

INDIRECT INFERENCE IN SPATIAL AUTOREGRESSION*

Maria Kyriacou[†] Peter C. B. Phillips[‡] Francesca Rossi[§]

December 6, 2016

Abstract

Ordinary least squares (OLS) is well known to produce an inconsistent estimator of the spatial parameter in pure spatial autoregression (SAR). This paper explores the potential of indirect inference to correct the inconsistency of OLS. Under broad conditions, it is shown that indirect inference (II) based on OLS produces consistent and asymptotically normal estimates in pure SAR regression. The II estimator used here is robust to departures from normal disturbances and is computationally straightforward compared with quasi maximum likelihood (QML). Monte Carlo experiments based on various specifications of the weight matrix show that: (i) the indirect inference estimator displays little bias even in very small samples and gives overall performance that is comparable to the QML while raising variance in some cases; (ii) indirect inference applied to QML also enjoys good finite sample properties; and (iii) indirect inference shows robust performance in the presence of heavy tailed error distributions.

Keywords: Bias, Binding function, Inconsistency, Indirect Inference, Spatial autoregression, Weight matrix.

JEL classification: C21, C23

*We thank the CoEditor and two referees for helpful comments and suggestions on the original version of this paper. Phillips acknowledges research support from the NSF under Grant No. SES 12-58258.

[†]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.

1 Introduction

Cross-section correlation poses a considerable challenge in econometric work that affects modelling, estimation, and inference. Correlation across spatial data is typically ubiquitous, arising from multiple sources such as competition, regulatory practices, spillover and aggregation effects, and the influence of macroeconomic factors on individual decision making. Spatial correlation can be transmitted in an econometric model via observed variables or unobserved disturbances. Parsimonious models such as the spatial autoregression (SAR) of Cliff and Ord (1981) have become increasingly popular in practical work. These models offer a useful and easily implemented framework for describing irregularly-spaced correlated spatial data, where space can be interpreted in general terms as a network and correlation may depend on various forms of economic distance, include physical distance as a special case. A central advantage of SAR models is the fact that exact empirical knowledge of location is not required. Instead, location effects, wider economic distance effects, and irregularly-spaced data effects may all be embodied in an $n \times n$ weight matrix (where n is the size of the dataset) that can be constructed by the practitioner using all available relevant information.

Given an n -vector of spatial observations y we consider the following simple (pure) SAR model

$$y = \lambda_0 W y + \epsilon, \quad (1.1)$$

where λ_0 denotes the spatial parameter, and ϵ is a vector of independent and identically distributed (iid) disturbances with mean zero and unknown variance σ_0^2 . The weight matrix W carries spatial correlation effects, is exogenously specified, and satisfies certain restrictions that facilitate asymptotic analysis. So elements of W typically depend on n and are likely to change as n increases. Thus, the components $W = W_n$, $y = y_n$ and $\epsilon = \epsilon_n$ are, in fact, triangular arrays, even though the subscript n is often omitted for notational simplicity.

Asymptotic properties of various parametric estimators of λ_0 in (1.1) and more general SAR models that include exogenous regressors have been extensively studied in recent years. In particular, under certain conditions on the behaviour of W as n increases, Lee (2004) derived asymptotic properties of the Gaussian maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators of λ_0 . Lee (2002) showed that the OLS estimator of λ_0 in (1.1) is inconsistent, while OLS applied to a more general SAR model with exogenous regressors can be consistent and asymptotically normal under stronger conditions on W . Estimates of SAR models based on generalized methods of moments (GMM) have been studied by Lee (2001), Lee (2007) and Liu et al. (2010), and they have been extended by Lin and Lee (2010) and Kelejian and Prucha (2010) to accommodate unobserved heterogeneity in the disturbances.

While asymptotic properties are generally favourable, small sample performance of SAR parameter estimates can be poor. Poor performance is particularly serious in the pure SAR model (1.1) since rates of convergence to the true value may be slower than usual \sqrt{n} parametric rates depending on the limit behaviour of W . Correspondingly, statistical tests about the spatial parameter that are

based on asymptotic theory can also be unreliable. Much Monte Carlo work has been conducted to study the finite sample performance of SAR estimates and tests (e.g. Anselin and Florax (1995), Das et al. (2003) and Egger et al. (2009)). But finite sample theory and analytic bias corrections are at a much earlier stage of development, in comparison to related work in areas such as panel data modeling. Recently, Bao and Ullah (2007) derived second-order bias and mean squared error formulae for the ML estimator of λ_0 in (1.1) using Nagar moment expansions, and Bao (2013) extended these results to a more general model that includes exogenous regressors and possibly non-normal disturbances. The literature about finite sample corrections for tests is now developing and includes both the derivation of finite sample corrections for t-type of tests (Robinson and Rossi (2014b)) and refinements for Moran I/LM statistics (e.g. Cliff and Ord (1981), Robinson (2008), Baltagi and Yang (2013) and Robinson and Rossi (2014a)).

The present paper uses indirect inference (II) methods to derive a new OLS-based estimation procedure that shows good performance and involves simpler computations than QML estimation of λ_0 in (1.1). Our use of indirect inference involves a mechanism to deliver an indirect bias correction that involves simulations or the indirect use of asymptotic approximations, as in Phillips (2012). The II estimator of λ_0 is consistent, asymptotically normal, and enjoys good finite sample behavior. II methods were originally introduced by Gouriéroux et al. (1993) and Smith (1993) to deal with models with intractable objective functions. The methods have also achieved success in bias correction under various time series settings (e.g. Gouriéroux et al. (2000)). Applications of II to obtain improved finite sample inference have been discussed in Phillips and Yu (2009) in a contingent claims pricing context, where II estimates display virtually no bias and often smaller variance compared to standard ML. Also, Gouriéroux et al. (2010) use II to accomplish bias reduction in dynamic panels and Phillips (2012) shows that II delivers improved estimation, even asymptotically, in a first order autoregression with potential nonstationarity. But these methods have so far never been applied to spatial data.

Given the novelty of II methodology in the spatial literature, this paper explores its use and develops the corresponding limit theory within the pure SAR model (1.1) with homogeneous disturbances. Our main result demonstrates the power of indirect inference in achieving corrections, showing how simple OLS estimation can be transformed to produce a consistent and asymptotically normal estimate of the spatial parameter. Extensions of our new method presented in this paper to SAR models with heterogeneous disturbances and/or a set of exogenous regressors is under investigation in separate work, results appear promising, and some findings are reported here. Furthermore, extensions of this method to models in which the spatial lag enters nonlinearly are possible due to the flexibility of II and more generally of simulation-based techniques. By means of a set of Monte Carlo simulations we also show how the II methodology can be applied to the standard QML estimator of λ_0 in order to reduce its small sample bias. Although such QML-based estimation of λ_0 performs very well in the case of model (1.1), we stress the importance of the OLS-based method for its degree of generality. In fact, it is well known that QML is not, in general, robust to the presence of heteroskedasticity of unknown form and thus the good performance of a simple QML-based II es-

timator may be lost in practical applications due to heterogeneity. Also, even though the analytical bias expansions developed in Bao (2013) would provide the groundwork to construct an II estimator based on QML when a set of regressors is included, its practical implementation and the need to substitute unknown higher order moments with their estimates would pose some computational challenges. Here the simplicity of the OLS-based procedure is an advantage and the approach can be extended to neatly accommodate heterogeneity and the presence of exogenous regressors.

The approach is defined and discussed in the next section, together with the main assumptions used in the asymptotic development. Section 3 provides the main results relating to the asymptotic distribution of the II estimator, and Section 4 reports simulation findings concerning finite sample performance for different forms of the spatial weight matrix W . In Section 5 we report some further examples of weight matrices that are amenable to exact analysis and comparison with the ML estimate of λ_0 in (1.1), while in Section 6 some extensions of our results are presented and discussed. Section 7 has concluding remarks and some discussion of extensions of the II methodology in spatial models. Proofs are given in the Appendix and in the Online Supplement to this paper (Kyriacou et al, 2016), which also provides further simulation findings.

Throughout the paper, λ_0 and σ_0^2 denote true values of these parameters while λ and σ^2 denote admissible values. We write $S_n(x) = S(x) = I - xW$, where I denotes the $n \times n$ identity matrix, and $G_n(x) = G(x) = WS^{-1}(x)$. We set $G = G(\lambda_0)$ and use A_{ij} to signify the ij 'th element of the matrix A . We denote by $\bar{\eta}(\cdot)$ the spectral radius, $\|\cdot\|$ and $\|\cdot\|_\infty$ the spectral norm and uniform absolute row sum norm, respectively, and K represents an arbitrary finite, positive constant whose value may change in each location. The notation $f^{(i)}(\cdot)$ denotes the i 'th derivative of the function $f(\cdot)$.

2 Indirect Inference in the Pure SAR Model

We consider model (1.1) whose reduced form is

$$y = S^{-1}(\lambda_0)\epsilon, \quad (2.1)$$

under assumed invertibility of $S(\lambda_0)$. We use the following assumptions:

Assumption 1 For all n , the elements of $\epsilon \sim_{iid} (0, \sigma_0^2)$ with unknown variance σ_0^2 and, for some $\delta > 0$, $\mathbb{E}(\epsilon_i)^{4+\delta} \leq K$.

