Online Supplement to "Indirect Inference in Spatial Autoregression"

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

December 6, 2016

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

S.1 Simulated vs approximate binding function results

In this section we report some additional results to support the construction of the II estimator using the approximate binding function $b^*(\lambda)$ given in (2.20) in the paper rather than its simulated/exact version from (2.19). Figures S1 and S2 in the Supplement depict the extended binding plots for $\lambda \in (-1, 1.5]$ in order to provide additional information concerning behavior outside the usual parameter bound of $\lambda \in (-1, 1)$. Figures S1 and S2 also include blow up plots on the exact binding functions with Gaussian, Student-t and standard Cauchy innovations. These indicate that the Gaussian and Student-t exact binding functions are very similar (but not identical), while Cauchy ones have minor differences only noticeable when zooming well into the plots.

[Figures S1-S2 about here]

While the QML binding functions becomes undefined at $\lambda = \pm 1$ no such discontinuity occurs in the OLS binding function, although depending on the structure of W the function can become nearly flat in the vicinity of unity. This suggests the OLS-based II can, in principle, be implemented beyond the bound of $\lambda \in (-1, 1)$, although in some cases there may be ambiguities induced by failure in monotonicity.

*University of Southampton. Corresponding author. E-mail address: M.Kyriacou@soton.ac.uk.
†Yale University, University of Auckland, University of Southampton, Singapore Management University
‡University of Verona.

S.2 Possible extensions and additional simulation results

This section reports additional simulation results to support the discussion in Section 5 of the paper. Tables S1 and S2 report bias and MSE of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$, $\hat{\lambda}_{II,QML}$ and $\hat{\lambda}_{QML,BC}$ when ϵ_i, $i = 1, \ldots, n$, are generated from a $t-$distribution with 3 degrees of freedom and W is chosen as a circulant and an asymmetric Toeplitz respectively, following the same design as in Section 4 of the paper. Table S3 instead displays bias and MSE of the same set of estimators when the choice of W is empirically motivated. For this purpose we construct an $n \times n$ randomly generated weight matrix composed of zeros and ones and re-scaled by its spectral norm. We restrict the number of ones to be 20% of the total number of elements in W. The fact that W is randomly generated for each sample size n does not violate any of the stated assumptions in the paper. Such a randomly generated matrix can reflect, for instance, an empirical weight matrix where neighbours are defined according to a certain contiguity criterion leading to the specification $w_{ij} = 1$ if units i and j are neighbours and zero otherwise.

We do not report similar results when disturbances are generated from a standard Cauchy distribution, as results are expected to be very similar to those obtained from student t and Gaussian innovation. This feature is confirmed by looking at binding plots in the main paper (Figures 1 and 4) and extended binding plots (Figures S1 and S2 in Supplement) which depict, inter alia, $b_n(\lambda)$ from Gaussian and standard Cauchy errors.

[Tables S1- S3 about here]

The pattern of results in Tables S1 and S2 is similar to that of Tables 1 and 2 in the paper. II-OLS and II-QML offer (expected) significant improvements over OLS and QML, respectively, in terms of bias. For most values of λ_0 and almost all sample sizes the bias of $\hat{\lambda}_{II}$ is comparable or even smaller than that of $\hat{\lambda}_{QML}$, with a modest increase in MSE. As somewhat expected, in the simple case of pure SAR with homogeneous disturbances, QML-based techniques such as II-QML and QML-BC offer the best results. The latter is true also in case of the randomly generated W, as shown in Table S3. However, in this case the QML performs generally worse than II-OLS and, surprisingly, even worse than standard OLS, especially for positive values of λ_0.

Table S4 presents simulation results obtained by an extended version of our new II-OLS estimator to a spatial autoregression with exogenous regressors and unknown heteroskedasticity in the error term. The theoretical development of this methodology is under investigation and will be reported in full in separate work. The simulation setting is based on the standard model

$$y = \lambda_0 W y + X\beta_0 + \epsilon, \quad \epsilon_i = \sigma_i z_i, \quad i = 1, ...n, \tag{S.1}$$

where X is the $n \times K$ matrix of exogenous regressors, β_0 the $K \times 1$ exogenous parameters, z_i for $i = 1, ...n$ are a set of i.i.d. standard normal random variables, and all other quantities are defined similarly to model (1.1) in the paper. For this limited simulation experiment we only consider one form of heteroskedasticity, where the σ_i, $i = 1, ...n$, are the square roots of a set of independent random variables generated from a χ^2 distribution with 5 degrees of freedom. The number of exogenous regressors is set at $K = 3$, the first one being the intercept and the other two drawn randomly from a standard uniform distribution. The spatial parameter values are $\lambda_0 = -0.5, 0.3, 0.5, 0.8$, $\beta_0 = (0.3, 0.5, -0.5)$ and the sample sizes are $n = 30, 50, 100, 200$. The weight matrix W is randomly generated as an $n \times n$ matrix of zeros and ones and then re-scaled with its spectral norm, as described above in the context of Table S3. Under this choice of W, in the presence of unknown heteroskedasticity, the ML/QML does not return consistent estimators (Lin and Lee (2010)). In addition to considering the standard OLS and ML estimators, we also compare our extended version of II estimation with the robust generalised method of moments (RGMM) methodology developed in Lin and Lee (2010).

Entries in Table S4 display bias and MSE of the OLS, ML, RGMM and II-OLS estimators based on 10^3 replications. The RGMM estimator is constructed in the same manner as in Liu and Lee (2010) where $P_n = (G_n - diag(G_n)$ and the IV matrix is $(G_n X_n \beta, X_n)$, where as usual $G = G(\lambda_0) = WS^{-1}(\lambda_0)$.

[Table S4 about here]

Several interesting features become evident throughout Table S4. First, under a randomly generated W the ML is shown to be severely biased for all values of the spatial parameter and the bias persists as n becomes larger. The OLS estimator's performance is comparable to that of ML here. Unlike the results discussed in Liu and Lee (2010), the RGMM performs poorly here in both bias and MSE, even worse than its ML counterparts. The II estimator evidently outperforms OLS, ML and RGMM in terms of bias reduction in all cases examined,and also manages to reduce MSE compared to the ML when λ_0 is positive.

