INDIRECT INFERENCE IN SPATIAL AUTOREGRESSION*

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

September 23, 2014

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 is robust to departures from normal disturbances and is computationally straightforward compared with pseudo Gaussian maximum likelihood (PML). Monte Carlo experiments based on various specifications of the weighting matrix confirm that the indirect inference estimator displays little bias even in very small samples and gives overall performance that is comparable to the Gaussian PML.

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

*Phillips acknowledges research support from the NSF under Grant No. SES 12-58258.

[†]University of Southampton

[‡]Yale University, University of Auckland, University of Southampton, Singapore Management University

[§]University of Southampton. Corresponding author. E-mail address: F.Rossi@soton.ac.uk. Tel: +44 (0)23 8059 3847.

JEL classification: C21, C23

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

$$y = \lambda_0 W y + \epsilon, \tag{1.1}$$

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 pseudo-maximum likelihood (PML) 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 of 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 favorable, 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 much simpler computations than PML estimation of λ_0 in (1.1). The II estimator 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 within the pure SAR model (1.1) with homogeneous disturbances. Our main result demonstrates the power of the indirect inference, showing how simple OLS estimation can be transformed to produce a consistent and asymptotically normal estimate of the spatial parameter. Extensions of this approach to ML estimation, to SAR models with heterogeneous disturbances, and to models in which the spatial lag enters nonlinearly are possible and appear promising, due to the flexibility of II and more generally of simulation-based techniques.

The new 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. Some further examples of weight matrices that are amenable to exact analysis and comparison with the ML estimate of λ_0 in (1.1) are presented in Section 5. Section 6 has concluding remarks and some discussion of extensions of the II methodology in spatial models. Proofs are given in the Appendix.

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 use $||.||$ and $||.||_\infty$ to indicate the spectral norm and uniform absolute row sum norm, respectively, and K represents an arbitrary finite, positive constant. The notation $f^{(i)}(.)$ denotes the i'th derivative of the function $f(.)$.

2 Indirect Inference in the Pure SAR Model

We consider model (1.1) whose reduced form is

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

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

Assumption 1 *For all* n*, the elements of* $\epsilon \sim_{iid} \left(0, \sigma_0^2\right)$ *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* $i, j = 1, ..., n$, $W_{ij} = O(1/h)$*, where* $h = h_n$ *is bounded away from zero for all* n *and* $h/n \to 0$ *as* $n \to \infty$.

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*

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

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

all exist and

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

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)). The choice of the parameter space in Assumption 2 together with 3(ii) seems natural in most applications since they are sufficient to guarantee existence of $S^{-1}(\lambda)$ and its power series representation, which in turn implies that $\forall\lambda \in \Lambda$

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

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 obviously 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 $h/n \to 0$ 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. Assumption 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,

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

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

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

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 $n \to \infty$ 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 $n \to \infty$. Condition (2.4) ensures nonsingularity of the asymptotic variance in our main theorem, since by the Cauchy inequality

$$\begin{aligned} 0 &< \left(\frac{h}{n}\right)^2 (\mathrm{tr}((G+G')G'G)^2 < \left(\frac{h}{n}\right)^2 \mathrm{tr}((G+G')^2)\mathrm{tr}((G'G)^2) \\ &< 2\left(\frac{h}{n}\right)^2 \mathrm{tr}(G'G)\mathrm{tr}((G'G)^2). \end{aligned} \tag{2.8}$$

The OLS estimator of λ_0 is given by the ratio

$$\hat{\lambda} = \frac{y'W'y}{y'W'Wy}, \tag{2.9}$$

and by a standard argument as $n \to \infty$

$$\hat{\lambda} - \lambda_0 \to_p \lim_{n\to\infty} \frac{h\mathrm{tr}G/n}{h\mathrm{tr}(G'G)/n}. \tag{2.10}$$

As $n \to \infty$ $\lim_{n\to\infty} h\mathrm{tr}(G'G)/n \neq 0$ under Assumption 5 and (2.8), and the limit in (2.10) exists and is bounded. But unless W is restricted to very specific choices, it is difficult to calculate the right side limit of (2.10) 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 $y^b = (y_1^b, y_2^b, ..., y_n^b)'$, $b = 1, 2, ..., B$ from the true model (under assumed Gaussianity of ϵ) and for each pseudo-data set the

OLS estimator of λ is computed as

$$\hat{\lambda}^b = \hat{\lambda}^b(\lambda) = \frac{y^b(\lambda)'W'y^b(\lambda)}{y^b(\lambda)'W'Wy^b(\lambda)} = \lambda + \frac{y^b(\lambda)'W'\epsilon^b}{y^b(\lambda)'W'Wy^b(\lambda)}, \quad b = 1, ..., B. \tag{2.11}$$

The II estimator of λ_0 is then defined by the extremum problem

$$\hat{\lambda}_{II} = \underset{\lambda}{\text{argmin}} |\hat{\lambda} - \frac{1}{B}\sum_{b=1}^{B} \hat{\lambda}^b(\lambda)|, \tag{2.12}$$

that produces an estimator that aligns the sample mean of the simulations to the observed $\hat{\lambda}$. As $B \to \infty$, (2.12) becomes

$$\hat{\lambda}_{II} = \underset{\lambda}{\text{argmin}} |\hat{\lambda} - \mathbb{E}_b(\hat{\lambda}^b(\lambda))|, \tag{2.13}$$

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

We define the binding function as

$$b_n(\lambda) = \mathbb{E}_b(\hat{\lambda}^b(\lambda)) = \lambda + \mathbb{E}_b\left(\frac{\epsilon'^b G(\lambda)'\epsilon^b}{\epsilon'^b G(\lambda)'G(\lambda)\epsilon^b}\right), \tag{2.14}$$

and introduce the next condition.

Assumption 6

(i) *For all n, the binding function $b_n(\lambda)$ is continuous and strictly increasing for all $\lambda \in \Lambda$.*

(ii) $\lim_{n\to\infty} b_n^{(1)}(\lambda_0)$ *exists and is positive.*

It would be useful to establish primitive conditions on W or, possibly, on the parameter space Λ and W under which Assumption 6 is satisfied. But such conditions are likely possible only in special cases. As is usual practice, we rely on numerical methods to check the validity of the assumption. Some examples are described in Section 5.

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

$$\mathbb{E}_b\left(\frac{\epsilon'^b G(\lambda)'\epsilon^b}{\epsilon'^b G(\lambda)'G(\lambda)\epsilon^b}\right) = \frac{\mathbb{E}_b(\epsilon'^b G(\lambda)'\epsilon^b)}{\mathbb{E}_b(\epsilon'^b G(\lambda)'G(\lambda)\epsilon^b)} + \theta_{1n} + \theta_{2n} + \theta_{3n} + ..., \quad (2.15)$$

where

$$\theta_{1n} = \frac{\mathbb{E}_b(\epsilon'^b G(\lambda)'\epsilon^b)\mathrm{cum}_2}{(\mathbb{E}_b(\epsilon'^b G(\lambda)'G(\lambda)\epsilon^b))^3} - \frac{\mathrm{cum}_{11}}{(\mathbb{E}_b(\epsilon'^b G(\lambda)'G(\lambda)\epsilon^b))^2}, \quad (2.16)$$

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_i 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 $n \to \infty$, under Assumptions 3, 4, 6 and by (2.7) the leading term in (2.15) is $O(1)$, and $\theta_1 = O(h/n)$.

By observing that higher-order terms in (2.15) are of increasingly smaller order (the computation is tedious and is not reported here), we write a formal expansion for $b_n(\lambda)$ as

$$b_n(\lambda) = \lambda + \frac{\mathrm{tr}(G(\lambda))}{\mathrm{tr}(G(\lambda)'G(\lambda))} + O\left(\frac{h}{n}\right). \quad (2.17)$$

An advantage of Lieberman's result is the fact that (2.15) and (2.17) 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

$$\hat{\lambda}_{II} = b_n^{-1}(\hat{\lambda}). \quad (2.18)$$

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

$$\frac{1}{B}\sum_{b=1}^{B}\hat{\lambda}^b(\lambda). \tag{2.19}$$

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

$$b_n^*(\lambda) = \lambda + \frac{\mathrm{tr}(G(\lambda))}{\mathrm{tr}(G(\lambda)'G(\lambda))}, \tag{2.20}$$

which holds more generally under Assumption 1. We will show that $\hat{\lambda}_{II}$ obtained by (2.20) is consistent and asymptotically normal without any additional distributional assumption, unlike $\hat{\lambda}$ which is not only biased in small samples, but also inconsistent (Lee (2002)). The generality offered by an implementation based on (2.20) offsets the potential gain of an estimator with an even smaller bias, which might be achieved by using the simulation based binding function (2.19) for B sufficiently large.