Assumption 2 $\lambda_0 \in \Lambda$, where Λ is a closed subset in $(-1, 1)$.

Assumption 3

- (i) For all n , $W_{ii} = 0$.

(ii) For all n , $\|W\| \leq 1$.

(iii) For all sufficiently large n , $\|W\|_\infty + \|W'\|_\infty \leq K$.

(iv) For all sufficiently large n , uniformly in $i, j = 1, \dots, n$, $W_{ij} = O(1/h)$, where $h = h_n$ is bounded away from zero for all n and $h/n \rightarrow 0$ as $n \rightarrow \infty$.

Assumption 4 For all sufficiently large n , $\sup_{\lambda \in \Lambda} \|S^{-1}(\lambda)\|_\infty + \|S^{-1}(\lambda)'\|_\infty \leq K$.

Assumption 5 The limits

$$\lim_{n \rightarrow \infty} \frac{h}{n} \text{tr}(G' G^j) \quad \text{with } 1 \leq i + j \leq 3, \quad \lim_{n \rightarrow \infty} \frac{h}{n} \text{tr}((G' G)^2), \quad (2.2)$$

$$\lim_{n \rightarrow \infty} \frac{h}{n} \sum_i (G_{ii})^2, \quad \lim_{n \rightarrow \infty} \frac{h}{n} \sum_i (G' G)_{ii}^2, \quad \lim_{n \rightarrow \infty} \frac{h}{n} \sum_i G_{ii} (G' G)_{ii} \quad (2.3)$$

all exist and

$$\lim_{n \rightarrow \infty} \frac{h}{n} \text{tr}((G + G') G' G) \neq 0. \quad (2.4)$$

Assumptions 2 and 3(ii), or some other related conditions are common in the SAR literature to ensure existence of a reduced form and define the likelihood function (e.g. Lee (2004)). Although not the only possibility, the set defined in Assumption 2 together with 3(ii) seems natural in most applications since existence of $S^{-1}(\lambda)$ is assured and its power series representation (which will be extensively used in Section 6) holds, so that for all $\lambda \in \Lambda$

$$\|S^{-1}(\lambda)\| = \left\| \sum_{s=0}^{\infty} \lambda^s W^s \right\| \leq \sum_{s=0}^{\infty} |\lambda|^s \|W\|^s \leq (1 - |\lambda|)^{-1} \leq K. \quad (2.5)$$

Detailed discussion on the choice of the parameter space of λ and further restrictions to guarantee existence of the reduced form (2.1) are given in Kelejian and Prucha (2010). In fact, QML estimation relies on the existence of a reduced form and $S^{-1}(\lambda)$ for $\lambda \in \Lambda$ under Assumptions 2 and 3(ii), while OLS estimation does not rely on such restrictions and so Indirect Inference implemented on OLS can be defined for values of λ beyond the $(-1, 1)$ space. Assumption 3(ii) is not particularly restrictive, since any W can be rescaled by its spectral norm so that $\|W\| \leq 1$ is trivially satisfied. Assumption 3(iii) (Kelejian and Prucha (1998)) rules out strong spatial dependence and it is evidently satisfied when each unit has a finite number of neighbours as n increases. When $W_{ij} = O(1/h)$, which is common practice when dealing with SAR models (e.g. Lee (2004)), then we impose $h/n \rightarrow 0$ along with Assumption 4 to establish a central limit theorem for quadratic forms (e.g. Robinson (2008)). From a practical perspective, Assumptions 3(iii) together with 3(iv) rule out the case in which a unit is related to all other units as n increases. Assumptions 3(iii) and 4 are satisfied, for instance, when W is row normalised so that $Wl = l$, where l indicates an $n \times 1$ column of ones, symmetric and with positive entries.

By a standard argument, under Assumption 3,

$$\frac{h}{n} \text{tr}(W^p W'^q) = O(1), \quad \forall p, q \text{ s.t. } p + q > 1, \quad (2.6)$$

as $n \rightarrow \infty$. Also, under Assumptions 3 and 4 as $n \rightarrow \infty$,

$$\frac{h}{n} \text{tr}(G(\lambda)^p G(\lambda)'^q) = O(1), \quad \forall p, q \text{ s.t. } p + q \geq 1, \quad (2.7)$$

since $\|S^{-1}(\lambda)\|_\infty + \|S^{-1}(\lambda)'\|_\infty \leq K$ uniformly in λ . Assumption 5 is required to impose existence and nonsingularity of limits of certain sequences that figure in the asymptotic development. The sequences in (2.2) are bounded as $n \rightarrow \infty$ according to (2.7) and converge under Assumption 5. Sequences in (2.3) are $O(1/h)$ and vanish as n increases when h is a divergent sequence and (2.3) ensures that limits are well defined also in case $h = O(1)$ as $n \rightarrow \infty$. Condition (2.4) ensures nonsingularity of the asymptotic variance in our main theorem, since by the Cauchy inequality

$$0 < \left(\frac{h}{n}\right)^2 (\text{tr}((G+G')G'G)^2) < \left(\frac{h}{n}\right)^2 \text{tr}((G+G')^2) \text{tr}((G'G)^2) < 2 \left(\frac{h}{n}\right)^2 \text{tr}(G'G) \text{tr}((G'G)^2). \quad (2.8)$$

The OLS estimator of λ_0 is given by the ratio $\hat{\lambda} = y'W'y/y'W'Wy$, and by a standard argument as $n \rightarrow \infty$

$$\hat{\lambda} - \lambda_0 \rightarrow_p \lim_{n \rightarrow \infty} \frac{h \text{tr}G/n}{h \text{tr}(G'G)/n}. \quad (2.9)$$

As $n \rightarrow \infty$ $\lim_{n \rightarrow \infty} h \text{tr}(G'G)/n \neq 0$ under Assumption 5 and (2.8), the limit in (2.9) exists and is bounded. However, unless W is restricted to very specific choices, it is difficult to calculate the limit on the right side of (2.9) and give an analytic expression as a function of λ_0 .

According to the usual indirect inference calculations, for any $\lambda \in \Lambda$ we can generate B sets of pseudo-data $y^b = (y_1^b, y_2^b, \dots, y_n^b)'$, $b = 1, 2, \dots, B$ from the true model (under assumed Gaussianity of ϵ) and for each pseudo-data set b the OLS estimator of λ is computed as

$$\hat{\lambda}^b = \hat{\lambda}^b(\lambda) = \frac{y^b(\lambda)'W'y^b(\lambda)}{y^b(\lambda)'W'Wy^b(\lambda)} = \lambda + \frac{y^b(\lambda)'W'\epsilon^b}{y^b(\lambda)'W'Wy^b(\lambda)}, \quad b = 1, \dots, B. \quad (2.10)$$

The II estimator of λ_0 , $\hat{\lambda}_{II}$, is then defined by the extremum problem

$$\hat{\lambda}_{II} = \underset{\lambda}{\text{argmin}} \left| \hat{\lambda} - \frac{1}{B} \sum_{b=1}^B \hat{\lambda}^b(\lambda) \right|, \quad (2.11)$$

that produces an estimator that aligns the sample mean of the simulations to the observed $\hat{\lambda}$. As $B \rightarrow \infty$, (2.11) becomes

$$\hat{\lambda}_{II} = \underset{\lambda}{\text{argmin}} |\hat{\lambda} - \mathbb{E}_b(\hat{\lambda}^b(\lambda))|, \quad (2.12)$$

where the expectation operator \mathbb{E}_b is interpreted with respect to the pseudo-variate ϵ^b .

We define the binding function as

$$b_n(\lambda) = \mathbb{E}_b(\hat{\lambda}^b(\lambda)) = \lambda + \mathbb{E}_b\left(\frac{\epsilon'^b G(\lambda)' \epsilon^b}{\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b}\right), \quad (2.13)$$

and introduce the following condition.

Assumption 6

- (i) For all n , the binding function $b_n(\lambda)$ is continuous and strictly increasing for all $\lambda \in \Lambda$.
- (ii) $\lim_{n \rightarrow \infty} b_n^{(1)}(\lambda_0)$ exists and is positive.

Assumption 6 reveals the key conditions under which II converts an inconsistent estimator, $\hat{\lambda}$ here, into a consistent estimator, which is a central contribution of the paper. Section 5 shows analytically that for various choices of W Assumption 6 is satisfied; but as the boundary value $\lambda = 1$ is approached, the binding function becomes nearly flat, which indicates that near the bound the II method is less effective due to reduced information in the approximate binding function, $b^*(\lambda)$, about the true value of λ . It would be useful to establish general primitive conditions on W or, possibly, on the parameter space Λ and W under which Assumption 6 is satisfied. However, derivation of such conditions is likely possible only in special cases. As is usual practice, therefore, we rely in general on numerical methods to verify the validity of this assumption.