S.3 Proofs of the Theorems

Proof of Theorem 1. (a):

Let ψ_{ij} be the vector $\psi_{ij} = (\ \psi_{1ij} \quad \psi_{2ij}\ )' = (\ (G + G')_{ij}/2 \quad (G'G)_{ij}\ )'$, and define

$$u_i = (\ u_{1i} \quad u_{2i}\ )' = (\epsilon_i^2 - \sigma^2)\psi_{ii} + 2\epsilon_i \sum_{j<i} \psi_{ij}\epsilon_j, \tag{S.2}$$

so that $U_n = \sum_{i=1}^n u_i$. We note that $\{u_i, 1 \le i \le n, n = 1, 2, ...\}$ is a triangular array of martingale differences with respect to the filtration formed by the σ-field generated by $\{\epsilon_j; j < i\}$. Define

$$A = Var\left(\sum_{i=1}^n u_i\right) = (\mu^{(4)} - \sigma^4)\sum_{i=1}^n \psi_{ii}\psi_{ii}' + 4\sigma^4\sum_{i=1}^n\sum_{j<i}\psi_{ij}\psi_{ij}', \tag{S.3}$$

and let $z_{in} = \eta' A^{-1/2} u_i$, 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:

$$\sum_{i=1}^n \mathbb{E}(z_{in}^2 | \epsilon_j; j < i) \xrightarrow{p} 1, \tag{S.4}$$

and

$$\sum_{i=1}^n \mathbb{E}(z_{in}^2 1(|z_{in} > \xi|)) \to 0 \quad \forall \xi > 0. \tag{S.5}$$

Now, (S.4) is equivalent to

$$\sum_{i=1}^n \mathbb{E}(z_{in}^2 | \epsilon_j; j < i) - \eta' A^{-1/2} A A^{-1/2} \eta \xrightarrow{p} 0, \tag{S.6}$$

which is

$$\begin{aligned} &\eta' A^{-1/2}\left(4\sigma^2 \sum_{i=1}^n \left(\sum_{j<i}\psi_{ij}\epsilon_j\right)\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)' - 4\sigma^4 \sum_{i=1}^n\sum_{j<i}\psi_{ij}\psi_{ij}'\right) A^{-1/2}\eta \\ &+ 4\eta' A^{-1/2}\mu^{(3)}\sum_{i=1}^n \psi_{ii}\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)' A^{-1/2}\eta \xrightarrow{p} 0, \end{aligned} \tag{S.7}$$

where $\mu^{(3)} = \mathbb{E}(\epsilon_i)^3$. From standard matrix algebra, A is positive definite for all n and satisfies $(hA/n) \to V > 0$ as $n \to \infty$, where

$$\begin{aligned} V &= \lim_{n\to\infty}\left(\begin{pmatrix} \sigma^4(g_{20} + g_{11}) & 2\sigma^4 g_{21} \\ 2\sigma^4 g_{21} & 2\sigma^4 g \end{pmatrix} + \begin{pmatrix} \frac{h}{n}\kappa_4\sum_i G_{ii}^2 & \frac{h}{n}\kappa_4\sum_i G_{ii}(G'G)_{ii} \\ \frac{h}{n}\kappa_4\sum_i G_{ii}(G'G)_{ii} & \frac{h}{n}\kappa_4\sum_i (G'G)_{ii}^2 \end{pmatrix}\right) \\ &= \Sigma + \Omega. \end{aligned} \tag{S.8}$$

Positivity of the smallest eigenvalue of Σ and existence of V is guaranteed by the Cauchy inequality and Assumption 5 since

$$\left(\frac{h}{n}\right)^2 (\mathrm{tr}((G+G')G'G))^2 < \left(\frac{h}{n}\right)^2 \mathrm{tr}((G+G')^2)\mathrm{tr}((G'G)^2). \tag{S.9}$$

Under Assumptions 3 and 4 the elements of Σ are bounded, while Ω has elements of order $O(1/h)$ that vanish in case h is a divergent sequence. Also, $\Omega = 0$ when $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Rather than (S.7), we can equivalently show

$$\frac{h}{n}\left(4\sigma^2 \sum_{i=1}^{n}\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)' - 4\sigma^4\sum_{i=1}^{n}\sum_{j<i}\psi_{ij}\psi_{ij}'\right) \xrightarrow{p} 0, \tag{S.10}$$

and

$$\frac{h}{n}\mu^{(3)}\sum_{i=1}^{n}\psi_{ii}\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)' \xrightarrow{p} 0. \tag{S.11}$$

Consider the following typical elements of the left side of (S.10)

$$4\sigma^2\frac{h}{n}\left(\sum_{i=1}^{n}\sum_{j<i}\psi_{sij}^2(\epsilon_j^2-\sigma^2) + \sum_{i=1}^{n}\sum_{\substack{j,k<i\\ j\neq k}}\psi_{sij}\psi_{sik}\epsilon_j\epsilon_k\right) \quad s=1,2, \tag{S.12}$$

and

$$4\sigma^2\frac{h}{n}\left(\sum_{i=1}^{n}\sum_{j<i}\psi_{sij}\psi_{tij}(\epsilon_j^2-\sigma^2) + \sum_{i=1}^{n}\sum_{\substack{j,k<i\\ j\neq k}}\psi_{sij}\psi_{tik}\epsilon_j\epsilon_k\right) \quad s,t=1,2, \quad s\neq t. \tag{S.13}$$

The first term in (S.12) has mean zero and variance bounded by

$$\begin{aligned}&\left(\frac{h}{n}\right)^2 K\sum_{i}\sum_{k}\sum_{j<i,k}\psi_{sij}^2\psi_{skj}^2 \leq \left(\frac{h}{n}\right)^2 K\sum_{i}\sum_{k}\sum_{j}\psi_{sij}^2\psi_{skj}^2\\ &\leq \left(\frac{h}{n}\right)^2 K\left(\max_j\sum_i\psi_{sij}^2\right)\sum_{k,j}\psi_{skj}^2 = O\left(\frac{h}{n}\right).\end{aligned} \tag{S.14}$$

The last equality in (S.14) follows because $\sum_{h,k}\psi_{shk}^2$ equals either $\mathrm{tr}((G'G)^2) = O(n/h)$ or $\mathrm{tr}(((G+G')/2)^2) = O(n/h)$, and, denoting by Ψ_s the matrix whose $ij-$th element is ψ_{sij} and e_j the $n\times 1$ vector with 1 in the $j-$th position and zero otherwise,

$$\sum_i\psi_{sij}^2 = e_j'\Psi_s^2 e_j \leq ||\Psi_s||^2 \leq K, \tag{S.15}$$

where the last inequality follows from Assumption 3(ii) after observing that Ψ_s equals either $(G+G')/2$ or $G'G$ for $s=1$ and $s=2$, respectively. The second term of (S.12) has mean zero and variance bounded by

