ADAPTIVE INFERENCE ON PURE SPATIAL MODELS

Jungyoon Lee[†] and Peter M. Robinson[‡]

Abstract. In a general class of semiparametric pure spatial models (having no explanatory variables) allowing nonlinearity in the parameter and the weight matrix, we propose adaptive tests and estimates which are asymptotically efficient in the presence of unknown, nonparametric distributional form. Feasibility of adaptive estimation is verified and its efficiency improvement over Gaussian pseudo maximum likelihood is shown to be either less than, or more than, for models with explanatory variables, depending on properties of the spatial weight matrix. An adaptive Lagrange Multiplier testing procedure for lack of spatial dependence is proposed and this, and our adaptive parameter estimate, are extended to cover regression with spatially correlated errors.

JEL classifications: C12, C13, C14, C21.

Keywords: Efficient test; adaptive estimation; spatial models.

....

[†] *Corresponding author.* Economics Department, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. Tel:+44 (0)1784 276506, Fax:+44 (0)1784 439534, E-mail: jungyoon.lee@rhul.ac.uk

[‡] Economics Department, London School of Economics, Houghton Street, London WC2A 2AE, UK. Tel:+44 (0)20 7955 7516, E-mail: p.m.robinson@lse.ac.uk

1. Introduction

Spatial autoregressive (SAR) models have been extensively developed in the econometric literature. As introduced by Cliff and Ord (1968), they modeled spatial correlation, without explanatory variables, echoing earlier work of Moran (1950) on testing for spatial correlation. We call these pure spatial models, and they are already known to lead to rather different statistical properties from those of models with explanatory variables. In particular the least squares estimate (LSE) of a pure spatial model is inconsistent, and with instrumental variables being unavailable, leading alternatives, the Gaussian pseudo maximum likelihood estimate (PMLE) and generalized method of moments estimate (GMME), may converge more slowly than at the parametric rate. Here we consider a quite general class of first-order pure spatial models which involves a known but possibly nonlinear transformation of the spatial dependence parameter and of a user-specified weight matrix, but a disturbance distribution of unknown, and thus possibly non-Gaussian, form. The latter aspect motivates us to develop adaptive estimates, and also tests for lack of spatial dependence, which are asymptotically as efficient as those based on correctly specified parametric distributions. Adaptive estimation was considered for spatial autoregressions with explanatory variables by Robinson (2010). Somewhat surprisingly, our setting leads to a different efficiency gain.

While Wald statistics based on our adaptive estimates have greater asymptotic local power compared to those based on less efficient estimates, we also provide adaptive Lagrange Multiplier (LM) tests which share these statistical properties but also have the computational advantage of being based on the restricted model only. Our adaptive LM tests achieve the same asymptotic power as LM tests based on the correctly specified error distributions. Many authors including

Cliff and Ord (1972), Burridge (1980), Kelejian and Prucha (2001), Robinson (2008), Baltagi and Yang (2013), Robinson and Rossi (2014) and Yang (2015) have considered Gaussian LM tests for lack of spatial dependence in the SAR model, extending Moran (1950). Although these tests enjoy the same robustness property as the Gaussian PMLE, there is room for further efficiency improvement which our adaptive LM tests achieve; our Monte Carlo study finds that their power improvement is often substantial.

A class of spatial models for a vector $y = (y_1, ..., y_n)^T$ of observations, with the same (unknown) mean $E(y_i) = \mu_0$, and ^{T} denoting transposition, is given by

$$Q(\lambda_0)(y - \mu_0 1_n) = \sigma_0 \varepsilon, \tag{1.1}$$

where 1_n is the $n \times 1$ vector of 1's, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^T$ is a vector of independent identically distributed (i.i.d.) random variables with zero mean and unit variance, and σ_0 and λ_0 are unknown scalar parameters. The $n \times n$ matrix $Q(\lambda_0)$ is described as follows.

Introduce the $n \times n$ weight matrix, $W = W_n$ with known real-valued (i, j)-th element w_{ij} such that $w_{ii} \equiv 0$. Letting $\|\cdot\|$ denote the spectral norm of a matrix, we impose the normalization $\|W\| = 1$. The paper develops asymptotic statistical theory with n diverging, and the individual $w_{ij}, 1 \leq i, j \leq n$ may change as n increases, but as with y, ε, Q and other quantities we suppress reference to n in our notation. The following are three special cases of (1.1).

(1) SAR(1) (spatial autoregression of degree 1, see e.g. Arbia(2006))

$$Q(\lambda_0) = I - \lambda_0 W, \quad \lambda_0 \in \Lambda, \tag{1.2}$$

where I is the $n \times n$ identity matrix and Λ is a closed subset in (-1,1).

(2) SMA(1) (spatial moving average of degree 1, see e.g. Anselin(2003))

$$(1.3) \qquad Q(\lambda_0) = (I - \lambda_0 W)^{-1}, \quad \lambda_0 \in \Lambda,$$

where Λ is a closed subset in (-1,1).

(3) MESS(1) (matrix spatial exponential model, see LeSage and Pace (2007))

$$(1.4) \qquad Q(\lambda_0) = \exp(\lambda_0 W).$$

When ε, and thus y, is Gaussian, (1.1) can be thought of as primarily describing the covariance matrix of y, since this, and μ_0, describe the distribution of y completely. The parameter vector $\theta_0 = \left(\lambda_0, \mu_0, \sigma_0\right)^T$ can be asymptotically efficiently estimated by the maximum likelihood estimate (MLE) $\tilde{\theta} = \left(\tilde{\lambda}, \tilde{\mu}, \tilde{\sigma}\right)^T$. Lee (2004) showed that for pure SAR, under some regularity conditions $\tilde{\theta}$ is consistent and asymptotically normal. In fact, he showed that these properties hold over a much wider class of distributions of the ε_i, for which $\tilde{\theta}$ is termed a (Gaussian) PMLE. Such robustness is also shared by the LM test for $H_0 : \lambda_0 = 0$ based on Gaussianity, see e.g. Baltagi and Yang (2013).