For each $\lambda \in \Lambda$ we have the formal moment expansion (Lieberman (1994))

$$\mathbb{E}_b\left(\frac{\epsilon'^b G(\lambda)' \epsilon^b}{\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b}\right) = \frac{\mathbb{E}_b(\epsilon'^b G(\lambda)' \epsilon^b)}{\mathbb{E}_b(\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b)} + \theta_{1n} + \theta_{2n} + \theta_{3n} + \dots, \quad (2.14)$$

where

$$\theta_{1n} = \frac{\mathbb{E}_b(\epsilon'^b G(\lambda)' \epsilon^b) \text{cum}_2}{(\mathbb{E}_b(\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b))^3} - \frac{\text{cum}_{11}}{(\mathbb{E}_b(\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b))^2}, \quad (2.15)$$

cum_p is the p 'th cumulant of $\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b$, cum_{1p} is the p 'th generalised cumulant of the product of $\epsilon'^b G(\lambda)' \epsilon^b$ and $\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b$ (e.g. McCullagh (1987)), while θ_{in} for $i > 1$ are functions of cum_p , cum_{1p} , and moments of $\epsilon'^b G(\lambda)' G(\lambda) \epsilon^b$ and $\epsilon'^b G(\lambda)' \epsilon^b$. As $n \rightarrow \infty$, under Assumptions 3, 4, 6 and by (2.7) the leading term in (2.14) is $O(1)$, and $\theta_{1n} = O(h/n)$.

By observing that higher-order terms in (2.14) are of increasingly smaller order (the computation is tedious and is not reported here), we may have the formal expansion for the binding function

$$b_n(\lambda) = \lambda + \frac{\text{tr}(G(\lambda))}{\text{tr}(G(\lambda)' G(\lambda))} + O\left(\frac{h}{n}\right). \quad (2.16)$$

An advantage of Lieberman's result is the fact that (2.14) and (2.16) do not rely on the normality of ϵ^b , so that procedures based on them should have some invariance properties with respect to the underlying data distribution.

Since we restrict our analysis to the class of W matrices such that Assumption 6 holds, we have the simple inverse function formulation

$$\hat{\lambda}_{II} = b_n^{-1}(\hat{\lambda}). \quad (2.17)$$

In practice we can construct $\hat{\lambda}_{II}$ by generating a large number B of pseudo-data to approximate the binding function by

$$\frac{1}{B} \sum_{b=1}^B \hat{\lambda}^b(\lambda). \quad (2.18)$$

However, distributional assumptions are required to generate the pseudo-data and since the asymptotic variance of $\hat{\lambda}$ depends on the fourth cumulant of the ϵ_i , as we will show, this mechanism is not fully robust to distributional misspecification. Instead, we construct $\hat{\lambda}_{II}$ using the approximate analytic form of the binding function

$$b_n^*(\lambda) = \lambda + \frac{\text{tr}(G(\lambda))}{\text{tr}(G(\lambda)'G(\lambda))}, \quad (2.19)$$

which holds more generally under Assumption 1. We will show that $\hat{\lambda}_{II}$ obtained by (2.19) is consistent and asymptotically normal without any additional distributional assumption, unlike the estimator $\hat{\lambda}$ which is biased in finite samples and also inconsistent (Lee (2002)). The generality offered by an implementation based on (2.19) offsets the potential gain of an estimator with possibly smaller finite sample bias, which might be achieved by using the binding function (2.18) based on simulations for sufficiently large B . Derivation of higher order bias adjustment also seems unnecessary not only because $b_n^*(\lambda)$ is close to $b_n(\lambda)$, but also because $b_n(\lambda)$ is found to be stable across different distributions, a feature which conforms to already well-known results, such as those in Andrews (1993). The binding function plots shown in Figures 1 and 4 verify that the approximate binding functions $b_n^*(\lambda)$ across different specifications of W offer very good approximations to the “exact” binding functions based on (2.18) from either Gaussian or standard Cauchy innovations.

Whereas the term “indirect inference”, as originally developed in econometrics, refers to methods where the parameters are estimated indirectly via a mechanism that involves simulations from another related model, the terminology may also be sensibly used in bias correction problems where simulations are employed to compute the binding function either within the procedure itself or separately in extensive simulation exercises or even via analytics. In such cases, simulations or analytic expansion formulae such as (2.16) enable indirect or implicit correction of the estimator, which captures the main idea of indirect inference, the key notion being indirect rather than explicit direct correction of the bias. Importantly in the present case, the implicit bias correction leads also to correction of inconsistency.

3 Limit Distribution of $\hat{\lambda}_{II}$

In the notation that follows some quantities are given an affix (subscript) n to emphasize their n -dependence. Let $g_{ij} = \text{htr}(G^i G^{j'})/n$, and $g = \text{htr}((G'G)^2)/n$. Define the centering quantity $\bar{\lambda}_n = \lambda_0 + \frac{g_{10}}{g_{11}}$. By a standard delta argument,

$$\hat{\lambda} - \bar{\lambda}_n = \left(\frac{h}{n}\right)^{1/2} f'_n \left(\frac{h}{n}\right)^{1/2} U_n + o_p \left(\left(\frac{h}{n}\right)^{1/2} \right), \quad (3.1)$$

where

$$U_n = (y'W\epsilon - \text{tr}(G)\sigma_0^2; \quad y'W'Wy - \text{tr}(G'G)\sigma_0^2)' \quad (3.2)$$

and

$$f_n = \left(\left(\frac{h}{n} y'W'Wy \right)^{-1}, \quad - \left(\frac{h}{n} y'W'Wy \right)^{-2} \left(\frac{h}{n} y'W\epsilon \right) \right)'. \quad (3.3)$$

Theorem 1

(a) Under (1.1) and Assumptions 1-5

$$\left(\frac{n}{h}\right)^{1/2} (\hat{\lambda} - \bar{\lambda}_n) \xrightarrow{d} \mathcal{N}(0, \omega), \quad (3.4)$$

where

$$\omega = \lim_{n \rightarrow \infty} \left(\frac{g_{20} + g_{11}}{g_{11}^2} - \frac{4g_{10}g_{21}}{g_{11}^3} + \frac{2g_{10}^2g}{g_{11}^4} + \frac{h}{n} \frac{\kappa_4}{\sigma_0^4 g_{11}^2} \sum_{i=1}^n (G_{ii} - g_{10}g_{11}^{-1}(G'G)_{ii})^2 \right) \quad (3.5)$$

and $\kappa_4 = \mathbb{E}(\epsilon_i^4) - 3\sigma_0^4$.

(b) Under (1.1) and Assumptions 1-6

$$\left(\frac{n}{h}\right)^{1/2} (\hat{\lambda}_{II} - \lambda_0) \rightarrow_d N(0, \omega^*), \quad (3.6)$$

where

$$\begin{aligned} \omega^* = \lim_{n \rightarrow \infty} & (g_{11} + g_{20})^{-1} \left(1 - \frac{2g_{10}g_{21}}{g_{11}(g_{20} + g_{11})} \right)^{-2} \left(1 - \frac{4g_{21}g_{10}}{g_{11}(g_{11} + g_{20})} + \frac{2gg_{10}^2}{g_{11}^2(g_{11} + g_{20})} \right. \\ & \left. + \frac{h}{n} \frac{\kappa_4}{\sigma_0^4(g_{11} + g_{20})} \sum_{i=1}^n (G_{ii} - g_{10}g_{11}^{-1}(G'G)_{ii})^2 \right). \end{aligned} \quad (3.7)$$

The proof is sketched in the Appendix and full details are reported in the Online Supplement. The limits on the right sides of (3.5) and (3.7) exist and are strictly positive under Assumptions 5 and 6.

Theorem 1 enables a comparison between $\hat{\lambda}_{II}$ and the Gaussian maximum likelihood estimator $\hat{\lambda}_{ML}$. When $\epsilon_i \sim_{iid} \mathcal{N}(0, \sigma^2)$, we have $\kappa_4 = 0$ and then, from Lee (2004),

$$\left(\frac{n}{h}\right)^{1/2} (\hat{\lambda}_{ML} - \lambda_0) \xrightarrow{d} N(0, V_{ML}), \quad (3.8)$$

where

$$V_{ML} = \lim_{n \rightarrow \infty} \left(g_{20} + g_{11} - \frac{2}{h} g_{10}^2 \right)^{-1}. \quad (3.9)$$

For $\lambda_0 = 0$, a case that is especially relevant in testing, $\text{tr}(G) = 0$ and $\omega^* = V_{ML}$, so that indirect inference and maximum likelihood have equivalent limit distributions. On the other hand, from Robinson and Rossi (2014b), when $\lambda_0 = 0$

$$\left(\frac{n}{h}\right)^{1/2} \hat{\lambda} \xrightarrow{d} N(0, V_{OLS}), \quad (3.10)$$

where $V_{OLS} = (g_{11}^2 / (g_{11} + g_{20}))^{-1}$. Furthermore, since $\hat{\lambda}$ is inconsistent when $\lambda_0 \neq 0$, a Wald test based on $\hat{\lambda}$ may not be reliable for all values of λ_0 . By contrast, a Wald test based on $\hat{\lambda}_{II}$ is equivalent to one based on the MLE and is consistent against any alternative value for λ_0 .