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 $g_{ij} = h\mathrm{tr}(G^i G^{j'})/n$, and $g = h\mathrm{tr}((G'G)^2)/n$. Define the centering quantity

$$\bar{\lambda}_n = \lambda_0 + \frac{g_{10}}{g_{11}}, \tag{3.1}$$

and by a standard delta argument,

$$\hat{\lambda} - \bar{\lambda}_n = \frac{h}{n} f_n' U_n + o_p\left(\left(\frac{h}{n}\right)^{1/2}\right), \tag{3.2}$$

where

$$U_n = \left(y'W\epsilon - \mathrm{tr}(G)\sigma_0^2; \quad y'W'Wy - \mathrm{tr}(G'G)\sigma_0^2\right)' \tag{3.3}$$

and

$$f_n = \left(\left(\frac{h}{n} y'W'Wy\right)^{-1}; \quad -\left(\frac{h}{n} y'W'Wy\right)^{-2}\left(\frac{h}{n} y'W\epsilon\right)\right)'. \tag{3.4}$$

Theorem 1

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

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda} - \bar{\lambda}_n) \underset{d}{\rightarrow} \mathcal{N}(0, \omega), \tag{3.5}$$

where

$$\omega = \lim_{n\to\infty}\left(\frac{g_{20} + g_{11}}{g_{11}^2} - \frac{4g_{10}g_{21}}{g_{11}^3} + \frac{2g_{10}^2 g}{g_{11}^4} + \frac{h}{n}\frac{\kappa_4}{\sigma_0^4 g_{11}^2}\sum_{i=1}^{n}(G_{ii} - g_{10}g_{11}^{-1}(G'G)_{ii})^2\right) \tag{3.6}$$

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

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

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda}_{II} - \lambda_0) \rightarrow_d N(0, \omega^*), \tag{3.7}$$

where

$$\omega^* = \lim_{n\to\infty} (g_{11}+g_{20})^{-1}\left(1-\frac{2g_{10}g_{21}}{g_{11}(g_{20}+g_{11})}\right)^{-2}\left(1-\frac{4g_{21}g_{10}}{g_{11}(g_{11}+g_{20})}+\frac{2gg_{10}^2}{g_{11}^2(g_{11}+g_{20})}\right. \tag{3.8}$$

$$\left.+\frac{h}{n}\frac{\kappa_4}{\sigma_0^4(g_{11}+g_{20})}\sum_{i=1}^{n}(G_{ii}-g_{10}g_{11}^{-1}(G'G)_{ii})^2\right).$$

The proof is given in the Appendix. The limits on the right sides of (3.6) and (3.8) 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 $\hat{\lambda}_{MLE}$. When $\epsilon_i \sim_{iid} \mathcal{N}(0,\sigma^2)$, we have $\kappa_4 = 0$ and then, from Lee (2004),

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda}_{MLE}-\lambda_0) \underset{d}{\to} N(0, V_{MLE}), \tag{3.9}$$

where

$$V_{MLE} = \lim_{n\to\infty}\left(g_{20}+g_{11}-\frac{2}{h}g_{10}^2\right)^{-1}. \tag{3.10}$$

For $\lambda_0 = 0$, a case that is especially relevant in testing, $\mathrm{tr}(G) = 0$ and $\omega^* = V_{MLE}$. Instead, from Robinson and Rossi (2014b), when $\lambda_0 = 0$

$$\left(\frac{n}{h}\right)^{1/2}\hat{\lambda} \underset{d}{\to} N(0, V_{OLS}), \tag{3.11}$$

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 be inconsistent. 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.

A result similar to Theorem 1 holds for the SAR model with unknown intercept μ_0

$$y = \mu_0 l + \lambda_0 W y + \epsilon, \tag{3.12}$$

where l in an n-vector of ones and W is row normalized, so that $Wl = l$. The OLS estimator of λ_0 in (3.12) is

$$\tilde{\lambda} = \frac{y'W'Py}{y'W'PWy}, \tag{3.13}$$

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 (3.12) is

$$y = S^{-1}(\lambda_0)(\mu_0 l + \epsilon) = \frac{\mu_0}{1-\lambda_0} l + S^{-1}(\lambda_0)\epsilon. \tag{3.14}$$

Thus, by standard algebra and observing that $l'Gl/n = O(1)$ under Assumptions 3 and 4, we conclude that (2.10) holds with $\hat{\lambda}$ replaced by $\tilde{\lambda}$ and the formal expansion for b_n in (2.17) is still appropriate so that we can define the II estimator of λ_0 in (3.12) 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 (3.12) 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 present paper focuses on using II to convert an inconsistent OLS estimator into a consistent estimator, we do not further pursue the case of model (3.12) with non-row normalized W.

Theorem 1 is 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*

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

where C 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 $\delta > 0$

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

If either $1/h + c/h \to 0$ or $h = O(1)$ and $c = O(1)$ as $n \to \infty$ the probability limit in (2.10), the formal expansion for $b_n(\lambda)$ in (2.17) and the asymptotic distribution in Theorem 1 still holds. The case of general heteroskedasticity may also be considered and is under investigation in other work.

4 Simulations

Simulations were conducted to assess the finite sample performance of $\hat{\lambda}_{II}$ in relation to $\hat{\lambda}$ and $\hat{\lambda}_{MLE}$. Three weight matrix specifications were used: a circulant matrix, an asymmetric Toeplitz matrix, and an ‘empirical-based’ matrix. Bias and mean square error (MSE) were computed for values of $\lambda \in \{-0.5, 0, 0.5, 0.8\}$ using 10^4 replications.

Case (i): Circulant weights

We take the case of a weight matrix W with a circulant structure similar to the one used by Kelejian and Prucha (1999) defined as

$$W_C = \frac{1}{||A_C||} A_C, \tag{4.1}$$

where A_C is a circulant matrix with leading row $(0,1,1,0,....,0,1,1)$, i.e.

$$A_C = \begin{pmatrix} 0 & 1 & 1 & 0 & \ldots & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}. \tag{4.2}$$

In (4.1) W_C 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 $n \to \infty$. The disturbances $\epsilon_i \sim_{iid} N(0,1)$ and sample sizes are $n \in \{30, 50, 100, 200\}$.

We implement indirect inference using the approximate binding function $b_n^*(.)$ in (2.20) to obtain $\hat{\lambda}_{II}$. Simulation results suggest that $b_n^*(.)$ closely approximates the true value $\mathbb{E}(\hat{\lambda})$, which can be computed for some simple choices of W. Figure 1 graphs the binding function and shows that $b_n^*(.)$ is invertible for $-1 < \lambda < 0.85$ but becomes flat as λ approaches unity and $b_n^*(.)$ does not vary with n.

[Figure 1 about here]

Table 1 gives the bias and MSE of the OLS, ML and II estimators of λ. The entries in the top panel reveal that the OLS estimator $\hat{\lambda}$ suffers from substantial bias for all values of λ_0. Consistent with asymptotic theory (Lee (2002)), the bias does not vanish as n increases. In fact, for a given $\lambda \neq 0$, the bias seems to increase with n and becomes particularly severe when λ_0 is negative. The entries in the last two panels of Table 1 indicate that $\hat{\lambda}_{II}$ outperforms $\hat{\lambda}_{MLE}$ in terms of bias reduction in many cases, but at the cost of a slight increase in the variance (and hence MSE). While the MSE increase of $\hat{\lambda}_{II}$ is often negligible, it becomes stronger when λ is close to unity as expected from the shape of the binding function $b_n^*(.)$ which becomes flat as λ approaches unity.

To shed light on their distributional characteristics, Figure 2 plots the simulated density functions of $\hat{\lambda}$, $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ for $n = 100$ when $\lambda_0 = 0.5$. The distribution of the OLS estimator $\hat{\lambda}$ is seen to be severely upward biased (centred around 0.85 rather than 0.5), whereas both $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ appear almost unbiased. All three estimators seem to have similar dispersion.