$$\begin{aligned}
&\left(\frac{h}{n}\right)^2 K|\sum_i\sum_h\sum_{j<i,h}\sum_{k<i,h}\psi_{sij}\psi_{sik}\psi_{shj}\psi_{shk}| \\
&\leq \left(\frac{h}{n}\right)^2 K\left(\sum_i\sum_h\sum_j\sum_k|\psi_{sij}\psi_{sik}\psi_{shj}\psi_{shk}|\right) \leq \left(\frac{h}{n}\right)^2 K\sum_i\sum_h\sum_j\sum_k|\psi_{sij}\psi_{sik}|(\psi_{shj}^2+\psi_{shk}^2) \\
&\leq \left(\frac{h}{n}\right)^2 K\left(\left(\max_j\sum_i|\psi_{sij}|\right)\left(\max_i\sum_k|\psi_{sik}|\right)\sum_{h,j}\psi_{shj}^2+\left(\max_i\sum_j|\psi_{sij}|\right)\left(\max_k\sum_i|\psi_{sik}|\right)\sum_{h,k}\psi_{shk}^2\right) \\
&= O\left(\frac{h}{n}\right), \qquad \text{(S.16)}
\end{aligned}$$

where the last equality follows from the argument above and Assumptions 3(iii) and 4. Similarly, the first and second terms on the left hand side (LHS) of (S.13) have mean zero and variance bounded by

$$\begin{aligned}
&\left(\frac{h}{n}\right)^2 K\sum_i\sum_k\sum_{j<i,k}|\psi_{sij}\psi_{tij}\psi_{skj}\psi_{tkj}| \leq \left(\frac{h}{n}\right)^2 K\sum_i\sum_k\sum_j|\psi_{sij}\psi_{tij}|(\psi_{skj}^2+\psi_{tkj}^2) \\
&\leq \left(\frac{h}{n}\right)^2 K\max_j\sum_i|\psi_{sij}|\max_i\sum_j|\psi_{tij}|\max_j\sum_k(\psi_{skj}^2+\psi_{tkj}^2)=o(1), \qquad \text{(S.17)}
\end{aligned}$$

and

$$\begin{aligned}
&\left(\frac{h}{n}\right)^2 K|\sum_i\sum_h\sum_{j<i,h}\sum_{k<i,h}\psi_{sij}\psi_{tik}\psi_{shj}\psi_{thk}| \\
&\leq \left(\frac{h}{n}\right)^2 K\left(\sum_i\sum_h\sum_j\sum_k|\psi_{sij}\psi_{tik}\psi_{shj}\psi_{thk}|\right) \leq \left(\frac{h}{n}\right)^2 K\sum_i\sum_h\sum_j\sum_k|\psi_{sij}\psi_{tik}|(\psi_{shj}^2+\psi_{thk}^2) \\
&\leq \left(\frac{h}{n}\right)^2 K\left(\left(\max_j\sum_i|\psi_{sij}|\right)\left(\max_i\sum_k|\psi_{tik}|\right)\sum_{h,j}\psi_{shj}^2+\left(\max_i\sum_j|\psi_{sij}|\right)\left(\max_k\sum_i|\psi_{tik}|\right)\sum_{h,k}\psi_{thk}^2\right) \\
&= o(1), \qquad \text{(S.18)}
\end{aligned}$$

The typical element on the LHS of (S.11) is

$$\frac{h}{n}\mu^{(3)}\sum_i\psi_{sii}\sum_{j<i}\psi_{tij}\epsilon_j, \qquad s,t=1,2, \qquad \text{(S.19)}$$

which has mean zero and variance bounded by

$$K\left(\frac{h}{n}\right)^2\sum_i\sum_k\sum_{j<i,k}|\psi_{sii}\psi_{skk}\psi_{tij}\psi_{tkj}|\leq K\left(\frac{h}{n}\right)^2\sum_i\sum_k\sum_j|\psi_{tij}||\psi_{tkj}|(\psi_{sii}^2+\psi_{skk}^2)$$
$$\leq K\left(\frac{h}{n}\right)^2\left(\max_i\sum_j|\psi_{tij}|\max_j\sum_k|\psi_{tkj}|\sum_i\psi_{sii}^2+\max_j\sum_i|\psi_{tij}|\max_k\sum_j|\psi_{tkj}|\sum_k\psi_{skk}^2\right)=o(1)\quad\text{(S.20)}$$

under Assumptions 3(iii) and 4 and since

$$\sum_i\psi_{sii}^2\leq\sum_{i,j}\psi_{sij}^2=O\left(\frac{n}{h}\right).\tag{S.21}$$

We prove (S.5) by verifying the sufficient Lyapunov condition

$$\sum_{i=1}^n\mathbb{E}|z_{in}|^{2+\delta}\to 0,\tag{S.22}$$

and we proceed by considering a typical standardized element of u_i, i.e. $\sum_i\mathbb{E}|(h/n)^{1/2}u_{si}|^{2+\delta}$ 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,

$$\left(\frac{h}{n}\right)^{1+\delta/2}\sum_i\mathbb{E}|u_{si}|^{2+\delta}\leq\left(\frac{h}{n}\right)^{1+\delta/2}K\sum_i|\psi_{sii}|^{2+\delta}+\left(\frac{h}{n}\right)^{1+\delta/2}K\sum_i\mathbb{E}|\sum_{j<i}\psi_{sij}\epsilon_j|^{2+\delta}.\tag{S.23}$$

The first term in the latter expression is

$$\left(\frac{h}{n}\right)^{1+\delta/2}K\left(\max_i\ |\psi_{sii}|^\delta\right)\sum_i\psi_{sii}^2=o(1),\tag{S.24}$$

by (S.21) and since for all i

$$|\psi_{sii}|\leq||\Psi_s||_\infty\leq K\tag{S.25}$$

under Assumptions 3(iii) and 4. The second term in (S.23) by the Burkholder and von Bahr/Esseen inequalities is bounded by

$$\begin{aligned}
&\left(\frac{h}{n}\right)^{1+\delta/2} K\sum_i \mathbb{E}|\sum_{j<i}\psi_{sij}^2\epsilon_j^2|^{1+\delta/2} \\
&\leq \left(\frac{h}{n}\right)^{1+\delta/2} K\sum_i\sum_{j<i}|\psi_{sij}|^{2+\delta} \leq \left(\frac{h}{n}\right)^{1+\delta/2} K\sum_i\left(\sum_{j<i}\psi_{sij}^2\right)^{1+\delta/2} \\
&\leq K\left(\frac{h}{n}\right)^{1+\delta/2}\left(\max_i\sum_j\psi_{sij}^2\right)^{\delta/2}\sum_i\sum_j\psi_{sij}^2 \\
&\leq K\left(\frac{h}{n}\right)^{\delta/2}\left(\max_i\sum_j\psi_{sij}^2\right)^{\delta/2}, \qquad \text{(S.26)}
\end{aligned}$$

which is $O((h/n)^{\delta/2})$ by (S.15).

