2}{(\tilde{\sigma}^2)^2}X'\left[V_n'(\tilde{\tau})V_n(\tilde{\tau})-V_n'(\tau_0)V_n(\tau_0)\right](\tilde{\boldsymbol{u}}-\boldsymbol{u}_0)\\
A_{43}&=-2X'V_n'(\tau_0)V_n(\tau_0)\boldsymbol{u}_0\left[\frac{1}{(\tilde{\sigma}^2)}-\frac{1}{(\sigma_0^2)^2}\right]
\end{aligned}$$

Second derivatives

$$\frac{1}{n}\left(\frac{\partial^2 Q\left(\tilde{\boldsymbol{\theta}}_n\right)}{\partial\tau^2}-\frac{\partial^2 Q(\boldsymbol{\theta}_0)}{\partial\tau^2}\right)=\frac{1}{n}(A_{51}+A_{52}+A_{53}+A_{54})$$

where

$$A_{51} = -2\frac{\partial^2}{\partial\tau^2}\left[\log|V_n(\tilde{\tau})| - \log|V_n(\tau_0)|\right]$$

$$A_{52} = 2\,\tilde{\boldsymbol{u}}'\left[\frac{\partial^2 V_n'(\tilde{\tau})}{\partial\tau^2}V_n(\tilde{\tau}) - \frac{\partial^2 V_n'(\tau_0)}{\partial\tau^2}V_n(\tau_0) + \frac{\partial V_n'(\tilde{\tau})}{\partial\tau}\frac{\partial V_n(\tilde{\tau})}{\partial\tau} - \frac{\partial V_n'(\tau_0)}{\partial\tau}\frac{\partial V_n(\tau_0)}{\partial\tau}\right]\tilde{\boldsymbol{u}}$$

$$A_{53} = \frac{2}{\tilde{\sigma}^2}\,(\tilde{\boldsymbol{u}} + \boldsymbol{u}_0)'\left[\frac{\partial^2 V_n'(\tau_0)}{\partial\tau^2}V_n(\tau_0) + \frac{\partial V_n'(\tau_0)}{\partial\tau}\frac{\partial V_n(\tau_0)}{\partial\tau}\right](\tilde{\boldsymbol{u}} - \boldsymbol{u}_0)$$

$$A_{54} = 2\boldsymbol{u}_0\left[\frac{\partial^2 V_n'(\tau_0)}{\partial\tau^2}V_n(\tau_0) + \frac{\partial V_n'(\tau_0)}{\partial\tau}\frac{\partial V_n(\tau_0)}{\partial\tau}\right]\boldsymbol{u}_0\left(\frac{1}{\tilde{\sigma}^2} - \frac{1}{\sigma_0^2}\right)$$

Next one

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\tau}\partial\rho} - \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\tau}\partial\rho}\right] = \frac{1}{n}\left(A_{61} + A_{62} + A_{63} + A_{64}\right)$$

where

$$A_{61} = \frac{2}{\tilde{\sigma}^2}\,\tilde{\boldsymbol{u}}'\left(\frac{\partial V_n'(\tilde{\tau})}{\partial\tau}V_n\left(\tilde{\tau}\right)\right)^s\left[\frac{\partial M_n(\tilde{\rho})}{\partial\rho} - \frac{\partial M_n(\rho_0)}{\partial\rho}\right]\boldsymbol{y}$$

$$A_{62} = \frac{2}{\tilde{\sigma}^2}\,\tilde{\boldsymbol{u}}'\left[\left(\frac{\partial V_n'(\tilde{\tau})}{\partial\tau}V_n\left(\tilde{\tau}\right)\right)^s - \left(\frac{\partial V_n'(\tau_0)}{\partial\tau}V_n\left(\tau_0\right)\right)^s\right]\frac{\partial M_n(\rho_0)}{\partial\rho}\boldsymbol{y}$$

$$A_{63} = \frac{2}{\tilde{\sigma}^2}\,(\tilde{\boldsymbol{u}} - \boldsymbol{u}_0)'\left(\frac{\partial V_n'(\tau_0)}{\partial\tau}V_n\left(\tau_0\right)\right)^s\frac{\partial M_n(\rho_0)}{\partial\rho}\boldsymbol{y}$$

$$A_{64} = \frac{\sigma_0^2}{n}\,\frac{\partial^2 Q_n\left(\boldsymbol{\theta}_0\right)}{\partial\tau\partial\rho}\left(\frac{1}{\tilde{\sigma}^2} - \frac{1}{\sigma_0^2}\right)$$

Next one

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\tau}\partial\sigma^2} - \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\tau}\partial\sigma^2}\right] = \frac{1}{n}\left(A_{71} + A_{72} + A_{73}\right)$$

where

$$A_{71} = -\frac{1}{\left(\tilde{\sigma}^2\right)^2}\,\tilde{\boldsymbol{u}}'\left[\left(\frac{\partial V_n'(\tilde{\tau})}{\partial\tau}V_n\left(\tilde{\tau}\right)\right)^s - \left(\frac{\partial V_n'(\tau_0)}{\partial\tau}V_n\left(\tau_0\right)\right)^s\right]\tilde{\boldsymbol{u}}$$

$$A_{72} = -\frac{1}{\left(\tilde{\sigma}^2\right)^2}\,(\tilde{\boldsymbol{u}} + \boldsymbol{u}_0)'\left(\frac{\partial V_n'(\tau_0)}{\partial\tau}V_n\left(\tau_0\right)\right)^s(\tilde{\boldsymbol{u}} + \boldsymbol{u}_0)$$

$$A_{73} = -\boldsymbol{u}_0'\left(\tilde{\boldsymbol{u}} + \boldsymbol{u}_0\right)\boldsymbol{u}_0\left(\frac{1}{\left(\tilde{\sigma}^2\right)^2} - \frac{1}{\left(\sigma_0^2\right)^2}\right)$$

Then

$$\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\rho^2} - \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\rho^2}\right] = \frac{1}{n}\sum_{i=1}^{8}A_{8i}$$

where

$$
\begin{aligned}
A_{81} &= -2\frac{\partial^2}{\partial\rho^2}\left(\log|M_n(\tilde\rho)| - \log|M_n(\rho_0)|\right)\\
A_{82} &= \frac{2}{\tilde\sigma^2}\,\tilde{\boldsymbol{u}}'V_n'(\tilde\tau)V_n(\tilde\tau)\left[\frac{\partial^2 M_n(\tilde\rho)}{\partial\rho^2} - \frac{\partial^2 M_n(\rho_0)}{\partial\rho^2}\right]\boldsymbol{y}\\
A_{83} &= \frac{2}{\tilde\sigma^2}\,\tilde{\boldsymbol{u}}'\left(V_n'(\tilde\tau)V_n(\tilde\tau) - V_n'(\tau_0)V_n(\tau_0)\right)\frac{\partial^2 M_n(\rho_0)}{\partial\rho^2}\boldsymbol{y}\\
A_{84} &= \frac{2}{\tilde\sigma^2}\,(\tilde{\boldsymbol{u}} - \boldsymbol{u}_0)'\,V_n'(\tau_0)V_n(\tau_0)\frac{\partial^2 M_n(\rho_0)}{\partial\rho^2}\boldsymbol{y}\\
A_{85} &= 2\boldsymbol{u}_0'V_n'(\tau_0)V_n(\tau_0)\frac{\partial^2 M_n(\rho_0)}{\partial\rho^2}\boldsymbol{y}\left(\frac{1}{\tilde\sigma^2} - \frac{1}{\sigma_0^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
A_{86} &= \frac{2}{\tilde\sigma^2}\,\boldsymbol{y}'\frac{\partial M_n'(\tilde\rho)}{\partial\rho}\left(V_n'(\tilde\tau)V_n(\tilde\tau) - V_n'(\tau_0)V_n(\tau_0)\right)\frac{\partial M_n(\tilde\rho)}{\partial\rho}\boldsymbol{y}\\
A_{87} &= \frac{2}{\tilde\sigma^2}\,\boldsymbol{y}'\left(\frac{\partial M_n(\tilde\rho)}{\partial\rho} + \frac{\partial M_n(\rho_0)}{\partial\rho}\right)'V_n'(\tau_0)V_n(\tau_0)\left(\frac{\partial M_n(\tilde\rho)}{\partial\rho} + \frac{\partial M_n(\rho_0)}{\partial\rho}\right)\boldsymbol{y}\\
A_{88} &= 2\boldsymbol{y}'\frac{\partial M_n'(\rho_0)}{\partial\rho}V_n'(\tau_0)V_n(\tau_0)\frac{\partial M_n(\rho_0)}{\partial\rho}\boldsymbol{y}\left(\frac{1}{\tilde\sigma^2} - \frac{1}{\sigma_0^2}\right)
\end{aligned}
$$