[Figure 2 about here]

Direct analytic comparison of the variances is difficult since (3.8) and (3.10) are complicated non-linear functions of the weight matrix. Figure 3 shows how the finite sample variances (ω^*, V_{MLE}) of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{MLE}$ vary with λ_0 for $n = 100$. The variances are close for small-moderate spatial autocorrelation, but as $|\lambda_0|$ increases ω^* becomes larger than V_{MLE} and increases rapidly as λ_0 tends to unity. The rise in variance is associated with the non-invertibility of the binding function $b_n(\lambda)^*$ as λ approaches the boundaries of the support.

[Figure 3 about here]

Case (ii): Asymmetric Toeplitz weights

We next consider the case of an asymmetric Toeplitz weight matrix W_{AT}. Working from the circulant matrix A_C, we introduce asymmetry by removing the neighbourhood effect of the $(n-1)$'th unit on the first unit in (4.1). This produces a three element neighbourhood effect in each row rather than four. Specifically, we define

$$W_{AT} = \frac{1}{\|A_{AT}\|} A_{AT}$$

where

$$A_{AT} = \begin{pmatrix} 0 & 1 & 1 & 0 & \dots & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \tag{4.3}$$

The weight matrix is again normalised so that Assumption 3 is satisfied. Figure 4 depicts the approximate binding function for $n = 100$, showing that $b^*(.)$ is monotonic. For $\lambda > 0.8$ the binding function tends to flatten out although not as markedly as in the symmetric case.

[Figure 4 about here]

The simulation results reported in Table 2 confirm that both the ML and the II estimators provide substantial reductions in both the bias and MSE of OLS. For most configurations, ML and II display similar performance. The II estimator generally outperforms ML in terms of bias reduction when $\lambda > 0$, without increasing MSE by much, and for $n = 200$ largely reproduces the performance characteristics of ML. Similar conclusions follow from the proximity of the empirical densities of II and ML shown for $n = 100$ and $\lambda = 0.5$ in Figure 5 .

[Figure 5 about here]

Figure 6 plots values of ω^* and V_{MLE} under (4.3) over $\lambda_0 \in (-1, 1)$, showing the variances are close for most admissible values of λ, with discrepancies emerging as $|\lambda_0|$ increases, but again not as severe as in the circulant weight matrix case. These results indicate that indirect inference delivers broader performance gains for asymmetric weight matrix structures.

[Figure 6 about here]

Case (iii): Empirical-based weights

The final simulation exercise uses an 'empirical-based' weight matrix W to illustrate how indirect inference performs in a setting that is relevant to practical work. We consider a sample of 43 European countries and construct W according to a contiguity criterion – see, for example, Chapter 2 of Arbia (2006) for various definitions of spatial contiguity that are used in empirical work. Countries i and j are said to be neighbours if they share a border, which leads to the specification

$$W_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ share a border (where } i \neq j) \\ 0 & \text{otherwise} \end{cases}. \tag{4.4}$$

As usual, $W_{ii} = 0$. The resulting matrix is then re-scaled by its spectral norm, so that Assumption 3(ii) is satisfied. Figure 7 shows the binding function $b_n^*(.)$ in this case, which is monotonic over $\lambda \in (-1, 1)$, so the II estimator appears well-defined for all admissible values of λ.

[Figure 7 about here]

Table 3 summarises the results for the bias and MSE of $\hat{\lambda}$, $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ for different values of λ_0. Again, $\hat{\lambda}$ is severely biased, while both $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ provide major improvements. More specifically, $\hat{\lambda}_{II}$ outperforms $\hat{\lambda}_{MLE}$ in terms of bias reduction at $\lambda_0 = 0.5, 0.8$, with only a slight increase in its MSE. Figure 8 plots the simulated densities of $\hat{\lambda}$, $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ for $\lambda_0 = 0.5$. These graphs reveal that the finite sample densities of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{MLE}$ are almost identical and are well centred at the true parameter value, whereas the OLS density appears mislocated with a larger spread. Overall, these results suggest that $\hat{\lambda}_{II}$ performs well even when W has a less-restrictive and more practical structure than that of the formal structures in (4.1) or (4.3).

[Figure 8 about here]

Finally, in Figure 9 we report a plot of the finite sample versions of ω^* and V_{MLE} in (3.8) and (3.10), respectively. In line with Figures 3 and 6, the finite sample versions of ω^* and V_{MLE} appear to be very close for small/moderate values of $|\lambda|$. As $|\lambda|$ approaches unity, ω^* tends to increase, but not as much as in case of the circulant W. This behaviour is, therefore, consistent with the plot of $b_n^*(.)$ in Figure 7.

[Figure 9 about here]

These simulations provide information about the finite sample performance of indirect inference under several different specifications of the weight matrix. The results collectively suggest that the II estimator substantially reduces the bias and MSE of the OLS estimator and can outperform the ML estimator. While the results in Tables 1-3 were obtained under normally distributed errors, we have verified that the reported performance of the II estimator is robust to nonnormal errors, specifically under mixed-normal distributions and a t distribution with 5 degrees of freedom. Those results are available on request.

5 Examples

In this section we consider a few examples for which we may assess analytically whether the binding function $b_n(\lambda)$ in (2.17) is invertible, at least as $n \to \infty$, rather than relying on numerical work, as in the plots of Figures 1, 4 and 7. Occasionally, an analytic comparison between the performance of $\hat{\lambda}_{II}$ and $\hat{\lambda}_{MLE}$ is also possible.

Example (i): The Districts Model

The simplest choice of W that is amenable to analysis and facilitates a comparison between (3.8) and (3.10) is the block diagonal 'districts model' weight matrix W (Case (1991)) which is defined as

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

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.

[Figure 10 about here]

The approximate binding function $b_n^*(.)$ in (2.20) appears invertible for $\lambda \in (-1, 1)$ and for all sample sizes, as shown in Figure 10. We derive the following.

Theorem 2 *Let W defined as in (5.1).*

(a) *As $n \to \infty$ the binding function b_n in (2.17) is strictly increasing for all $\lambda \in \Lambda$.*

(b) *If $1/m + 1/r \to 0$, $\omega^* = V_{MLE}$, where V_{MLE} is defined in (3.10).*

The proof of Theorem 2 is in the Appendix. The condition in part (b) of Theorem 2 corresponds to a case of divergent h and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Example (ii): Circulant Weight Matrix Model

As another example we can consider the simple circulant matrix C with leading row $(0, 1, 0, ...0, 1)$, i.e.

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{5.2}$$

and

$$W = \frac{1}{2} C, \tag{5.3}$$

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

[Figure 11 about here]

From Figure 11, the approximate binding function $b_n^*(\lambda)$ in (2.20) seems to be strictly monotonic for $\lambda \in (-0.7, 0.7)$ but becomes almost flat (and even decreases slightly) as $\lambda \to 1$, with related behavior as $\lambda \to -1$. Similar behavior was found in simulations for the case where W was chosen as in (4.1). We have the following analytic result.

Theorem 3 *Define W as in (5.3). As $n \to \infty$, $b_n(\lambda)$ in (2.17) is strictly increasing for all $\lambda \in \Lambda$, where* Λ is *any closed subset of* $(-\sqrt{3}/2, \sqrt{3}/2)$.

The proof of Theorem 3 is in the Appendix. 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 (3.8) and the results reported in Appendix (viz., (A.49), (A.53) and (A.57)) we also conclude that $\omega^* \to \infty$ as $\lambda_0 \to \pm\frac{\sqrt{3}}{2}$, since

$$1 - \frac{2g_{10}g_{21}}{g_{11}(g_{20} + g_{11})} \to 0 \quad \text{as} \quad \lambda \to \pm\frac{\sqrt{3}}{2}. \tag{5.4}$$

This result, even though it is derived under the simpler circulant weight matrix (5.3), is consistent with the Monte Carlo results based on the weight matrix W defined in (4.1). 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 and performs well as an estimator for small through moderate values of λ_0.

6 Conclusions

Our main result shows how indirect inference methodology can be used in pure spatial autoregession 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 MLE 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. The results of the present paper, although novel for spatial regression, 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.