However, the Gaussian PMLE and LM test are asymptotically inefficient under non-Gaussianity. Given a (non-Gaussian) parametric specification of the distribution of ε_1, we can construct a (non-Gaussian) MLE and LM statistics as follows. Let $f(x;\zeta_0) = \mathbb{R}^{1+q} \to \mathbb{R}^1$ be the probability density function of ε_1, a given function of all its arguments, with ζ_0 being an unknown $q \times 1$ parameter vector. Set $\tau_0 = \left(\lambda_0, \mu_0, \sigma_0, \zeta_0^T\right)^T$, and denote by $\tau = \left(\lambda, \mu, \sigma^2, \zeta^T\right)^T$ any admissible value of τ_0. Write the corresponding log likelihood as

$$(1.5) \qquad L(\tau) = \sum_{i=1}^{n} \log f\left(\frac{Q_i^T(\lambda)(y - \mu 1)}{\sigma}; \zeta\right) + \log\det\{Q(\lambda)\} - \frac{n}{2}\log\sigma^2,$$

where $Q_i^T(\lambda)$ denotes the ith row of $Q(\lambda)$. The MLE $\bar{\tau} = \left(\bar{\lambda}, \bar{\mu}, \bar{\sigma}^2, \bar{\zeta}^T\right)^T$ of θ_0 maximizes (1.5) over a suitable compact set, and can be expected to be asymptotically efficient. The LM statistic can be constructed from the first and second derivatives of $L(\theta)$ evaluated at $\lambda = 0$. However there are rarely strong prior grounds for specifying f, and misspecification of a non-Gaussian probability density f in general leads to inconsistent estimation and tests.

In practice, λ_0 is often the main feature of interest, with μ_0 and σ_0 being nuisance parameters (and our results on inference on λ_0 are unaffected if μ_0 is known *a priori*). In this paper we establish an estimate $\hat{\lambda}$ of λ_0 that achieves the same asymptotic distribution as the MLE $\bar{\lambda}$, in the presence of only nonparametric assumptions on the distribution of ε_1. Specifically, the adaptive estimate $\hat{\lambda}$ takes a Newton step from a consistent preliminary estimate such as the Gaussian PMLE $\tilde{\lambda}$, using nonparametric (series) estimation of the score function. In similar vein, our adaptive LM statistic based on the nonparametrically-estimated score function is shown to achieve the same efficiency as that based on the (unknown) true score function.

This kind of "adaptive" property was previously established in a spatial context by Robinson (2010), for the mixed regressive SAR(1) (MRSAR)

$$(I - \lambda_0 W)\, y = \mu_0 1_n + X\beta_0 + \sigma_0 \varepsilon, \tag{1.6}$$

where X is a $n \times k$ matrix of observed explanatory variables and β_0 is a vector of unknown parameters. Although it may seem that (1.2) is a special case of (1.6) with $\beta_0 = 0$, the asymptotic behaviors of estimates of λ_0 under the two models can differ, even their convergence rates (see Lee (2004)). Consequently, the feasibility and implementation of such adaptive estimation in pure spatial models need to be established separately. Both Robinson (2010) and the current paper find that adaptive inference is feasible in spatial models under certain conditions on W, which in pure spatial models lead to a slower-than-$\sqrt{n}$ rate of convergence, unlike in (1.6). Our adaptive estimation offers efficiency improvements, but not a faster rate of convergence.

Another notable difference of the current paper from Robinson (2010) is that while the asymptotic variance matrix of his estimate of $(\lambda_0, \beta_0^T)^T$ corresponds to that found in the classical adaptive estimation literature, achieving the parametric Cramer-Rao bound, ours differs from the classical one. In particular, the efficiency gain of the improved $\hat{\lambda}$ over the preliminary $\tilde{\lambda}$ can be either less or more than in the classical outcome, depending on the symmetry and signs of elements of the matrix W. This is a unique finding in the adaptive estimation literature and could be explored further in other settings where dependence is not one-directional, unlike in time series.

Pure spatial models are often used to model spatial correlation in regression disturbances. All our results on adaptive estimation and testing also cover this setting, in which testing for lack of error spatial dependence is of particular practical interest, in order to gauge the need for spatial-correlation-robust standard errors. Kelejian and Prucha (1999) proposed method-of-moments estimates (MME) for the spatial parameters in such models, while numerous papers study LM testing for absence of spatial error dependence see e.g. Robinson (2008), Baltagi and Yang (2013), Robinson and Rossi (2014) and Yang (2015). In a broader model that allows SAR in both the regressors and disturbances, Liu *et al.* (2010) proposed a best GMME (BGMME) based on certain linear and quadratic moments. As a special case of interest, they consider regression models with SAR error terms in their Corollary 3, which is of relevance to the current paper. In the online appendix, we show (as is intuitively expected) that our adaptive estimate is more efficient than Liu *et al.*'s (2010) BGMME under our conditions, apart from the Gaussian case when they are equally efficient. The online appendix additionally contains a small Monte Carlo study, a summary of which can also be found in Subsection 5.1 below.

