Online Supplement to "Continuously Updated Indirect Inference in Heteroskedastic Spatial Models "

Maria Kyriacou* Peter C. B. Phillips† Francesca Rossi‡

June 11, 2021

This supplement provides additional technical material, expanded proofs for the main paper, and further simulation results.

S.1 Derivation of bias expressions for MLE/QMLE

In this section we report the derivation of the bias function displayed in Figure 1 of the manuscript. To assist in the bias calculation we derive the following explicit moment expressions

$$\mathbb{E}(l^{(1)}(\lambda_0)) = \frac{tr(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} - \frac{1}{n}tr(G) + o(1), \tag{S.1.1}$$

$$\mathbb{E}(l^{(2)}(\lambda_0)) = -\frac{\beta_0' X' G' M G X \beta_0 + tr(G' G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} + \frac{2tr^2(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} - \frac{1}{n}tr(G^2) + o(1), \tag{S.1.2}$$

$$\begin{aligned}\mathbb{E}(l^{(2)}(\lambda_0)l^{(1)}(\lambda_0)) = &- \frac{tr(G\Omega_0(\gamma))\,(tr(G'G\Omega_0(\gamma)) + \beta_0 X'G'MGX\beta_0)}{tr^2(\Omega_0(\gamma))} - \frac{1}{n}tr(G^2)\frac{tr(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} \\ &+ \frac{1}{n}tr(G)\frac{tr(G'G\Omega_0(\gamma)) + \beta_0 X'G'MGX\beta_0}{tr(\Omega_0(\gamma))} - \frac{2}{n}\frac{tr^2(G\Omega_0(\gamma))}{tr^2(\Omega_0(\gamma))} + 2\frac{tr^3(G\Omega_0(\gamma))}{tr^3(\Omega_0(\gamma))} \\ &+ \frac{1}{n^2}tr(G)tr(G^2) + o(1),\end{aligned} \tag{S.1.3}$$

*Corresponding author. University of Southampton. e-mail: m.kyriacou@soton.ac.uk

†Yale University, University of Auckland, University of Southampton, Singapore Management University. e-mail: peter.phillips@yale.edu

‡Department of Economics, University of Verona. e-mail: francesca.rossi_02@univr.it.

$$
\begin{aligned}
\mathbb{E}(l^{(3)}(\lambda_0)) = & -6\frac{tr(G\Omega_0(\gamma))\,(\beta_0'X'G'MGX\beta_0 + tr(G'G\Omega_0(\gamma)))}{tr^2(\Omega_0(\gamma))} + \frac{8tr^3(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} \\
& -\frac{2}{n}tr(G^3) + o(1)
\end{aligned}
\tag{S.1.4}
$$

and

$$
\mathbb{E}(l^{(1)}(\lambda_0)^2) = \frac{tr^2(G\Omega_0(\gamma))}{tr^2(\Omega_0(\gamma))} + \frac{1}{n^2}tr^2(G) - \frac{2}{n}tr(G)\frac{tr(G\Omega_0(\gamma))}{tr(\Omega_0(\gamma))} + o(1). \tag{S.1.5}
$$

Let $B(\gamma, \lambda_0) = \mathbb{E}(\hat{\lambda}_{QML}) - \lambda_0$. From these calculations and Bao (2013), we deduce the following result.

Corollary S1 *Let ϵ be a vector of n independent random variables, normally distributed and such that $\mathbb{E}(\epsilon\epsilon') = \Omega_0(\gamma)$, where $\Omega_0(\gamma)$ is defined in (2.7) in the manuscript with $\sigma^2 = 1$. Let Assumptions 2-4, reported in the manuscript, hold. The leading term of $B(\gamma, \lambda_0)$ is given by*

$$
\begin{aligned}
B(\gamma, \lambda_0) = & -2\left(\mathbb{E}(l^{(2)}(\lambda_0))\right)^{-1}\mathbb{E}(l^{(1)}(\lambda_0)) + \left(\mathbb{E}(l^{(2)}(\lambda_0))\right)^{-2}\mathbb{E}(l^{(2)}(\lambda_0)l^{(1)}(\lambda_0)) \\
& -\frac{1}{2}\left(\mathbb{E}(l^{(2)}(\lambda_0))\right)^{-3}\mathbb{E}(l^{(3)}(\lambda_0))\mathbb{E}(l^{(1)}(\lambda_0)^2).
\end{aligned}
\tag{S.1.6}
$$

Under $\Omega_0(\gamma)$ in (2.7), terms in (S.1.1), (S.1.3) and (S.1.5) do not vanish as n increases, unless $\gamma = 0$ (i.e. the homoskedastic case) and/or some specific structure of W is imposed which ensures that a condition related to (2.8) in the manuscript holds. Given the likelihood function (2.3) in the manuscript, the calculation of (S.1.1)-(S.1.4) is based on the explicit computation of moments of ratio of quadratic form. Most of the moments of ratios involved are indeed exactly ratio of moments, as ratios of the form $\epsilon'A\epsilon/\epsilon'M_X\epsilon$ for a generic $n \times n$ matrix A are independent of $\epsilon'M_X\epsilon$[1]. However, since we are only interested in the leading terms of (S.1.6), we can approximate moments of ratios as ratios of moments even when the independence conditions fails. The computation of moments is standard (Bao and Ullah (2007)) and details are omitted here.

S.2 Proofs of the Theorems

Proof of Theorem 1:

Proof of part (i). Let ψ_{ij} and $\tilde{\psi}_{ij}$ be the 2×1 vectors defined as $\psi_{ij} = (\ \psi_{1ij} \quad \psi_{2ij}\ )' = (\ (P + P')_{ij}/2 \quad (Q'Q)_{ij}\ )'$ and $\tilde{\psi}_{ij} = (\ \tilde{\psi}_{1ij} \quad \tilde{\psi}_{2ij}\ )' = (\ (M_XP')_{ij} \quad (M_XQ'Q)_{ij}\ )'$, respectively.

After showing

$$
U_n = \frac{1}{\sqrt{n}}\begin{pmatrix} \epsilon'P\epsilon - tr(P\tilde{\Omega}) + 2\beta_0'X'PM_X\epsilon \\ \epsilon'Q'Q\epsilon - tr(Q'Q\tilde{\Omega}) + 2\beta_0'X'Q'QM_X\epsilon \end{pmatrix} + o_p(1), \tag{S.2.1}
$$

[1]See, for example, Conniffe and Spencer (2001), for an analysis and history of this result on ratios of quadratic forms and other moments.

as reported in the manuscript, the rest of the proof is similar to KPR (2017). In order to avoid repetition we refer to their proof when steps follow in a similar way.

Define

$$u_i = \begin{pmatrix} u_{1i} & u_{2i} \end{pmatrix}' = 2\epsilon_i \sum_j \tilde{\psi}_{ij} X_j' \beta_0 + 2\epsilon_i \sum_{j<i} \psi_{ij}\epsilon_j, \tag{S.2.2}$$

so that $\sqrt{n}U_n = \sum_{i=1}^n u_i + o_p(1)$, according to (S.2.1). The $\{u_i, 1 \leq i \leq n, n = 1, 2, ...\}$ form a triangular array of martingale differences with respect to the filtration formed by the σ-field generated by $\{\epsilon_j; j < i\}$. Let

$$A = Var\left(\sum_{i=1}^n u_i\right) = 4\sum_{i=1}^n \sigma_i^2 \sum_{j=1}^n \sum_{t=1}^n \tilde{\psi}_{ij} X_j' \beta_0 \beta_0' X_t \tilde{\psi}_{it}' + 4\sum_{i=1}^n \sum_{j<i} \sigma_i^2 \sigma_j^2 \psi_{ij} \psi_{ij}'. \tag{S.2.3}$$

Define $z_{in} = \eta' A^{-1/2} u_i$, where η is a 2×1 vector satisfying $\eta'\eta = 1$. By Theorem 2 of Scott (1973) $\sum_i^n z_{in} \to_d \mathcal{N}(0, 1)$ if the following stability and Lindeberg conditions hold:

$$\sum_{i=1}^n \mathbb{E}(z_{in}^2 | \epsilon_j; j < i) \xrightarrow{p} 1, \tag{S.2.4}$$

and

$$\sum_{i=1}^n \mathbb{E}(z_{in}^2 1(|z_{in} > \xi|)) \to 0 \quad \forall \xi > 0. \tag{S.2.5}$$

As $n \to \infty$,

$$A/n \to \lim_{n\to\infty} V_n, \tag{S.2.6}$$

where

$$\begin{aligned} V_n = {} & \frac{4}{n} \begin{pmatrix} \beta_0' X' P M_X \Omega_0 M_X P' X \beta_0 & \beta_0' X' P M_X \Omega_0 M_X Q' Q X \beta_0 \\ \beta_0' X' Q' Q M_X \Omega_0 M_X P' X \beta_0 & \beta_0' X' Q' Q M_X \Omega_0 M_X Q' Q X \beta_0 \end{pmatrix} \\ & + \frac{4}{n} \sum_i \sum_{j<i} \sigma_i^2 \sigma_j^2 \begin{pmatrix} \frac{(P+P')_{ij}^2}{4} & \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} \\ \frac{(P+P')_{ij}(Q'Q)_{ij}}{2} & (Q'Q)_{ij}^2 \end{pmatrix} \\ = {} & C_1 + C_2, . \end{aligned} \tag{S.2.7}$$

where C_1 and C_2 contain the first and second terms in (S.2.7), respectively. All terms in C_1 are $O(1)$, while those in C_2 are bounded by $O(1/h)$ under Assumptions 3 and 4, and by standard algebra. Existence of limits in (S.2.7) is guaranteed under Assumption 7, and non singularity of C_1 is ensured by Assumptions 2, 3(ii) and 5. Thus, we can replace A by n when showing (S.2.4) and (S.2.5).