The present approach complements earlier work on analytic bias corrections of ML or PML estimators (Bao and Ullah (2007); Bao (2013)) and offers an alternative mechanism of improving finite sample performance. While our focus has been on OLS, the II methodology can equally well be applied to other estimators, like the MLE, which are consistent but suffer from finite sample bias. The methodology can also be extended to more complex settings, due to 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 PML fail to be consistent when the disturbances are heterogeneously distributed. Extensions of the indirect inference methodology to SAR models with unknown heteroskedasticity seems promising and is currently under investigation.

Appendix

Proof of Theorem 1

The proof of part (a) is carried out in a similar way to Robinson (2008). Let ψ_{ij} be the vector $\psi_{ij} = \begin{pmatrix} \psi_{1ij} & \psi_{2ij} \end{pmatrix}' = \begin{pmatrix} (G+G')_{ij}/2 & (G'G)_{ij} \end{pmatrix}'$, and define

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

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

$$A = Var\left(\sum_{i=1}^n u_i\right) = (\mu^{(4)} - \sigma^4)\sum_{i=1}^n \psi_{ii}\psi_{ii}' + 4\sigma^4 \sum_{i=1}^n \sum_{j<i} \psi_{ij}\psi_{ij}', \tag{A.2}$$

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

$$\sum_{i=1}^{n} \mathbb{E}(z_{in}^2 | \epsilon_j; j < i) \xrightarrow{p} 1 \tag{A.3}$$

and

$$\sum_{i=1}^{n} \mathbb{E}(z_{in}^2 1(|z_{in} > \xi|)) \to 0 \quad \forall \xi > 0. \tag{A.4}$$

Now, (A.3) is equivalent to

$$\sum_{i=1}^{n} \mathbb{E}(z_{in}^2 | \epsilon_j; j < i) - \eta' A^{-1/2} A A^{-1/2} \eta \xrightarrow{p} 0, \tag{A.5}$$

which is

$$\begin{aligned} &\eta' A^{-1/2} \left(4\sigma^2 \sum_{i=1}^{n} \left(\sum_{j<i} \psi_{ij} \epsilon_j \right) \left(\sum_{j<i} \psi_{ij} \epsilon_j \right)' - 4\sigma^4 \sum_{i=1}^{n} \sum_{j<i} \psi_{ij} \psi_{ij}' \right) A^{-1/2} \eta \\ &+ 4\eta' A^{-1/2} \mu^{(3)} \sum_{i=1}^{n} \psi_{ii} \left(\sum_{j<i} \psi_{ij} \epsilon_j \right)' A^{-1/2} \eta \xrightarrow{p} 0, \end{aligned} \tag{A.6}$$

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

$$\begin{aligned} V = &\lim_{n\to\infty} \left(\begin{pmatrix} \sigma^4(g_{20} + g_{11}) & 2\sigma^4 g_{21} \\ 2\sigma^4 g_{21} & 2\sigma^4 g \end{pmatrix} + \begin{pmatrix} \frac{h}{n}\kappa_4 \sum_i G_{ii}^2 & \frac{h}{n}\kappa_4 \sum_i G_{ii}(G'G)_{ii} \\ \frac{h}{n}\kappa_4 \sum_i G_{ii}(G'G)_{ii} & \frac{h}{n}\kappa_4 \sum_i (G'G)_{ii}^2 \end{pmatrix} \right) \\ =&\Sigma + \Omega. \end{aligned} \tag{A.7}$$

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

$$\left(\frac{h}{n}\right)^2 (\mathrm{tr}((G + G')G'G))^2 < \left(\frac{h}{n}\right)^2 \mathrm{tr}((G + G')^2)\mathrm{tr}((G'G)^2). \tag{A.8}$$

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. $\Omega = 0$ when $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Rather than (A.6), we can equivalently show

$$\frac{h}{n}\left(4\sigma^2\sum_{i=1}^{n}\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)' - 4\sigma^4\sum_{i=1}^{n}\sum_{j<i}\psi_{ij}\psi_{ij}'\right) \xrightarrow{p} 0, \tag{A.9}$$

and

$$\frac{h}{n}\mu^{(3)}\sum_{i=1}^{n}\psi_{ii}\left(\sum_{j<i}\psi_{ij}\epsilon_j\right)' \xrightarrow{p} 0. \tag{A.10}$$

Consider the following typical elements of the left side of (A.9)

$$4\sigma^2\frac{h}{n}\left(\sum_{i=1}^{n}\sum_{j<i}\psi_{sij}^2(\epsilon_j^2 - \sigma^2) + \sum_{i=1}^{n}\sum_{\substack{j,k<i\\ j\neq k}}\psi_{sij}\psi_{sik}\epsilon_j\epsilon_k\right) \quad s = 1, 2, \tag{A.11}$$

and

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

The first term in (A.11) has mean zero and variance bounded by

$$\begin{aligned}&\left(\frac{h}{n}\right)^2 K\sum_{i}\sum_{k}\sum_{j<i,k}\psi_{sij}^2\psi_{skj}^2 \leq \left(\frac{h}{n}\right)^2 K\sum_{i}\sum_{k}\sum_{j}\psi_{sij}^2\psi_{skj}^2\\ &\leq \left(\frac{h}{n}\right)^2 K\left(\max_{j}\sum_{i}\psi_{sij}^2\right)\sum_{k,j}\psi_{skj}^2 = O\left(\frac{h}{n}\right).\end{aligned} \tag{A.13}$$

The last equality in (A.13) 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,

$$\sum_i \psi_{sij}^2 = e_j' \Psi_s^2 e_j \leq ||\Psi_s||^2 \leq K, \tag{A.14}$$

where the last inequality follows from Assumption 3(ii) and (2.5) after observing that Ψ_s equals either $(G+G')/2$ or $G'G$ for $s=1$ and $s=2$, respectively. The second term of (A.11) has mean zero and variance bounded by

$$\begin{aligned}
&\left(\frac{h}{n}\right)^2 K |\sum_i \sum_h \sum_{j<i,h} \sum_{k<i,h} \psi_{sij}\psi_{sik}\psi_{shj}\psi_{shk}| \\
&\leq \left(\frac{h}{n}\right)^2 K \left(\sum_i \sum_h \sum_j \sum_k |\psi_{sij}\psi_{sik}\psi_{shj}\psi_{shk}|\right) \leq \left(\frac{h}{n}\right)^2 K \sum_i \sum_h \sum_j \sum_k |\psi_{sij}\psi_{sik}|(\psi_{shj}^2 + \psi_{shk}^2) \\
&\leq \left(\frac{h}{n}\right)^2 K \left(\left(\max_j \sum_i |\psi_{sij}|\right)\left(\max_i \sum_k |\psi_{sik}|\right) \sum_{h,j} \psi_{shj}^2 + \left(\max_i \sum_j |\psi_{sij}|\right)\left(\max_k \sum_i |\psi_{sik}|\right) \sum_{h,k} \psi_{shk}^2\right) \\
&= O\left(\frac{h}{n}\right),
\end{aligned} \tag{A.15}$$

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 (A.12) have mean zero and variance bounded by

$$\begin{aligned}
&\left(\frac{h}{n}\right)^2 K \sum_i \sum_k \sum_{j<i,k} |\psi_{sij}\psi_{tij}\psi_{skj}\psi_{tkj}| \leq \left(\frac{h}{n}\right)^2 K \sum_i \sum_k \sum_j |\psi_{sij}\psi_{tij}|(\psi_{skj}^2 + \psi_{tkj}^2) \\
&\leq \left(\frac{h}{n}\right)^2 K \max_j \sum_i |\psi_{sij}| \max_i \sum_j |\psi_{tij}| \max_j \sum_k (\psi_{skj}^2 + \psi_{tkj}^2) = o(1),
\end{aligned} \tag{A.16}$$

and

$$\left(\frac{h}{n}\right)^2 K|\sum_i\sum_h\sum_{j<i,h}\sum_{k<i,h}\psi_{sij}\psi_{tik}\psi_{shj}\psi_{thk}|$$
$$\leq \left(\frac{h}{n}\right)^2 K\left(\sum_i\sum_h\sum_j\sum_k|\psi_{sij}\psi_{tik}\psi_{shj}\psi_{thk}|\right) \leq \left(\frac{h}{n}\right)^2 K\sum_i\sum_h\sum_j\sum_k|\psi_{sij}\psi_{tik}|(\psi^2_{shj}+\psi^2_{thk})$$
$$\leq \left(\frac{h}{n}\right)^2 K\left(\left(\max_j\sum_i|\psi_{sij}|\right)\left(\max_i\sum_k|\psi_{tik}|\right)\sum_{h,j}\psi^2_{shj}+\left(\max_i\sum_j|\psi_{sij}|\right)\left(\max_k\sum_i|\psi_{tik}|\right)\sum_{h,k}\psi^2_{thk}\right)$$
$$= o\left(1\right), \tag{A.17}$$