4 Simulations

Simulations were conducted to assess the finite sample performance of $\hat{\lambda}_{II}$ in (2.17) in relation to $\hat{\lambda}$ and the QML estimator of λ_0 , $\hat{\lambda}_{QML}$. We consider two different specifications of the weight matrix W : a circulant matrix and an asymmetric toeplitz matrix. Bias and mean square error (MSE) were computed for values of $\lambda_0 \in \{-0.8, -0.5, 0, 0.5, 0.8\}$ and the sample size range is $n \in \{30, 50, 100, 200\}$. In the reported results, the disturbances ϵ_i are generated from a t -distribution with 5 degrees of freedom with 10^4 replications. Results based on Gaussian errors appear to follow a very similar pattern and are available from the authors upon request. We implement OLS-based indirect inference using the approximate binding function $b_n^*(\cdot)$ in (2.19) to obtain $\hat{\lambda}_{II}$. Simulation results suggest that $b_n^*(\cdot)$ closely approximates the true value $\mathbb{E}(\hat{\lambda})$, a feature which is also verified analytically for some simple choices of W shown in Section 5.

Although the main idea explored in this paper is the application of the II methodology to $\hat{\lambda}$ to restore its consistency, an II estimator based on QML (rather than OLS), $\hat{\lambda}_{II, QML}$, can also be constructed based on $\hat{\lambda}_{QML}$ following the same principles discussed in previous sections. Although consistent, $\hat{\lambda}_{QML}$ does suffer from finite-sample bias, as evidenced in the findings of Bao and Ullah (2007) and Bao (2013). We construct $\hat{\lambda}_{II, QML}$ using the analytical bias expansion of Bao (2013), expression (2.15), p. 78, to obtain an approximate QML binding function, $b_{n, QML}^*(\lambda)$. The explicit form of $b_{n, QML}^*(\lambda)$ is omitted here to avoid introducing further notation. Provided monotonicity

requirements similar to Assumption 6 are satisfied, then

$$\hat{\lambda}_{II,QML} = b_{n,QML}^{*-1}(\hat{\lambda}_{QML}). \quad (4.1)$$

Both the QML binding function and estimator $\hat{\lambda}_{II,QML}$ itself depend on the unknown excess kurtosis parameter γ_2 , which satisfies $E(\epsilon_i^4) = (\gamma_2 + 3)\sigma_0^4$. We use the QML residuals \hat{e}_i , $i = 1, \dots, n$ to estimate γ_2 with $\hat{\gamma}_2 = \frac{\hat{\mu}_4}{\hat{\mu}_2} - 3$, where $\hat{\mu}_4 = (\sum_{i=1}^n \hat{e}_i^4)/n$ and $\hat{\mu}_2 = (\sum_{i=1}^n \hat{e}_i^2)/n$. Although $\hat{\gamma}_2$ is a consistent estimator of γ_2 , using it for calculating the QML binding function may result in an increase of the variance of the II-QML estimator and may affect its asymptotic theory. For each choice of W the entries in Tables 1-3 display the bias and mean squared error (MSE) of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$ and $\hat{\lambda}_{II,QML}$, labelled as OLS, II-OLS, QML and II-QML respectively. For comparison purposes, the last rows of Tables 1 and 2 report bias and MSE of the bias-corrected QML estimator using the analytical bias expression of Bao (2013), $\hat{\lambda}_{BC,QML}$ (denoted by BC-QML in Tables). This estimator is calculated by bias-correcting $\hat{\lambda}_{QML}$ using the analytical expression for the leading bias term as in expression (2.15) in Bao (2013) and by estimating the excess kurtosis parameter term γ_2 as discussed above. Additional simulation results based on a different weight matrix structure can be found in the Supplementary Material.

Case (i): Circulant weights

Our first example is a row-normalised weight matrix with a circulant structure, W_C , similar to the one used by Kelejian and Prucha (1999) defined as

$$W_C = \frac{1}{\|A_C\|} A_C, \quad (4.2)$$

where A_C is a circulant matrix with leading row $(0, 1, 1, 0, \dots, 0, 1, 1)$, W_C in (4.2) is normalised with respect to its spectral norm so that $\|W_C\| = 1$. Assumptions 3 – 5 are readily verified with $h = \|A_C\|$, which in this case remains fixed as $n \rightarrow \infty$.

Figure 1 depicts the “exact” and “approximate” OLS binding functions b_n^* as well as the QML approximate binding function $b_{n,QML}^*$ all with $n = 100$. For the latter, we employ the true value of the excess kurtosis parameter γ_2 under the t distribution with 5 degrees of freedom.¹ Some key features of the binding functions are immediately apparent from the plots in Figure 1. The simulated binding functions following (2.18) (denoted as “exact” henceforth) are based on $B = 50000$ pseudo-datasets drawn from either Gaussian or standard Cauchy innovations, while the “approximate” binding OLS binding function, $b_n^*(\lambda)$ is based on (2.19). The exact and approximate binding functions are indistinguishable which confirms that $b_n^*(\lambda)$ serves as a valuable approximation at very little computational cost without relying on a restrictive distributional assumption on ϵ . The (approximate) QML binding function $b_{n,QML}^*$ is noticeably monotonic and increasing over the full

¹This could be replaced with $\hat{\gamma}_2$ since $\hat{\lambda}_{QML}$ is consistent.

domain $-1 < \lambda < 1$, whereas the OLS binding functions $b_n(\cdot)$ and $b_n^*(\cdot)$ are monotonic and decreasing with a slope that becomes steeper as $\lambda \rightarrow 1^2$, so that the inverse binding function is also monotonic but becomes flatter as $\lambda \rightarrow 1$.

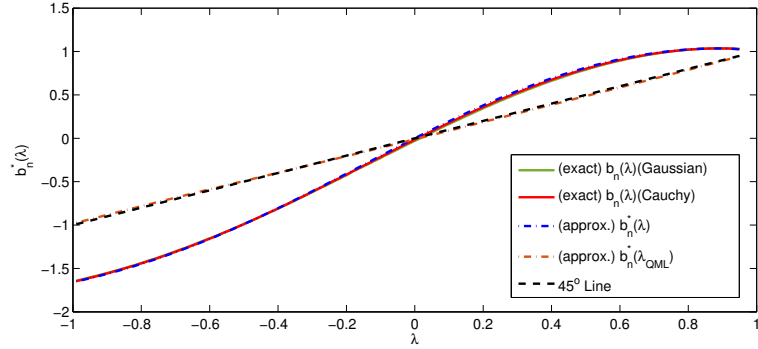


Figure 1: Exact OLS binding functions $b_n(\cdot)$ based on Gaussian and standard Cauchy innovations ($B = 50000$), approximate OLS binding functions, $b_n^*(\cdot)$ and approximate QML binding functions, $b_{n,QML}(\cdot)$, for $\lambda \in (-1, 1)$ when W is (4.2) at $n = 100$.

Table 1 summarises the bias and MSE of OLS, II-OLS, QML, II-QML and BC-QML when W is chosen to have a circulant structure as in (4.2). The entries in the top panel of Table 1 reveal that the OLS estimator $\hat{\lambda}$ suffers from substantial bias for all values of λ_0 . In accordance with asymptotic theory (Lee (2002)), the bias does not vanish as n increases. In fact, for a given $\lambda_0 \neq 0$, the bias seems to increase with n and becomes particularly severe when λ_0 is negative. The entries in the last four panels of Table 1 reveal that the II-OLS, QML, II-QML and BC-QML provide substantial reductions in the OLS bias and MSE. In moderately large sample sizes ($n = 100, 200$) II-OLS often outperforms QML in bias reduction without much compromise in the MSE. The MSE of $\hat{\lambda}_{II}$ is comparable to the MSE of $\hat{\lambda}_{QML}$ in most cases other than when λ_0 is close to unity, as might be expected from the shape of the OLS binding function $b_n^*(\cdot)$ which becomes flat as λ approaches unity. The entries in the last panel of Table 1 indicate that Indirect Inference applied to $\hat{\lambda}_{QML}$ achieves the best results both in terms of bias and MSE reduction. In most cases II-QML outperforms BC-QML, although their respective performance is in general very satisfactory. Importantly, the MSE of $\hat{\lambda}_{II,QML}$ does not suffer any deterioration as λ approaches unity, consonant with the form of the binding function $b_{n,QML}^*$ over the full domain of λ .

To shed light on their distributional characteristics, Figure 2 plots the simulated density functions of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$, $\hat{\lambda}_{II,QML}$ and $\hat{\lambda}_{BC,QML}$ for $n = 100$ when $\lambda_0 = 0.5$. The distribution of $\hat{\lambda}$ is seen to be severely upward biased (centred around 0.85 rather than its true value of 0.5), whereas the II-OLS, QML, II-QML and BC-QML estimators appear to be almost unbiased on the scale of this

²Extended binding plots for values of λ beyond unity are shown in the Supplementary Material.

figure. All five estimators seem to have similar dispersion in this case.