Our allowance for a class of functional forms Q in (1.1) is unusual in the spatial econometric literature. We focus on i.i.d. ε_i in (1.1), for which the Gaussian PMLE, as well as MME's of Kelejian and Prucha (1999) and Liu *et al.* (2010) are consistent, all of which can serve as the preliminary estimate in our adaptive estimation. When the ε_i are independent but heteroscedastic, these estimates are inconsistent, and Lin and Lee (2010) and Kyriacou, Phillips and Rossi (2019) offer alternative methods of estimation based on BGMME and indirect inference, respectively. Adapting to heteroscedasticity of unknown form has been pursued in other models (see e.g., Robinson (1987)) but is motivated by the Gauss-Markov, not the Cramer-Rao, bound.

Section 2 presents the information matrix corresponding to the MLE based on (1.5), its form suggesting both potential for adapting to unknown distributional form of ε_1 in the estimation of λ_0, and the scope for efficiency gains described earlier. Section 3 describes our estimate $\hat{\lambda}$ and its asymptotic distribution, also when (1.1) models the unobserved errors in linear regression. The nonparametric estimate of the score function for ε_1 introduced in Section 3 is used in Section 4 to construct an adaptive LM test for lack of spatial dependence. Section 5 presents results of a small Monte Carlo study of finite sample performance of our adaptive estimate and LM testing procedure. Both estimation and testing lead to substantial efficiency gains compared to the Gaussian PMLE, and our LM testing tends to avoid the substantial oversizing of Wald tests (also reported in a panel data setting in Robinson and Rossi (2015)). Section 6 applies our methods to an economic dataset on crime rates across Italian provinces, an appendix includes the proof of one of our results, and the remainder of the proofs and other supplementary materials appear in an online appendix.

2. Block-diagonality of the information matrix

The feasibility of adaptive estimation of λ_0 is shown by establishing block-diagonality of the information matrix. Denote $M(\lambda) := -dQ(\lambda)/d\lambda$ and $M = M(\lambda_0) = (m_{ij})$. For SAR (1.2) $M(\lambda) = W$, for SMA (1.3) $M(\lambda) = -(I - \lambda W)^{-1}W(I - \lambda W)^{-1}$ and for MESS (1.4) $M(\lambda) = -W\exp(\lambda W)$.

Assumption 1. (i) *For all sufficiently large n, $M = (m_{ij})_{i,j=1,\cdots,n}$ is uniformly bounded in both row and column sums, i.e. as $n \to \infty$,*

$$\max_{1\leq i\leq n}\sum_{j=1}^{n}|m_{ij}| + \max_{1\leq j\leq n}\sum_{i=1}^{n}|m_{ij}| = O(1).$$

(ii) *For a sequence $h = h_n$ such that $h^{-1} + h/n \to 0$ as $n \to \infty$, $\max_{1\leq i,j\leq n}|m_{ij}| = O\big(h^{-1}\big)$.* (iii) *Uniformly over λ in a neighbourhood of λ_0, $Q(\lambda)$ is non-singular and $Q^{-1}(\lambda)$ is uniformly bounded in both row and column sums, for all sufficiently large n.*

The sequence h is important in asymptotic analysis, defining the rate of convergence of estimates of λ_0. Assumption 1 (i), (iii) is often assumed for SAR (1.2) with $M = W$ and $Q = I - \lambda_0 W$ (see e.g. Assumptions 2-5, 7 of Lee (2004)), and one can show that Assumption 1 (i), (iii) will then hold also for SMA and MESS models with the same W. Assumption 1(ii) acknowledges that feasibility of adaptive inference rests crucially on divergence of h as n increases. This also implies that estimates of λ_0 are only $\sqrt{n/h}$-consistent, and the efficiency improvement offered by adaptive estimation is of particular value in light of this.

Assumption 2. *The limits*

$$\omega_1 := \lim_{n\to\infty} \frac{h}{n}\text{tr}(MQ^{-1}Q^{-1T}M^T), \quad \omega_2 := \lim_{n\to\infty} \frac{h}{n}\text{tr}(MQ^{-1}MQ^{-1}),$$
$$\omega_3 := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(Q_i^T 1_n)^2, \quad \omega_4 := \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(Q_i^T 1_n)$$

exist and are finite, and $\omega_1 > 0$, $\omega_2 \neq 0$, $\omega_3 > 0$.

Similar assumptions on ω_1 and ω_2 are imposed and discussed in Robinson and Rossi (2015, Assumption 5) and Kyriacou, Phillips and Rossi (2017, Assumption 5) for SAR (1.2) where, with row-normalized W, i.e. $\sum_{j=1}^n w_{ij} \equiv 1$ for all $i = 1, ..., n$, we obtain $\omega_3 = (1-\lambda_0)^2$ and $\omega_4 = 1 - \lambda_0$.