The typical element on the LHS of (A.10) is

$$\frac{h}{n}\mu^{(3)}\sum_i\psi_{sii}\sum_{j<i}\psi_{tij}\epsilon_j, \qquad s,t=1,2, \tag{A.18}$$

and has mean zero and variance bounded by

$$K\left(\frac{h}{n}\right)^2\sum_i\sum_k\sum_{j<i,k}|\psi_{sii}\psi_{skk}\psi_{tij}\psi_{tkj}| \leq K\left(\frac{h}{n}\right)^2\sum_i\sum_k\sum_j|\psi_{tij}||\psi_{tkj}|(\psi^2_{sii}+\psi^2_{skk})$$
$$\leq K\left(\frac{h}{n}\right)^2\left(\max_i\sum_j|\psi_{tij}|\max_j\sum_k|\psi_{tkj}|\sum_i\psi^2_{sii}+\max_j\sum_i|\psi_{tij}|\max_k\sum_j|\psi_{tkj}|\sum_k\psi^2_{skk}\right) = o(1) \tag{A.19}$$

under Assumptions 3(iii) and 4 and since

$$\sum_i\psi^2_{sii} \leq \sum_{i,j}\psi^2_{sij} = O\left(\frac{n}{h}\right). \tag{A.20}$$

We prove (A.4) by verifying the sufficient Lyapunov condition

$$\sum_{i=1}^n\mathbb{E}|z_{in}|^{2+\delta} \to 0, \tag{A.21}$$

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,

$$\left(\frac{h}{n}\right)^{1+\delta/2} \sum_i \mathbb{E}|u_{si}|^{2+\delta} \leq \left(\frac{h}{n}\right)^{1+\delta/2} K \sum_i |\psi_{sii}|^{2+\delta} + \left(\frac{h}{n}\right)^{1+\delta/2} K \sum_i \mathbb{E}|\sum_{j<i} \psi_{sij}\epsilon_j|^{2+\delta}. \tag{A.22}$$

The first term in the latter expression is

$$\left(\frac{h}{n}\right)^{1+\delta/2} K \left(\max_i \ |\psi_{sii}|^{\delta}\right) \sum_i \psi_{sii}^2 = o(1), \tag{A.23}$$

by (A.20) and since for all i

$$|\psi_{sii}| \leq ||\Psi_s||_\infty \leq K \tag{A.24}$$

under Assumptions 3(iii) and 4. The second term in (A.22) by the Burkholder and von Bahr/Esseen inequalities is bounded by

$$\begin{aligned}
&\left(\frac{h}{n}\right)^{1+\delta/2} K \sum_i \mathbb{E}|\sum_{j<i} \psi_{sij}^2 \epsilon_j^2|^{1+\delta/2} \\
&\leq \left(\frac{h}{n}\right)^{1+\delta/2} K \sum_i \sum_{j<i} |\psi_{sij}|^{2+\delta} \leq \left(\frac{h}{n}\right)^{1+\delta/2} K \sum_i \left(\sum_{j<i} \psi_{sij}^2\right)^{1+\delta/2} \\
&\leq K \left(\frac{h}{n}\right)^{1+\delta/2} \left(\max_i \sum_j \psi_{sij}^2\right)^{\delta/2} \sum_i \sum_j \psi_{sij}^2 \\
&\leq K \left(\frac{h}{n}\right)^{\delta/2} \left(\max_i \sum_j \psi_{sij}^2\right)^{\delta/2},
\end{aligned} \tag{A.25}$$

which is $O((h/n)^{\delta/2})$ by (A.14).

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

$$\left(\frac{h}{n}\right)^{1/2}\sum_i u_i \underset{d}{\rightarrow} \mathcal{N}(0,V), \tag{A.26}$$

where V is defined in (A.7). (3.5) follows trivially since

$$f_n'\left(\frac{h}{n}\right)^{1/2}U_n = \bar{f}'\left(\frac{h}{n}\right)^{1/2}U_n + o_p(1), \tag{A.27}$$

where

$$\bar{f} = \lim_{n\to\infty}\left(g_{11}^{-1}\sigma_0^{-2} \quad - g_{11}^{-2}g_{10}\sigma_0^{-2}\right)', \tag{A.28}$$

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

In order 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

$$b_n^{(1)}(x) = 1 + \frac{\mathrm{tr}(G(x)^2)\mathrm{tr}(G'(x)G(x)) - 2\mathrm{tr}G(x)\mathrm{tr}(G'(x)G(x)^2)}{(\mathrm{tr}(G'(x)G(x)))^2} + O\left(\frac{h}{n}\right), \tag{A.29}$$

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

$$b_n^{-1(1)}(x)|_{x=b_n(\lambda_0)} = (b_n^{(1)}(q))^{-1}|_{q=b_n^{-1}(b_n(\lambda_0))=\lambda_0}. \tag{A.30}$$

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

$$\begin{aligned} b_n^{-1}(\bar{\lambda}_n) &= b_n^{-1}\left(b_n(\lambda_0) + O\left(\frac{h}{n}\right)\right) \\ &= b_n^{-1}(b_n(\lambda_0)) + (b_n^{(1)}(x))^{-1}|_{x=\lambda_0}O\left(\frac{h}{n}\right) + ... = \lambda_0 + O\left(\frac{h}{n}\right) \end{aligned} \tag{A.31}$$

and thus

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda}_{II} - \lambda_0) = \left(\frac{n}{h}\right)^{1/2}(b_n^{-1}(\hat{\lambda}) - b_n^{-1}(\bar{\lambda}_n)) + o(1). \tag{A.32}$$

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

$$\left|\frac{b_n^{(1)}(\lambda_0) - b_n^{(1)}(r)}{b_n^{(1)}(r)}\right| \to 0 \tag{A.33}$$

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

$$\begin{aligned}\left|\frac{b_n^{(1)}(\lambda_0) - b_n^{(1)}(r)}{b_n^{(1)}(r)}\right| &\leq K|b_n^{(1)}(\lambda_0) - b_n^{(1)}(r)| \\ &\leq K\left(\left|\frac{g_{20}}{g_{11}} - \frac{h\mathrm{tr}(G(r)^2)/n}{h\mathrm{tr}(G'(r)G(r))/n}\right| + \left|\frac{g_{10}g_{21}}{g_{11}^2} - \frac{h^2\mathrm{tr}(G(r))\mathrm{tr}(G(r)^2G'(r))/n^2}{(h\mathrm{tr}(G'(r)G(r))/n)^2}\right|\right).\end{aligned} \tag{A.34}$$

The first term of the latter expression is bounded by

$$\begin{aligned}&K\left(\left|g_{20} - \frac{h}{n}\mathrm{tr}(G(r)^2)\right| + \left|g_{11} - \frac{h}{n}\mathrm{tr}(G(r)'G(r))\right|\right) \\ &= K\left(\left|\frac{h}{n}\mathrm{tr}(G(\lambda^*)^2)(\lambda_0 - r)\right| + \left|\frac{h}{n}\mathrm{tr}(G(\lambda^*)'G(\lambda^*))(\lambda_0 - r)\right|\right) \\ &\leq K|\lambda_0 - r| \leq s^{-1}\delta\end{aligned} \tag{A.35}$$

as $n \to \infty$, 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.34) can be dealt with in a similar fashion.