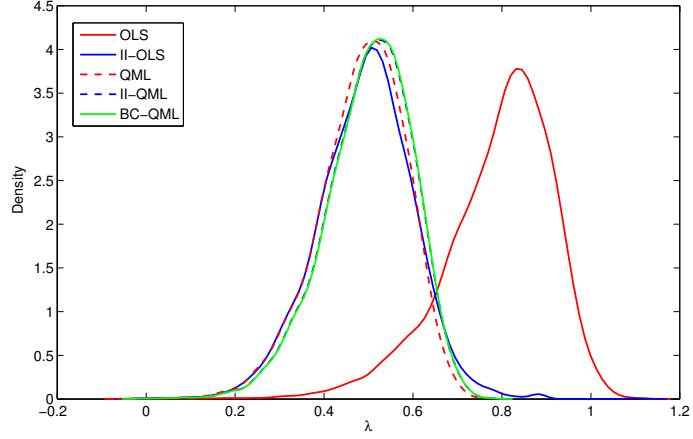


Figure 2: Empirical densities of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$, $\hat{\lambda}_{II,QML}$ and $\hat{\lambda}_{BC,QML}$ for $\lambda_0 = 0.5$ when W is chosen as in (4.2) at $n = 100$.

Direct analytic comparison of the variances of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{QML}$ is difficult since (3.7) and the asymptotic variance of the QML estimator (Lee (2004)) are highly complicated non-linear functions of the weight matrix. Figure 3 shows how the finite sample variances of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{QML}$ vary with λ_0 at $n = 100$. The variances are close for small to moderate spatial autocorrelation, but as $|\lambda_0|$ increases ω^* in (3.7) increases rapidly as λ_0 tends to unity. This variance increase can be attributed to flatness in the binding function $b_n(\lambda)^*$ as λ approaches the boundaries of the support in this case.

Case (ii): Asymmetric Toeplitz weights

We next consider an asymmetric toeplitz weight matrix W_{AT} . Using the circulant matrix A_C as a starting point, we introduce asymmetry in the weight matrix structure by removing the neighbourhood effect of the $(n - 1)$ 'th unit on the first unit in (4.2). This produces a three-element neighbourhood effect in each row rather than four. Specifically, we define

$$W_{AT} = \frac{1}{\|A_{AT}\|} A_{AT} \quad (4.3)$$

where the leading row of A_{AT} is $(0, 1, 1, 0, \dots, 0, 0, 0, 1)$.

The weight matrix is again row-normalised, in accordance with Assumption 3. Figure 4 depicts the approximate binding function for $n = 100$, verifying that both $b_n^*(\cdot)$ and $b_{n,QML}^*(\cdot)$ are monotonic over a large subset of $(-1, 1)$. For $\lambda > 0.8$ the OLS binding function flattens out although not as markedly as in the symmetric case examined previously.

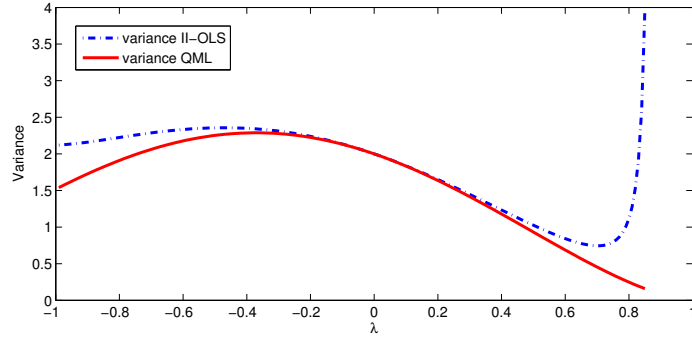


Figure 3: Finite sample variances of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{QML}$ for $\lambda \in (-1, 1)$ when W is chosen as in (4.2) at $n = 100$.

The simulation results reported in Table 2 confirm that QML, II-OLS, BC-QML and II-QML estimators provide substantial reductions to both the bias and MSE of OLS. Interestingly, for most configurations, QML and II-OLS display similar performance: II-OLS generally has smaller bias than QML when $\lambda_0 > 0$, without any evident increase in MSE and for $n = 200$ largely reproduces the performance characteristics of QML. These results along with the entries of Table 1 indicate that once symmetry is removed from the weight matrix structure, the QML performance starts to weaken. Similarly to the circulant W case, II-QML outperforms both II-OLS and QML in terms of bias and often outperforms QML in MSE reduction when λ_0 is positive. The entries of the last block of rows displayed in Table 2 show that in this case BC-QML is very effective in reducing the bias of $\hat{\lambda}_{QML}$ and its performance is comparable to that of II-QML. Indeed, this is expected as the bias of $\hat{\lambda}_{BC, QMLE}$ (and also $\hat{\lambda}_{II, QMLE}$) is $o(h/n)$, while the bias of $\hat{\lambda}_{II, OLS}$ is $O(h/n)$. The bias reduction of II-OLS, QML, II-QML and BC-QML compared to OLS is clearly confirmed by the plots of the empirical densities reported in Figure 5 for the parameter setting $n = 100$ and $\lambda_0 = 0.5$.

Figure 6 displays comparisons of the finite sample variances of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{QML}$ over $\lambda_0 \in (-1, 1)$, revealing that the plots virtually overlap for most admissible values of λ_0 , with discrepancies emerging as $|\lambda_0|$ tends to unity.

5 Examples

In this section we consider a few examples for which we may analyze whether the binding function $b_n(\lambda)$ in (2.16) is invertible, at least as $n \rightarrow \infty$, rather than relying on numerical work, as in the

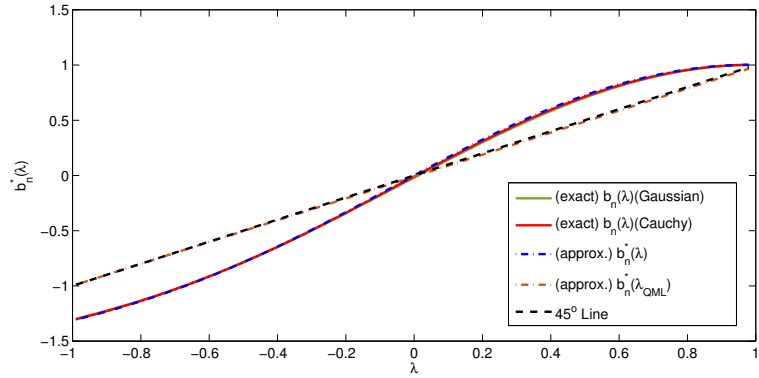


Figure 4: Exact OLS binding functions $b_n(\cdot)$ based on Gaussian and standard Cauchy innovations ($B = 50000$), approximate OLS binding functions, $b_n^*(\cdot)$ and approximate QML binding functions, $b_{n,QML}(\cdot)$, for $\lambda \in (-1, 1)$ when W is (4.3) at $n = 100$.

plots of Figures 1 and 4. In some cases, an analytic comparison between the performance of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{QML}$ is also possible.

Example (i): The Districts Model

The simplest choice of W that is amenable to analysis and facilitates a comparison between (3.7) and (3.9) is the block diagonal ‘districts model’ weight matrix W (Case (1991)) which is defined as

$$W_n = I_r \otimes B_m, \quad B_m = \frac{1}{m-1}(l_m l_m' - I_m), \quad (5.1)$$

where I_s is the $s \times s$ identity matrix, l_m is an m -vector of 1’s, and \otimes is the Kronecker product. It is easy to verify that W in (5.1) satisfies Assumptions 3 and 4 with $n = mr$ and $h = m - 1$. The specification (5.1) indicates that within a particular district (block) the spatial dependence has the same form, whereas it is zero between blocks.

The approximate binding function $b_n^*(\cdot)$ shown in Figure 3 of the Online Supplement is invertible for $\lambda \in (-1, 1)$ and for all sample sizes but it flattens considerably as λ approaches unity.

Theorem 2 Let W defined as in (5.1).

- (a) As $n \rightarrow \infty$ the binding function b_n in (2.16) is strictly increasing for all $\lambda \in \Lambda$.
- (b) If $1/m + 1/r \rightarrow 0$ $\hat{\lambda}_{II}$ is asymptotically equivalent to $\hat{\lambda}_{QML}$.

A sketch of the proof of Theorem 2 is given in the Appendix and a full development is in the Online Supplement. The condition in part (b) of Theorem 2 corresponds to a case of divergent h .

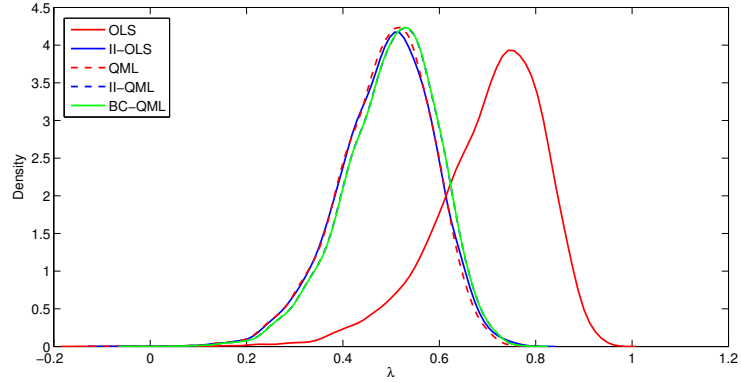


Figure 5: Empirical densities of $\hat{\lambda}$, $\hat{\lambda}_{II}$, $\hat{\lambda}_{QML}$, $\hat{\lambda}_{II,QML}$ and $\hat{\lambda}_{BC,QML}$ for $\lambda_0 = 0.5$ when W is chosen as in (4.3) at $n = 100$.