Introduce:

$$\begin{aligned}
\psi_i &:= -\frac{\partial}{\partial \varepsilon}\log f(\varepsilon_i;\zeta_0), \quad \chi_i := -\frac{\partial}{\partial \zeta}\log f(\varepsilon_i;\zeta_0), \quad i \geq 1, \\
\mathcal{J} &:= E\psi_i^2, \quad d := \text{diag}\big\{(n/h)^{\frac{1}{2}}, \quad n^{\frac{1}{2}} I_{d+2}\big\}.
\end{aligned}$$

Proposition 1. *Under Assumptions 1-2,* $\Xi := \lim_{n\to\infty} d^{-1}E\big(-\frac{d^2L(\tau_0)}{d\tau d\tau^T}\big)d^{-1}$ *exists, where*

$$\Xi = \begin{pmatrix} \mathcal{J}\omega_1 + \omega_2 & & & \\ 0 & \frac{\mathcal{J}}{\sigma_0^2}\omega_3 & & \\ 0 & \frac{E(\varepsilon_i\psi_i^2)}{2\sigma_0^3}\omega_4 & \frac{1}{4\sigma_0^4}E(\varepsilon_i^2\psi_i^2 - 1) & \\ 0 & 0 & -\frac{1}{2\sigma_0^2}E(\varepsilon_i\psi_i\chi_i) & E(\chi_i\chi_i^T) \end{pmatrix}.$$

Noting the zero non-diagonal elements of the first column, the feasibility of adaptive estimation of λ_0 is established. The proof of Proposition 1 is given in the Appendix and the online appendix, which also includes a discussion of models with regressors when their coefficients are zero, e.g. MRSAR (1.6) when $\beta_0 = 0$.

3. Adaptive Estimation and its Asymptotic Properties

With f, f' respectively denoting the nonparametric density and derivative-of-density of ε_1, its score function is given by $\psi(s) = -f'(s)/f(s)$, when $f(s) \neq 0$. The nonparametric estimate of ψ we use is a series one following those proposed in Beran (1976), Newey (1988) and Robinson (2005) in other contexts, whose advantages over kernel estimation are discussed in Robinson (2010). Denote by $\tilde{\psi}_{iL}$ the series estimate of $\psi(\varepsilon_i)$ where the integer $L = L_n$, regarded as slowly increasing with n, represents the number of series functions. Construction of $\tilde{\psi}_{iL}, i = 1, ..., n$, using a sequence of smooth series functions $\phi_\ell(s)$, $\ell = 1, 2, ...$, is the same as in Robinson (2010) and to save space, details can be found therein, or in our online appendix.

Introduce the estimate of the information $\mathcal{J}$, $\tilde{\mathcal{J}}_L = \sum_{i=1}^n \tilde{\psi}_{iL}^2/n$, and define $P(\lambda) := M(\lambda)Q^{-1}(\lambda), P = P(\lambda_0) = (p_{ij})$. For SAR (1.2) $P(\lambda) = W(I - \lambda W)^{-1}$, for SMA (1.3) $P(\lambda) = -(I - \lambda W)^{-1}W$ and for MESS (1.4) $P(\lambda) = -W$.

Our adaptive estimate of λ_0 is given by:

$$\hat{\lambda} = \tilde{\lambda} + \left(\tilde{\mathcal{J}}_L \cdot \operatorname{tr}\left\{P(\tilde{\lambda})P(\tilde{\lambda})^T\right\} + \operatorname{tr}\left\{P(\tilde{\lambda})^2\right\}\right)^{-1} \left(\frac{1}{\tilde{\sigma}}(\tilde{\psi}_{1L}, \cdots, \tilde{\psi}_{nL})M(\tilde{\lambda})Hy - \operatorname{tr}\left\{P(\tilde{\lambda})\right\}\right). \tag{3.1}$$

For the proxy $\tilde{\varepsilon}$ to be used for ε in the above construction, define $e(\lambda) = (e_1(\lambda), \cdots, e_n(\lambda))^T := Q(\lambda)y$. For given λ, sample mean-adjusted residuals are $\epsilon_i(\lambda) := e_i(\lambda) - n^{-1}\sum_{j=1}^n e_j(\lambda)$. Using the $n \times n$ matrix $H := I - n^{-1}1_n 1_n^T$, we can write

$$\epsilon(\lambda) = (\epsilon_1(\lambda), \cdots, \epsilon_n(\lambda))^T = HQ(\lambda)y. \tag{3.2}$$

Given an estimate $\tilde{\lambda}$ of λ_0, we estimate σ_0^2 by $\tilde{\sigma}^2(\tilde{\lambda}) := \epsilon(\tilde{\lambda})^T\epsilon(\tilde{\lambda})/n$. Our proxy $\tilde{\varepsilon}$ for ε is then $\tilde{\varepsilon} := \epsilon(\tilde{\lambda})/\tilde{\sigma}$.

The following assumptions are introduced for our asymptotic theory.

Assumption 3. $\{\varepsilon_i\}$ *is a sequence of i.i.d. random variables with zero mean, unit variance and twice differentiable probability density function* $f(\cdot)$ *such that* $sf'(s) \to 0$ *and* $s^2f''(s) \to 0$ *as* $|s| \to \infty$ *and, for some* $\delta > 0$,

$$E(\psi^4(\varepsilon_1)) + E|\varepsilon_1\psi(\varepsilon_1)|^{2+\delta} < \infty.$$

In Robinson (2010), for the MRSAR (1.6), symmetry of $f(\cdot)$ and/or W could be exploited in a bias-corrected adaptive estimate. In pure spatial models, such bias-correction is not relevant and our adaptive estimate and test take the same form, regardless of the (a)symmetry of $f(\cdot)$ and W.

Assumption 4. *The series functions satisfying* $\phi_\ell(s) = \phi(s)^\ell$, $l = 1, ..., L$, *where* $\phi(s)$ *is a strictly increasing and thrice differentiable function such that for some* $\kappa \geq 0$, $K > 0$,

$$(3.3)\quad |\phi(s)| \leq 1 + |s|^\kappa, \quad |\phi'(s)| + |\phi''(s)| + |\phi'''(s)| \leq C(1 + |\phi(s)|^K), \quad s \in \mathbb{R}.$$

Assumption 4 is the same as in Robinson (2010), where it is discussed.