Next one

$$
\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{\partial\rho\partial\sigma^2} - \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{\partial\rho\partial\sigma^2}\right] = A_{91} + A_{92} + A_{93} + A_{94}
$$

where

$$
\begin{aligned}
A_{91} &= -\frac{2}{(\tilde\sigma^2)^2}\,\tilde{\boldsymbol{u}}'V_n'(\tilde\tau)V_n(\tilde\tau)\left[\frac{\partial M_n(\tilde\rho)}{\partial\rho} - \frac{\partial M_n(\rho_0)}{\partial\rho}\right]\boldsymbol{y}\\
A_{92} &= -\frac{2}{(\tilde\sigma^2)^2}\,\tilde{\boldsymbol{u}}'\left(V_n'(\tilde\tau)V_n(\tilde\tau) - V_n'(\tau_0)V_n(\tau_0)\right)\frac{\partial M_n(\rho_0)}{\partial\rho}\boldsymbol{y}\\
A_{93} &= -\frac{2}{(\tilde\sigma^2)^2}\,(\tilde{\boldsymbol{u}} - \boldsymbol{u}_0)'\,V_n'(\tau_0)V_n(\tau_0)\frac{\partial^2 M_n(\rho_0)}{\partial\rho}\boldsymbol{y}\\
A_{94} &= 2\boldsymbol{u}_0'V_n'(\tau_0)V_n(\tau_0)\frac{\partial M_n(\rho_0)}{\partial\rho}\boldsymbol{y}\left(\frac{1}{(\tilde\sigma^2)^2} - \frac{1}{(\sigma_0^2)^2}\right)
\end{aligned}
$$

Finally,

$$
\frac{1}{n}\left[\frac{\partial^2 Q_n(\tilde{\boldsymbol{\theta}})}{(\partial\sigma^2)^2} - \frac{\partial^2 Q_n(\boldsymbol{\theta}_0)}{(\partial\sigma^2)^2}\right] = A_{10,1} + A_{10,2} + A_{10,3} + A_{10,4}
$$

where

$$
\begin{aligned}
A_{10,1} &= \frac{1}{\tilde{\sigma}^2} - \frac{1}{\sigma_0^2} \\
A_{10,2} &= -\frac{1}{\left(\tilde{\sigma}^2\right)^2}\, \tilde{\boldsymbol{u}}' \left(V_n'(\tilde{\tau})V_n(\tilde{\tau}) - V_n'(\tau_0)V_n(\tau_0)\right) \tilde{\boldsymbol{u}} \\
A_{10,3} &= -\frac{1}{\left(\tilde{\sigma}^2\right)^2}\, (\tilde{\boldsymbol{u}} + \boldsymbol{u}_0)' \, V_n'(\tau_0)V_n(\tau_0)\,(\tilde{\boldsymbol{u}} - \boldsymbol{u}_0) \\
A_{10,4} &= -\boldsymbol{u}_0' V_n'(\tau_0)V_n(\tau_0)\boldsymbol{u}_0 \left(\frac{1}{\left(\tilde{\sigma}^2\right)^2} - \frac{1}{\left(\sigma_0^2\right)^2} \right)
\end{aligned}
$$

From Assumptions 1–6 and applying the lemmas in Section S.3, we have all terms $n^{-1}A_{i,j} = o_p(1)$, for all $i = 1, \ldots, 10$, $j = 1, \ldots, n_i$.

□

S.4.8 Other proofs

S.4.8.1 Proof of proposition 2

For every $\varepsilon > 0$,

$$
\begin{aligned}
\|M_n(\rho) - M_n(\rho_0)\| &= \left\| \sum_k a_k W^k (\rho^k - \rho_0^k) \right\| \\
&\leq |\rho - \rho_0| \left\| \sum_{k=1}^{\infty} a_k W^k \sum_{j=0}^{k-1} \rho^j \rho_0^{k-1-j} \right\| \\
&\leq |\rho - \rho_0| \sum_k |a_k| \|W^k\| \max_{\{\rho \leq r_G/\|W\|_1\}} |\rho|^{k-1} \\
&\leq \frac{|\rho - \rho_0|}{\gamma} \sum_k |a_k| \gamma^k < \varepsilon
\end{aligned}
$$

whenever $|\rho - \rho_0| < \varepsilon\gamma/M$, where $M = \sum_k |a_k|\gamma^k < \infty$, because the series $\sum_k |a_k u_k|$ is convergent for $|u| \leq \gamma$.

S.4.8.2 Proposition 3

We note that under the condition of Proposition 2, for all ρ we have $\|M_n(\rho)\| \leq 1 - a_0 + c\sum_k |a_k| < 1 - a_0 + c\,M$ is bounded, both in row and column sum norm.

Further, the derivative

$$
\frac{\partial M_n(\rho)}{\partial \rho} = -\sum_{k=1}^{\infty} k a_k \rho^{k-1} W^k = -W g_n(\rho W)
$$

where g_n is the matrix version of the derivative $g(x) = dG/dx$, and it is continuous if G is continuously differentiable. The same is true for $\frac{\partial^2 M_n(\rho)}{\partial \rho^2} = -W^2 g_n'(\rho W)$.

Now, we focus on the invertibility conditions, to study this problem we consider truncated form of analytical functions: we truncate the power series M_n and we denote by M_n^K the truncated series at the K-th term. As one could expect, the invertibility of M_n^K depends on the zeros of the associated scalar polynomial, of order K,

$$\zeta(x) := 1 - g(x) = (1 - a_0) - \sum_{k=1}^{K} a_k x^k.$$

By fundamental theorem of algebra, ζ has K roots, that we denote $\gamma_k, k = 1, \ldots, \tilde{K}$, where $\tilde{K}$ is the number of distinct roots of the polynomial, then we can factorize ζ as

$$\zeta(x) = \gamma_0 \prod_{k=1}^{\tilde{K}} (x - \gamma_k)^{c_k},$$

where $\gamma_0 = (1 - a_0)(-1)^K / \prod_{i=1}^{\tilde{K}} \gamma_i^{c_i}$ and c_k is the multiplicity of the k–th root. This operation is essential also in the estimation algorithm, for this reason we remark that the computational effort for this factorization is very low, indeed we have only to calculate the roots of a polynomial of order K in one variable.

Applying this representation to the matrix version of a truncated polynomial, we get

$$M_n^K(\rho) = \gamma_0 \prod_{k=1}^{\tilde{K}} (\rho W - I_n \gamma_k)^{c_k},$$

and this function will be invertible for all ρ such that $|M_n(\rho)| \neq 0$. Its determinant is given by

$$\left|M_n^K(\rho)\right| = \gamma_0^n \prod_{k=1}^{\tilde{K}} |\rho W - I_n \gamma_k|^{c_k} = \gamma_0^n \rho^{nK} \prod_{k=1}^{\tilde{K}} \left| W - \frac{\gamma_k}{\rho} I_n \right|^{c_k}. \tag{S.18}$$

To calculate $\left| W - \frac{\gamma_k}{\rho} I_n \right|$ we can proceed as in the SAR model calculating their eigenvalues: the characteristic equation is

$$0 = \left| W - \frac{\gamma_k}{\rho} I_n - \omega I_n \right| = \left| W - \left(\frac{\gamma_k}{\rho} + \omega \right) I_n \right|$$

so, denoting by λ_j the eigenvalues of W and ω_i^k the i–th eigenvalues of the above matrix, we have

$$\lambda_i = \frac{\gamma_k}{\rho} + \omega_i^k \Leftrightarrow \omega_i^k = \lambda_i - \frac{\gamma_k}{\rho}.$$

Finally, substituting in (S.18) we get

$$\left|M_n^K(\rho)\right| = \gamma_0^n \rho^K \prod_{k=1}^{\tilde{K}} \prod_{i=1}^{n} \left(\lambda_i - \frac{\gamma_k}{\rho}\right)^{c_k} = \prod_{i=1}^{n} \gamma_0 \prod_{k=1}^{\tilde{K}} (\rho\lambda_i - \gamma_k)^{c_k} = \prod_{i=1}^{n} \zeta(\rho\lambda_i) \tag{S.19}$$

then we can characterize the invertible set of ρ in terms of W eigenvalues, i.e. this is the set

$$\Phi \cap \{\rho : \zeta(\rho\lambda_i) \neq 0, i = 1, \ldots, n\}.$$

Furthermore, we remark that $\lambda_i \in \mathbb{C}$, then the set on which the function is invertible is given by Φ minus the set of zeros of ζ lying in the complex ball. If the zeros are outside the complex ball of radius one, then the invertibility condition is satisfied and the assumptions holds.