We start by establishing (S.2.4), which can equivalently be written as

$$\sum_i \mathbb{E}\left(z_{in}^2|\epsilon_j, j<i\right) - \eta' A^{-1/2} A A^{-1/2}\eta \underset{p}{\rightarrow} 0. \tag{S.2.8}$$

The latter, by standard manipulations and (S.2.6), is equivalent to showing

$$\frac{4}{n}\eta'\left(\sum_i \sigma_i^2\left(\sum_{j<i}\epsilon_j\psi_{ij}\right)\left(\sum_{j<i}\epsilon_j\psi_{ij}\right)' - \sum_i\sum_{j<i}\sigma_i^2\sigma_j^2\psi_{ij}\psi_{ij}'\right)\eta \underset{p}{\rightarrow} 0, \tag{S.2.9}$$

and

$$\frac{4}{n}\eta'\left(\sum_i\sum_j\sum_{t<i}\sigma_i^2\beta_0' X_j\left(\tilde{\psi}_{ij}\psi_{it}' + \psi_{it}\tilde{\psi}_{ij}'\right)\epsilon_t\right)\eta \underset{p}{\rightarrow} 0 \tag{S.2.10}$$

as $n \to \infty$.

In order to avoid replications, we omit the proof of (S.2.9), referring to KPR and observing that

$$||P||_\infty + ||P'||_\infty < K, \quad ||Q||_\infty + ||Q'||_\infty < \infty \tag{S.2.11}$$

and both P_{ij} and Q_{ij}, for $i,j = 1,....,n$, are uniformly bounded by $O(1/h)$, so that ψ_{1ij} and ψ_{2ij} have, respectively, similar asymptotic properties to $(G+G')_{ij}/2$ and $(G'G)_{ij}$ appearing in the proof of Theorem 1 in KPR. We verify (S.2.10) by examining the convergence of each typical element, i.e. by showing

$$\frac{1}{n}\sum_i\sum_j\sum_{t<i}\sigma_i^2\beta_0' X_j\tilde{\psi}_{sij}\psi_{vit}\epsilon_t \underset{p}{\rightarrow} 0 \tag{S.2.12}$$

for each $s, v = 1, 2$. Under Assumption 5, i.e. for uniformly bounded X_{ij} for $i,j = 1,....,n$, the left hand side (LHS) of (S.2.12) has mean zero and variance bounded by

$$\begin{aligned}&\frac{1}{n^2}K|\sum_i\sum_j\sum_u\sum_h\sum_{t<i,u}\tilde{\psi}_{sij}\tilde{\psi}_{suh}\psi_{vit}\psi_{vut}| \leq \frac{1}{n^2}K\sum_i\sum_j\sum_u\sum_h\sum_t|\tilde{\psi}_{sij}\tilde{\psi}_{suh}\psi_{vit}\psi_{vut}| \\ &\frac{1}{n}K\sup_{0<i\leq n}\sum_j|\tilde{\psi}_{sij}|\sup_{0<u\leq n}\sum_h|\tilde{\psi}_{shu}|\sup_{0<t\leq n}\sum_i|\psi_{vit}|\sup_{0<u\leq n}\sum_t|\psi_{vut}| = O\left(\frac{1}{n}\right),\end{aligned} \tag{S.2.13}$$

since (S.2.11) holds and

$$||M_X P||_\infty + ||P' M_X||_\infty < K, \quad ||M_X Q'Q||_\infty + ||Q'QM_X||_\infty < \infty. \tag{S.2.14}$$

In order to prove (S.2.5) we verify the sufficient Lyapunov condition

$$\sum_{i=1}^{n} \mathbb{E}|z_{in}|^{2+\delta} \to 0 \tag{S.2.15}$$

by considering a typical standardized element of u_i, i.e. $\sum_i \mathbb{E}|(1/n)^{1/2}u_{si}|^{2+\delta}$ for $s = 1, 2$. Under Assumption 1, using $\sum_i \mathbb{E}|u_{si}|^{2+\delta} = \sum_i \mathbb{E}(\mathbb{E}|u_{si}|^{2+\delta}|\epsilon_j, j < i))$ and the c_r inequality,

$$\left(\frac{1}{n}\right)^{1+\delta/2} \sum_i \mathbb{E}|u_{si}|^{2+\delta} \leq \left(\frac{1}{n}\right)^{1+\delta/2} K \sum_i \mathbb{E}|\sum_{j<i} \psi_{sij}\epsilon_j|^{2+\delta} + \left(\frac{1}{n}\right)^{1+\delta/2} K \sum_i |\sum_j \beta_0' X_j \tilde{\psi}_{sij}|^{2+\delta}. \tag{S.2.16}$$

Convergence to zero of the first term on the right hand side (RHS) of (S.2.16) can be shown as in KPR. Convergence of the third term on the RHS of (S.2.16) can be shown after observing that

$$|\sum_j \beta_0' X_j \tilde{\psi}_{sij}|^{2+\delta} \leq K \sup_{0<j\leq n} |\beta_0' X_j|^{2+\delta} \sum_j |\tilde{\psi}_{sij}|^{2+\delta}, \tag{S.2.17}$$

where $\beta_0' X_j$ is uniformly bounded under Assumption 5. Thus, the second term on the RHS of (S.2.16) is bounded by

$$\begin{aligned}
\left(\frac{1}{n}\right)^{1+\delta/2} K \sum_i \sum_j |\tilde{\psi}_{sij}|^{2+\delta} &\leq \left(\frac{1}{n}\right)^{1+\delta/2} K \sum_i \left(\sum_j \tilde{\psi}_{sij}^2\right)^{1+\delta/2} \\
\leq \left(\frac{1}{n}\right)^{1+\delta/2} K \left(\sup_i \sum_j \tilde{\psi}_{sij}\right)^{\delta/2} &\sum_i \sum_j \tilde{\psi}_{sij}^2 = O\left(\frac{1}{n}\right)^{\delta/2}
\end{aligned} \tag{S.2.18}$$

similarly to KPR, under Assumptions 3-5.

Thus, $A^{-1/2} \sum_i u_i \underset{d}{\to} \mathcal{N}(0, I)$, and the statement in Theorem 1(i) follows by standard delta arguments.

Proof of part (ii). Again, we proceed similarly to KPR and we refer to their proof to avoid repetitions. We rewrite the binding function $\tau_n(\lambda)$ as

$$\begin{aligned}
\tau_n(\lambda, \Omega_\lambda, \hat{\beta}(\lambda)) &= \frac{tr(P(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)' X' P(\lambda) X \hat{\beta}(\lambda)}{tr(Q(\lambda)' Q(\lambda)\Omega_\lambda) + \hat{\beta}(\lambda)' X' Q(\lambda)' Q(\lambda) X \hat{\beta}(\lambda)} + O_p\left(\frac{1}{n}\right) \\
&= \frac{a(\lambda) + b(\lambda)}{c(\lambda) + d(\lambda)} + O_p\left(\frac{1}{n}\right),
\end{aligned} \tag{S.2.19}$$

where

$$a(\lambda) = \frac{1}{n}tr(P(\lambda)\Omega_\lambda), \quad b(\lambda) = \frac{1}{n}\hat{\beta}(\lambda)'X'P(\lambda)X\hat{\beta}(\lambda), \quad c(\lambda) = \frac{1}{n}tr(Q(\lambda)'Q(\lambda)\Omega_\lambda),$$
$$d = \frac{1}{n}\hat{\beta}(\lambda)'X'Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda). \tag{S.2.20}$$

We write

$$\tau_n^{(1)}(\lambda) = \frac{a^{(1)}(\lambda) + b^{(1)}(\lambda)}{c(\lambda) + d(\lambda)} - \frac{(c^{(1)}(\lambda) + d^{(1)}(\lambda))(a(\lambda) + b(\lambda))}{(c(\lambda) + d(\lambda))^2} + O\left(\frac{1}{n}\right), \tag{S.2.21}$$

where

$$\begin{aligned}
a^{(1)}(\lambda) &= \frac{1}{n}tr(G'(\lambda)P(\lambda)\Omega_\lambda) + \frac{1}{n}tr(P(\lambda)G(\lambda)\Omega_\lambda) + \frac{1}{n}tr(P\Omega_\lambda^{(1)}),\\
b^{(1)}(\lambda) &= -\frac{2}{n}y'W'(I_n - M_X)P(\lambda)X\hat{\beta}(\lambda) + \frac{1}{n}\hat{\beta}(\lambda)'X'G(\lambda)'P(\lambda)X\hat{\beta}(\lambda) + \frac{1}{n}\hat{\beta}(\lambda)'X'P(\lambda)G(\lambda)X\hat{\beta}(\lambda),\\
c^{(1)}(\lambda) &= \frac{2}{n}tr(G(\lambda)'Q(\lambda)'Q(\lambda)\Omega_\lambda) + \frac{1}{n}tr(Q(\lambda)'Q(\lambda)\Omega_\lambda^{(1)}),\\
d^{(1)}(\lambda) &= -\frac{2}{n}y'W'(I - M_X)Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda) + \frac{2}{n}\hat{\beta}(\lambda)'X'G(\lambda)'Q(\lambda)'Q(\lambda)X\hat{\beta}(\lambda)
\end{aligned} \tag{S.2.22}$$

and

$$\Omega_\lambda^{(1)} = -2\text{diag}(M_X W y \epsilon(\lambda)'). \tag{S.2.23}$$

Since

$$\hat{\lambda}_{CUII} - \lambda_0 = \tau_n^{-1}(\hat{\lambda}) - \tau_n^{-1}(\tau_n(\lambda_0)), \tag{S.2.24}$$

we can derive the limit distribution of $\sqrt{n}(\hat{\lambda}_{CUII} - \lambda_0)$ by the delta method, as long as the asymptotic local relative equicontinuity condition (Phillips, 2012) holds. Thus, similar to KPR, we need to show

$$\left| \frac{\tau_n^{(1)}(\lambda_0) - \tau_n^{(1)}(r)}{\tau_n^{(1)}(r)} \right| \underset{p}{\rightarrow} 0 \tag{S.2.25}$$

as $n \to \infty$, uniformly in $\mathcal{N}_\delta = \{r \in \Re : |s(r - \lambda_0)| < \delta, \ \ \delta > 0\}$, $s = s_n \to \infty$ and $s(1/n)^{1/2} \to 0$. Under Assumption 6(ii), the expression on the LHS of (S.2.25) is bounded by

$$K \left| \tau_n^{(1)}(\lambda_0) - \tau_n^{(1)}(r) \right|, \tag{S.2.26}$$

which by the mean value theorem is in turn bounded by

$$K\left|\tau_n^{(2)}(\lambda^*)(\lambda_0 - r)\right|, \tag{S.2.27}$$

where λ^* is an intermediate point between λ_0 and r. The expression in (S.2.27) is $O_p(|\lambda_0 - r|) = O_p(s^{-1})$ as long as

$$\tau_n^{(2)}(\lambda^*) = O_p(1), \tag{S.2.28}$$

which holds under Assumptions 3-5, a derivation of which will be supplied on request.