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

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda}_{II} - \lambda_0) \to_d N(0, \omega^*), \tag{A.36}$$

where

$$\omega^* = \lim_{n\to\infty} (g_{11}+g_{20})^{-1}\left(1-\frac{2g_{10}g_{21}}{g_{11}(g_{20}+g_{11})}\right)^{-2}\left(1-\frac{4g_{21}g_{10}}{g_{11}(g_{11}+g_{20})}+\frac{2gg_{10}^2}{g_{11}^2(g_{11}+g_{20})}\right.$$
$$\left.+\frac{h}{n}\frac{\kappa_4}{\sigma_0^4(g_{11}+g_{20})}\sum_{i=1}^{n}(G_{ii}-g_{10}g_{11}^{-1}(G'G)_{ii})^2\right). \tag{A.37}$$

Proof of Theorem 2

We have

$$\mathrm{tr}(G) = \mathrm{tr}\left(\sum_{i=0}^{\infty}\lambda_0^i W^{i+1}\right) = r\sum_{i=0}^{\infty}\lambda_0^i \mathrm{tr}(B_m^{i+1}), \tag{A.38}$$

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

$$\mathrm{tr}(B_m^{i+1}) = 1+(m-1)\left(\frac{-1}{m-1}\right)^{i+1}. \tag{A.39}$$

Thus

$$\frac{h}{n}\mathrm{tr}(G) = \left(\frac{m-1}{mr}\right)\left(r\sum_{i=0}^{\infty}\lambda_0^i\left(1-\left(\frac{-1}{m-1}\right)^i\right)\right) = \left(\frac{m-1}{mr}\right)\left(\frac{r}{1-\lambda_0}-\frac{r}{1+\frac{\lambda_0}{m-1}}\right)$$
$$=\frac{\lambda_0}{1-\lambda_0}\frac{(m-1)}{m-1+\lambda_0} \tag{A.40}$$

and, for $s \geq 2$,

$$\frac{h}{n}\mathrm{tr}(G^s) = \frac{m-1}{m}\frac{1}{(1-\lambda_0)^s}+(-1)^s\frac{(m-1)^2}{m(m-1+\lambda_0)^s}. \tag{A.41}$$

To show part (a), from (2.17)

$$\frac{db_n(\lambda)}{d\lambda} = 2-\frac{2h^2\mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3)/n^2}{h^2(\mathrm{tr}(G(\lambda)^2))^2/n^2}+O\left(\frac{h}{n}\right). \tag{A.42}$$

As $n \to \infty$, the sign of the right hand side (RHS) of (A.42) depends on

$h^2((\mathrm{tr}(G(\lambda)^2))^2 - \mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3))/n^2$. The condition $h/n \to 0$ as $n \to \infty$ is satisfied when $r \to \infty$, whether $m \to \infty$ or $m = O(1)$ as $n \to \infty$.

When $m = O(1)$ as $n \to \infty$, collecting (A.42), (A.40), (A.41) and by some straightforward algebra,

$$\left(\frac{h}{n}\right)^2 ((\mathrm{tr}(G(\lambda)^2))^2 - \mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3))$$
$$= \frac{(m-1)^2}{m}\left(\frac{(m-1)(1-\lambda)}{(1-\lambda)^4(m-1+\lambda)m} + \frac{2(m-1)}{m(1-\lambda)^2(m-1+\lambda)^2} + \frac{(m-1)}{m(m-1+\lambda)^3(1-\lambda)}\right)$$
$$= \frac{(m-1)^3}{m^2(m-1+\lambda)(1-\lambda)}\left(\frac{1}{1-\lambda} + \frac{1}{m-1+\lambda}\right)^2, \qquad \text{(A.43)}$$

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

$$\left(\frac{h}{n}\right)^2 ((\mathrm{tr}(G(\lambda)^2))^2 - \mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3)) \to \frac{1}{(1-\lambda)^2}, \qquad \text{(A.44)}$$

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 $db_n(\lambda)/d\lambda \to 0$, consistently with Figure 11.

To show part (b) we notice that as $m \to \infty$ and $r \to \infty$,

$$\lim_{n\to\infty} \frac{h}{n}\mathrm{tr}(G) = \frac{\lambda_0}{(1-\lambda_0)} \quad \lim_{n\to\infty} \frac{h}{n}\mathrm{tr}(G^s) = \frac{1}{(1-\lambda_0)^s} \qquad \text{(A.45)}$$

Hence, from (3.8), (A.45) and standard algebra

$$\lim_{n\to\infty}\left(\frac{2\mathrm{tr}G\mathrm{tr}(G^2G')}{\mathrm{tr}(G'G)\mathrm{tr}(G^2+G'G)}\right)^2 = \lim_{n\to\infty}\frac{2\mathrm{tr}((G'G)^2)(\mathrm{tr}G)^2}{(\mathrm{tr}(G'G))^2\mathrm{tr}(G^2+G'G)} = \lambda_0^2, \quad \text{(A.46)}$$

so that

$$\omega^* = V_{MLE} = \frac{(1-\lambda_0)^2}{2}. \qquad \text{(A.47)}$$

Proof of Theorem 3

As $n \to \infty$,

$$\frac{1}{n}\text{tr}(G(\lambda)) = \frac{1}{n}\sum_{s=0}^{\infty}\lambda^s \text{tr}(W^{s+1}) \to \sum_{s=0}^{\infty}\lambda^s \frac{1}{2^{s+1}}\frac{1}{2\pi}\int_0^{2\pi}(2cosx)^{s+1}dx$$
$$= \frac{1}{\lambda}\sum_{s=1}^{\infty}\lambda^s \frac{1}{2\pi}\int_0^{2\pi}(cosx)^s dx. \qquad \text{(A.48)}$$

Since $\int_0^{2\pi}(cosx)^s dx = 0$ for odd s, the last expression in (A.48) can be written as

$$\frac{1}{\lambda}\sum_{p=1}^{\infty}\lambda^{2p}\frac{1}{2\pi}\int_0^{2\pi}(cosx)^{2p}dx = \frac{1}{\lambda}\sum_{p=1}^{\infty}\lambda^{2p}\frac{(2p-1)!!}{(2p)!!} = \frac{1}{\lambda}\sum_{p=1}^{\infty}\lambda^{2p}\frac{(2p)!}{2^{2p}(p!)^2}$$
$$= \frac{1}{\lambda}\sum_{p=0}^{\infty}\left(\frac{\lambda^2}{4}\right)^p \binom{2p}{p} - \frac{1}{\lambda} = \frac{1}{\lambda}((1-\lambda^2)^{-1/2} - 1). \qquad \text{(A.49)}$$

Similarly,

$$\frac{1}{n}\text{tr}(G(\lambda)^2) = \frac{1}{n}\sum_{s,t=0}^{\infty}\lambda^{s+t}\text{tr}(W^{s+t+2}) \to \frac{1}{2\pi}\sum_{s,t=0}^{\infty}\lambda^{s+t}\int_0^{2\pi}(cosx)^{s+t+2}dx. \qquad \text{(A.50)}$$

Since $\int_0^{2\pi}(cosx)^{s+t+2}dx \neq 0$ only when $s+t+2 = 2p$, for $s,t = 0,....\infty$ and $p = 1,....\infty$,

$$\frac{1}{2\pi}\sum_{s,t=0}^{\infty}\lambda^{s+t}\int_0^{2\pi}(cosx)^{s+t+2}dx = \frac{1}{\lambda^2}\sum_{p=1}^{\infty}\lambda^{2p}(2p-1)\frac{(2p-1)!!}{(2p)!!}, \qquad \text{(A.51)}$$

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

$$\sum_{p=1}^{\infty}px^p\binom{2p}{p} = 2x(1-4x)^{-3/2} \qquad |x| < \frac{1}{4}, \qquad \text{(A.52)}$$

$$\frac{1}{n}\mathrm{tr}(G(\lambda)^2) \to \frac{1}{\lambda^2}\left(2\sum_{p=1}^{\infty} p\left(\frac{\lambda^2}{4}\right)^p \binom{2p}{p} - \sum_{p=1}^{\infty}\left(\frac{\lambda^2}{4}\right)^p \binom{2p}{p}\right)$$
$$= \frac{1}{\lambda^2}\left(\lambda^2(1-\lambda^2)^{-3/2} - ((1-\lambda^2)^{-1/2} - 1)\right). \tag{A.53}$$

Along the same lines,

$$\frac{1}{n}\mathrm{tr}(G(\lambda)^3) = \frac{1}{n}\sum_{s,t,q=0}^{\infty} \lambda^{s+t+q}\mathrm{tr}(W^{s+t+q+3}) \to \frac{1}{2\pi}\sum_{s,t,q=0}^{\infty}\lambda^{s+t+q}\int_0^{2\pi}(cosx)^{s+t+q+3}dx. \tag{A.54}$$