Thus, $A^{-1/2}\sum_i u_i \underset{d}{\rightarrow} \mathcal{N}(0, I)$, or equivalently

$$\left(\frac{h}{n}\right)^{1/2}\sum_i u_i \underset{d}{\rightarrow} \mathcal{N}(0, V), \qquad \text{(S.27)}$$

where V is defined in (S.8). The statements (3.5) and (3.6) in the paper follow trivially since

$$f_n'\left(\frac{h}{n}\right)^{1/2} U_n = \bar{f}'\left(\frac{h}{n}\right)^{1/2} U_n + o_p(1), \qquad \text{(S.28)}$$

where

$$\bar{f} = \lim_{n\to\infty}\left(g_{11}^{-1}\sigma_0^{-2} \quad - g_{11}^{-2}g_{10}\sigma_0^{-2}\right)', \qquad \text{(S.29)}$$

which is non-zero and finite under Assumption 5 and Cauchy inequality.

Proof of Theorem 2.

We show parts (a) and (b) using formulae (A.10)-(A.13) given in the paper. To show part (a), given the binding function analytical expression (2.17), we obtain the derivative

$$\frac{db_n(\lambda)}{d\lambda} = 2 - \frac{2h^2\text{tr}(G(\lambda))\text{tr}(G(\lambda)^3)/n^2}{h^2(\text{tr}(G(\lambda)^2))^2/n^2} + O\left(\frac{h}{n}\right). \qquad \text{(S.30)}$$

As $n \rightarrow \infty$, the sign of the right hand side (RHS) of (S.30) depends on $h^2((\text{tr}(G(\lambda)^2))^2 - \text{tr}(G(\lambda))\text{tr}(G(\lambda)^3))/n^2$. The condition $h/n \rightarrow 0$ as $n \rightarrow \infty$ is satisfied when $r \rightarrow \infty$, whether $m \rightarrow \infty$ or $m = O(1)$ as $n \rightarrow \infty$.

When $m = O(1)$ as $n \to \infty$,

$$\left(\frac{h}{n}\right)^2 ((\mathrm{tr}(G(\lambda)^2))^2 - \mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3))$$
$$= \frac{(m-1)^2}{m}\left(\frac{(m-1)(1-\lambda)}{(1-\lambda)^4(m-1+\lambda)m} + \frac{2(m-1)}{m(1-\lambda)^2(m-1+\lambda)^2} + \frac{(m-1)}{m(m-1+\lambda)^3(1-\lambda)}\right)$$
$$= \frac{(m-1)^3}{m^2(m-1+\lambda)(1-\lambda)}\left(\frac{1}{1-\lambda} + \frac{1}{m-1+\lambda}\right)^2, \tag{S.31}$$

which is stricly positive for $\lambda < 1$ and $m \geq 2$. As $m \to \infty$,

$$\left(\frac{h}{n}\right)^2 ((\mathrm{tr}(G(\lambda)^2))^2 - \mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3)) \to \frac{1}{(1-\lambda)^3}, \tag{S.32}$$

which, again, is strictly positive for $\lambda < 1$. As $\lambda \to 1$, for both $m = O(1)$ and $m \to \infty$ as $n \to \infty$, it is easy to see that $db_n(\lambda)/d\lambda \to 0$.

To show part (b) we notice that under Assumptions 3(iii), 3(iv) and 4 each element of both G and $G'G$ is uniformly bounded by $1/h$. The asymptotic variance of $\hat{\lambda}_{QML}$ is

$$V_{QML} = \lim_{n\to\infty}\left(g_{20} + g_{11} - \frac{2}{h}g_{10}^2\right)^{-2}\left(g_{20} + g_{11} - \frac{2}{h}g_{10}^2 + \frac{\kappa_4}{\sigma_0^4}\frac{h}{n}\sum_i\left(G_{ii} - \frac{g_{10}}{h}\right)^2\right), \tag{S.33}$$

and thus when $h = m - 1$ increases without bound as $n \to \infty$

$$V_{QML} = \lim_{n\to\infty}(g_{20} + g_{11})^{-1} \tag{S.34}$$

and

$$\omega^* = \lim_{n\to\infty}(g_{11} + g_{20})^{-1}\left(1 - \frac{2g_{10}g_{21}}{g_{11}(g_{20} + g_{11})}\right)^{-2}\left(1 - \frac{4g_{21}g_{10}}{g_{11}(g_{11} + g_{20})} + \frac{2gg_{10}^2}{g_{11}^2(g_{11} + g_{20})}\right) \tag{S.35}$$

As $m \to \infty$ and $r \to \infty$,

$$\lim_{n\to\infty}\frac{h}{n}\mathrm{tr}(G) = \frac{\lambda_0}{(1-\lambda_0)} \quad \lim_{n\to\infty}\frac{h}{n}\mathrm{tr}(G^s) = \frac{1}{(1-\lambda_0)^s} \tag{S.36}$$

Hence, from (S.34), (S.36) and standard algebra

$$\lim_{n\to\infty}\left(\frac{2\mathrm{tr}G\mathrm{tr}(G^2G')}{\mathrm{tr}(G'G)\mathrm{tr}(G^2 + G'G)}\right)^2 = \lim_{n\to\infty}\frac{2\mathrm{tr}((G'G)^2)(\mathrm{tr}G)^2}{(\mathrm{tr}(G'G))^2\mathrm{tr}(G^2 + G'G)} = \lambda_0^2, \tag{S.37}$$

so that

$$\omega^* = V_{QML} = \frac{(1-\lambda_0)^2}{2}. \tag{S.38}$$

Proof of Theorem 3:

Since $\int_0^{2\pi}(cosx)^s dx = 0$ for odd s, the RHS of (A.14) in the paper can be written as

$$\frac{1}{\lambda}\sum_{p=1}^{\infty}\lambda^{2p}\frac{1}{2\pi}\int_0^{2\pi}(cosx)^{2p}dx = \frac{1}{\lambda}\sum_{p=1}^{\infty}\lambda^{2p}\frac{(2p-1)!!}{(2p)!!} = \frac{1}{\lambda}\sum_{p=1}^{\infty}\lambda^{2p}\frac{(2p)!}{2^{2p}(p!)^2}$$
$$=\frac{1}{\lambda}\sum_{p=0}^{\infty}\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p} - \frac{1}{\lambda} = \frac{1}{\lambda}((1-\lambda^2)^{-1/2}-1). \tag{S.39}$$