Therefore, by a delta argument we conclude that

$$\sqrt{n}\tau_n^{(1)}(\hat{\lambda}_{CUII} - \lambda_0) \underset{d}{\rightarrow} \mathcal{N}(0, \bar{f}' \lim_{n\to\infty} V_n \bar{f}), \tag{S.2.29}$$

where V_n and $\bar{f}_n$ are defined in (4.4) and (4.11), respectively. The statement in Theorem 1 follows by standard algebra once we write

$$\bar{\tau}^{(1)} = \bar{\tau}^{(1)}(\lambda_0) = \text{p} \quad \lim_{n\to\infty} \tau_n^{(1)}(\lambda_0), \tag{S.2.30}$$

in terms of $\bar{a}^{(1)}$, $\bar{b}^{(1)}$, $\bar{c}^{(1)}$ and $\bar{d}^{(1)}$. $\bar{\tau}$ exists and is non singular under Assumption 7(ii).

Proof of Theorem 2:

In order to prove (A.8) in the manuscript, we need to show

$$\frac{1}{n}\sum_i \sum_{j<i} (\epsilon_i^2\epsilon_j^2 - \sigma_i^2\sigma_j^2)\psi_{sij}\psi_{tij} = o_p(1), \tag{S.2.31}$$

$$\frac{1}{n}\sum_i \sum_{j<i} (\hat{\epsilon}_i^2\hat{\epsilon}_j^2 - \epsilon_i^2\epsilon_j^2)\psi_{sij}\psi_{tij} = o_p(1) \tag{S.2.32}$$

and

$$\frac{1}{n}\sum_i \sum_{j<i} \hat{\epsilon}_i^2\hat{\epsilon}_j^2(\hat{\psi}_{sij}\hat{\psi}_{tij} - \psi_{sij}\psi_{tij}) = o_p(1). \tag{S.2.33}$$

We start by (S.2.31). We have, for $s, t = 1, 2$

$$\begin{aligned}\frac{1}{n}\sum_i \sum_{j<i} (\epsilon_i^2\epsilon_j^2 - \sigma_i^2\sigma_j^2)\psi_{sij}\psi_{tij} &= \frac{1}{n}\sum_i \sum_{j<i} \psi_{sij}\psi_{tij}(\epsilon_i^2 - \sigma_i^2)(\epsilon_j^2 - \sigma_j^2) + \frac{1}{n}\sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\sigma_i^2(\epsilon_j^2 - \sigma_j^2) \\ &+ \frac{1}{n}\sum_i \sum_{j<i} \psi_{sij}\psi_{tij}\sigma_j^2(\epsilon_i^2 - \sigma_i^2). \end{aligned} \tag{S.2.34}$$

The first term on the RHS of (S.2.34) has mean zero and variance bounded by

$$\frac{C}{n^2}\sum_i\sum_{j<i}\psi_{sij}^2\psi_{tij}^2 \le \frac{C}{n^2}\sum_i\sum_j\psi_{sij}^2\psi_{tij}^2 \le \frac{C}{n^2h^2}\sum_i\sum_j\psi_{tij}^2 = O\left(\frac{1}{nh^3}\right) \tag{S.2.35}$$

since

$$\sum_i\sum_j\psi_{tij}^2 = tr(\Psi_t^2) = O\left(\frac{n}{h}\right)$$

for $t = 1, 2$. The second term on the RHS of (S.2.34) has mean zero and variance bounded by

$$\begin{aligned}
&\frac{C}{n^2}\sum_i\sum_j\sum_u|\psi_{sij}\psi_{tij}\psi_{suj}\psi_{tuj}| \le \frac{C}{n^2h^2}\sum_i\sum_j\sum_u|\psi_{sij}||\psi_{tuj}| \\
&\le \frac{C}{nh^2}\sup_j\sum_i|\psi_{sij}|\sup_u\sum_j|\psi_{sij}| = O\left(\frac{1}{nh^2}\right).
\end{aligned} \tag{S.2.36}$$

Similarly, we can show that the third term on the RHS of (S.2.34) converges to zero in quadratic mean. By Markov's inequality (S.2.31) follows.

In order to show (S.2.32) we write

$$\hat{\epsilon}_i = \epsilon_i - \sum_j B_{ij}\epsilon_j - (\hat{\lambda}_{CUII} - \lambda_0)Q_i'X\beta - (\hat{\lambda}_{CUII} - \lambda_0)Q_i'\epsilon, \tag{S.2.37}$$

where Q_i' is the $1 \times n$ vector displaying the $i-$th row of Q and $B_{ij} = X_i'(X'X)^{-1}X_j$, as defined at the beginning of the proof of Theorem 1. By standard arguments, we can show that the last two terms on the RHS of (S.2.37) are bounded in probability by $1/\sqrt{n}$, uniformly in i. Let

$$\hat{v}_i = \hat{\epsilon}_i - \epsilon_i = -\sum_k B_{ik}\epsilon_k + O_p\left(\frac{1}{\sqrt{n}}\right). \tag{S.2.38}$$

Thus, (S.2.32) is equivalent to

$$\frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}(\hat{v}_i\hat{v}_j + \epsilon_i\hat{v}_j + \epsilon_j\hat{v}_i)(\hat{v}_i\hat{v}_j + \hat{v}_i\epsilon_j + \epsilon_i\hat{v}_j + 2\epsilon_i\epsilon_j) = o_p(1), \tag{S.2.39}$$

as $n \to \infty$. We therefore need to show, as $n \to \infty$, that

$$\frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}\hat{v}_i^2\hat{v}_j^2 = o_p(1), \tag{S.2.40}$$

$$\frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}\hat{v}_i^2\hat{v}_j\epsilon_j = o_p(1), \tag{S.2.41}$$

$$\frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}\hat{v}_i\hat{v}_j\epsilon_i\epsilon_j = o_p(1), \tag{S.2.42}$$

$$\frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}\hat{v}_i^2\epsilon_j^2 = o_p(1), \tag{S.2.43}$$

$$\frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}\hat{v}_j\epsilon_i^2\epsilon_j = o_p(1). \tag{S.2.44}$$

We only consider the leading term in $\hat{v}_i$ in (S.2.38) when showing (S.2.40)- (S.2.48), but similar routine arguments can be applied to deal with higher order terms.

The modulus of the left hand side (LHS) of (S.2.40) has expectation bounded by

$$\begin{aligned}&\frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}||\psi_{tij}|\mathbb{E}\left(\hat{v}_i^4\right)^{1/2}\mathbb{E}\left(\hat{v}_j^4\right)^{1/2} \leq \frac{C}{n}\sum_i\sum_j|\psi_{sij}||\psi_{tij}|\left(\sum_v B_{iv}^2\right)\left(\sum_h B_{jh}^2\right)\\ &\leq \frac{C}{n}\sum_i\sum_j|\psi_{sij}||\psi_{tij}|B_{ii}B_{jj} \leq \frac{C}{nh^2}\sum_i\sum_j B_{ii}B_{jj} = O\left(\frac{1}{h^2n}\right).\end{aligned} \tag{S.2.45}$$

Similarly, the modulus of the LHS of (S.2.41) has expectation bounded by

$$\begin{aligned}&\frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}||\psi_{tij}|\left(\mathbb{E}\hat{v}_j^4\right)^{1/4}\left(\mathbb{E}\hat{v}_i^4\right)^{1/2}\left(\mathbb{E}\epsilon_j^4\right)^{1/4} \leq \frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}||\psi_{tij}|\left(\sum_v B_{jv}^2\right)^{1/2}\left(\sum_h B_{ih}^2\right)\\ &\leq \frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}||\psi_{tij}|B_{jj}^{1/2}B_{ii} \leq \frac{C}{nh}\sum_i\sum_j|\psi_{sij}|B_{ii} \leq \frac{C}{nh}\sup_i\sum_j|\psi_{sij}|\sum_i B_{ii} = O\left(\frac{1}{nh}\right),\end{aligned} \tag{S.2.46}$$

as $B_{jj}^{1/2} < 1$. The modulus of the LHS of (S.2.42) has expectation bounded by

$$\begin{aligned}&\frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}||\psi_{tij}|\left(\mathbb{E}\hat{v}_i^4\right)^{1/4}\left(\mathbb{E}\hat{v}_j^4\right)^{1/4}\left(\mathbb{E}\epsilon_j^4\right)^{1/4}\left(\mathbb{E}\epsilon_i^4\right)^{1/4} \leq \frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}||\psi_{tij}|B_{ii}^{1/2}B_{jj}^{1/2}\\ &\frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}||\psi_{tij}|(B_{ii}+B_{jj}) \leq \frac{C}{nh}\left(\sup_i\sum_j|\psi_{sij}|\sum_i B_{ii} + \sup_j\sum_i|\psi_{sij}|\sum_j B_{jj}\right) = O\left(\frac{1}{nh}\right).\end{aligned} \tag{S.2.47}$$