Again, $\int_0^{2\pi}(cosx)^{s+t+q+3}dx \neq 0$ only when $s+t+q+3 = 2p$, $s,t,q = 0,....,\infty$ and $p = 2,....,\infty$. Thus,

$$\frac{1}{2\pi}\sum_{s,t,q=0}^{\infty}\lambda^{s+t+q}\int_0^{2\pi}(cosx)^{s+t+q+3}dx = \frac{1}{2\pi\lambda^3}\sum_{p=2}^{\infty}\lambda^{2p}(p-1)(2p-1)\int_0^{2\pi}(cosx)^{2p}dx, \tag{A.55}$$

where the factor $(p-1)(2p-1)$ takes into account the number of combinations of s,t,q s.t. $s+t+q = 2p-3$, for $s,t,q = 0,....,\infty$ and $p = 1,....,\infty$. By

$$\sum_{p=1}^{\infty} p^2 x^p \binom{2p}{p} = \frac{2x(2x+1)}{(1-4x)^{5/2}} \qquad |x| < \frac{1}{4}, \tag{A.56}$$

we deduce

$$\frac{1}{n}\mathrm{tr}(G(\lambda)^3) \to \frac{1}{\lambda^3}\sum_{p=1}^{\infty}\lambda^{2p}(p-1)(2p-1)\frac{(2p-1)!!}{(2p)!!}$$
$$= \frac{1}{\lambda^3}\left(2\sum_{p=1}^{\infty}p^2\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p} - 3\sum_{p=1}^{\infty}p\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p} + \sum_{p=1}^{\infty}\left(\frac{\lambda^2}{4}\right)^p\binom{2p}{p}\right)$$
$$= \frac{1}{\lambda^3}\left(\lambda^2\left(\frac{\lambda^2}{2}+1\right)(1-\lambda^2)^{-5/2} - \frac{3}{2}\lambda^2(1-\lambda^2)^{-3/2} + (1-\lambda^2)^{-1/2} - 1\right). \tag{A.57}$$

Collecting (A.49), (A.53) and (A.57), we can show that (A.42) is strictly positive for any $\lambda \in (-\sqrt{3}/2, \sqrt{3}/2)$ (and $\lambda \neq 0$) as $n \to \infty$, since

$$\frac{1}{n^2}((\mathrm{tr}(G(\lambda)^2))^2 - \mathrm{tr}(G(\lambda))\mathrm{tr}(G(\lambda)^3)) \to \frac{1}{\lambda^4}(\frac{\lambda^4}{2}(1-\lambda^2)^{-3}(1+(1-\lambda^2)^{1/2})$$
$$-\frac{\lambda^2}{2}(1-\lambda^2)^{-2}(1-(1-\lambda^2)^{1/2}) - \lambda^2(1-\lambda^2)^{-3}(1-(1-\lambda^2)^{1/2})) \tag{A.58}$$

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 (A.58) is strictly positive when

$$2z^2 - 3z + 1 < 0, \tag{A.59}$$

which is satisfied for $z \in (1/2, 1)$. Solving for λ, we obtain that the RHS of (A.58) is strictly positive for $\lambda \in (-\sqrt{3}/2, \sqrt{3}/2)$, somehow consistently with the plot in Figure 12. From (A.42), (A.49), (A.53) and (A.57) it is easy to see that as $n \to \infty$, for $\lambda \to \pm 1$, $db_n(\lambda)/d\lambda \to -1$.

n		30		50		100		200	
OLS	λ	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.5	-0.485	0.469	-0.481	0.376	-0.488	0.312	-0.500	0.276
	0.0	-0.075	0.265	-0.061	0.162	-0.035	0.082	-0.016	0.040
	0.5	0.229	0.136	0.268	0.111	0.291	0.100	0.302	0.098
	0.8	0.207	0.052	0.216	0.050	0.221	0.050	0.223	0.005
ML	λ	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.5	-0.004	0.071	-0.001	0.044	-0.001	0.022	-0.001	0.001
	0.0	-0.028	0.066	-0.024	0.040	-0.014	0.020	-0.006	0.010
	0.5	-0.052	0.041	-0.029	0.022	-0.015	0.010	-0.007	0.005
	0.8	-0.041	0.016	-0.025	0.008	-0.011	0.003	-0.006	0.001
II	λ	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.5	-0.038	0.103	-0.017	0.054	-0.009	0.025	-0.004	0.012
	0.0	-0.030	0.076	-0.025	0.043	-0.015	0.021	-0.006	0.010
	0.5	-0.020	0.058	-0.010	0.029	0.007	0.012	-0.004	0.006
	0.8	0.004	0.035	0.001	0.017	0.005	0.008	0.006	0.004

Table 1: Bias and Mean Square Error (MSE) of the OLS, MLE and II estimators at $n = 30, 50, 100, 200$ for $\lambda = 0.5, 0, 0.5, 0.8$ when W is given by (4.1) (10^4 repl. and $\epsilon \sim N(0, 1)$).

n		30		50		100		200	
OLS	λ	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.5	-0.279	0.220	-0.283	0.165	-0.288	0.126	-0.286	0.103
	0.0	-0.061	0.170	-0.038	0.100	-0.016	0.049	-0.010	0.025
	0.5	0.153	0.082	0.182	0.064	0.205	0.055	0.215	0.052
	0.8	0.128	0.026	0.142	0.024	0.152	0.024	0.156	0.025
ML	λ	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.5	-0.024	0.087	-0.015	0.049	-0.001	0.024	-0.003	0.012
	0.0	-0.037	0.068	-0.022	0.038	-0.009	0.018	-0.006	0.009
	0.5	-0.045	0.038	-0.027	0.021	-0.013	0.010	-0.007	0.005
	0.8	-0.042	0.019	-0.027	0.010	-0.013	0.004	-0.007	0.002
II	λ	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
	-0.5	-0.031	0.092	-0.019	0.051	-0.011	0.025	-0.004	0.012
	0.0	-0.037	0.071	-0.022	0.039	-0.009	0.018	-0.006	0.009
	0.5	-0.035	0.041	-0.022	0.022	-0.010	0.010	-0.005	0.005
	0.8	-0.011	0.027	-0.007	0.014	-0.004	0.006	-0.003	0.002

Table 2: Bias and Mean Square Error (MSE) of the OLS, MLE and II estimators at $n = 30, 50, 100, 200$ for $\lambda = 0.5, 0, 0.5, 0.8$ when W is given by (4.3) (10^4 repl. and $\epsilon \sim N(0, 1)$).

		BIAS	MSE
$\lambda = -0.5$		BIAS	MSE
	OLS	-0.377	0.305
	ML	0.007	0.055
	II	-0.014	0.069
$\lambda = 0$		BIAS	MSE
	OLS	-0.038	0.187
	ML	-0.016	0.056
	II	-0.016	0.062
$\lambda = 0.5$		BIAS	MSE
	OLS	0.258	0.150
	ML	-0.030	0.039
	II	0.019	0.076
$\lambda = 0.8$		BIAS	MSE
	OLS	0.290	0.111
	ML	-0.037	0.023
	II	0.097	0.074

Table 3: Bias and Mean Square Error (MSE) of the OLS, MLE and II estimators at $n = 43$ for $\lambda = 0.5, 0, 0.5, 0.8$ when W has an Empirical-based structure (10^4 repl. and $\epsilon \sim N(0, 1)$).