Define $\eta := 1 + \sqrt{2}$ and $\varphi := (1 + |\phi(s_1)|)/\{\phi(s_2) - \phi(s_1)\}$, with $[s_1, s_2]$ being an interval on which $f(s)$ is bounded away from zero.

Assumption 5. *The sequences* h *and* L *satisfy one of the following conditions with* κ *as in (3.3). (i)* $\kappa = 0$, $E(\varepsilon_1^4) < \infty$, *and for some* $A > \eta \max(\varphi, 1)$, $L \leq \log h/8 \log A$, $n \to \infty$. *(ii)* $\kappa > 0$, *for some* $\omega > 0$ *and* $t > 0$, $E\left(e^{t|\varepsilon_i|^\omega}\right) < \infty$, *and for some* $B > 8\kappa \max(1, \frac{1}{\omega})$, $L \log L \leq \log h/B$, $n \to \infty$. *(iii)* $\kappa > 0$, ε_1 *is almost surely bounded, and for some* $C > 4\kappa$, $L \log L \leq \log h/C$, $n \to \infty$.

Assumption 5 is an amended version of Assumption 5 of Robinson (2010), capturing the trade-offs in the choice of series functions and restrictions imposed on the ε_i's, L and h. For bounded ϕ and $E\varepsilon_i^4 < \infty$ only, Assumption 5 (i) entails a relatively modest upper bound on the rate of growth of L.

Define $\phi^{(L)}(s) = (\phi_1(s), \cdots, \phi_L(s))^T$, $\bar{\phi}^{(L)}(s) = \phi^{(L)}(s) - E\left\{\phi^{(L)}(\varepsilon_i)\right\}$, $\phi'^{(L)}(s) = (\phi_1'(s), \cdots, \phi_L'(s))^T$, and $a^{(L)} = \left[E\left\{\bar{\phi}^{(L)}(\varepsilon_i)\bar{\phi}^{(L)}(\varepsilon_i)^T\right\}\right]^{-1} E\left\{\phi'^{(L)}(\varepsilon_i)\right\}$, where $\psi(s)$ is approximated by $\bar{\phi}^{(L)}(s)^T a^{(L)}$.

Assumption 6. *As $n \to \infty$, $h = O(\sqrt{n})$ and*

$$E\left\{\bar{\phi}^{(L)}(\varepsilon)^T a^{(L)} - \psi(\varepsilon_i)\right\}^2 = o\left(h/n\right).$$

Assumption 6 requires the series functions approximate $\psi(\cdot)$ sufficiently well as n increases, a typical condition imposed in the series estimation literature. Assumption 6 is stronger than Assumption 7 of Robinson (2010), necessitated by the slower rate of convergence of estimates of λ_0 in pure spatial models.

Assumption 7. *As $n \to \infty$,*

$$\tilde{\lambda} - \lambda_0 = O_p((h/n)^{1/2}), \quad \tilde{\sigma} - \sigma_0 = O_p(n^{-1/2}).$$

The Gaussian PMLE satisfies Assumption 7. Method-of-moments estimates such as those of Kelejian and Prucha (1999) and Liu *et al.* (2010) also satisfy Assumption 7.

We can extend our setting to linear regression with spatially dependent errors;

$$z = X\beta_0 + y, \tag{3.4}$$

with $n \times 1$ vector of dependent variables z, $n \times k$ matrix of regressors $X = (X_1, ..., X_n)^T$, and the (now unobservable) error y following (1.1) with $\mu_0 = 0$.

The model (3.4) along with (1.1) is called the Spatial Error model (SEM) by Anselin (1988). Under Assumption 1, dependence in the error y is weak and estimates of β_0 such as the LSE are $\sqrt{n}$-consistent. Denoting such estimates $\tilde{\beta}$, we replace y by $\tilde{y} := Z - X\tilde{\beta}$ in constructing adaptive estimates.

Assumption 8. *The X_i are i.i.d. random variables with $E\|X_i\|^4 < \infty$, which is independent of $\{\varepsilon_i\}$. In addition, $\tilde{\beta} - \beta = O_p(n^{-1/2})$.*

The following theorem states asymptotic normality of the adaptive estimate $\hat{\lambda}$ in (3.1).

Theorem 1.

(i) Let y satisfy (1.1) and Assumptions 1 - 7. Then, as $n \to \infty$,

$$\sqrt{\frac{n}{h}}\left(\hat{\lambda} - \lambda_0\right) \to_d N(0, \{\mathcal{J}\omega_1 + \omega_2\}^{-1}).$$

(ii) Let y satisfy (3.4) and (1.1), y be replaced by $\tilde{y}$ in $\hat{\lambda}$, and Assumptions 1-8 be satisfied. Then, as $n \to \infty$,

$$\sqrt{\frac{n}{h}}\left(\hat{\lambda} - \lambda_0\right) \to_d N(0, \{\mathcal{J}\omega_1 + \omega_2\}^{-1}).$$

3.1. **Efficiency comparison of adaptive estimate with PMLE.** Lee (2004) showed that for pure SAR

$$\sqrt{\frac{n}{h}}(\tilde{\lambda} - \lambda_0) \to_d N\left(0, \{\omega_1 + \omega_2\}^{-1}\right).$$

It is of interest to compare the asymptotic variance of $\tilde{\lambda}$ to that of $\hat{\lambda}$ given in Theorem 1, and see how the efficiency improvement attained via adaptive estimation in our spatial setting contrasts with that in other settings.