Similarly, since $\int_0^{2\pi}(cosx)^{s+t+2}dx \neq 0$ only when $s+t+2 = 2p$, for $s,t = 0,....\infty$ and $p = 1,....\infty$, the RHS of (A.15) becomes

$$\frac{1}{2\pi}\sum_{s,t=0}^{\infty}\lambda^{s+t}\int_0^{2\pi}(cosx)^{s+t+2}dx = \frac{1}{\lambda^2}\sum_{p=1}^{\infty}\lambda^{2p}(2p-1)\frac{(2p-1)!!}{(2p)!!}, \tag{S.40}$$

where the factor $(2p-1)$ takes into account all the combinations of $s,t = 0,.....,\infty$ s.t. $s+t = 2p-2$, for $p = 1,....\infty$. Since

$$\sum_{p=1}^{\infty}px^p\binom{2p}{p} = 2x(1-4x)^{-3/2} \quad |x| < \frac{1}{4}, \tag{S.41}$$

$$\frac{1}{n}\mathrm{tr}(G(\lambda)^2) \to \frac{1}{\lambda^2}\left(2\sum_{p=1}^{\infty}p\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p} - \sum_{p=1}^{\infty}\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p}\right)$$
$$=\frac{1}{\lambda^2}\left(\lambda^2(1-\lambda^2)^{-3/2} - ((1-\lambda^2)^{-1/2}-1)\right). \tag{S.42}$$

Along the same lines, the RHS of (A.16) is

$$\frac{1}{2\pi}\sum_{s,t,q=0}^{\infty}\lambda^{s+t+q}\int_0^{2\pi}(cosx)^{s+t+q+3}dx = \frac{1}{2\pi\lambda^3}\sum_{p=2}^{\infty}\lambda^{2p}(p-1)(2p-1)\int_0^{2\pi}(cosx)^{2p}dx, \tag{S.43}$$

since $\int_0^{2\pi}(cosx)^{s+t+q+3}dx \neq 0$ only when $s+t+q+3 = 2p$, $s,t,q = 0,....,\infty$ and $p = 2,....,\infty$, and where the factor $(p-1)(2p-1)$ takes into account the number of combinations of s,t,q s.t. $s+t+q = 2p-3$. By

$$\sum_{p=1}^{\infty}p^2x^p\binom{2p}{p} = \frac{2x(2x+1)}{(1-4x)^{5/2}} \quad |x| < \frac{1}{4}, \tag{S.44}$$

we deduce

$$
\begin{aligned}
\frac{1}{n}\mathrm{tr}(G(\lambda)^3) \to & \frac{1}{\lambda^3}\sum_{p=1}^{\infty}\lambda^{2p}(p-1)(2p-1)\frac{(2p-1)!!}{(2p)!!} \\
= & \frac{1}{\lambda^3}\left(2\sum_{p=1}^{\infty}p^2\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p} - 3\sum_{p=1}^{\infty}p\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p} + \sum_{p=1}^{\infty}\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p}\right) \\
= & \frac{1}{\lambda^3}\left(\lambda^2\left(\frac{\lambda^2}{2}+1\right)(1-\lambda^2)^{-5/2} - \frac{3}{2}\lambda^2(1-\lambda^2)^{-3/2} + (1-\lambda^2)^{-1/2} - 1\right). \qquad \text{(S.45)}
\end{aligned}
$$

Collecting (S.39), (S.42) and (S.45), we can show that (S.30) is strictly positive for any $\lambda \in (-\sqrt{3}/2, \sqrt{3}/2)$ (and $\lambda \neq 0$) as $n \to \infty$, since

$$
\begin{aligned}
&\frac{1}{n^2}((\mathrm{tr}(G(\lambda)^2))^2 - \mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3)) \to \frac{1}{\lambda^4}(\frac{\lambda^4}{2}(1-\lambda^2)^{-3}(1+(1-\lambda^2)^{1/2}) \\
&-\frac{\lambda^2}{2}(1-\lambda^2)^{-2}(1-(1-\lambda^2)^{1/2}) - \lambda^2(1-\lambda^2)^{-3}(1-(1-\lambda^2)^{1/2})) \qquad \text{(S.46)}
\end{aligned}
$$

as $n \to \infty$. By setting $z = (1-\lambda^2)^{1/2}$ and by some algebraic manipulation, for $\lambda \in (-1, 1)$ and $\lambda \neq 0$ the RHS of (S.46) is strictly positive when

$$2z^2 - 3z + 1 < 0, \qquad \text{(S.47)}$$

which is satisfied for $z \in (1/2, 1)$. Solving for λ, we obtain that the RHS of (S.46) is strictly positive for $\lambda \in (-\sqrt{3}/2, \sqrt{3}/2)$.

References

- Lin, X., Lee, L. F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157, 34-52.
- 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.