S.4.8.3 Proposition 4

We prove that, if the limit Λ is degenerate, then Assumption 8 is violated. Since $B_0 = \frac{\partial M_n}{\partial \rho}(\rho_0) M_n(\rho_0) = \frac{\partial}{\partial \rho} \log(M_n) = \frac{\partial}{\partial \rho} \log(I - G_n(\rho_0 W))$, it can be written as a formal series. If $a_0 = 0$, using $\log(1 - G(x)) = -\sum_{k\geq 1} \frac{G(x)^k}{k}$ and Fàa di Bruno formula, as above:

$$\log(1 - G(x)) = -\sum_{n\geq 1} \frac{G(x)^n}{n} = -\sum_{n\geq 1} d_n x^n$$

$$\frac{\partial \log(1 - G(x))}{\partial x} = -\sum_{n\geq 1} n\, d_n x^{n-1}$$

where $d_n = \sum_{k=1}^{n} k^{-1} \hat{B}_{n,k}(a_1, \ldots, a_{n-k+1})$, and then

$$B_0 = \frac{\partial}{\partial \rho} \log(I - G_n(\rho_0 W)) = -W \sum_{k\geq 1} k d_k (\rho_0 W)^{k-1}.$$

If in particular W is diagonalizable with real eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$, then B_0 has eigenvalues $\eta_i := \lambda_{n:i}^{B_0} = -\lambda_i \sum_{k\geq 1} k d_k (\rho_0 \lambda_i)^{k-1}$ and assumption 8.2 writes

$$\begin{aligned}
\lim_n n^{-1} \sum_{i\leq n} \eta_i^2 - \left(n^{-1} \sum_i \eta_i\right)^2 &= \lim_n n^{-1} \sum_i \left(\sum_{k\geq 1} k d_k \rho_0^{k-1} \lambda_i^k\right)^2 - \left(n^{-1} \sum_i \sum_{k\geq 1} k d_k \rho_0^{k-1} \lambda_i^k\right)^2 \\
&= \lim_n \sum_{h\geq 1} \sum_{k\geq 1} h\, k\, d_h d_k \rho_0^{h+k-2} \left(n^{-1} \sum_{i\leq n} \lambda_i^{k+h} - n^{-2} \sum_{i\leq n} \lambda_i^h \sum_j \lambda_j^k\right) \\
&= \lim_n \sum_{h\geq 1} \sum_{k\geq 1} h\, k\, d_h d_k \rho_0^{h+k-2} cov(\Lambda_n^h, \Lambda_n^k) > 0.
\end{aligned}$$

It then follows that, if the empirical distribution of Λ_n tends to concentrate on a single point, all the above covariances $cov(\Lambda_n^h, \Lambda_n^k)$ tend to zero and assumption 8.2 is violated, independently on the functional form chosen for $G_n(\rho W)$.

For similar reasons, if Λ_n is composed by m_n distinct eigenvalues $(\lambda_1, \ldots, \lambda_n)$, with multiplicities $N_1 + N_2 + \ldots + N_{m_n} = n$, that satisfy (w.l.o.g.) $n^{-1}N_1 \to 1$, then

$$n^{-1}\mathrm{tr}(\boldsymbol{S}_n) = \sum_{j=1}^{m_n} \frac{N_j}{n} \left(\frac{1 - G(\rho\lambda_j)}{1 - G(\rho_0\lambda_j)} \right)^2 \to \left(\frac{1 - G(\rho\lambda_1)}{1 - G(\rho_0\lambda_1)} \right)^2$$

and

$$|\boldsymbol{S}_n|^{1/n} = \left(\frac{1 - G(\rho\lambda_1)}{1 - G(\rho_0\lambda_1)} \right)^{2N_1/n} \sqrt[n]{\prod_{j=2}^{m_n} \left(\frac{1 - G(\rho\lambda_j)}{1 - G(\rho_0\lambda_j)} \right)^{2N_j}} \to \left(\frac{1 - G(\rho\lambda_1)}{1 - G(\rho_0\lambda_1)} \right)^2 = \lim_n n^{-1}\mathrm{tr}(\boldsymbol{S}_n),$$

that violates Assumption 7.2.

S.5 Additional tables and figures

Table S.2: Family of link functions considered in model selection (both in simulations and the application).

	a	b	c	p	q	
1	1	1	0	1	0.5	
2	1	2	1	1	0.5	
3	1	1	0	3	0.5	
4	1	2	1	3	0.5	
5	1	1	0	5	0.5	
6	1	2	1	5	0.5	
7	1	1	0	1	1	⇐ case (A)
8	1	2	1	1	1	⇐ case (C)
9	1	1	0	3	1	⇐ case (B)
10	1	2	1	3	1	
11	1	1.5	0	1	1.500	
12	1	2	1	1	1.500	
13	1	2	0	1	2	
14	1	2	1	1	2	
15	6	1	NA	0.167	∞	
16	5	1	NA	0.200	∞	
17	4	1	NA	0.250	∞	
18	3	1	NA	0.333	∞	
19	2	1	NA	0.500	∞	⇐ case (D)
20	6	1	NA	0.500	∞	
21	1	1	NA	1	∞	⇐ case (E)
22	3	1	NA	1	∞	⇐ case (F)
23	5	1	NA	1	∞	
24	1	1	NA	3	∞	

Table S.3: Estimation results of simulations based on Texas weight matrix, Gaussian errors. Estimates of the parameter vector $(\beta_1, \beta_2, \varrho)$ under the two scenarios: (i) when the correct model is known; (ii) when the correct model is not known.