Example (ii): Circulant Weight Matrix Model

As another example we consider the simple circulant matrix C with leading row $(0, 1, 0, \dots, 0, 1)$. and

$$W = \frac{1}{2}C, \quad (5.2)$$

so that $\|W\| = 1$ and $h = 2$ for all n .

From Figure S4 in the Online Supplement where $n = 100$, the approximate binding function $b_n^*(\lambda)$ in this case is strictly monotonic for $\lambda \in (-0.7, 0.7)$ but becomes almost flat (and even decreases slightly) as $\lambda \rightarrow 1$, with related behavior as $\lambda \rightarrow -1$. Similar behavior was found in simulations for the case where W was chosen as in (4.2). We have the following analytic result.

Theorem 3 Define W as in (5.2). As $n \rightarrow \infty$, $b_n(\lambda)$ in (2.16) is strictly increasing for all $\lambda \in \Lambda$, where Λ is any closed subset of $(-\sqrt{3}/2, \sqrt{3}/2)$.

A sketch of the proof of Theorem 3 is given in the Appendix and details are in the Online Supplement. In principle we can extend the argument below to any choice of W with a toeplitz structure, and thus to circulants with more than “one behind and one ahead” neighbors. However, this would require numerical solutions of integrals and is beyond the scope of the present example.

From (3.7) and the results reported in the Supplement (viz., (S.39), (S.42) and (S.45)) we also conclude that $\omega^* \rightarrow \infty$ as $\lambda_0 \rightarrow \pm \frac{\sqrt{3}}{2}$, since

$$1 - \frac{2g_{10}g_{21}}{g_{11}(g_{20} + g_{11})} \rightarrow 0 \quad \text{as} \quad \lambda \rightarrow \pm \frac{\sqrt{3}}{2}. \quad (5.3)$$

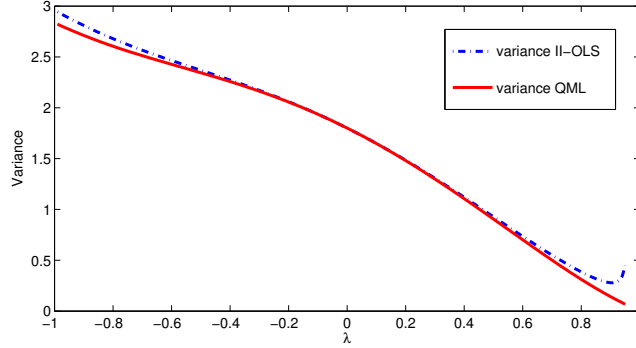


Figure 6: Finite sample variances of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{QML}$ for λ in $(-1, 1)$ when W is chosen as in (4.3) at $n = 100$.

This result, even though it is derived under the simpler circulant weight matrix (5.2), is consistent with the Monte Carlo results based on the weight matrix W defined in (4.2). Hence both analytic and simulation findings reveal that for circulant weight matrices W the indirect inference estimator $\hat{\lambda}_{II}$ can be obtained by inversion of the binding function for small through moderate values of λ_0 and performs well as an estimator over this domain.

6 Extensions

The simplest generalisation of results derived in Section 3 involves the inclusion of an unknown intercept μ_0 in the SAR model, so that

$$y = \mu_0 l + \lambda_0 W y + \epsilon, \quad (6.1)$$

where l is an n -vector of ones, W is row normalized, so that $Wl = l$, y is the n -dimensional observation vector and ϵ is a vector of iid disturbances.

The OLS estimator of λ_0 in (6.1) is

$$\tilde{\lambda} = \frac{y' W' P y}{y' W' P W y}, \quad (6.2)$$

where $P = I - ll'/n$. When W is row normalized, it is easy to verify by a series expansion of $S^{-1}(\lambda_0)$

that the reduced form of (6.1) is

$$y = S^{-1}(\lambda_0)(\mu_0 l + \epsilon) = \frac{\mu_0}{1 - \lambda_0} l + S^{-1}(\lambda_0)\epsilon. \quad (6.3)$$

Thus, by standard algebra and observing that $l'Gl/n = O(1)$ under Assumptions 3 and 4, we conclude that (2.9) holds with $\hat{\lambda}$ replaced by $\tilde{\lambda}$ and the formal expansion for b_n in (2.16) is still appropriate so that we can define the II estimator of λ_0 in (6.1) as $\tilde{\lambda}_{II} = b_n^{-1}(\tilde{\lambda})$. Thus, Theorem 1 holds with $\hat{\lambda}$ replaced by $\tilde{\lambda}$ and $\hat{\lambda}_{II}$ replaced by $\tilde{\lambda}_{II}$. When W is not row normalized, the asymptotic theory for the OLS of λ_0 in (6.1) would be different, as $\tilde{\lambda}$ may be consistent and asymptotically normal with a standard \sqrt{n} rate under some additional conditions on the behaviour of W in the limit (see Lee (2002)). Since the focus of the present work is on using II to convert an inconsistent OLS estimator into a consistent estimator, we do not further pursue the case of model (6.1) with non-row normalized W .

Theorem 1 is also robust to mild forms of unobserved heterogeneity, such as the following.

Assumption 1' *For all n , the elements of ϵ are independent with mean zero and*

$$\mathbb{E}(\epsilon\epsilon') = D > 0, \quad \text{with } D = \sigma_0^2 I + C,$$

where $C = (C_{ii})$ is an $n \times n$ diagonal matrix with rank $c = c_n$, where c_n is a positive sequence satisfying $c_n = o(n)$, and uniformly in i and n $|C_{ii}| \leq K$. For some $\delta > 0$

$$\sup_{1 \leq i \leq n, n \geq 1} \mathbb{E}(\epsilon_i)^{4+\delta} \leq K.$$

If either $1/h + c/h \rightarrow 0$ or $h = O(1)$ and $c = O(1)$ as $n \rightarrow \infty$ the probability limit in (2.9), the formal expansion for $b_n(\lambda)$ in (2.16) and the asymptotic distribution in Theorem 1 still holds.

Although our theoretical results have been derived under Assumption 1, i.e. for error terms with finite moments up to order $4 + \delta$, $\delta > 0$, the Supplementary Material reports Monte Carlo results for SAR estimation under heavy tailed errors. Specifically, Tables S1-S3 in Section S.3 illustrate the behaviour of II-based estimators compared to OLS, QML and BC-QML when the errors are generated from a Student t distribution with 3 degrees of freedom across different W specifications. These results confirm that II estimators continue to enjoy good robustness properties in all cases with heavy tailed errors.

Cases of general heteroskedasticity and the extension of our method to a SAR models with exogenous regressors (SARX) is under investigation in ongoing work. Dealing with error heterogeneity involves new theory and technical challenges, so we do not report details on the implementation of our methodology to SARX under unknown heteroskedasticity here. But to illustrate the potential of the methods some Monte Carlo results are tabulated in Section S.3 of the Online Supplement.

These findings are promising and reveal, in particular, that II estimation in its extended version that accommodates error heterogeneity dominates both standard QML, as might be expected in view of its inconsistency, and the robust GMM estimator which is known to be consistent (Lin and Lee, 2010).

7 Conclusions

Indirect inference methodology can be used in pure spatial autoregression to convert the inconsistent OLS estimator of the spatial parameter into a consistent and asymptotically normal estimator. The method is simple to implement and its performance characteristics are broadly comparable to the QML and can be superior in terms of bias reduction, although variance typically increases when the binding function flattens out towards the boundary of the domain of definition of λ . In addition, the II methodology can be straightforwardly applied to other estimation techniques. In particular, Monte Carlo simulations show that II adaptations of QML successfully remove much of the finite sample bias of QML. In fact, II-QML estimation enjoys the best finite sample behaviour in many cases, especially those where there is some random organization of the elements of the weight matrix. In this latter case, simulations confirm that standard QML has finite sample performance characteristics that are sensitive to the weight matrix structure.

The present approach complements earlier work on analytic bias corrections of ML or QML estimators (Bao and Ullah, 2007; Bao, 2013) and offers an alternative mechanism of improving finite sample performance. The results of the present paper, although novel for spatial regressions, are limited by the restrictive assumptions implied by the pure SAR model (1.1), viz. a single spatial lag (and thus a single weight matrix W), a linear functional form for the spatial lag, and homoskedastic disturbances. Subject to this limitation, the II methodology enjoys the advantages of the flexibility of simulation based methods, in comparison to analytic expansions for bias functions and densities.

Allowance for heterogeneity is of particular importance in practical work. It is well known (Lin and Lee (2010)) that ML or QML fail to be consistent when the disturbances are heterogeneously distributed. Extensions of the indirect inference methodology based on OLS to SAR models with unknown heteroskedasticity and a set of exogenous regressors seems promising. As indicated above, some preliminary results of this extended methodology are given in S.3 of the Supplementary Material. A full development of this extension will be reported in subsequent work.