Tables

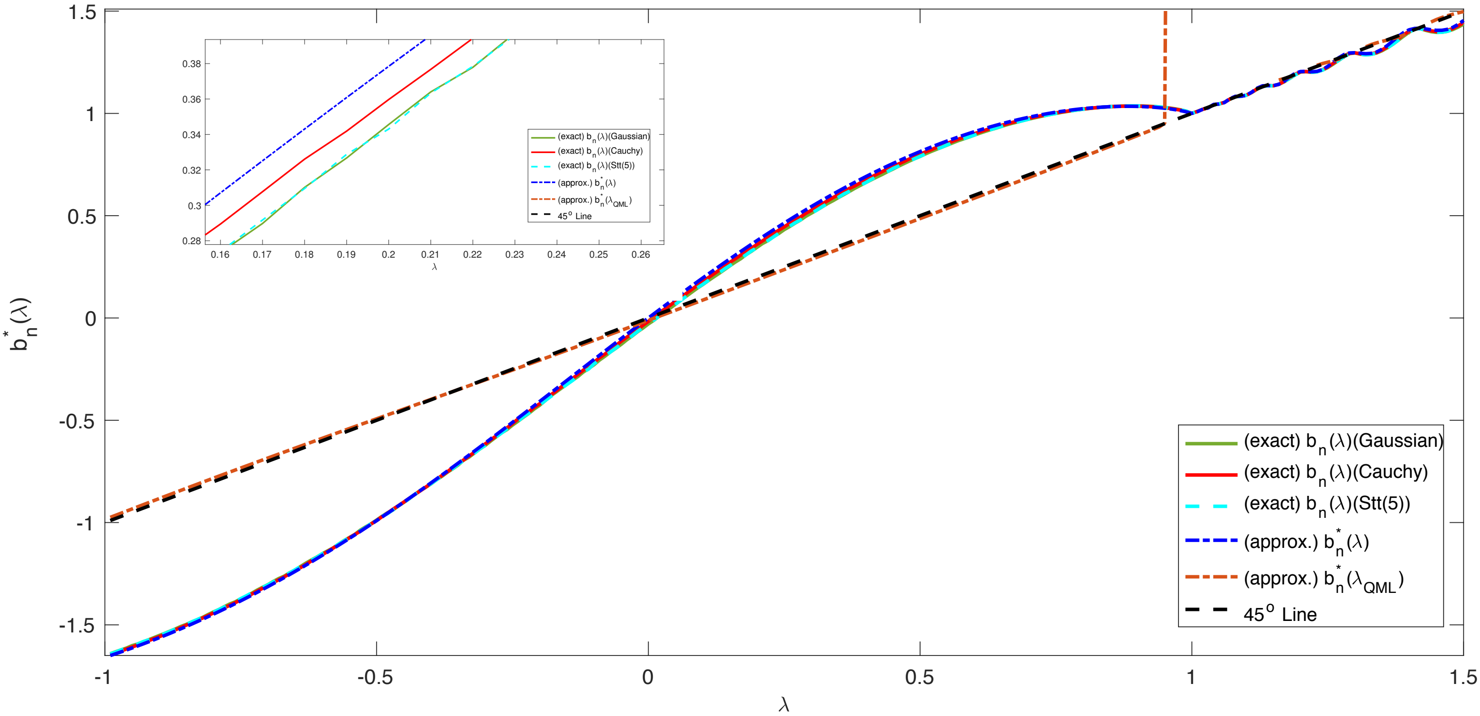

Figure S1: Exact OLS binding functions $b_n(.)$ based on Gaussian, Student-t (5 $d.f.$) and standard Cauchy innovations ($B = 50000$), approximate OLS binding functions, $b_n^*(.)$ and approximate QML binding functions, $b_{n,QML}(.)$, for $\lambda \in (-1, 1.5]$ when W has a circulant structure at $n = 100$. Inserts on the top-left show the blow up plots of exact binding functions under Gaussian, Student-t and standard Cauchy innovations.

S.4 Additional Figures

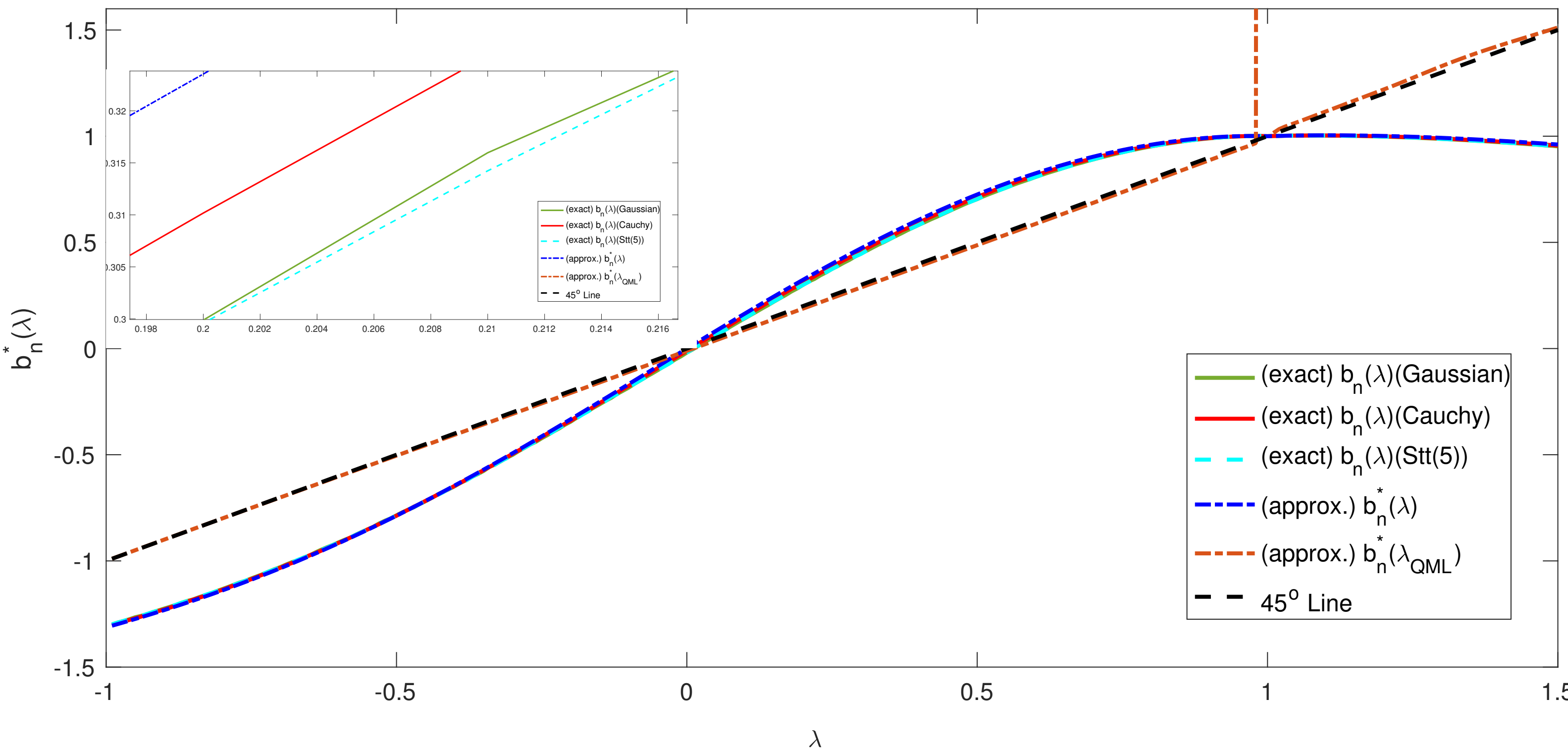

Figure S2: Exact OLS binding functions $b_n(.)$ based on Gaussian Student-t (5 $d.f.$) and standard Cauchy innovations ($B = 50000$), approximate OLS binding functions, $b_n^*(.)$ and approximate QML binding functions, $b_{n,QML}(.)$, for $\lambda \in (-1, 1.5]$ when W has an Asymmetric Toeplitz structure at $n = 100$. Inserts on the top-left show the blow up plots of exact binding functions under Gaussian, Student-t and standard Cauchy innovations.

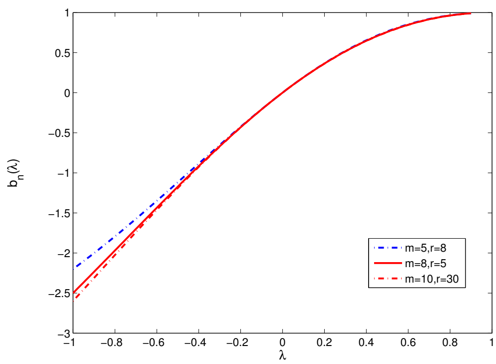

Figure S3: Approximate binding functions, $b_n^*(.)$, at various sample sizes when W is block diagonal (districts model). See Section 5 of main paper for discussion.