For SAR, $P = G := W(I - \lambda_0 W)^{-1}$, with, from Assumption 2,

$$\omega_1 = \lim_{n\to\infty} \frac{h}{n} tr(GG^T), \quad \omega_2 = \lim_{n\to\infty} \frac{h}{n} tr(G^2).$$

There exist W that imply $\mathrm{tr}(G^2) < 0$, $\omega_2 < 0$. However, if all elements of G are non-negative, which is implied if $w_{ij} \geq 0$ for all i, j and $\lambda_0 \geq 0$, or if W is symmetric, then $\omega_2 > 0$. In any case, it is possible to show that $\mathrm{tr}\big(G(G+G^T)\big) > 0$, so since $\mathrm{tr}(GG^T) \geq 0$ also, we have $\omega_1 > 0$ and $\omega_1 + \omega_2 > 0$, implying

$$\mathcal{J}\omega_1 + \omega_2 \geq \omega_1 + \omega_2 > 0, \quad \text{because} \quad \mathcal{J} \geq 1.$$

This shows that $\hat{\lambda}$ is better than $\tilde{\lambda}$. The relative efficiency of $\hat{\lambda}$ to $\tilde{\lambda}$ is given by

$$\frac{\omega_1 + \omega_2}{\mathcal{J}\omega_1 + \omega_2} = \frac{1 + \omega_2/\omega_1}{\mathcal{J} + \omega_2/\omega_1}.$$

In the autoregressive time series setting, where W is a lower triangular matrix, $\omega_2 = 0$, and therefore the relative efficiency is $1/\mathcal{J}$. Thus when $\omega_2 > 0$ in our setting the efficiency improvement achieved by our adaptive estimate is less than in the time series case or other setting studied in the literature. For example if W is symmetric, the relative efficiency is $2/(\mathcal{J} + 1)$. On the contrary, $\omega_2 < 0$ yields greater efficiency improvement than under the time series setting and many others including MRSAR (1.6).

4. Testing for lack of spatial dependence

One can construct an "adaptive" LM statistic based on the series estimation of the score function given in Section 3 in order to test $H_0 : \lambda_0 = 0$ against $H_1 : \lambda_0 \neq 0$ in (1.1). The LM statistic has the usual advantage of requiring estimation only of the null model. It is based on the following standardized residuals from the null model denoted $\tilde{\varepsilon}_i^{(r)}$:

$$\tilde{\varepsilon}_i^{(r)} = \tilde{\epsilon}_i^{(r)}/\tilde{\sigma}_{(r)}, \text{ where } \quad \tilde{\epsilon}_i^{(r)} = y_i - \bar{y}, \; \tilde{\sigma}_{(r)}^2 = \tilde{\epsilon}^{(r)T}\tilde{\epsilon}^{(r)}/n. \tag{4.1}$$

For SAR, the Gaussian LM statistic LM_{SAR}^G takes the form:

$$LM_{SAR}^G = \frac{\left(\tilde{\varepsilon}^{(r)T}W\tilde{\varepsilon}^{(r)}\right)^2}{\mathrm{tr}(\mathrm{WW^T}) + \mathrm{tr}(\mathrm{W}^2)}. \tag{4.2}$$

Burridge (1980) noted that LM_{SAR}^G is also the Gaussian LM statistic for SMA. Whilst the Gaussian LM test shares the robustness of the Gaussian PMLE, one expects power to improve under a correctly specified error distribution. To build a LM statistic which adapts to distribution of nonparametric form, note that under H_0:

$$\frac{\partial L(\theta_0)}{\partial \lambda}\Big|_{H_0} = \sum_{i=1}^{n} \frac{M_i^T(0)(y - \mu_0 1_n)}{\sigma_0}\psi\Big(\frac{y - \mu_0 1_n}{\sigma_0}\Big) - \mathrm{tr}\{P(0)\},$$

$$\lim_{n\to\infty} \frac{h}{n} E(\frac{\partial^2 L(\theta_0)}{\partial \lambda^2}\Big|_{H_0}) = \mathcal{J}\omega_{10} + \omega_{20},$$

with ω_{10} and ω_{20} evaluated at $\lambda_0 = 0$.

We estimate $\psi((y_i - \mu_0)/\sigma_0)$ and $\mathcal{J}$ by $\tilde{\psi}_{iL}^{(r)} := \Phi^L(\tilde{\varepsilon}_i^{(r)})^T \tilde{a}^L(\tilde{\varepsilon}^{(r)})$ and $\tilde{\mathcal{J}}_L^{(r)} = \sum_{i=1}^n (\tilde{\psi}_{iL}^{(r)})^2/n$. Our adaptive LM statistic is given by

$$LM^A = \frac{\big(\sum_{i=1}^{n} M_i^T(0)\tilde{\varepsilon}^{(r)} \cdot \tilde{\psi}_{iL}^{(r)} - \mathrm{tr}(P(0))\big)^2}{\tilde{\mathcal{J}}_L^{(r)}\mathrm{tr}(P(0)P^T(0)) + \mathrm{tr}(P(0)^2)}.$$

For SAR we have $M(0) = W = P(0)$, while for SMA and MESS $M(0) = -W = P(0)$. Hence in SAR:

$$\frac{\partial L(\theta_0)}{\partial \lambda}\Big|_{H_0} = \sum_{i=1}^{n} \frac{W_i^T(y-\mu_0 1_n)}{\sigma_0}\psi\Big(\frac{y_i-\mu_0}{\sigma_0}\Big),$$

$$\lim_{n\to\infty}\frac{h}{n}E(\frac{\partial^2 L(\theta_0)}{\partial \lambda^2}\Big|_{H_0}) = \lim_{n\to\infty}\mathcal{J}\frac{h}{n}\mathrm{tr}(WW^T) + \frac{h}{n}\mathrm{tr}(W^2).$$

The adaptive LM statistics under SAR, SMA and MESS turn out to take the the identical form

$$LM_{SAR}^A = \frac{\big(\sum_{i=1}^{n} W_i^T\tilde{\varepsilon}^{(r)}\cdot\tilde{\psi}_{iL}^{(r)}\big)^2}{\tilde{\mathcal{J}}_L^{(r)}\mathrm{tr}(WW^T) + \mathrm{tr}(W^2)}. \tag{4.3}$$

In the SEM model (3.4) with (1.1), $\tilde{\epsilon}^{(r)} = H\tilde{y}$.