(a, b, c, p, q)		Estimates from correct model						Estimates after model selection					
		$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\varrho})$	sd$(\hat{\beta}_1)$	sd$(\hat{\beta}_2)$	sd$(\hat{\varrho})$	$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\varrho})$	sd$(\hat{\beta}_1)$	sd$(\hat{\beta}_2)$	sd$(\hat{\varrho})$
$(1,1,0,1,1)$	$\varrho=-0.4$	1.001	0.997	-0.402	0.026	0.100	0.042	1.004	0.998	-0.410	0.030	0.100	0.053
	$\varrho=0.2$	1.001	0.998	0.198	0.028	0.095	0.027	1.004	1.001	0.197	0.028	0.096	0.028
	$\varrho=0.5$	1.001	1.000	0.499	0.029	0.098	0.020	1.003	1.001	0.498	0.030	0.099	0.021
	$\varrho=0.9$	1.001	0.997	0.900	0.026	0.100	0.004	0.999	0.996	0.900	0.028	0.100	0.004
$(1,1,0,3,1)$	$\varrho=-0.4$	0.999	1.001	-0.399	0.028	0.100	0.052	0.999	1.002	-0.403	0.028	0.101	0.053
	$\varrho=0.2$	0.999	1.002	0.200	0.028	0.093	0.032	1.001	1.003	0.199	0.029	0.093	0.032
	$\varrho=0.5$	1.000	1.003	0.501	0.027	0.094	0.020	1.001	1.003	0.500	0.031	0.094	0.020
	$\varrho=0.9$	1.000	0.998	0.900	0.023	0.093	0.004	1.000	0.998	0.900	0.023	0.093	0.004
$(1,2,1,1,1)$	$\varrho=-0.4$	1.001	0.996	-0.401	0.029	0.096	0.052	0.998	0.993	-0.398	0.031	0.097	0.057
	$\varrho=0.2$	1.000	0.992	0.200	0.026	0.097	0.026	0.997	0.991	0.204	0.028	0.097	0.028
	$\varrho=0.5$	1.001	0.994	0.499	0.022	0.098	0.016	0.988	0.986	0.505	0.028	0.098	0.017
	$\varrho=0.9$	1.000	0.995	0.900	0.022	0.101	0.004	1.000	0.995	0.900	0.022	0.101	0.004
$(2,1,\text{na},0.5,\infty)$	$\varrho=-0.4$	0.999	0.998	-0.399	0.025	0.099	0.039	1.001	0.998	-0.406	0.030	0.100	0.053
	$\varrho=0.2$	1.000	0.996	0.199	0.028	0.097	0.029	1.002	0.999	0.198	0.028	0.097	0.031
	$\varrho=0.5$	1.000	1.005	0.500	0.028	0.101	0.018	1.005	1.009	0.498	0.029	0.102	0.020
	$\varrho=0.9$	0.999	0.996	0.900	0.031	0.095	0.004	1.009	1.005	0.900	0.037	0.099	0.004
$(1,1,\text{na},1,\infty)$	$\varrho=-0.4$	1.001	1.003	-0.401	0.028	0.099	0.047	1.000	1.002	-0.402	0.032	0.099	0.058
	$\varrho=0.2$	1.001	1.003	0.200	0.026	0.099	0.026	1.001	1.003	0.202	0.027	0.100	0.028
	$\varrho=0.5$	1.002	1.002	0.498	0.028	0.105	0.017	1.004	1.005	0.497	0.034	0.107	0.023
	$\varrho=0.9$	1.001	1.001	0.900	0.028	0.096	0.003	1.001	1.001	0.900	0.028	0.096	0.003
$(3,1,\text{na},1,\infty)$	$\varrho=-0.4$	1.001	1.004	-0.404	0.027	0.092	0.053	1.001	1.005	-0.407	0.028	0.091	0.053
	$\varrho=0.2$	1.002	0.994	0.197	0.029	0.095	0.033	1.004	0.995	0.196	0.030	0.095	0.033
	$\varrho=0.5$	1.001	1.001	0.499	0.028	0.100	0.020	1.004	1.002	0.499	0.030	0.100	0.020
	$\varrho=0.9$	1.000	0.994	0.900	0.028	0.093	0.004	1.000	0.994	0.900	0.028	0.093	0.004

Table S.4: Estimation results of simulations based on Texas weight matrix, Gaussian errors. Direct effects, true (first column) and estimated. Estimated effects are obtained from: estimation of the correct model (col. 2,3); estimation after model selection from beta and gamma families (col. 4,5)

(a, b, c, p, q)		True direct	Estimated - true model		Estimated- selected model	
			X_1	X_2	X_1	X_2
(1,1,0,1,1)	$\varrho = -0.4$	1.022	1.024	1.019	1.025	1.019
	$\varrho = 0.2$	1.006	1.007	1.005	1.008	1.005
	$\varrho = 0.5$	1.048	1.049	1.047	1.050	1.048
	$\varrho = 0.9$	1.294	1.296	1.290	1.295	1.292
(1,1,0,3,1)	$\varrho = -0.4$	0.988	0.987	0.989	0.987	0.990
	$\varrho = 0.2$	1.010	1.009	1.011	1.010	1.011
	$\varrho = 0.5$	1.033	1.033	1.036	1.034	1.036
	$\varrho = 0.9$	1.118	1.119	1.116	1.119	1.116
(1,2,1,1,1)	$\varrho = -0.4$	1	1.001	0.996	1.001	0.997
	$\varrho = 0.2$	1	1.000	0.992	0.998	0.991
	$\varrho = 0.5$	1.001	1.002	0.995	0.997	0.995
	$\varrho = 0.9$	1.012	1.012	1.006	1.012	1.006
(2,1,NA,.5,∞)	$\varrho = -0.4$	1.025	1.024	1.022	1.025	1.022
	$\varrho = 0.2$	1.006	1.006	1.003	1.006	1.003
	$\varrho = 0.5$	1.050	1.050	1.056	1.051	1.056
	$\varrho = 0.9$	1.488	1.488	1.484	1.487	1.480
(1,1,NA, 1, ∞)	$\varrho = -0.4$	1.008	1.010	1.012	1.010	1.012
	$\varrho = 0.2$	1.004	1.005	1.007	1.004	1.007
	$\varrho = 0.5$	1.039	1.040	1.041	1.041	1.041
	$\varrho = 0.9$	1.570	1.569	1.571	1.569	1.571
(3,1,NA,1, ∞)	$\varrho = -0.4$	0.988	0.989	0.992	0.989	0.993
	$\varrho = 0.2$	1.010	1.012	1.004	1.013	1.004
	$\varrho = 0.5$	1.037	1.038	1.039	1.040	1.039
	$\varrho = 0.9$	1.249	1.250	1.242	1.250	1.242

Table S.5: Estimation results of simulations based on Texas weight matrix, Gaussian errors. Indirect effects, true (first column) and estimated. Estimated effects are obtained from: estimation of the correct model (col. 2,3); estimation after model selection from beta and gamma families (col. 4,5)

(a, b, c, p, q)		True indirect	Estimated - true model		Estimated- selected model	
			X_1	X_2	X_1	X_2
(1,1,0,1,1)	$\varrho = -0.4$	-0.275	-0.276	-0.274	-0.277	-0.275
	$\varrho = 0.2$	0.206	0.204	0.204	0.203	0.204
	$\varrho = 0.5$	0.765	0.764	0.764	0.763	0.763
	$\varrho = 0.9$	5.618	5.620	5.597	5.622	5.609
(1,1,0,3,1)	$\varrho = -0.4$	-0.211	-0.210	-0.210	-0.210	-0.210
	$\varrho = 0.2$	0.173	0.173	0.174	0.172	0.173
	$\varrho = 0.5$	0.658	0.660	0.663	0.658	0.660
	$\varrho = 0.9$	3.883	3.875	3.864	3.875	3.864
(1,2,1,1,1)	$\varrho = -0.4$	-0.247	-0.247	-0.245	-0.248	-0.246
	$\varrho = 0.2$	0.216	0.217	0.216	0.219	0.218
	$\varrho = 0.5$	0.825	0.821	0.816	0.831	0.830
	$\varrho = 0.9$	1.693	1.693	1.683	1.693	1.683
(2,1,NA,.5,∞)	$\varrho = -0.4$	-0.278	-0.278	-0.277	-0.278	-0.277
	$\varrho = 0.2$	0.206	0.205	0.205	0.205	0.205
	$\varrho = 0.5$	0.773	0.772	0.778	0.771	0.776
	$\varrho = 0.9$	6.195	6.199	6.186	6.206	6.184
(1,1,NA, 1, ∞)	$\varrho = -0.4$	-0.258	-0.259	-0.259	-0.259	-0.259
	$\varrho = 0.2$	0.210	0.211	0.212	0.212	0.213
	$\varrho = 0.5$	0.802	0.799	0.801	0.798	0.799
	$\varrho = 0.9$	6.478	6.466	6.477	6.466	6.477
(3,1,NA,1, ∞)	$\varrho = -0.4$	-0.207	-0.208	-0.209	-0.208	-0.209
	$\varrho = 0.2$	0.175	0.173	0.173	0.172	0.171
	$\varrho = 0.5$	0.684	0.682	0.683	0.679	0.680
	$\varrho = 0.9$	5.653	5.652	5.623	5.652	5.623

Table S.6: Estimation results of simulations based on North Carolina weight matrix, gamma errors. Estimates of the parameter vector $(\beta_1, \beta_2, \varrho)$ under the two scenarios: (i) when the correct model is known; (ii) when the correct model is not known.