(S.2.43) can be shown by similar arguments as (S.2.40)-(S.2.42), while (S.2.48) can be written as

$$\frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}B_{ji}\epsilon_i^3\epsilon_j + \frac{1}{n}\sum_i\sum_{j<i}\psi_{sij}\psi_{tij}\epsilon_i^2\epsilon_j^2B_{jj} + \frac{1}{n}\sum_i\sum_{j<i}\sum_{u\neq j,i}\psi_{sij}\psi_{tij}\epsilon_i^2\epsilon_j\epsilon_uB_{ju}. \tag{S.2.48}$$

The modulus of the first term in the last displayed expression has expectation bounded by

$$\frac{C}{n}\sum_i\sum_{j<i}|\psi_{sij}|\psi_{tij}||B_{ij}| \leq \frac{C}{n}\sum_i\sum_j|\psi_{sij}|\psi_{tij}|(B_{ii}+B_{jj}) = O\left(\frac{1}{hn}\right), \tag{S.2.49}$$

as in previous calculations. Similarly, the second term in (S.2.48) is $O(1/nh)$, while the third term has mean zero and variance bounded by

$$\frac{C}{n^2}\sum_i\sum_j\sum_u\sum_l|\psi_{sij}\psi_{tij}\psi_{sil}\psi_{til}|B_{uj}^2+\frac{C}{n^2}\sum_i\sum_j\sum_k\sum_l|\psi_{sij}\psi_{tij}\psi_{skl}\psi_{tkl}|B_{lj}^2$$
$$\frac{C}{n^2}\sum_i\sum_j\sum_l|\psi_{sij}\psi_{tij}\psi_{sil}\psi_{til}|B_{jj}+\frac{C}{n^2}\sum_i\sum_j\sum_k\sum_l|\psi_{sij}\psi_{tij}\psi_{skl}\psi_{tkl}|B_{jl}^2. \qquad \text{(S.2.50)}$$

Proceeding as before, the first term in the last displayed expression is bounded by $O(1/n^2h^2)$, while the second one is bounded by $O(1/nh^2)$. By Markov's inequality, this concludes the proof of (S.2.32).

In order to show (S.2.33) we apply a standard mean value theorem argument, such as

$$\frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2(\hat{\psi}_{sij}\hat{\psi}_{tij}-\psi_{sij}\psi_{tij})=\frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2\left(\bar{\psi}_{sij}(\hat{\psi}_{tij}-\psi_{tij})+\bar{\psi}_{tij}(\hat{\psi}_{sij}-\psi_{sij})\right), \qquad \text{(S.2.51)}$$

where $\bar{\psi}_{sij}$ (or $\bar{\psi}_{tij}$) is an intermediate point between $\hat{\psi}_{sij}$ and ψ_{sij}. From Theorem 1, $\hat{\psi}_{sij}-\psi_{sij}=O_p(1/\sqrt{n})$ and thus $\bar{\psi}_{sij}-\psi_{sij}=o_p(1)$. Therefore, (S.2.51) is bounded by

$$\sup_{i,j}|\hat{\psi}_{sij}-\psi_{sij}|\frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2|\psi_{tij}|. \qquad \text{(S.2.52)}$$

By similar arguments to those applied to prove (S.2.31) and (S.2.32), we conclude that as $n\to\infty$

$$\frac{1}{n}\sum_i\sum_{j<i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2|\psi_{tij}|\underset{p}{\rightarrow}\lim\frac{1}{n}\sum_i\sum_{j<i}\sigma_i^2\sigma_j^2|\psi_{tij}|, \qquad \text{(S.2.53)}$$

which is $O(1)$ in the limit. Thus, (S.2.52) is $O_p(1/\sqrt{n})$, concluding the proof of (A.8).

S.3 Additional simulation results

This section reports additional simulation results to support the discussion in Section 7 of the paper. Results in Tables S1 and S2 have been obtained using a symmetric, randomly generated matrix of zeros and ones, where the number of ones is restricted to be 20% of the total entries. The resulting matrix is then normalized so that each row sums to 1. As discussed in the manuscript, W is generated once for each n and is kept fixed across scenarios. Table S1 contains results for σ_i generated as in (7.2) in the manuscript, while Table S2 displays values for σ_i generated from $\chi^2(5)$.

Tables S3 and S4 have been obtained by setting $\beta_0=(2,1.5,-1)$ and X being $n\times 3$, with the first regressor being an $n\times 1$ column of ones and the other two being randomly drawn from two independent uniform distributions on the support $[0,4]$. The rest of the design is identical to that described in

Section 7 in the main manuscript. In both S3 and S4 W is 'exponential', with S3 corresponding to σ_i generated as in (7.2) in the manuscript, with S4 displaying values for σ_i generated from $\chi^2(5)$.

Tables S5 and S6 report results for CUII, QML, MQML and RGMM when the true data generating process is a pure SAR, while the estimated model is a SARX with intercept and one exogenous regressor which is drawn from a uniform distribution on the support $[0, 1]$. In both S5 and S6 W is 'exponential', with S5 corresponding to σ_i generated as in (7.2) in the manuscript, with S6 displaying values for σ_i generated from $\chi^2(5)$. The rest of the design is identical to that described in Section 7 of the manuscript.

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0929	0.2956	-0.0476	0.1763	-0.0064	0.1307	-0.0156	0.1311
	0.3	0.0110	0.2437	0.0193	0.1760	0.0073	0.1376	0.0029	0.1333
	0.5	0.0474	0.2298	0.0419	0.1854	0.0477	0.1405	0.0061	0.1394
	0.8	0.1142	0.2000	0.0550	0.1526	0.0332	0.1230	0.0385	0.1235
ML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0322	0.1031	-0.0833	0.1068	-0.0686	0.1056	-0.0806	0.1162
	0.3	-0.1788	0.1403	-0.1713	0.1286	-0.1725	0.1166	-0.1680	0.1134
	0.5	-0.2266	0.1484	-0.1855	0.1202	-0.1839	0.1023	-0.2093	0.1191
	0.8	-0.2760	0.1486	-0.2629	0.1299	-0.2757	0.1235	-0.2686	0.1245
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.0508	0.1425	0.0127	0.1187	0.0165	0.1156	-0.0035	0.1244
	0.3	-0.0281	0.1423	-0.0073	0.1308	-0.0084	0.1181	-0.0084	0.1199
	0.5	-0.0261	0.1393	-0.0206	0.1283	0.0120	0.1109	-0.0127	0.1241
	0.8	-0.0136	0.1173	-0.0286	0.1093	-0.0205	0.1011	0.0060	0.1094
2SLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.6496	3.3561	-0.7360	6.4335	-0.7703	11.0633	-0.3523	17.3900
	0.3	-0.2990	3.6600	0.3778	4.4825	-0.1449	7.5171	0.0250	11.5254
	0.5	0.0666	3.7634	0.2094	4.2141	0.1665	6.2116	0.3013	10.6641
	0.8	0.3420	2.0216	0.2889	2.7892	0.2744	3.8288	0.1160	5.1442
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.3042	0.8991	-0.1418	0.2627	-0.0892	0.1509	-0.0956	0.1434
	0.3	-0.1103	0.6274	-0.0616	0.4327	-0.1353	0.5117	-0.1319	0.4633
	0.5	-0.0825	0.5744	-0.0103	0.9525	-0.1008	0.4841	-0.1327	0.6457
	0.8	0.0582	0.9081	0.0306	0.8375	-0.0524	0.8867	-0.0916	2.6146

Table S1: Bias & MSE of CUII, QML, MQML, 2SLS and RGMM estimators for 'random' W. The ϵ_is are defined as in (7.1) with $\zeta_i \sim iid \;\; t(5)$ and σ_i defined as in (7.2). The design corresponds to an artificially dense choice of W.

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0464	0.1597	-0.0079	0.1646	-0.0100	0.1352	-0.0106	0.1193
	0.3	-0.0181	0.1473	-0.0118	0.1411	0.0032	0.1315	0.0087	0.1349
	0.5	0.0234	0.1435	0.0126	0.1353	0.0094	0.1307	0.0240	0.1298
	0.8	0.0126	0.1401	0.0351	0.1329	0.0272	0.1226	-0.0026	0.1196
QML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0264	0.0806	-0.0200	0.1036	-0.0582	0.1073	-0.0757	0.1063
	0.3	-0.1866	0.1208	-0.1706	0.1144	-0.1679	0.1087	-0.1601	0.1130
	0.5	-0.1662	0.1092	-0.1911	0.1111	-0.2081	0.1135	-0.1909	0.1081
	0.8	-0.2536	0.1320	-0.2397	0.1114	-0.2690	0.1192	-0.2919	0.1344
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.0258	0.0967	0.0429	0.1219	0.0162	0.1187	0.0016	0.1133
	0.3	-0.0097	0.1092	-0.0240	0.1134	-0.0052	0.1140	0.0039	0.1249
	0.5	-0.0034	0.1076	-0.0167	0.1055	-0.0090	0.1115	0.0120	0.1166
	0.8	-0.0361	0.1007	-0.0166	0.1017	-0.0096	0.0996	-0.0257	0.1067
2SLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.2671	1.9420	-0.1351	5.3380	-0.9920	12.0616	-1.0292	22.5435
	0.3	-0.1673	2.3131	-0.0803	4.4500	-0.5362	8.1619	0.0281	25.9411
	0.5	0.0434	2.9366	0.3936	5.4701	0.1937	7.4490	0.2233	15.9209
	0.8	0.2173	1.0161	0.2689	1.9738	0.0910	6.4317	0.0224	8.4702
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1750	0.6055	-0.0583	0.2043	-0.0973	0.1515	-0.1020	0.1471
	0.3	-0.1162	0.5475	-0.1183	0.7414	-0.1641	0.2754	-0.1658	0.2963
	0.5	-0.0365	0.6129	-0.0125	0.8190	-0.1210	0.7283	-0.1509	0.6385
	0.8	0.0011	0.7205	0.0344	0.8222	-0.1000	1.1082	-0.1832	1.4971

Table S2: Bias & MSE of CUII, QML, MQML, 2SLS and RGMM estimators for 'random' W. The ϵ_is are defined as in (7.1) with $\zeta_i \sim iidN(0,1)$ and $\sigma_i \sim \chi^2(5)$. The design corresponds to an artificially dense choice of W.