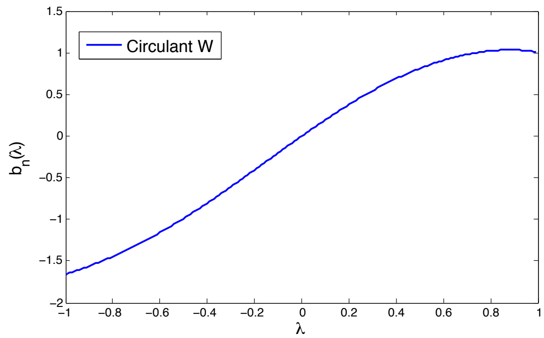

Figure 1: Approximate binding function, $b_n^*(.)$, for $\lambda \in (-1, 1)$ when W is chosen as in (4.1). $n = 100$

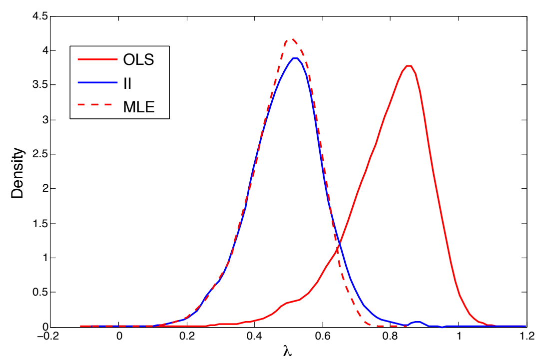

Figure 2: Empirical densities of $\hat{\lambda}$, $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ for $\lambda_0 = 0.5$ when W is chosen as in (4.1). $n = 100$.

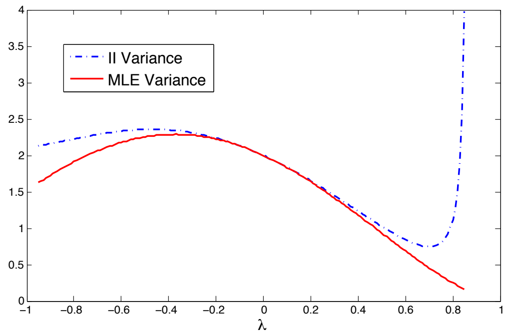

Figure 3: Finite sample (3.8) and (3.10) for $\lambda \in (-1, 1)$ when W is chosen as in (4.1). $n = 100$

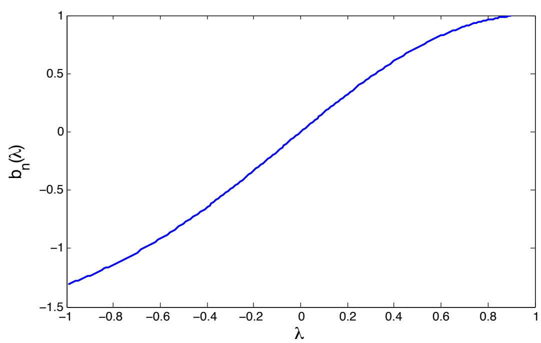

Figure 4: Approximate binding function, $b_n^*(.)$, for $\lambda \in (-1, 1)$ when W is chosen as in (4.3). $n = 100$

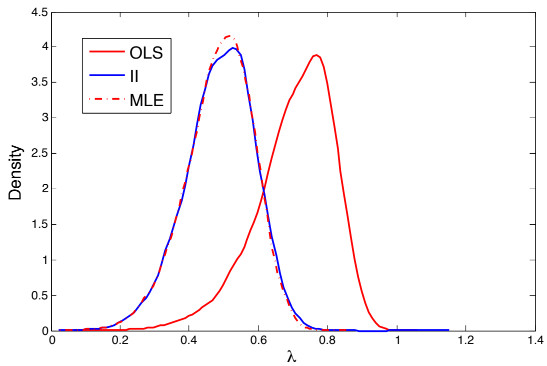

Figure 5: Empirical densities of $\hat{\lambda}$, $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ for $\lambda_0 = 0.5$ when W is chosen as in (4.3) at $n = 100$.

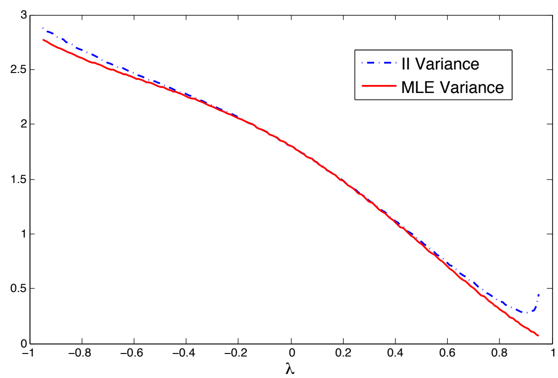

Figure 6: FInite sample (3.8) and (3.10) for $\lambda \in (-1, 1)$ when W is chosen as in (4.3). $n = 100$

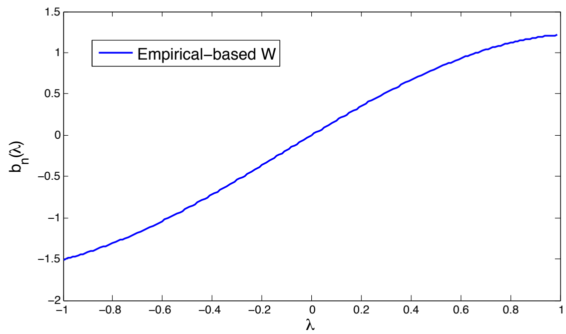

Figure 7: Approximate binding function, $b_n^*(.)$, for $\lambda \in (-1, 1)$ when W is chosen as in (4.4) and rescaled by its spectral norm. $n = 43$.

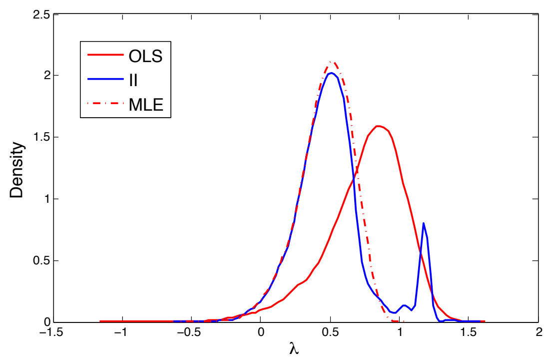

Figure 8: Empirical densities of $\hat{\lambda}$, $\hat{\lambda}_{MLE}$ and $\hat{\lambda}_{II}$ for $\lambda_0 = 0.5$ when W is chosen as in (4.4) and re-scaled by its spectral norm. $n = 43$.

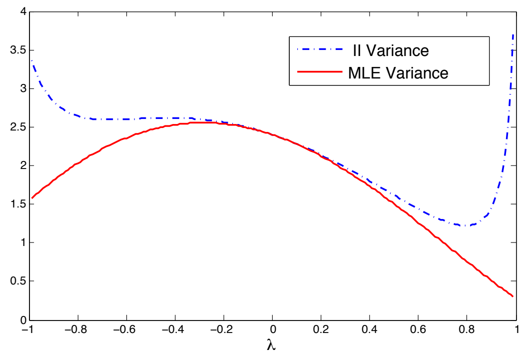

Figure 9: Finite sample (3.8) and (3.10) for $\lambda \in (-1, 1)$ when W is chosen as in (4.4) and re-scaled by its spectral norm. $n = 43$

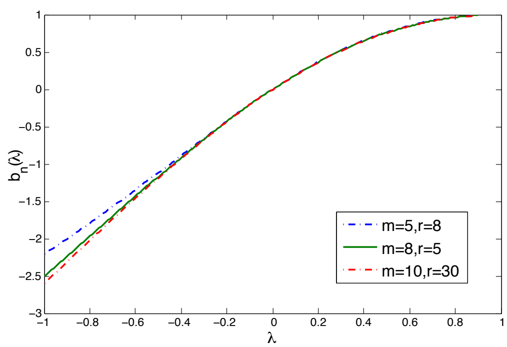

Figure 10: Approximate binding functions, $b_n^*(.)$, at various sample sizes when W is chosen as in (5.1).

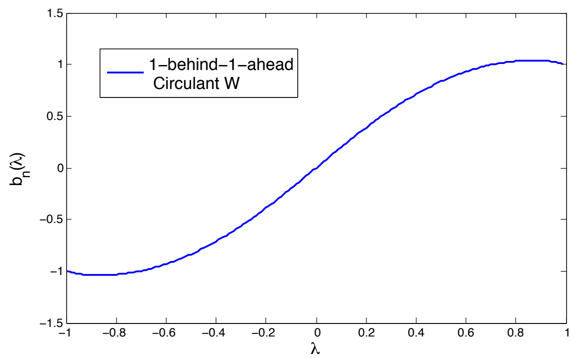

Figure 11: Approximate binding function, $b_n^*(.)$, for $\lambda \in (-1, 1)$ when W is chosen as in (5.3). $n = 100$

References

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.

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.