(a, b, c, p, q)		Estimates from correct model						Estimates after model selection					
		$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\varrho})$	sd$(\hat{\beta}_1)$	sd$(\hat{\beta}_2)$	sd$(\hat{\varrho})$	$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\varrho})$	sd$(\hat{\beta}_1)$	sd$(\hat{\beta}_2)$	sd$(\hat{\varrho})$
(1,1,0,1,1)	$\varrho = -0.4$	1.001	1.004	-0.401	0.038	0.151	0.063	1.005	1.005	-0.415	0.044	0.151	0.081
	$\varrho = 0.2$	0.998	0.994	0.199	0.041	0.156	0.044	1.000	0.995	0.200	0.041	0.158	0.045
	$\varrho = 0.5$	1.001	1.013	0.499	0.044	0.157	0.030	1.005	1.016	0.497	0.045	0.160	0.032
	$\varrho = 0.9$	1.004	0.988	0.899	0.038	0.159	0.006	0.998	0.987	0.899	0.039	0.160	0.006
(1,1,0,3,1)	$\varrho = -0.4$	1.005	0.993	-0.415	0.044	0.155	0.082	1.005	0.990	-0.416	0.045	0.156	0.084
	$\varrho = 0.2$	1.003	1.001	0.195	0.042	0.155	0.047	1.004	1.005	0.195	0.042	0.156	0.047
	$\varrho = 0.5$	1.002	0.990	0.495	0.039	0.157	0.028	1.005	0.990	0.495	0.043	0.160	0.029
	$\varrho = 0.9$	1.001	0.990	0.900	0.036	0.152	0.006	0.997	0.990	0.900	0.043	0.154	0.006
(1,2,1,1,1)	$\varrho = -0.4$	1.002	0.992	-0.404	0.044	0.155	0.080	0.998	0.987	-0.399	0.044	0.155	0.083
	$\varrho = 0.2$	0.999	1.003	0.200	0.042	0.158	0.041	0.995	1.001	0.209	0.044	0.159	0.045
	$\varrho = 0.5$	1.003	1.005	0.497	0.037	0.149	0.027	0.985	0.993	0.507	0.045	0.150	0.030
	$\varrho = 0.9$	1.003	0.991	0.900	0.037	0.150	0.005	1.003	0.991	0.900	0.038	0.150	0.005
(2,1,NA,.5,∞)	$\varrho = -0.4$	1.002	1.001	-0.404	0.038	0.154	0.059	1.009	1.005	-0.425	0.044	0.155	0.079
	$\varrho = 0.2$	1.001	1.010	0.199	0.041	0.153	0.043	1.003	1.012	0.202	0.041	0.155	0.044
	$\varrho = 0.5$	1.001	0.995	0.498	0.044	0.153	0.030	1.008	1.002	0.495	0.045	0.155	0.032
	$\varrho = 0.9$	1.000	0.996	0.900	0.045	0.151	0.006	1.014	1.006	0.899	0.051	0.153	0.006
(1,1,NA, 1, ∞)	$\varrho = -0.4$	0.998	0.997	-0.396	0.042	0.154	0.073	0.998	0.995	-0.404	0.044	0.154	0.084
	$\varrho = 0.2$	1.002	1.013	0.199	0.042	0.158	0.041	1.002	1.013	0.204	0.043	0.159	0.043
	$\varrho = 0.5$	1.005	0.991	0.497	0.043	0.157	0.026	1.008	0.994	0.497	0.048	0.161	0.033
	$\varrho = 0.9$	1.000	1.007	0.900	0.043	0.151	0.006	1.000	1.007	0.900	0.043	0.151	0.006
(3,1,NA,1, ∞)	$\varrho = -0.4$	1.006	1.001	-0.415	0.042	0.156	0.082	1.007	1.001	-0.418	0.045	0.156	0.084
	$\varrho = 0.2$	1.001	0.972	0.196	0.042	0.152	0.048	1.003	0.974	0.196	0.042	0.153	0.048
	$\varrho = 0.5$	1.002	1.004	0.498	0.044	0.159	0.031	1.005	1.004	0.498	0.047	0.159	0.031
	$\varrho = 0.9$	1.003	1.000	0.900	0.041	0.156	0.006	1.004	1.000	0.900	0.043	0.156	0.006

Table S.7: Estimation results of simulations based on North Carolina weight matrix, gamma errors. Direct effects, true (first column) and estimated. Estimated effects are obtained from: estimation of the correct model (col. 2,3); estimation after model selection from beta and gamma families (col. 4,5)

(a, b, c, p, q)		True direct	Estimated - true model		Estimated- selected model	
			X_1	X_2	X_1	X_2
(1,1,0,1,1)	$\varrho = -0.4$	1.020	1.022	1.025	1.024	1.024
	$\varrho = 0.2$	1.006	1.004	1.001	1.004	0.999
	$\varrho = 0.5$	1.046	1.047	1.060	1.048	1.059
	$\varrho = 0.9$	1.305	1.308	1.287	1.307	1.293
(1,1,0,3,1)	$\varrho = -0.4$	0.986	0.991	0.980	0.994	0.978
	$\varrho = 0.2$	1.010	1.013	1.011	1.012	1.013
	$\varrho = 0.5$	1.036	1.037	1.024	1.039	1.025
	$\varrho = 0.9$	1.142	1.143	1.131	1.141	1.133
(1,2,1,1,1)	$\varrho = -0.4$	1	1.002	0.992	1.004	0.993
	$\varrho = 0.2$	1	0.999	1.003	0.997	1.003
	$\varrho = 0.5$	1.001	1.005	1.007	0.998	1.006
	$\varrho = 0.9$	1.019	1.022	1.010	1.022	1.011
(2,1,NA,.5,∞)	$\varrho = -0.4$	1.023	1.026	1.024	1.028	1.024
	$\varrho = 0.2$	1.006	1.007	1.016	1.007	1.016
	$\varrho = 0.5$	1.048	1.049	1.043	1.050	1.044
	$\varrho = 0.9$	1.480	1.480	1.474	1.476	1.466
(1,1,NA, 1, ∞)	$\varrho = -0.4$	1.008	1.006	1.005	1.007	1.004
	$\varrho = 0.2$	1.004	1.006	1.017	1.005	1.017
	$\varrho = 0.5$	1.036	1.041	1.026	1.041	1.026
	$\varrho = 0.9$	1.550	1.549	1.562	1.549	1.562
(3,1,NA,1, ∞)	$\varrho = -0.4$	0.987	0.992	0.988	0.995	0.989
	$\varrho = 0.2$	1.011	1.012	0.982	1.011	0.982
	$\varrho = 0.5$	1.040	1.043	1.044	1.044	1.043
	$\varrho = 0.9$	1.272	1.275	1.272	1.275	1.271

Table S.8: Estimation results of simulations based on North Carolina weight matrix, gamma errors. Indirect effects, true (first column) and estimated. Estimated effects are obtained from: estimation of the correct model (col. 2,3); estimation after model selection from beta and gamma families (col. 4,5)