		$n=30$		$n=50$		$n=100$		$n=200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0293	0.0408	-0.0293	0.0451	-0.0176	0.0243	-0.0156	0.0161
	0.3	-0.0162	0.0119	-0.0195	0.0113	-0.0104	0.0083	-0.0149	0.0091
	0.5	-0.0140	0.0139	-0.0130	0.0060	-0.0061	0.0070	-0.0117	0.0056
	0.8	-0.0119	0.0036	-0.0114	0.0024	-0.0063	0.0018	-0.0044	0.0007
QML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0463	0.0416	-0.0475	0.0449	-0.0331	0.0254	-0.0242	0.0166
	0.3	-0.0254	0.0126	-0.0257	0.0118	-0.0149	0.0086	-0.0181	0.0093
	0.5	-0.0286	0.0148	-0.0185	0.0064	-0.0085	0.0073	-0.0129	0.0057
	0.8	-0.0175	0.0040	-0.0139	0.0026	-0.0070	0.0019	-0.0043	0.0007
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0108	0.0490	-0.0215	0.0440	-0.0141	0.0239	-0.0144	0.0161
	0.3	-0.0423	0.0815	-0.0275	0.0332	-0.0112	0.0083	-0.0153	0.0091
	0.5	-0.0205	0.0257	-0.0464	0.1466	-0.0077	0.0071	-0.0125	0.0056
	0.8	-0.1472	1.4401	-0.0132	0.0026	-0.0082	0.0019	-0.0047	0.0008
2SLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.0031	0.0563	0.0103	0.0654	0.0059	0.0359	0.0039	0.0227
	0.3	0.0031	0.0124	-0.0086	0.0131	0.0094	0.0097	-0.0019	0.0105
	0.5	0.0093	0.0165	-0.0002	0.0059	0.0106	0.0087	-0.0033	0.0062
	0.8	0.0043	0.0036	-0.0030	0.0025	0.0034	0.0022	-0.0001	0.0008
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0221	0.0439	-0.0266	0.0509	-0.0143	0.0273	-0.0095	0.0184
	0.3	-0.0121	0.0124	-0.0086	0.0131	-0.0074	0.0091	-0.0132	0.0100
	0.5	-0.0069	0.0151	-0.0123	0.0065	-0.0055	0.0083	-0.0116	0.0061
	0.8	-0.0110	0.0043	-0.0104	0.0030	-0.0045	0.0027	-0.0030	0.0007
CUGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0067	0.0940	-0.0063	0.0388	-0.0021	0.0246	-0.0101	0.0233
	0.3	-0.0063	0.0080	-0.0104	0.0177	-0.0088	0.0094	-0.0046	0.0073
	0.5	-0.0078	0.0067	-0.0081	0.0060	-0.0066	0.0039	-0.0086	0.0046
	0.8	-0.0033	0.0016	-0.0020	0.0009	-0.0037	0.0010	-0.0037	0.0009

Table S3: Bias & MSE of CUII, QML, MQML, 2SLS, RGMM and CUGMM estimators for 'exponential' W using 1000 Monte Carlo replications. The ϵ_is are defined as in (7.1) with $\zeta_i \sim iid \;\; t(5)$ and σ_i is defined as in (7.2). The design corresponds to a strong relevance of instruments.

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0668	0.1386	-0.0386	0.0791	-0.0220	0.0465	-0.0074	0.0342
	0.3	-0.0458	0.0540	-0.0246	0.0464	-0.0113	0.0165	-0.0146	0.0287
	0.5	-0.0427	0.0312	-0.0316	0.0298	-0.0093	0.0139	-0.0163	0.0125
	0.8	-0.0222	0.0091	-0.0155	0.0071	-0.0083	0.0050	-0.0077	0.0079
QML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0702	0.1064	-0.0564	0.0696	-0.0158	0.0404	-0.0202	0.0349
	0.3	-0.0793	0.0539	-0.0603	0.0435	-0.0298	0.0165	-0.0315	0.0281
	0.5	-0.0726	0.0344	-0.0685	0.0302	-0.0352	0.0140	-0.0284	0.0123
	0.8	-0.0472	0.0115	-0.0334	0.0084	-0.0289	0.0049	-0.0257	0.0066
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0473	0.1358	-0.0215	0.0742	-0.0140	0.0450	-0.0033	0.0336
	0.3	-0.0483	0.0510	-0.0303	0.0426	-0.0134	0.0161	-0.0179	0.0274
	0.5	-0.0453	0.0309	-0.0419	0.0271	-0.0137	0.0130	-0.0211	0.0119
	0.8	-0.0297	0.0094	-0.0220	0.0073	-0.0159	0.0042	-0.0213	0.0063
2SLS	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.1104	0.3900	0.0265	0.2288	0.0332	0.1513	0.0402	0.0806
	0.3	0.0421	0.0812	0.0351	0.1248	0.0148	0.0290	0.0401	0.0625
	0.5	0.0031	0.0412	0.0101	0.0582	0.0127	0.0270	-0.0138	0.0224
	0.8	0.0109	0.0113	0.0043	0.0114	0.0001	0.0074	0.0006	0.0082
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0588	0.2053	-0.0413	0.1093	-0.0199	0.0681	-0.0025	0.0502
	0.3	-0.0317	0.0619	-0.0073	0.0853	-0.0120	0.0219	-0.0108	0.0375
	0.5	-0.0361	0.0434	-0.03451	0.0741	-0.0085	0.0195	-0.0321	0.0208
	0.8	-0.0080	0.0144	-0.0108	0.0159	-0.0235	0.0223	-0.0149	0.0220
CUGMM		bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	0.0492	0.3059	0.0178	0.4043	-0.0210	0.0806	-0.0142	0.0568
	0.3	-0.0454	0.1164	-0.0585	0.0811	-0.0309	0.0443	-0.0145	0.0246
	0.5	-0.0332	0.0568	-0.0270	0.0247	-0.0159	0.0138	-0.0274	0.0270
	0.8	-0.0079	0.0046	-0.0155	0.0683	-0.0130	0.0053	-0.0226	0.0130

Table S4: Bias & MSE of CUII, QML, MQML, 2SLS, RGMM and CUGMM estimators for 'exponential' W using 1000 Monte Carlo replications. The ϵ_is are defined as in (7.1) with $\zeta_i \sim iid\, t(5)$ and $\sigma_i \sim \chi^2(5)$. The design corresponds to a strong relevance of instruments.

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.1099	0.1836	-0.0689	0.1035	-0.0219	0.0379	-0.0128	0.0167
	0.3	-0.0443	0.0819	-0.0334	0.0487	-0.0149	0.0197	-0.0072	0.0096
	0.5	-0.0273	0.0672	-0.0214	0.0338	-0.0118	0.0142	-0.0052	0.0073
	0.8	0.0413	-0.0937	0.0260	0.0233	0.0224	0.0113	0.0115	0.0060
QML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0384	0.0636	-0.0139	0.0448	0.0156	0.0224	0.0190	0.0122
	0.3	-0.1326	0.0763	-0.0943	0.0472	-0.0478	0.0200	-0.0305	0.0098
	0.5	-0.1402	0.0692	-0.0948	0.0364	-0.0546	0.0154	-0.0360	0.0078
	0.8	-0.0937	0.0316	-0.0643	0.0155	-0.0362	0.0061	-0.0247	0.0031
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0312	0.0867	-0.0257	0.0615	-0.0066	0.0300	-0.0052	0.0146
	0.3	-0.0536	0.0642	-0.0417	0.0423	-0.0188	0.0187	-0.0093	0.0093
	0.5	-0.0611	0.0509	-0.0406	0.0283	-0.0212	0.0128	-0.0109	0.0067
	0.8	0.0004	0.0842	0.0021	0.0283	0.0293	0.0299	-0.0005	0.0053
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0898	0.1885	-0.0681	0.1877	-0.0188	0.1372	-0.0109	0.1161
	0.3	0.0780	0.4495	0.2229	0.6669	0.4425	0.9531	0.2037	0.4949
	0.5	0.2137	0.5528	0.4398	0.9161	0.6847	1.1509	0.7826	1.3339
	0.8	0.2979	0.3711	0.4606	0.4411	0.4389	0.4054	0.5763	0.4810

Table S5: Bias & MSE of CUII, QML, MQML and RGMM estimators for 'exponential' W using 1000 Monte Carlo replications. The ϵ_is are defined as in (7.1) with $\zeta_i \sim iid\ \ t(5)$ and σ_i is defined as in (7.2). The design corresponds to a misspecification setting where the true data generating process is a pure SAR, while the fitted model includes an intercept and one exogenous regressor drawn from a uniform distribution on $[0, 1]$.