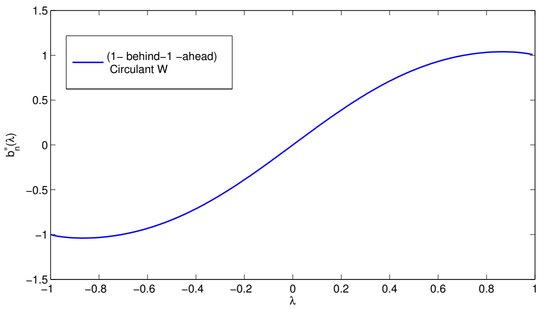

Figure S4: Approximate binding function, $b_n^*(.)$, for $\lambda \in (-1, 1)$ when W is Circulant with leading row $(0, 1, 0, \ldots, 0, 1)$ at $n = 100$. See Section 5 of main paper for discussion.

n		30		50		100		200	
OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	−0.8	−0.6101	0.4871	−0.6271	−0.4670	−0.6331	0.4437	−0.6428	0.4296
	−0.5	−0.4978	0.4540	−0.4759	0.3568	−0.4900	−3112	−0.4888	0.2729
	0.0	−0.1070	0.2468	−0.0516	0.1476	−0.0283	0.0795	−0.0136	0.0397
	0.5	0.2475	0.1302	0.2627	0.1044	0.2862	0.0970	0.3022	0.0979
	0.8	0.2108	0.0541	0.2142	0.0489	0.2193	0.0493	0.2212	0.0494
II-OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	−0.8	−0.0319	0.0936	−0.0235	0.0557	−0.0051	0.0221	−0.0044	0.0108
	−0.5	−0.0391	0.0859	−0.0112	0.0477	−0.0095	0.0241	−0.0036	0.0111
	0.0	−0.0480	0.0684	−0.0207	0.0391	−0.0110	0.0204	−0.0051	0.0101
	0.5	−0.0130	0.0468	−0.0194	0.0248	−0.0118	0.0114	−0.0045	0.0050
	0.8	0.1411	0.0229	0.1482	0.0251	0.1473	0.0244	0.1491	0.0249
QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	−0.8	0.0194	0.0549	0.0082	0.0363	0.0095	0.0172	0.0045	0.0091
	−0.5	−0.0139	0.0637	0.0026	0.0405	−0.0031	0.0216	0.0009	0.0105
	0.0	−0.0443	0.0597	−0.0202	0.0362	−0.0109	0.0198	−0.0051	0.0099
	0.5	−0.0386	0.0346	−0.0328	0.0204	−0.0194	0.0097	−0.0080	0.0046
	0.8	−0.0380	0.0143	−0.0279	0.0204	−0.0125	0.0032	−0.0081	0.0014
II-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	−0.8	−0.0075	0.0607	−0.0160	0.0387	−0.0090	0.0178	−0.0092	0.0093
	−0.5	−0.0213	0.0751	−0.0062	0.0456	−0.0120	0.0239	−0.0063	0.0113
	0.0	−0.0109	0.0650	0.0005	0.0385	−0.0002	0.0204	0.0002	0.0101
	0.5	0.0078	0.0332	−0.0046	0.0194	−0.0055	0.0094	−0.0013	0.0046
	0.8	−0.0014	0.0116	−0.0051	0.0060	−0.0009	0.0029	−0.0020	0.0013
BC-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	−0.8	0.0241	0.0603	0.0181	0.0387	0.0208	0.0176	0.0138	0.0093
	−0.5	0.0039	0.0687	0.0174	0.0422	0.0083	0.0215	0.0092	0.0104
	0.0	−0.0094	0.0634	0.0020	0.0372	0.0003	0.0202	0.0005	0.0100
	0.5	0.0104	0.0335	−0.0024	0.0193	−0.0036	0.0093	0.0001	0.0045
	0.8	0.0045	0.0315	−0.0052	0.0060	−0.0015	0.0029	−0.0027	0.0013

Table S1: Bias and Mean Square Error (MSE) of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$, $\hat{\lambda}_{II,QML}$ and $\hat{\lambda}_{QML,BC}$ at $n = 30, 50, 100, 200$ for $\lambda_0 = -0.8, -0.5, 0, 0.5, 0.8$ when W is Circulant (10^3 repl. and ϵ is generated from a t-distribution with 3 degrees of freedom).

n		30		50		100		200	
OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.3221	0.1897	-0.3213	0.1521	-0.3359	0.1378	-0.3389	-0.3389
	-0.5	-0.2719	0.1932	-0.2744	0.1542	-0.2821	0.1220	-0.2940	0.1056
	0.0	-0.0502	0.1494	-0.0336	0.0924	-0.0066	0.0480	-0.0129	0.0237
	0.5	0.1497	0.0807	0.1748	0.0598	0.2059	0.0542	0.2128	0.0516
	0.8	0.1310	0.0272	0.1454	0.0243	0.1522	0.0243	0.1552	0.0246
II-OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0323	0.0907	-0.0114	0.0508	-0.0127	0.0267	-0.0079	0.0132
	-0.5	-0.0192	0.0743	-0.0109	0.0478	-0.0066	0.0238	-0.0099	0.0108
	0.0	-0.0316	0.0643	-0.0196	0.0359	-0.0030	0.0179	-0.0073	0.0087
	0.5	-0.0403	0.0391	-0.0290	0.0215	-0.0100	0.0093	-0.0043	0.0050
	0.8	-0.0033	0.0277	-0.0048	0.0119	-0.0055	0.047	-0.0053	0.0021
QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0264	0.0874	-0.0053	0.0483	-0.0103	0.0262	-0.0079	0.0130
	-0.5	-0.0132	0.0709	-0.0060	0.0449	-0.0070	0.0234	-0.0093	0.0106
	0.0	-0.0312	0.0614	-0.0195	0.0351	-0.0031	0.0177	-0.0073	0.0086
	0.5	-0.0488	0.0363	-0.0339	0.0206	-0.0125	0.0089	-0.0089	0.0048
	0.8	-0.0409	0.0169	-0.0226	0.0075	-0.0132	0.0034	-0.0085	0.0018
II-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0112	0.0910	0.0004	0.0487	-0.0109	0.0263	-0.0098	0.0131
	-0.5	0.0092	0.0739	0.0051	0.0467	-0.0013	0.0242	-0.0090	0.0110
	0.0	0.0046	0.0633	0.0018	0.0358	0.0077	0.0180	-0.0019	0.0087
	0.5	-0.0049	0.0348	-0.0076	0.0197	0.0006	0.0088	-0.0024	0.0048
	0.8	0.0035	0.0151	0.0041	0.0069	0.0023	0.0032	-0.0016	0.0017
BC-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.0062	0.0898	0.0195	0.0481	0.0065	0.0255	0.0075	0.0124
	-0.5	0.0198	0.0709	0.0158	0.0440	0.0074	0.0229	-0.0014	0.0103
	0.0	0.0048	0.0628	0.0023	0.0356	0.0079	0.0180	-0.0038	0.0088
	0.5	-0.0065	0.0348	-0.0069	0.0197	0.0013	0.0087	-0.0038	0.0048
	0.8	0.0035	0.0152	0.0033	0.0069	-0.0010	0.0032	-0.0022	0.0017