Appendix

Proof of Theorem 1

The proof of part (a) is carried out in a similar way to Robinson (2008) and a detailed derivation is given in the Online Supplement. To prove part (b), let $q = b_n^{-1}(x)$ and, for any function $v(x)$

$dv^r(x)/dx^r = v^{(r)}(x)$. By standard algebra

$$b_n^{(1)}(x) = 1 + \frac{\text{tr}(G(x)^2)\text{tr}(G'(x)G(x)) - 2\text{tr}G(x)\text{tr}(G'(x)G(x)^2)}{(\text{tr}(G'(x)G(x)))^2} + O\left(\frac{h}{n}\right), \quad (\text{A.1})$$

which is non-zero under Assumption 6 and $O(1)$ under Assumptions 3 and 4. Also,

$$b_n^{-1(1)}(x)|_{x=b_n(\lambda_0)} = (b_n^{(1)}(q))^{-1}|_{q=b_n^{-1}(b_n(\lambda_0))=\lambda_0}. \quad (\text{A.2})$$

Since $\bar{\lambda}_n = b_n(\lambda_0) + O(h/n)$, by Taylor expansion,

$$\begin{aligned} b_n^{-1}(\bar{\lambda}_n) &= b_n^{-1}\left(b_n(\lambda_0) + O\left(\frac{h}{n}\right)\right) \\ &= b_n^{-1}(b_n(\lambda_0)) + (b_n^{(1)}(x))^{-1}|_{x=\lambda_0} O\left(\frac{h}{n}\right) + \dots = \lambda_0 + O\left(\frac{h}{n}\right) \end{aligned} \quad (\text{A.3})$$

and thus

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda}_{II} - \lambda_0) = \left(\frac{n}{h}\right)^{1/2}(b_n^{-1}(\hat{\lambda}) - b_n^{-1}(\bar{\lambda}_n)) + o(1). \quad (\text{A.4})$$

We can derive the asymptotic distribution of the latter by means of the extended Delta method (Phillips, 2012) if the derivative sequence $\{b_n^{-1(1)}(x)\}$ is asymptotically locally relatively equicontinuous, which in this case is equivalent to showing

$$\left| \frac{b_n^{(1)}(\lambda_0) - b_n^{(1)}(r)}{b_n^{(1)}(r)} \right| \rightarrow 0 \quad (\text{A.5})$$

as $n \rightarrow \infty$, uniformly in $\mathcal{N}_\delta = \{r \in \mathfrak{R} : |s(r - \lambda_0)| < \delta, \delta > 0\}$, $s = s_n \rightarrow \infty$ and $s(h/n)^{1/2} \rightarrow 0$. Under Assumptions 3, 4 and 6,

$$\begin{aligned} \left| \frac{b_n^{(1)}(\lambda_0) - b_n^{(1)}(r)}{b_n^{(1)}(r)} \right| &\leq K |b_n^{(1)}(\lambda_0) - b_n^{(1)}(r)| \\ &\leq K \left(\left| \frac{g_{20}}{g_{11}} - \frac{h\text{tr}(G(r)^2)/n}{h\text{tr}(G'(r)G(r))/n} \right| + \left| \frac{g_{10}g_{21}}{g_{11}^2} - \frac{h^2\text{tr}(G(r))\text{tr}(G(r)^2G'(r))/n^2}{(h\text{tr}(G'(r)G(r))/n)^2} \right| \right). \end{aligned} \quad (\text{A.6})$$

The first term of the latter expression is bounded by

$$\begin{aligned} &K \left(\left| g_{20} - \frac{h}{n}\text{tr}(G(r)^2) \right| + \left| g_{11} - \frac{h}{n}\text{tr}(G(r)'G(r)) \right| \right) \\ &= K \left(\left| \frac{h}{n}\text{tr}(G(\lambda^*)^2)(\lambda_0 - r) \right| + \left| \frac{h}{n}\text{tr}(G(\lambda^*)'G(\lambda^*))(\lambda_0 - r) \right| \right) \\ &\leq K|\lambda_0 - r| \leq s^{-1}\delta \end{aligned} \quad (\text{A.7})$$

as $n \rightarrow \infty$, where the first equality follows by the mean value theorem, λ^* indicating an intermediate point between λ_0 and r . The second term in (A.6) can be dealt with in a similar fashion.

Therefore, since $b_n^{-1(1)}(\bar{\lambda}_n) = (b_n^{(1)}(\lambda_0))^{-1} + O(h/n)$,

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda}_{II} - \lambda_0) \rightarrow_d N(0, \omega^*), \quad (\text{A.8})$$

where

$$\begin{aligned} \omega^* = & \lim_{n \rightarrow \infty} (g_{11} + g_{20})^{-1} \left(1 - \frac{2g_{10}g_{21}}{g_{11}(g_{20} + g_{11})}\right)^{-2} \left(1 - \frac{4g_{21}g_{10}}{g_{11}(g_{11} + g_{20})} + \frac{2gg_{10}^2}{g_{11}^2(g_{11} + g_{20})}\right) \\ & + \frac{h}{n} \frac{\kappa_4}{\sigma_0^4(g_{11} + g_{20})} \sum_{i=1}^n (G_{ii} - g_{10}g_{11}^{-1}(G'G)_{ii})^2, \end{aligned} \quad (\text{A.9})$$

giving the required result.

Proof of Theorem 2

From the block diagonal structure, we have

$$\text{tr}(G) = \text{tr} \left(\sum_{i=0}^{\infty} \lambda_0^i W^{i+1} \right) = r \sum_{i=0}^{\infty} \lambda_0^i \text{tr}(B_m^{i+1}), \quad (\text{A.10})$$

where B_m has one eigenvalue equal to 1 and the other $(m-1)$ eigenvalues equal to $-1/(m-1)$, so that

$$\text{tr}(B_m^{i+1}) = 1 + (m-1) \left(\frac{-1}{m-1} \right)^{i+1}. \quad (\text{A.11})$$

Thus

$$\frac{h}{n} \text{tr}(G) = \frac{\lambda_0}{1 - \lambda_0} \frac{(m-1)}{m-1 + \lambda_0}, \quad (\text{A.12})$$

and, for $s \geq 2$,

$$\frac{h}{n} \text{tr}(G^s) = \frac{m-1}{m} \frac{1}{(1 - \lambda_0)^s} + (-1)^s \frac{(m-1)^2}{m(m-1 + \lambda_0)^s}. \quad (\text{A.13})$$

Proofs of (a) and (b) involve now only routine calculations and are reported in the Supplementary Material. In particular, as $m \rightarrow \infty$, we find that $b'_n(\lambda) \rightarrow 2(1 - \lambda) > 0$.

Proof of Theorem 3

The proof is based on the following results involving well-known asymptotic formulae for traces of

toeplitz matrices, as $n \rightarrow \infty$,

$$\begin{aligned} \frac{1}{n} \text{tr}(G(\lambda)) &= \frac{1}{n} \sum_{s=0}^{\infty} \lambda^s \text{tr}(W^{s+1}) \rightarrow \sum_{s=0}^{\infty} \lambda^s \frac{1}{2^{s+1}} \frac{1}{2\pi} \int_0^{2\pi} (2\cos x)^{s+1} dx \\ &= \frac{1}{\lambda} \sum_{s=1}^{\infty} \lambda^s \frac{1}{2\pi} \int_0^{2\pi} (\cos x)^s dx. \end{aligned} \quad (\text{A.14})$$

$$\frac{1}{n} \text{tr}(G(\lambda)^2) = \frac{1}{n} \sum_{s,t=0}^{\infty} \lambda^{s+t} \text{tr}(W^{s+t+2}) \rightarrow \frac{1}{2\pi} \sum_{s,t=0}^{\infty} \lambda^{s+t} \int_0^{2\pi} (\cos x)^{s+t+2} dx, \quad (\text{A.15})$$

$$\frac{1}{n} \text{tr}(G(\lambda)^3) \rightarrow \frac{1}{2\pi} \sum_{s,t,q=0}^{\infty} \lambda^{s+t+q} \int_0^{2\pi} (\cos x)^{s+t+q+3} dx. \quad (\text{A.16})$$

These results may then be used to establish the theorem. Complete proofs are given in the Supplement.