		$n = 30$		$n = 50$		$n = 100$		$n = 200$	
CUII	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0886	0.1403	-0.0470	0.0648	-0.0150	0.0254	-0.0024	0.0141
	0.3	-0.0396	0.0675	-0.0210	0.0393	-0.0139	0.0168	-0.0031	0.0090
	0.5	-0.0255	0.0556	-0.0096	0.0286	-0.0054	0.0107	-0.0012	0.0068
	0.8	0.0129	0.0320	0.0139	0.0211	0.0115	0.0089	0.0124	0.0056
QML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0328	0.0541	-0.0365	0.0417	-0.0290	0.0206	-0.0135	0.0136
	0.3	-0.1269	0.0674	-0.0669	0.0385	-0.0327	0.0177	-0.0134	0.0090
	0.5	-0.1107	0.0547	-0.0611	0.0278	-0.0168	0.0105	-0.0125	0.0063
	0.8	-0.0901	0.0266	-0.0595	0.0148	-0.0148	0.0039	-0.0071	0.0023
MQML	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0175	0.0690	-0.0156	0.0457	0.0009	0.0202	0.0013	0.0132
	0.3	-0.0419	0.0557	-0.0263	0.0352	-0.0179	0.0160	-0.0054	0.0087
	0.5	-0.0430	0.0429	-0.0234	0.0240	-0.0109	0.0096	-0.0068	0.0060
	0.8	-0.0070	0.0610	0.0021	0.0452	0.0239	0.0269	0.0014	0.0055
RGMM	λ	bias	MSE	bias	MSE	bias	MSE	bias	MSE
	-0.5	-0.0940	0.1472	-0.0543	0.0833	-0.0382	0.0327	-0.0169	0.0158
	0.3	0.0380	0.2930	0.0845	0.2877	0.0683	0.1795	0.0327	0.0776
	0.5	0.1411	0.3761	0.2661	0.5124	0.3986	0.6379	0.3764	0.6269
	0.8	0.2016	0.2253	0.3360	0.3320	0.3735	0.2693	0.5598	0.4289

Table S6: Bias & MSE of CUII, QML, MQML and RGMM estimators for 'exponential' W using 1000 Monte Carlo replications. The ϵ_is are defined as in (7.1) with $\zeta_i \sim iid\ \ t(5)$ and $\sigma_i \sim \chi^2(5)$. The design corresponds to a misspecification setting where the true data generating process is a pure SAR, while the fitted model includes an intercept and one exogenous regressor drawn from a uniform distribution on $[0, 1]$.

S.4 Figures

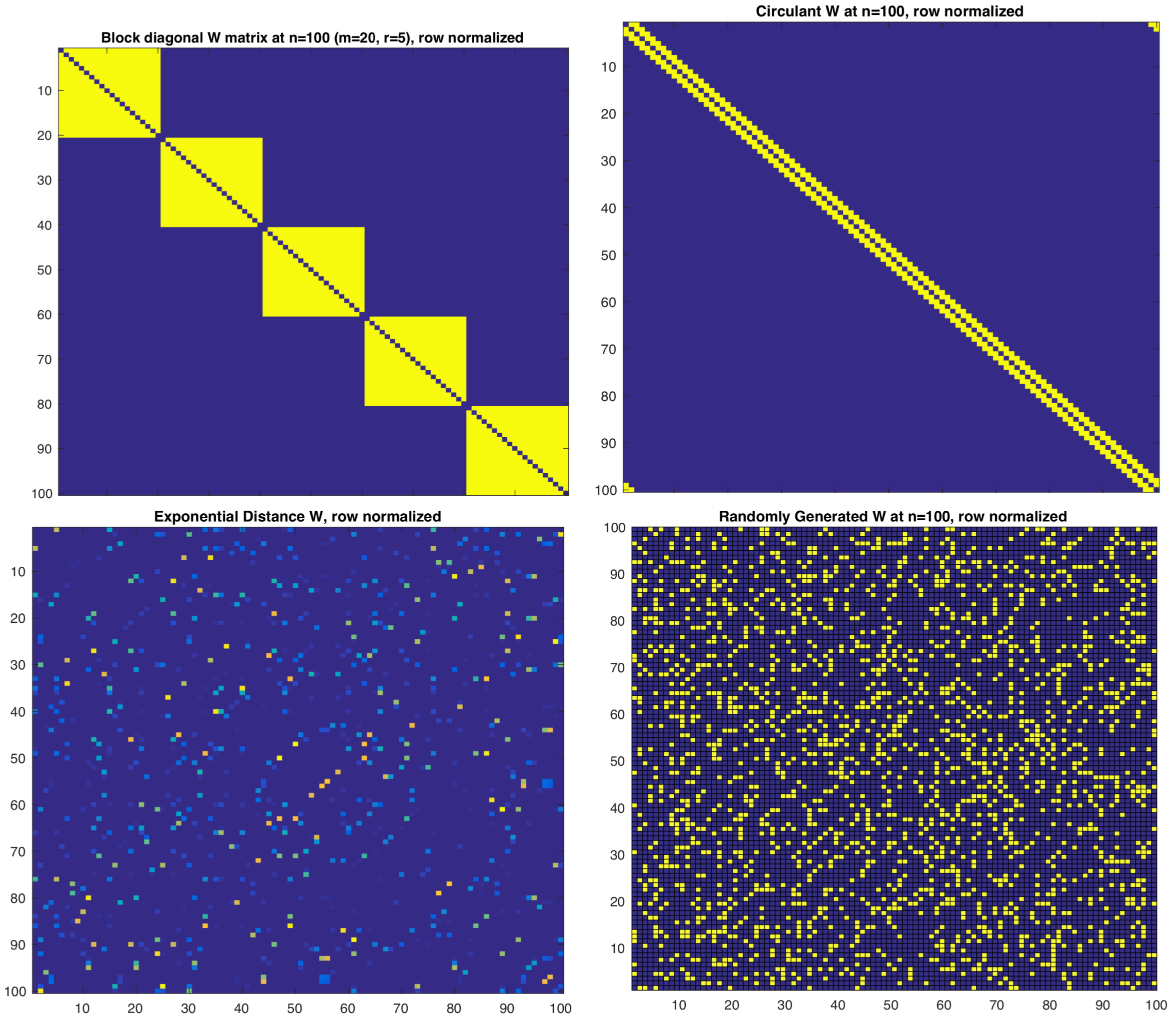

Figure S1: Weight Matrix structures. Top: (L) block diagonal W; (R) circulant, two ahead-two behind; Bottom: (L) 'exponential', (R) 'random'. $n = 100$.

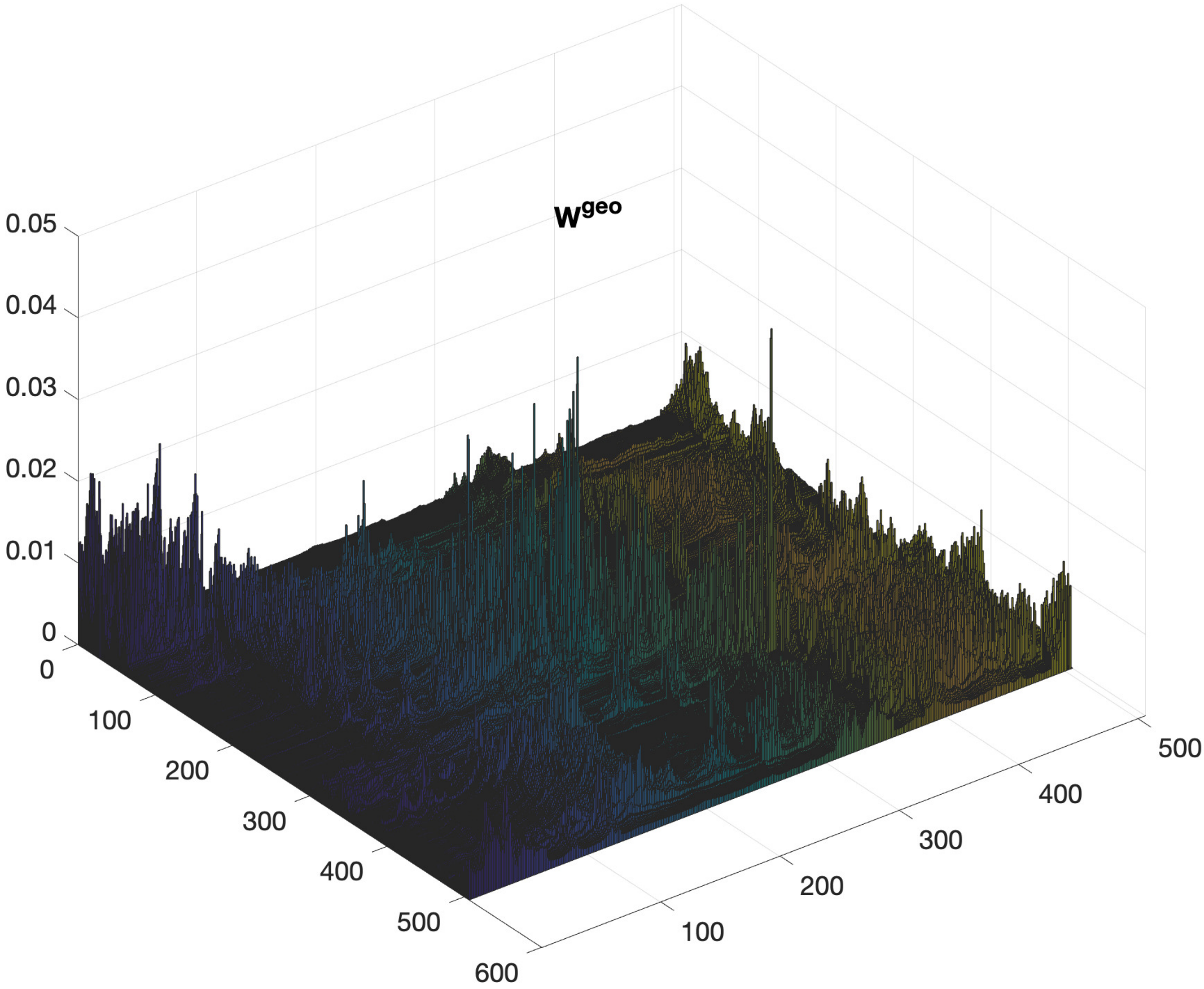

Figure S2: 3D plot of W^{geo}. W^{geo} is defined such that $w_{ij} = 1/geo_{ij}$, resulting in a non-sparse structure with weights that decay with Euclidean/geographical distance. $n = 506$.