Table S2: Bias and Mean Square Error (MSE) of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$, $\hat{\lambda}_{II,QML}$ and $\hat{\lambda}_{QML,BC}$ at $n = 30, 50, 100, 200$ for $\lambda_0 = -0.8, -0.5, 0, 0.5, 0.8$ when W is AT (10^3 repl. and ϵ is generated from a t-distribution with 3 degrees of freedom).

n		30		50		100		200	
OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0609	0.2697	-0.0811	0.2467	-0.1029	0.2328	-0.0963	0.2530
	-0.5	-0.0273	0.2861	-0.0578	0.2430	-0.1092	0.2343	-0.0906	0.2289
	0.0	-0.0976	0.3103	-0.0663	0.2229	-0.0356	0.2181	-0.0586	0.2276
	0.5	-0.0938	0.2579	0.0840	0.1784	-0.0912	0.1791	0.0833	0.1583
	0.8	0.0843	0.1662	0.0158	0.1066	0.0302	0.0922	0.0773	0.0920
II-OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.0051	0.2468	-0.0074	0.2147	-0.0170	0.1940	-0.0054	0.2083
	-0.5	-0.0294	0.2479	-0.0049	0.2042	-0.0459	0.1873	-0.0257	0.1793
	0.0	-0.0712	0.2480	-0.0679	0.1723	-0.0400	0.1651	-0.0584	0.1784
	0.5	-0.0767	0.2324	-0.0628	0.1757	-0.0641	0.1750	-0.0555	0.1545
	0.8	-0.0537	0.1947	-0.0465	0.1526	-0.0299	0.1383	-0.0333	0.1428
QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.0088	0.2370	-0.0037	0.2108	-0.0249	0.1930	-0.0036	0.2062
	-0.5	-0.0273	0.2328	-0.0033	0.1973	-0.0444	0.1850	-0.0249	0.1766
	0.0	-0.0797	0.2218	-0.0729	0.1610	-0.0490	0.1494	-0.0688	0.1560
	0.5	-0.1251	0.1744	-0.1182	0.1194	-0.1208	0.1192	-0.1099	0.1025
	0.8	-0.1497	0.1193	-0.1412	0.0938	-0.1333	0.0781	0.1428	0.0773
II-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.0015	0.2977	-0.0096	0.2590	-0.0219	0.2344	-0.0282	0.2706
	-0.5	0.0016	0.3023	0.0273	0.2532	-0.0189	0.2293	-0.0058	0.2291
	0.0	0.0211	0.2880	0.0227	0.2072	0.0477	0.1967	0.0185	0.1976
	0.5	0.0281	0.1920	0.0261	0.1230	0.0173	0.1202	0.0219	0.1021
	0.8	0.0017	0.0949	-0.0195	0.0657	-0.0178	0.0513	-0.0311	0.0497
BC-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.0092	0.2878	0.0092	0.2478	0.0187	0.2134	0.0212	0.2168
	-0.5	0.0027	0.2891	0.0331	0.2393	0.0098	0.2084	0.0382	0.1913
	0.0	0.0109	0.2803	0.0140	0.1993	0.0450	0.1868	0.0247	0.1847
	0.5	0.0276	0.2032	0.0290	0.1330	0.0264	0.1316	0.0360	0.1138
	0.8	0.0189	0.1085	0.0036	0.0743	0.0108	0.0592	0.0058	0.0563

Table S3: Bias and Mean Square Error (MSE) of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$, $\hat{\lambda}_{II,QML}$ and $\hat{\lambda}_{QML,BC}$ at $n = 30, 50, 100, 200$ for $\lambda_0 = -0.8, -0.5, 0, 0.5, 0.8$ when W is randomly generated (10^3 repl. and ϵ is generated from a t-distribution with 3 degrees of freedom).

		n=30		n=50		n=100		n=200	
	λ	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
OLS	0.5	-0.2409	0.3710	-0.2600	0.3581	-0.2247	0.3199	-0.2201	0.2759
	-0.5	-0.2194	0.3806	-0.2338	0.3350	-0.2133	0. 3110	-0.2580	0.3082
	0.3	-0.2169	0.3739	-0.2629	0.3597	-0.2435	0.3272	-0.2217	0.3049
	0.8	-0.1908	0.3806	-0.2446	0.2979	-0.1674	0.2253	-0.2203	0.2591
		BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
ML	0.5	-0.3296	0.2842	-0.3591	0.2818	-0.3315	0.2433	-0.3221	0.2184
	-0.5	-0.1584	0.3008	-0.1630	0.2599	-0.1366	0.2330	-0.1727	0.2280
	0.3	-0.2680	0.2785	-0.3103	0.2657	-0.2928	0.2393	-0.2747	0.2179
	0.8	-0.3845	0.2925	-0.4285	0.2816	-0.3706	0.2121	-0.4089	0.2486
		BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
II-OLS	0.5	-0.1088	0.3096	-0.0569	0.2709	-0.0291	0.2243	-0.0012	0.2114
	-0.5	0.0056	0.4066	-0.0058	0.3172	0.0148	0.2797	-0.0095	0.2445
	0.3	-0.0228	0.3252	-0.0106	0.2640	-0.0267	0.2365	0.0049	0.2429
	0.8	-0.0745	0.2836	-0.0739	0.1999	-0.0346	0.1501	-0.0495	0.2010
		BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
RGMM	0.5	-0.3558	0.4851	-0.3826	0.4141	-0.3386	0.4745	-0.3116	0.3852
	-0.5	-0.2262	0.4949	-0.1886	0.3416	-0.1527	0.3401	-0.1623	0.2459
	0.3	-0.3325	0.5306	-0.3195	0.3656	-0.2559	0.3512	-0.2906	0.3493
	0.8	-0.4222	0.5166	-0.3247	0.3520	-0.3468	0.3058	-0.3709	0.3328

Table S4: Bias and Mean Square Error (MSE) of OLS, ML, II-OLS and RGMM for model (S.1) with $K = 3$ at $n = 30, 50, 100, 200$ for $\lambda_0 = 0.5, -0.5, 0.3, 0.8$ when W is randomly generated (10^3 repl. and ϵ is generated according to (S.1)