(a, b, c, p, q)		True indirect	Estimated - true model		Estimated- selected model	
			X_1	X_2	X_1	X_2
(1,1,0,1,1)	$\varrho = -0.4$	-0.262	-0.263	-0.263	-0.265	-0.264
	$\varrho = 0.2$	0.197	0.195	0.198	0.196	0.198
	$\varrho = 0.5$	0.731	0.732	0.743	0.730	0.741
	$\varrho = 0.9$	5.510	5.491	5.408	5.499	5.446
(1,1,0,3,1)	$\varrho = -0.4$	-0.198	-0.204	-0.201	-0.207	-0.203
	$\varrho = 0.2$	0.165	0.161	0.162	0.163	0.165
	$\varrho = 0.5$	0.633	0.624	0.618	0.620	0.613
	$\varrho = 0.9$	3.910	3.901	3.864	3.923	3.902
(1,2,1,1,1)	$\varrho = -0.4$	-0.236	-0.238	-0.234	-0.239	-0.236
	$\varrho = 0.2$	0.207	0.207	0.209	0.210	0.213
	$\varrho = 0.5$	0.794	0.787	0.790	0.800	0.808
	$\varrho = 0.9$	1.699	1.703	1.684	1.707	1.688
(2,1,NA,.5,∞)	$\varrho = -0.4$	-0.265	-0.269	-0.267	-0.271	-0.268
	$\varrho = 0.2$	0.197	0.197	0.200	0.197	0.200
	$\varrho = 0.5$	0.739	0.736	0.734	0.735	0.732
	$\varrho = 0.9$	5.934	5.925	5.913	5.942	5.907
(1,1,NA, 1, ∞)	$\varrho = -0.4$	-0.246	-0.244	-0.243	-0.245	-0.244
	$\varrho = 0.2$	0.201	0.201	0.205	0.203	0.207
	$\varrho = 0.5$	0.767	0.764	0.755	0.764	0.756
	$\varrho = 0.9$	6.186	6.169	6.230	6.169	6.230
(3,1,NA,1, ∞)	$\varrho = -0.4$	-0.195	-0.201	-0.199	-0.203	-0.201
	$\varrho = 0.2$	0.166	0.164	0.160	0.165	0.161
	$\varrho = 0.5$	0.651	0.650	0.654	0.649	0.652
	$\varrho = 0.9$	5.501	5.502	5.492	5.503	5.486

Table S.9: Estimation results of simulations based on Texas weight matrix, gamma errors. Estimates of the parameter vector $(\beta_1, \beta_2, \varrho)$ under the two scenarios: (i) when the correct model is known; (ii) when the correct model is not known.

(a, b, c, p, q)		Estimates from correct model						Estimates after model selection					
		$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\varrho})$	$\mathrm{sd}(\hat{\beta}_1)$	$\mathrm{sd}(\hat{\beta}_2)$	$\mathrm{sd}(\hat{\varrho})$	$E(\hat{\beta}_1)$	$E(\hat{\beta}_2)$	$E(\hat{\varrho})$	$\mathrm{sd}(\hat{\beta}_1)$	$\mathrm{sd}(\hat{\beta}_2)$	$\mathrm{sd}(\hat{\varrho})$
(1,1,0,1,1)	$\varrho = -0.4$	1.000	0.997	-0.400	0.025	0.098	0.040	1.003	0.998	-0.408	0.031	0.099	0.055
	$\varrho = 0.2$	1.002	1.009	0.197	0.029	0.097	0.028	1.005	1.011	0.196	0.029	0.098	0.030
	$\varrho = 0.5$	1.000	0.997	0.500	0.028	0.096	0.019	1.002	0.998	0.498	0.029	0.096	0.020
	$\varrho = 0.9$	1.001	0.993	0.900	0.026	0.095	0.004	1.000	0.992	0.900	0.027	0.095	0.004
(1,1,0,3,1)	$\varrho = -0.4$	1.002	0.996	-0.405	0.029	0.094	0.054	1.002	0.996	-0.408	0.030	0.095	0.055
	$\varrho = 0.2$	1.000	0.998	0.199	0.027	0.091	0.032	1.003	0.999	0.198	0.028	0.091	0.033
	$\varrho = 0.5$	1.004	0.999	0.498	0.025	0.095	0.019	1.003	0.999	0.497	0.028	0.096	0.019
	$\varrho = 0.9$	1.000	0.996	0.900	0.022	0.095	0.004	1.000	0.996	0.900	0.022	0.095	0.004
(1,2,1,1,1)	$\varrho = -0.4$	1.001	1.000	-0.401	0.028	0.099	0.050	0.998	0.997	-0.396	0.030	0.100	0.057
	$\varrho = 0.2$	1.002	1.005	0.199	0.027	0.098	0.026	1.001	1.005	0.203	0.028	0.099	0.028
	$\varrho = 0.5$	1.000	1.002	0.499	0.024	0.099	0.018	0.988	0.994	0.505	0.031	0.100	0.020
	$\varrho = 0.9$	1.000	0.994	0.900	0.021	0.096	0.003	1.000	0.994	0.900	0.021	0.096	0.003
(2,1,NA,.5,∞)	$\varrho = -0.4$	1.000	1.004	-0.399	0.025	0.098	0.038	1.002	1.005	-0.406	0.030	0.099	0.052
	$\varrho = 0.2$	1.000	0.994	0.199	0.029	0.095	0.029	1.002	0.997	0.198	0.030	0.095	0.031
	$\varrho = 0.5$	1.001	1.003	0.499	0.030	0.092	0.019	1.007	1.008	0.497	0.032	0.093	0.022
	$\varrho = 0.9$	1.001	1.001	0.900	0.031	0.097	0.004	1.009	1.008	0.899	0.036	0.099	0.004
(1,1,NA, 1, ∞)	$\varrho = -0.4$	0.998	0.998	-0.399	0.028	0.094	0.049	0.997	0.996	-0.400	0.031	0.095	0.059
	$\varrho = 0.2$	1.001	1.000	0.201	0.027	0.097	0.026	1.000	1.001	0.203	0.029	0.097	0.029
	$\varrho = 0.5$	1.002	0.990	0.499	0.027	0.093	0.017	1.005	0.992	0.497	0.035	0.095	0.023
	$\varrho = 0.9$	0.999	1.002	0.900	0.028	0.100	0.003	0.999	1.002	0.900	0.028	0.100	0.003
(3,1,NA,1, ∞)	$\varrho = -0.4$	0.998	1.004	-0.398	0.029	0.096	0.058	0.998	1.004	-0.401	0.031	0.096	0.059
	$\varrho = 0.2$	1.001	1.009	0.197	0.027	0.093	0.030	1.003	1.009	0.197	0.028	0.093	0.031
	$\varrho = 0.5$	1.001	1.000	0.499	0.028	0.097	0.019	1.004	1.000	0.500	0.029	0.096	0.019
	$\varrho = 0.9$	0.999	1.005	0.900	0.026	0.092	0.004	0.999	1.005	0.900	0.026	0.092	0.004

Table S.10: Estimation results of simulations based on Texas weight matrix, gamma errors. Direct effects, true (first column) and estimated. Estimated effects are obtained from: estimation of the correct model (col. 2,3); estimation after model selection from beta and gamma families (col. 4,5)

(a, b, c, p, q)		True direct	Estimated - true model		Estimated- selected model	
			X_1	X_2	X_1	X_2
(1,1,0,1,1)	$\varrho = -0.4$	1.022	1.023	1.019	1.024	1.019
	$\varrho = 0.2$	1.006	1.009	1.015	1.009	1.015
	$\varrho = 0.5$	1.048	1.047	1.045	1.048	1.045
	$\varrho = 0.9$	1.294	1.296	1.285	1.295	1.286
(1,1,0,3,1)	$\varrho = -0.4$	0.988	0.990	0.984	0.989	0.984
	$\varrho = 0.2$	1.010	1.010	1.007	1.012	1.007
	$\varrho = 0.5$	1.033	1.036	1.032	1.036	1.032
	$\varrho = 0.9$	1.118	1.119	1.114	1.119	1.114
(1,2,1,1,1)	$\varrho = -0.4$	1	1.001	1.000	1.002	1.001
	$\varrho = 0.2$	1	1.002	1.005	1.001	1.005
	$\varrho = 0.5$	1.001	1.001	1.003	0.996	1.002
	$\varrho = 0.9$	1.012	1.012	1.005	1.012	1.005
(2,1,NA,.5,∞)	$\varrho = -0.4$	1.025	1.025	1.029	1.026	1.029
	$\varrho = 0.2$	1.006	1.006	1.001	1.006	1.001
	$\varrho = 0.5$	1.050	1.052	1.054	1.053	1.054
	$\varrho = 0.9$	1.488	1.488	1.489	1.487	1.486
(1,1,NA, 1, ∞)	$\varrho = -0.4$	1.008	1.007	1.006	1.007	1.006
	$\varrho = 0.2$	1.004	1.005	1.004	1.003	1.004
	$\varrho = 0.5$	1.039	1.040	1.028	1.041	1.028
	$\varrho = 0.9$	1.570	1.567	1.573	1.567	1.573
(3,1,NA,1, ∞)	$\varrho = -0.4$	0.988	0.986	0.992	0.986	0.992
	$\varrho = 0.2$	1.010	1.011	1.019	1.012	1.018
	$\varrho = 0.5$	1.037	1.038	1.038	1.041	1.037
	$\varrho = 0.9$	1.249	1.248	1.256	1.248	1.256