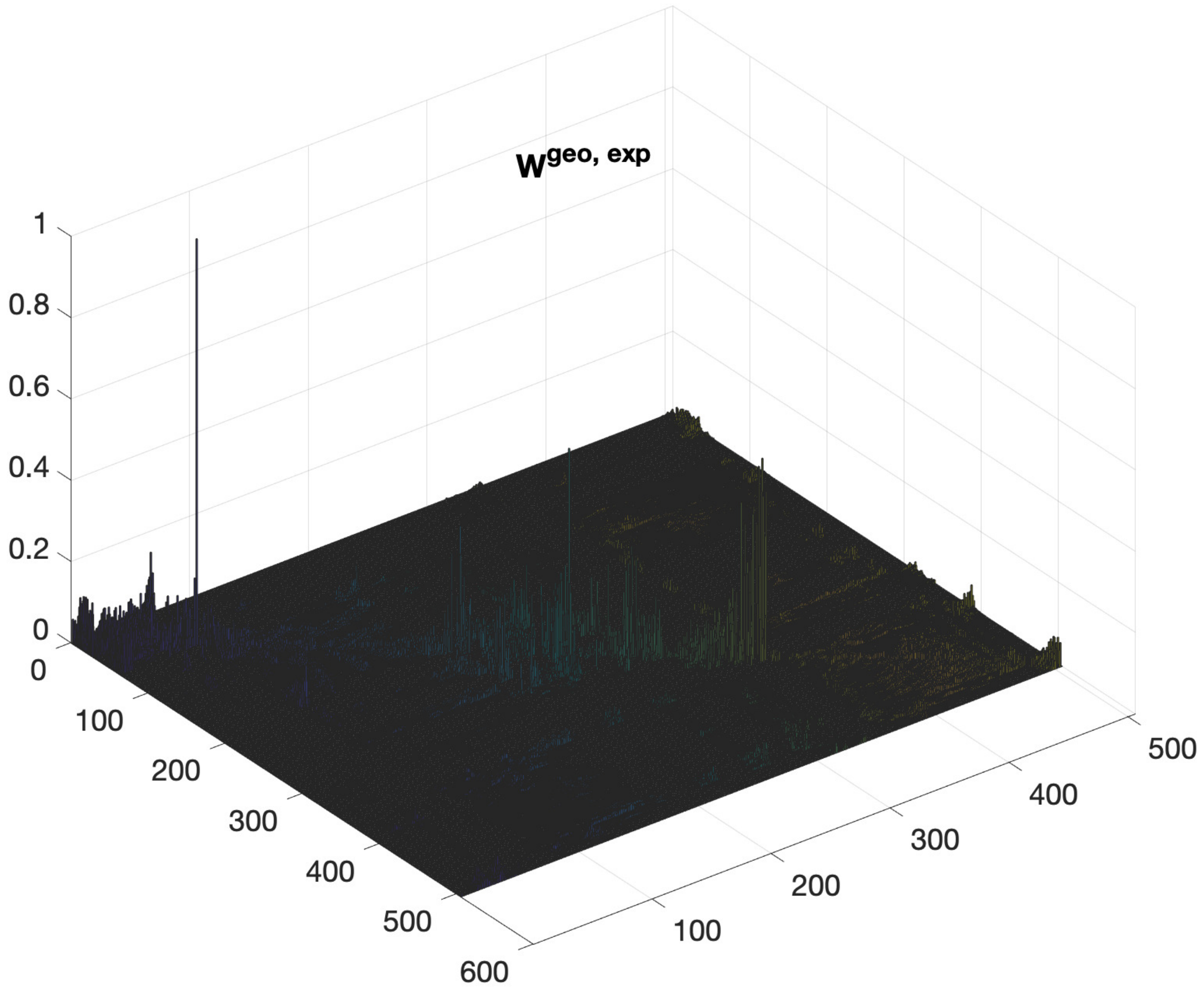

Figure S3: 3D plot of $W^{geo,exp}$. $W^{geo,exp}$ is defined such that $w_{ij} = exp\left(-|geo_{ij}|\right) \mathbb{1}(|geo_{ij}| < log(n))$, resulting in sparsity that amounts to about 37%. $n = 506$.

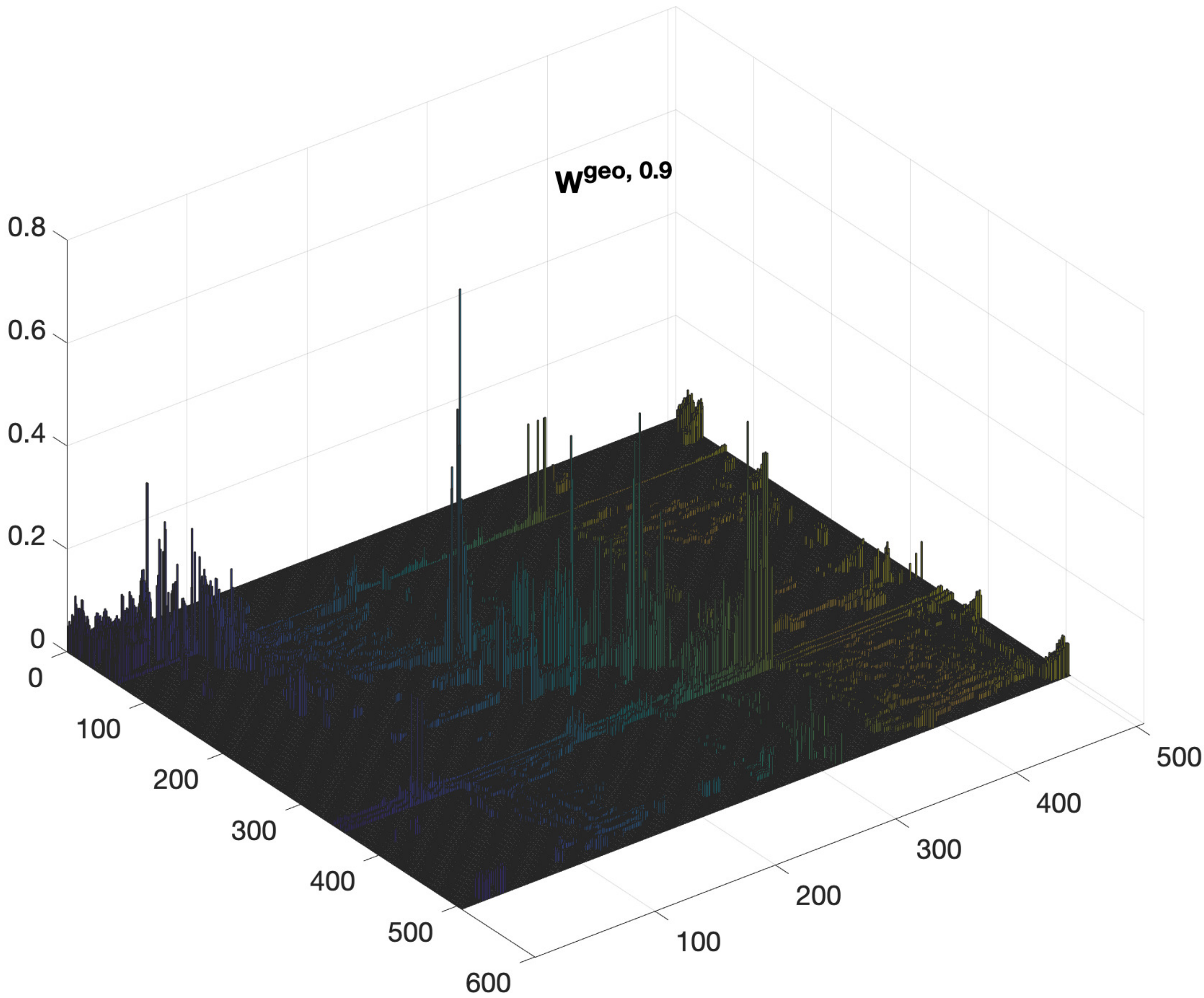

Figure S4: 3D plot of $W^{geo,0.9}$. $W^{geo,0.9}$ is defined such that $w_{ij} = \mathbb{1}(|geo_{ij}| < D^*)$, resulting in sparsity that amounts to about 9%. $n = 506$.