Theorem 2.

(i) Let y follow model (1.1) with Assumptions 1-7 satisfied. Under $H_0 : \lambda_0 = 0$, as $n \to \infty$, $LM^A \to_d \chi^2(1)$.

(ii) Let y satisfy (3.4) and (1.1), y be replaced by $\tilde{y}$ in $\hat{\lambda}$, and Assumptions 1-8 be satisfied. Then, as $n \to \infty$ under $H_0 : \lambda_0 = 0$, $LM^A \to_d \chi^2(1)$.

Robinson (2008), Baltagi and Yang (2013) and Yang (2015) note that the Gaussian LM statistic may suffer from size distortion and low power in finite samples, with the latter two works pointing out that both problems worsen with larger h. The first two references suggest various modifications to the LM statistic in order to improve finite sample performance, while the latter suggests bootstrap approximation of the critical values. In our Monte Carlo study of Section 5.2, the adaptive LM test often leads to improved size and power performance compared to the Gaussian LM one. Developing procedures for our adaptive LM statistic similar to those in the aforementioned works to further refine the test may be of value.

5. Monte Carlo Study of Finite Sample Performance

In this section, we report results from a small Monte Carlo study of finite sample performance of our adaptive estimate and test. We first study the efficiency improvement achieved by the adaptive $\hat{\lambda}$ relative to the preliminary PMLE $\tilde{\lambda}$ under differing error distributions, sample sizes, and magnitudes of spatial dependence, and then compare size and power performance.

We use the following block-diagonal weight matrix introduced in Case (1992), where 1_m denotes a $m \times 1$ vector of 1's and I_m is the $m \times m$ identity matrix:

$$W = \frac{1}{r-1} \begin{pmatrix} 1_m 1_m' - I_m & 0 & 0 & \ldots & 0 \\ 0 & 1_m 1_m' - I_m & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 1_m 1_m' - I_m \end{pmatrix}.$$

The sample size is $n = mr$ and we have $h = r - 1$. We take values of (m, r) as in the Monte Carlo study of Robinson (2010): $(m, r) = (12, 8), (18, 11)$ and $(28, 14)$ with the corresponding sample sizes $n = 96, 198$ and 392. To investigate effects of differing strength of spatial dependence, we consider $\lambda_0 = 0.2, 0.4, 0.8$ for SAR, $\lambda_0 = -0.2, -0.4, -0.8$ for SMA and $\lambda_0 = 1, 2, 3$ for MESS. The following five distributions of ε_i are used, the first four as in the Monte Carlo study of Robinson (2010), with the last added to include an asymmetric density, namely Gamma(2,1) taken from the Monte Carlo study of Liu *et al.* (2010) and standardized to have variance 1, to be comparable to the first four distributions.

(a) Unimodal mixture normal, $\varepsilon_i = u/\sqrt{2.2}$ where

$$f(u) = \frac{0.05}{\sqrt{50\pi}} exp\big(- \frac{u^2}{50}\big) + \frac{0.95}{\sqrt{2\pi}} exp\big(- \frac{u^2}{2}\big), \quad u \in \mathbb{R}.$$

(b) Bimodal mixture normal, $\varepsilon_i = u/\sqrt{10}$, where the pdf of u is

$$f(u) = \frac{0.5}{\sqrt{2\pi}} exp\big(- \frac{(u-3)^2}{2}\big) + \frac{0.5}{\sqrt{2\pi}} exp\big(- \frac{(u+3)^2}{2}\big), \quad u \in \mathbb{R}.$$

(c) Laplace, $\varepsilon_i = u$ where $f(u) = \exp(-|u|\sqrt{2})\sqrt{2}, \quad u \in \mathbb{R}$.

(d) Student t_5, $\varepsilon_i = u\sqrt{3/5}$, where $u \sim t_5$.

(e) Gamma, $\varepsilon_i = (u-2)/\sqrt{2}$, where $f(u) = ue^{-u}/\Gamma(2)$, $u > 0$; $= 0$, $u \leq 0$.

We report results with $L = 1, 3, 5$ for $n = 96$, $L = 3, 5, 7$ for $n = 198$ and $L = 4, 6, 8$ for $n = 392$ with $\phi_\ell(s) = \phi^\ell(s), \ell = 1, ..., L$ and two choices of $\phi(s)$:

$$(5.1) \qquad (i) \quad \phi(s) = s, \qquad (ii) \quad \phi(s) = \frac{s}{(1+s^2)^{1/2}}.$$

5.1. **Efficiency improvement in estimation.** Based on 1000 replications, the Monte Carlo variances of the two estimates of λ_0 were computed in each setting, and their ratios presented in Tables 1-3 for SAR, SMA and MESS, respectively. A ratio less than 1 indicates efficiency improvement. Substantial improvements are reported in cases (a) and (b) for all three models, as also observed in Robinson (2010). For error distributions (c) and (d) in SAR and SMA, relative variance is greater than 1 when $|\lambda_0| = 0.2$ (except for SMA with $n = 392$), and when $|\lambda_0| = 0.4, 0.8$ the ratio is less than 1 for some L but not dramatically so. In MESS the ratios are mostly less than 1 for all λ for (c) and (d) but not by much. In (e), the efficiency improvement is marked in SAR and MESS, but ratios are mostly greater than 1 in SMA, although they are somewhat smaller for $\lambda_0 = 0.8$.