References

- Andrews, D.W.K (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* **61**, 139-165.
- Anselin, L. and R. Florax (1995). Small sample properties of tests for spatial dependence in regression models: some further results. In L. Anselin and R. Florax (Eds.) *New Directions in Spatial Econometrics*. Springer-Verlag, Berlin, 21-74.
- Arbia, G. (2006). *Spatial econometrics: statistical foundation and applications to regional analysis*. Springer-Verlag, Berlin.
- Baltagi, B.H. and Z. Yang (2013). Standardized LM tests for spatial error dependence in linear or panel regressions. *Econometrics Journal* **16**, 103-134.
- Bao, Y. (2013) Finite-sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* **29**, 68-88.
- Bao, Y and A. Ullah (2007). Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics*, **137**, 396-413.
- Case, A.C. (1991). Spatial patterns in household demand. *Econometrica*, **59**, 953-965.
- Cliff, A. & J.K. Ord (1981) *Spatial processes: models & applications*. Pion.
- Das, D., H.H. Kelejian, and I. Prucha (2003). Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science*, **82**, 1-26.
- Egger, P., M. Larch, M. Pfaffermayr and J. Walde (2009). Small sample properties of maximum likelihood versus generalized method of moments based tests for spatially autocorrelated errors. *Regional Science and Urban Economics*, **39**, 670-78.
- Gouriéroux, C, A. Montfort and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics* **8**, S85-S118.
- Gouriéroux, C., P.C.B. Phillips and J. Yu (2010). Indirect inference for dynamic panels. *Journal of Econometrics* **157**, 68-77.
- Gouriéroux, C., Renault, E., Touzi, N. (2000). Calibration by simulation for small sample bias correction. In: Mariano, R.S., Schuermann, T., Weeks, M. (Eds.), *Simulation-Based Inference in Econometrics: Methods and Applications*. Cambridge University Press, 328-58.
- Kelejian, H.H. and I.R. Prucha (1998). A generalized spatial two-stages least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* **17**, 99-121.

- Kelejian, H.H. and I.R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* **40**, 509-533.
- Kelejian, H.H. and I.R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* **157**, 53-67.
- Kyriacou, M., P.C.B. Phillips and F. Rossi (2016). Online Supplement to “Indirect Inference in Spatial Autoregression”, Working paper.
- Lee, L.F. (2001). Generalized method of moments estimation of spatial autoregressive processes. Preprint.
- Lee, L.F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* **137**, 489-514.
- Lee, L.F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric theory* **18**, 252-277.
- Lee, L.F. (2004). Asymptotic distribution of quasi-maximum likelihood estimates for spatial autoregressive models. *Econometrica* **72**, 1899-1925.
- Liu, X., L.F. Lee and C.R. Bollinger (2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics* **159**, 303-319.
- Lin, X. and L.F. Lee (2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics* **157**, 34-52.
- Lieberman, O. (1994). A Laplace approximation to the moments of a ratio of quadratic forms. *Biometrika* **81**, 681-90.
- McCullagh, P. (1987). *Tensor methods in statistics*. Chapman and Hall, London.
- Phillips, P.C.B. (2012). Folklore theorems, implicit maps, and indirect inference. *Econometrica* **80**, 425-454.
- Phillips, P.C.B. and Yu, J. (2009). Simulation-based estimation of contingent-claims prices. *Review of Financial Studies* **22**, 3669-3705.
- Robinson, P.M. (2008). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics* **147**, 5-16.
- Robinson, P.M. and F. Rossi (2014a). Improved Lagrange multiplier tests for spatial autoregressions. *Econometrics Journal* **17**, 139-164.

- Robinson, P.M. and F. Rossi (2014b). Refined tests for spatial correlation. Forthcoming in *Econometric Theory* .
- 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.
- Smith, A.A. (1993). Estimating non-linear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8, S63-S84.

n		30		50		100		200	
OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.2278	0.1462	0.2690	0.1102	0.2897	0.0990	0.2988	0.0958
	-0.5	0.4869	0.4582	-0.4923	0.3783	-0.4867	0.3083	-0.4924	0.2780
	0.0	-0.0964	0.2627	-0.0575	0.1606	-0.0286	0.0803	-0.0148	0.0393
	0.5	-0.2250	0.1323	0.2665	0.1086	0.2918	0.1006	0.3021	0.0981
	0.8	0.2084	0.0525	0.2164	0.0501	0.2211	0.0500	0.2218	0.0497
II-OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0225	0.0562	-0.0113	0.0269	-0.0084	0.0116	-0.0074	0.0051
	-0.5	-0.0345	0.0902	-0.0399	-0.0399	-0.0075	0.0209	-0.0044	0.0133
	0.0	-0.0491	0.0804	-0.0234	0.0417	-0.0157	0.0204	-0.0073	0.0098
	0.5	-0.0240	0.0498	-0.0264	0.0297	-0.0115	0.0116	-0.0047	0.0054
	0.8	-0.0185	0.0251	-0.0084	0.0131	0.0054	0.0078	0.0100	0.0045
QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0496	0.0447	-0.0283	0.0214	-0.0156	0.0097	-0.0105	0.0045
	-0.5	-0.0048	0.0668	-0.0118	0.0687	-0.004	0.0190	0.088	0.0119
	0.0	-0.0452	0.0682	-0.0227	0.0389	-0.0154	0.0196	-0.0072	0.0096
	0.5	-0.0508	0.0383	-0.0430	0.0247	-0.0182	0.0102	-0.0087	0.0047
	0.8	-0.0400	0.0151	-0.0224	0.0070	-0.0080	0.0028	-0.0055	0.0013
II-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0031	0.0426	0.0003	0.0205	-0.0011	0.0095	-0.0032	0.0044
	-0.5	-0.005	0.0775	-0.0130	0.0798	-0.0029	0.0203	-0.0065	0.0124
	0.0	-0.0119	0.0743	-0.0015	0.0410	0.0046	0.0201	-0.0017	0.0097
	0.5	-0.0038	0.0359	-0.0143	0.0231	-0.0037	0.0099	-0.0014	0.0046
	0.8	-0.0038	0.0120	0.0023	0.0070	0.0016	0.0028	0.0022	0.0013
BC-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.0079	0.0623	0.0087	0.0387	0.0066	0.0191	0.0062	0.0098
	-0.5	0.0035	0.0751	0.0035	0.0450	0.0020	0.0213	0.0048	0.0104
	0.0	-0.0055	0.0715	-0.0048	0.0421	0.0046	0.0204	0.0045	0.0107
	0.5	-0.0032	0.0372	-0.0026	0.0214	-0.0049	0.0093	0.0046	0.0045
	0.8	-0.0017	0.0130	-0.0006	0.0063	0.0079	0.0027	-0.0023	0.0013

Table 1: 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 given by (4.2) (10^4 repl. and ϵ is generated from a t-distribution with 5 degrees of freedom).

n		30		50		100		200	
OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.3088	0.1843	-0.3296	0.1609	-0.3336	0.1363	-0.3417	0.1287
	-0.5	-0.2837	0.2122	-0.2921	0.1666	-0.2848	0.1231	-0.2858	0.1031
	0.0	-0.0681	0.1598	-0.0377	0.0979	-0.0182	0.0489	-0.0121	0.0246
	0.5	0.1503	0.0810	0.1829	0.0628	0.2046	0.0547	0.2158	0.0523
	0.8	0.1302	0.0254	0.1439	0.0241	0.1524	0.0244	0.1564	0.0249
II-OLS	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0209	0.0966	-0.0232	0.0585	-0.0104	0.0266	-0.0104	0.0127
	-0.5	-0.0292	0.0800	-0.0298	0.0506	-0.0135	0.0223	-0.0041	0.0129
	0.0	-0.0392	0.0646	-0.0194	0.0356	-0.0173	0.0193	-0.0025	0.0086
	0.5	-0.0320	0.0375	-0.0151	0.0203	-0.0092	0.0096	-0.0047	0.0048
	0.8	-0.0114	0.0277	-0.0085	0.0135	0.0002	0.0054	-0.0032	0.0022
QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0156	0.0930	-0.0187	0.0557	-0.0075	0.0256	-0.0088	0.0122
	-0.5	-0.0241	0.0766	-0.0256	0.0484	-0.0111	0.0215	-0.0032	0.0127
	0.0	-0.0390	0.0620	-0.0194	0.0350	-0.0172	0.0190	-0.0025	0.0085
	0.5	-0.0425	0.0334	-0.0203	0.0192	-0.0124	0.0092	-0.0063	0.0045
	0.8	-0.0453	0.0190	-0.0296	0.0087	-0.0105	0.0035	-0.0072	0.0018
II-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	-0.0031	0.0930	-0.0076	0.0566	-0.0029	0.0256	-0.0071	0.0123
	-0.5	0.0092	0.0788	-0.0114	0.0488	-0.0047	0.0217	-0.0023	0.0128
	0.0	-0.0033	0.0634	0.0021	0.0356	-0.0065	0.0190	0.0029	0.0086
	0.5	0.0016	0.0323	0.0064	0.0191	0.0093	0.0092	0.0037	0.0045
	0.8	-0.0011	0.0167	-0.0032	0.0078	0.0027	0.0034	-0.0059	0.0018
BC-QML	λ_0	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.8	0.0108	0.0983	-0.0030	0.0561	0.0033	0.0254	0.0042	0.0142
	-0.5	0.0095	0.0787	-0.0097	0.0466	-0.0093	0.0252	-0.0052	0.0113
	0.0	-0.0064	0.0661	0.0021	0.0373	0.0014	0.0177	-0.0011	-0.0090
	0.5	-0.0023	0.0366	-0.0029	0.0190	0.0022	0.0091	-0.0012	0.0044
	0.8	-0.0020	0.0083	0.0019	0.0078	0.0017	0.0032	0.0013	0.0046

Table 2: 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 given by AT (10^4 repl. and ϵ is generated from a t-distribution with 5 degrees of freedom).