Table S.11: Estimation results of simulations based on Texas weight matrix, gamma errors. Indirect effects, true (first column) and estimated. Estimated effects are obtained from: estimation of the correct model (col. 2,3); estimation after model selection from beta and gamma families (col. 4,5)

(a, b, c, p, q)		True indirect	Estimated - true model		Estimated- selected model	
			X_1	X_2	X_1	X_2
(1,1,0,1,1)	$\varrho = -0.4$	-0.275	-0.275	-0.273	-0.276	-0.274
	$\varrho = 0.2$	0.206	0.202	0.205	0.202	0.205
	$\varrho = 0.5$	0.765	0.765	0.764	0.763	0.762
	$\varrho = 0.9$	5.618	5.621	5.577	5.623	5.587
(1,1,0,3,1)	$\varrho = -0.4$	-0.211	-0.213	-0.211	-0.212	-0.211
	$\varrho = 0.2$	0.173	0.173	0.173	0.170	0.170
	$\varrho = 0.5$	0.658	0.655	0.653	0.655	0.654
	$\varrho = 0.9$	3.883	3.880	3.865	3.879	3.865
(1,2,1,1,1)	$\varrho = -0.4$	-0.247	-0.247	-0.247	-0.248	-0.247
	$\varrho = 0.2$	0.216	0.216	0.217	0.218	0.219
	$\varrho = 0.5$	0.825	0.822	0.824	0.831	0.838
	$\varrho = 0.9$	1.693	1.693	1.681	1.693	1.681
(2,1,NA,.5,∞)	$\varrho = -0.4$	-0.278	-0.277	-0.278	-0.279	-0.279
	$\varrho = 0.2$	0.206	0.205	0.205	0.205	0.205
	$\varrho = 0.5$	0.773	0.772	0.775	0.771	0.773
	$\varrho = 0.9$	6.195	6.178	6.186	6.185	6.183
(1,1,NA,1, ∞)	$\varrho = -0.4$	-0.258	-0.257	-0.256	-0.258	-0.257
	$\varrho = 0.2$	0.210	0.212	0.212	0.213	0.215
	$\varrho = 0.5$	0.802	0.801	0.792	0.800	0.791
	$\varrho = 0.9$	6.478	6.470	6.495	6.470	6.495
(3,1,NA,1, ∞)	$\varrho = -0.4$	-0.207	-0.206	-0.207	-0.205	-0.206
	$\varrho = 0.2$	0.175	0.173	0.175	0.171	0.173
	$\varrho = 0.5$	0.684	0.684	0.684	0.681	0.680
	$\varrho = 0.9$	5.653	5.642	5.685	5.642	5.685

Table S.12: Direct and indirect effects the SAR, MESS and selected model specification. The selected model corresponds to a generalized beta with parameters $Beta(1, 2, 1, 1, 1)$.

	Direct effects			Indirect Effects		
	SAR	MESS	$B(1, 2, 1, 1, 1)$	SAR	MESS	$B(1, 2, 1, 1, 1)$
const	-0.40544	-0.39472	-0.40264	-0.23597	-0.24649	-0.23820
repub92	0.02089	0.02018	0.02051	0.01216	0.01260	0.01213
repub96	-0.02925	-0.02827	-0.02913	-0.01702	-0.01766	-0.01724
perot92	0.02640	0.02545	0.02590	0.01537	0.01590	0.01532
perot96	-0.01090	-0.01081	-0.01149	-0.00634	-0.00675	-0.00680
income00	0.00040	0.00148	0.00153	0.00023	0.00093	0.00090
dincome90to00	0.01143	0.01111	0.01122	0.00665	0.00694	0.00664
urate00	0.01548	0.01494	0.01495	0.00901	0.00933	0.00884
urate00_urate90	0.00621	0.00604	0.00585	0.00362	0.00377	0.00346
female_male	0.02699	0.02750	0.02762	0.01571	0.01717	0.01634
black_pop	0.00561	0.00541	0.00467	0.00326	0.00338	0.00276
asian_pop	0.00587	0.00571	0.00547	0.00342	0.00356	0.00324
hispanic_pop	0.02110	0.02182	0.02119	0.01228	0.01363	0.01254
femhhwchild_pop	0.00307	0.00335	0.00369	0.00179	0.00209	0.00218
owner_occupied	0.03673	0.03656	0.03664	0.02138	0.02283	0.02167
highschool	0.03193	0.03052	0.03022	0.01858	0.01906	0.01788
college	0.05386	0.05433	0.05470	0.03135	0.03393	0.03236
nevermarried	-0.01839	-0.01901	-0.01906	-0.01071	-0.01187	-0.01128
divorced	0.00906	0.00878	0.00854	0.00528	0.00548	0.00505
widowed	0.00196	0.00178	0.00156	0.00114	0.00111	0.00092
samehouse	0.04081	0.04025	0.04017	0.02375	0.02513	0.02376
foreignborn	-0.01826	-0.01833	-0.01807	-0.01063	-0.01144	-0.01069
language	-0.01599	-0.01717	-0.01778	-0.00931	-0.01072	-0.01052
military	-0.00814	-0.00786	-0.00773	-0.00474	-0.00491	-0.00457
fem_emp_females	-0.00012	-0.00001	-0.00002	-0.00007	-0.00001	-0.00001
work_home	0.03415	0.03364	0.03376	0.01988	0.02101	0.01997
traveltime	0.00576	0.00597	0.00543	0.00335	0.00373	0.00321
poverty	-0.03100	-0.03044	-0.03094	-0.01804	-0.01901	-0.01831
log_hvalue	0.02437	0.02391	0.02377	0.01418	0.01493	0.01406
log_rent	-0.03127	-0.03197	-0.03251	-0.01820	-0.01996	-0.01923
log_mortgage	-0.01254	-0.01275	-0.01367	-0.00730	-0.00796	-0.00809
govt_workers_work	0.01078	0.01086	0.01056	0.00627	0.00678	0.00625
manuf_workers_work	-0.01340	-0.01285	-0.01295	-0.00780	-0.00802	-0.00766
arts_workers_work	0.00125	0.00106	0.00083	0.00073	0.00066	0.00049