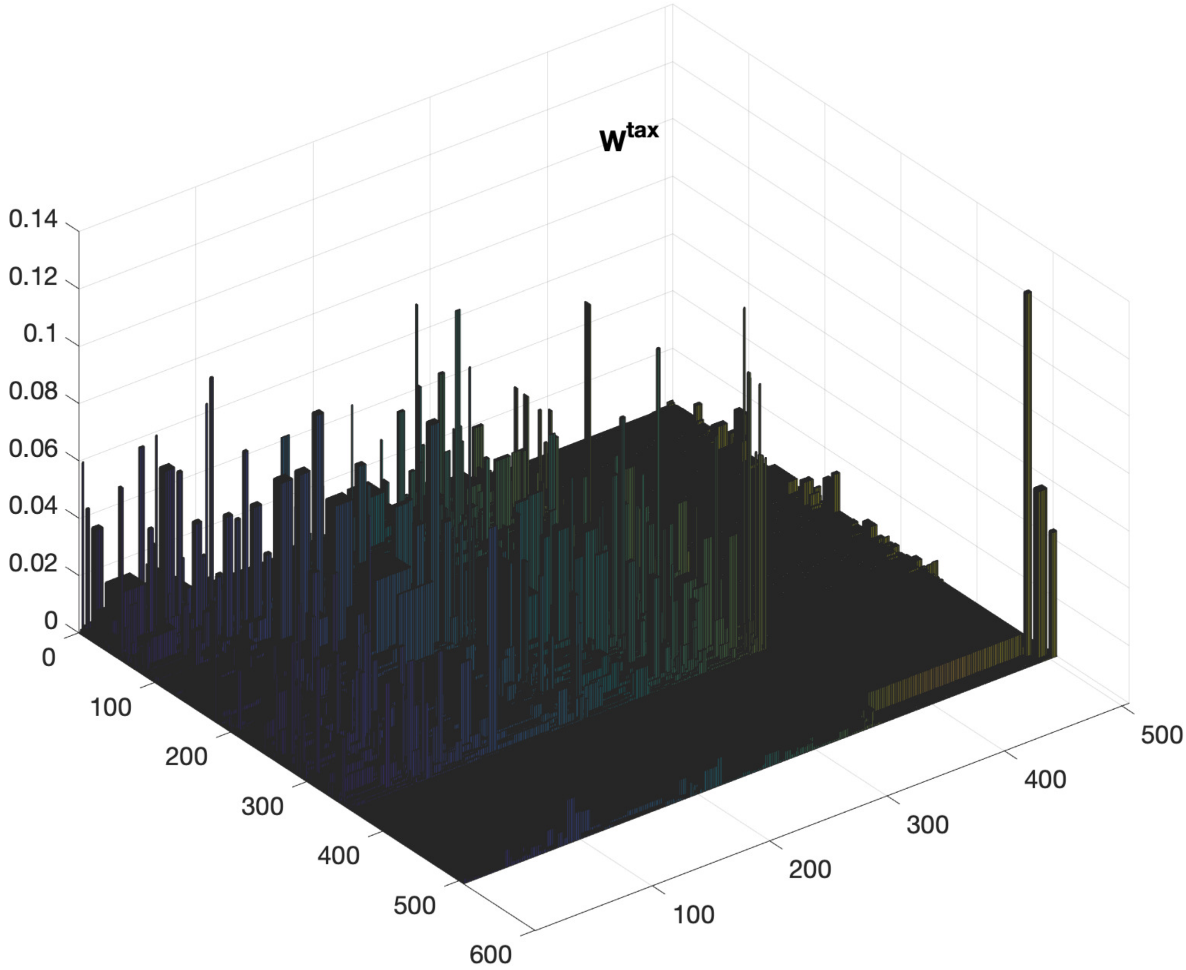

Figure S5: 3D plot of weight matrix W^{tax}. W^{tax} is defined such that $w_{ij} = 1/|tax_i - tax_j|$, resulting in a non-sparse structure with weights that decay with an economic distance driven by tax similarity. $n = 506$.

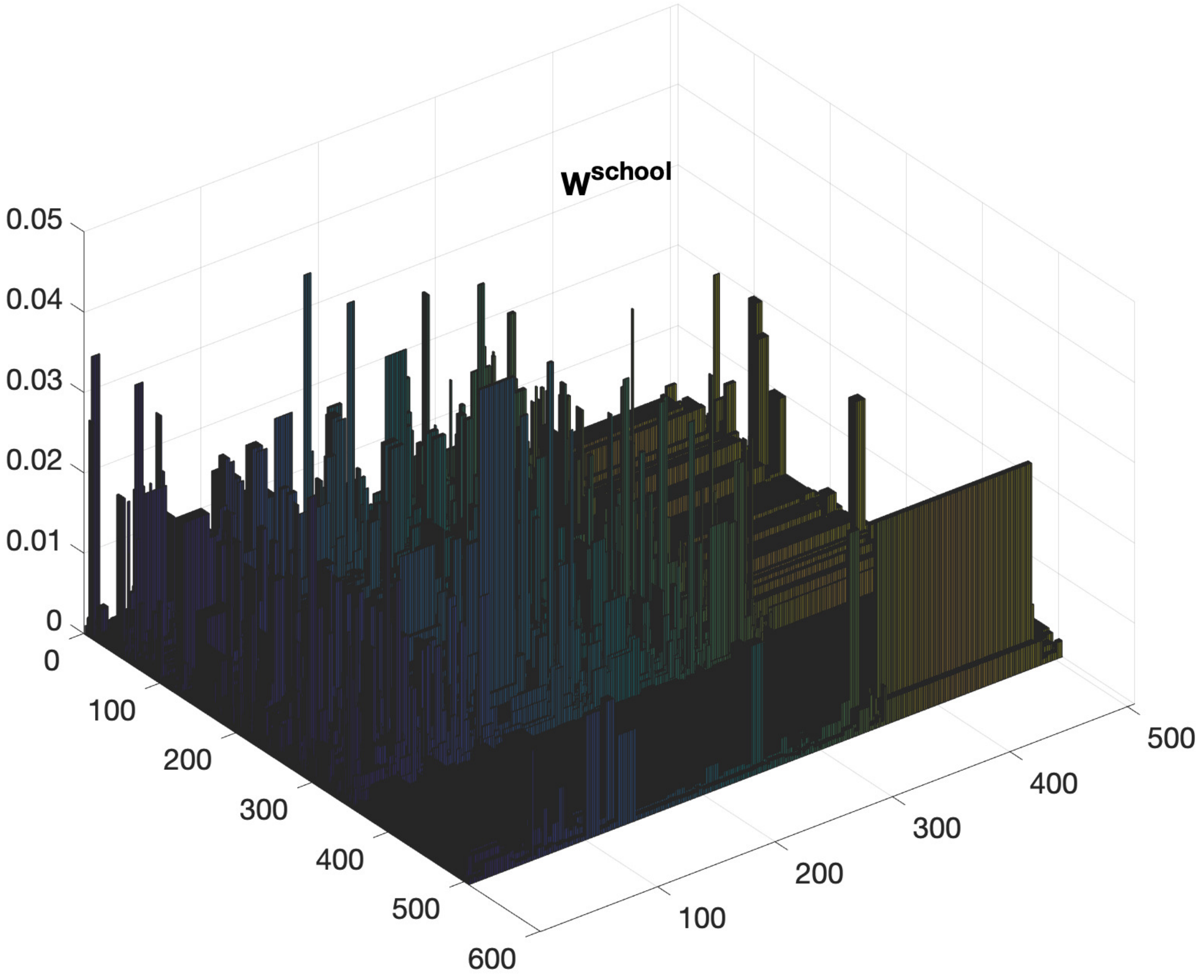

Figure S6: 3D plot of weight matrix W^{school}. W^{school} is defined such that $w_{ij} = 1/|school_i - school_j|$, resulting in a non-sparse structure with weights that decay with an economic distance driven by socio-economic similarity. $n = 506$.

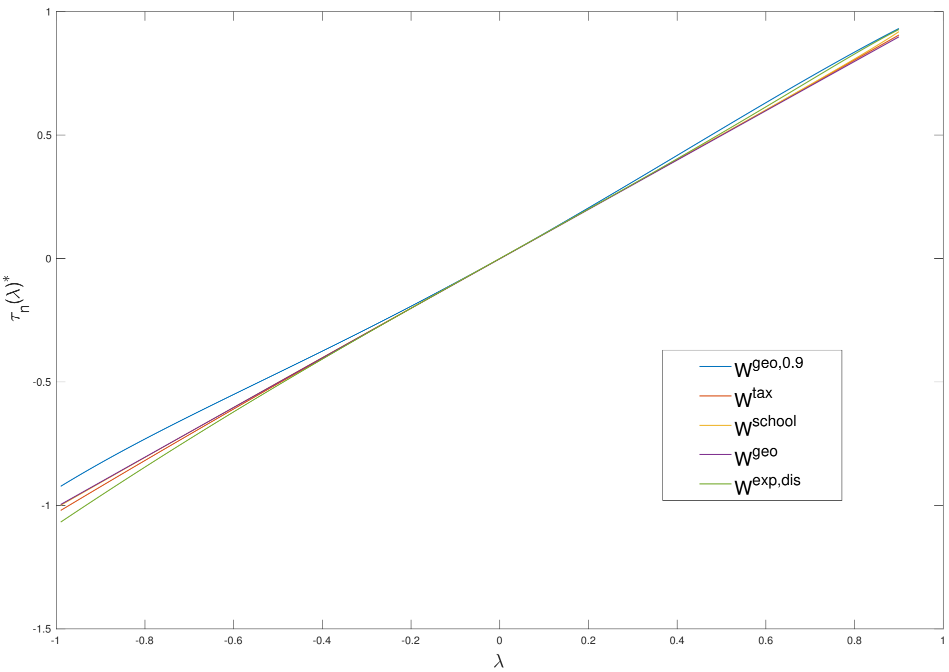

Figure S7: Approximate binding functions for W^{geo}, $W^{exp,dis}$, $W^{geo.0.9}$, W^{tax} and W^{school}. $n = 506$.

References

Bao, Y., (2013). Finite-sample bias of the QMLE in spatial autoregressive models. Econometric Theory 29(1), 68-88.

Bao, Y. and Ullah, A., (2007). Finite sample moments of maximum likelihood estimator in spatial models. Journal of Econometrics 137, 396-413.

Conniffe, D. and Spencer J.E., (2001). When moments of ratios are ratios of moments. The Statistician 50(2), 161-168.

Kyriacou, M, Phillips P.C.B. and Rossi F., (2017). Indirect inference in spatial autoregressions. The Econometrics Journal 20, 168-189.

Scott, D. J., (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability* 5 (1), 119-37.