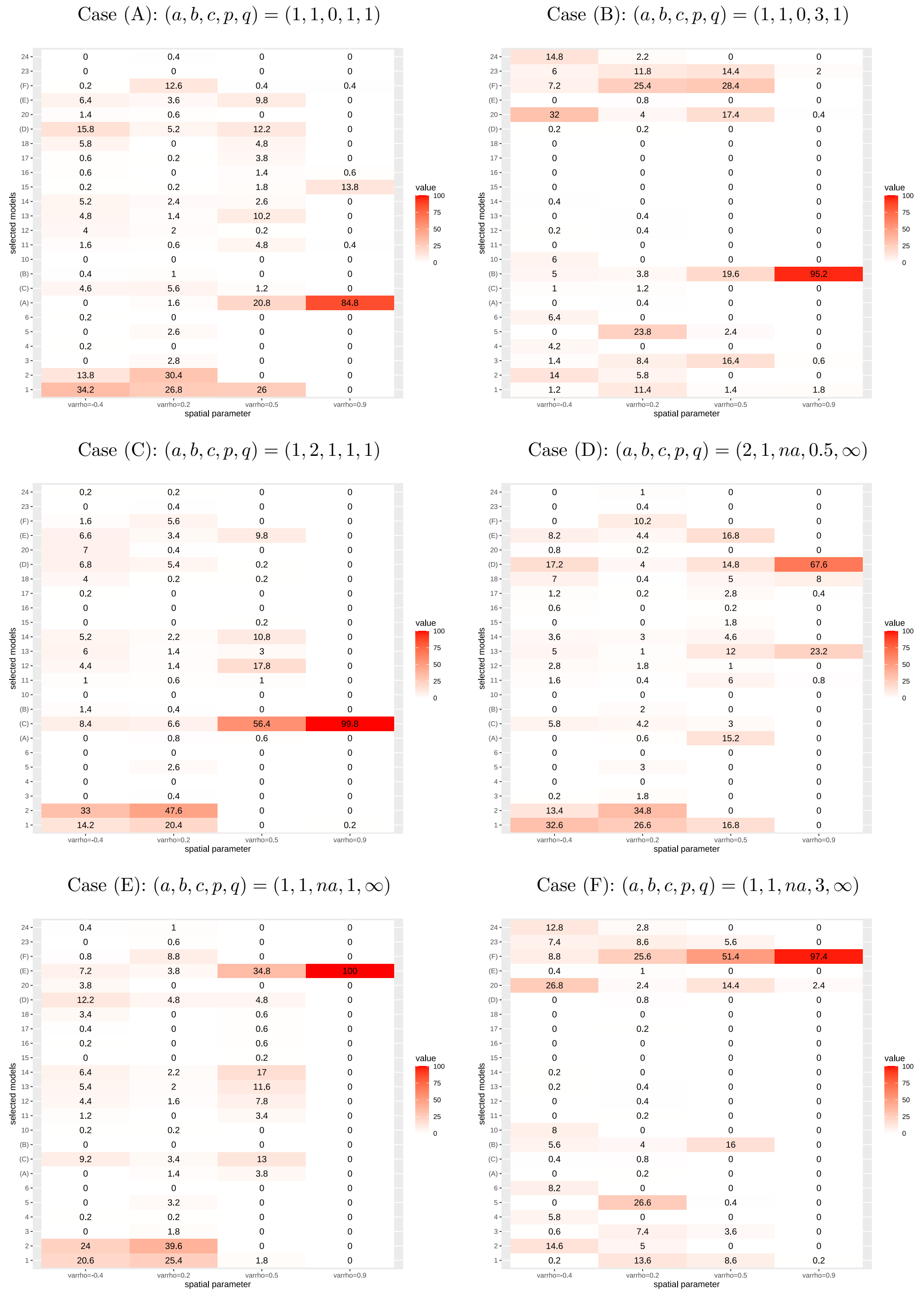
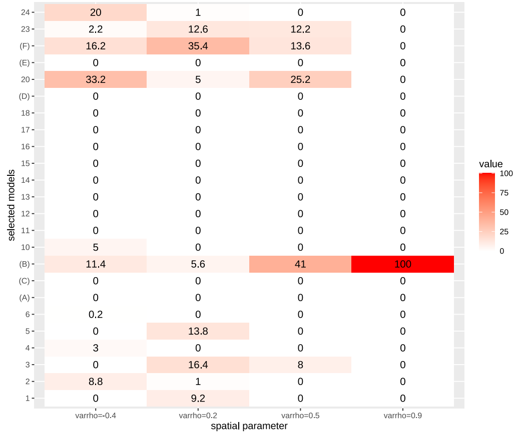


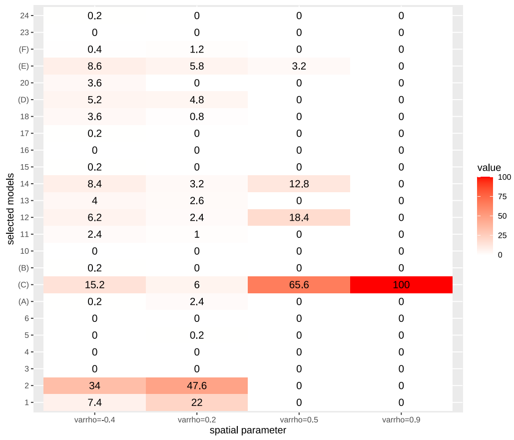
Figure S.1: Heatmap of the percentages of the selections of each candidate model (in rows) and for all values of ϱ (in columns), from the simulations based on North Carolina weight matrix in the 6 cases (panels from (A) to (F)). The proposed models that coincide with the specifications (A)–(F) are labeled with the corresponding letter.

Case (A): $(a, b, c, p, q) = (1, 1, 0, 1, 1)$

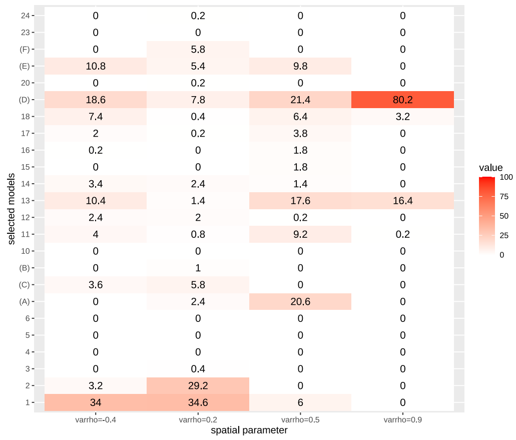
Case (B): $(a, b, c, p, q) = (1, 1, 0, 3, 1)$



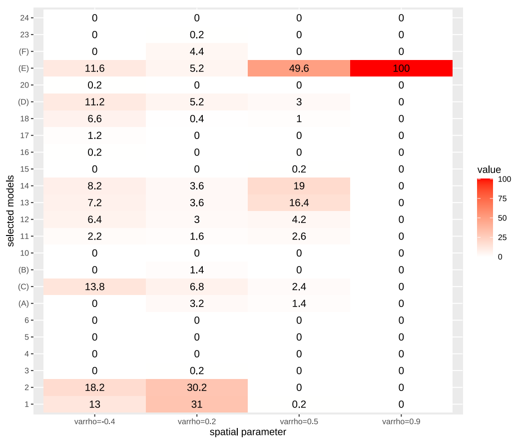
Case (C): $(a, b, c, p, q) = (1, 2, 1, 1, 1)$



Case (D): $(a, b, c, p, q) = (2, 1, na, 0.5, \infty)$



Case (E): $(a, b, c, p, q) = (1, 1, na, 1, \infty)$



Case (F): $(a, b, c, p, q) = (1, 1, na, 3, \infty)$

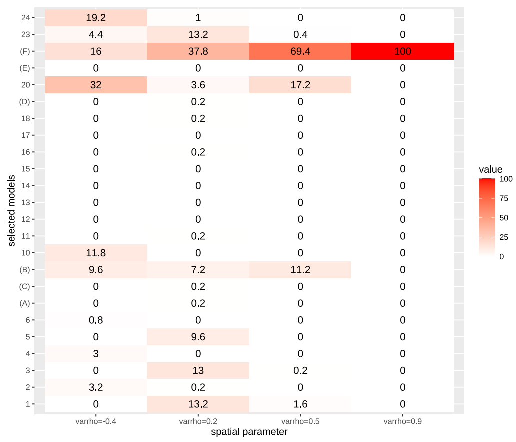


Figure S.2: Heatmap of the percentages of the selections of each candidate model (in rows) and for all values of ϱ (in columns), from the simulations based on Texas weight matrix in the 6 cases (panels from (A) to (F)). The proposed models that coincide with the specifications (A)–(F) are labeled with the corresponding letter.

References

Birmajer, D., Gil, J. B., and Weiner, M. D. (2012). Some convolution identities and an inverse relation involving partial bell polynomials. *Electr. J. Comb.*, 19:P34.

Debarsy, N., Jin, F., and fei Lee, L. (2015a). Large sample properties of the matrix exponential spatial specification with an application to fdi. *Journal of Econometrics*, 188:1–21.

Debarsy, N., Jin, F., and fei Lee, L. (2015b). Supplement to "large sample properties of the matrix exponential spatial specification with an application to fdi". *Journal of Econometrics*, 188.

Gradshtein, I. S., Jeffrey, A., and Ryzhik, I. M. (1965). Table of integrals, series, and products.

Lee, L.-f. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72:1899–1925.

Petz, D. (1994). A survey of certain trace inequalities. *Journal Banach Center Publications*, 30:287 – 298.

Tung, S. H. (1975). On lower and upper bounds of the difference between the arithmetic and the geometric mean. *Mathematics of Computation*, 29(131):834–836.

Wang, W. and Wang, T. (2009). General identities on bell polynomials. *Computers & Mathematics with Applications*, 58(1):104 – 118.