In most settings, the efficiency improvement increases with n, and with the choice (ii) of ϕ over (i). The best choice of L differs across models, error distributions and ϕ and λ. Apart from case (a), there is little discernible pattern in the best L except that it increases with n in almost all settings. With (a), across all three models the best L is $5, 7, 7/8$ for $n = 96, 198, 392$, respectively.

Table 4 reports the relative MSE to ascertain whether bias is adversely affected by adaptive estimation for choice (ii) of ϕ. In fact, relative MSE often exhibited greater improvement than relative variance, suggesting bias has been also reduced. In the online appendix, corresponding results can be found for $\phi = (i)$ in Table A1 which reports also that bias is often reduced with adaptive estimation.

In Tables 1 and 4, a distinctive contrast to results obtained in MRSAR (1.6) of Robinson (2010) is that the efficiency improvement is greater for larger λ_0. For SMA (MESS), the efficiency improvement is greater for $\lambda_0 = -0.4(2)$ than $\lambda_0 = -0.2(1)$ but the pattern is less clear between $\lambda_0 = -0.4(2)$ and $\lambda_0 = -0.8(3)$.

[Tables 1-4 about here]

A referee suggested that we compare our adaptive estimate with the best generalized method of moments estimate (BGMME) of Liu *et al.* (2010) based on linear and quadratic moment conditions. Under their conditions (which do not include our requirement that $h \to \infty$), not only is their BGMME never asymptotically less efficient than the PMLE (which of course is based on second moments), but as well as having equal efficiency under Gaussianity they found it can be more efficient in non-Gaussian settings. In the online appendix we establish that under our conditions, including $h \to \infty$ (which is crucial to the adaptive property of our estimate), the BGMME always has equal efficiency to the PMLE, and thus is less efficient than our adaptive estimate except in the Gaussian case $\mathcal{J} = 1$ when it is equally efficient. The online appendix also includes a Monte Carlo comparison of finite sample performance in which our adaptive estimate was generally more efficient than the BGMME across various choices of n, error distribution, ϕ and L.

5.2. **Test of $H_0 : \lambda_0 = 0$.** We now compare the finite sample size and power properties of tests of lack of spatial dependence based on 4 different test statistics: the Gaussian LM LM_{SAR}^G in (4.2), the adaptive LM LM_{SAR}^A in (4.3), abbreviated LM^G and LM^A below and Wald statistics based on Gaussian PMLE $\tilde{\lambda}$ and our adaptive estimate $\hat{\lambda}$:

$$W^G = \tilde{\lambda}\sqrt{\mathrm{tr}(G(\tilde{\lambda})G^T(\tilde{\lambda})) + \mathrm{tr}(G^2(\tilde{\lambda}))}, \quad W^A = \hat{\lambda}\sqrt{\tilde{\mathcal{J}}_L \mathrm{tr}(G(\hat{\lambda})G^T(\hat{\lambda})) + \mathrm{tr}(G^2(\hat{\lambda}))}.$$

We report results with the same choices of L as in the previous subsection: $(L_1, L_2, L_3) = (1,3,5), (3,5,7), (4,6,8)$ for $n = 96, 198, 392$, respectively. All results are based on 1000 iterations and the data generating process (DGP) stays unchanged from the previous subsection.

In Table 5, we report Monte Carlo size, for nominal size $s = 0.1, 0.05, 0.01$. For the Wald statistic W^G, undersizing is severe and does not improve with increasing n, frequently getting worse with larger n across all five distributions. This is in line with what Robinson and Rossi (2015) observed in a panel data setting, notably for Gaussian data. Their Table 2 reported severe undersizing for the Gaussian PMLE, albeit for smaller sample sizes $n = 12, 15, 20, 40$. Their Figures 1 and 2 demonstrated how the Gaussian cumulative distribution function (cdf) poorly approximated the cdf of the Gaussian PMLE, even when $\lambda_0 = 0$. Our adaptive Wald statistic improves matters except when $L = 1, n = 96$ in (b), and the extent of improvement increases with n and L in (a), (b) and (e). In (c) and (d), sizes for W^A do not necessarily improve with larger n, although they do improve with increasing L for given n. But the size based on W^A is still unsatisfactory across the five distributions. Size results based on LM statistics are much more encouraging, with LM^G better than W^G and W^A in all five distributions. Our adaptive LM^A improves size even further, with the exception of $n = 96$ in (c) and (e), and for $L = 1$ in (b) and (a) for $n = 96$. In (b) and (d), sizes tend to improve with increasing n for all LM tests, while there is no clear pattern in (a), (c) and (e). In (a) and (b), sizes are best for larger L, and for (c) and (d) it is often best for smaller L, with the exception of (c) $n = 198$. In all cases but two ((e) $n = 96$, (d) $n = 392$), our adaptive LM^A had the best size of the four.

[Table 5 about here]

In Tables 6 and 7, we report Monte Carlo power for nominal sizes $s = 0.1, 0.05, 0.01$ when there is mild spatial correlation $\lambda_0 = 0.1, 0.2$, respectively. In Table 6, W^G has worst power, which improves only slightly with increasing n. Our adaptive estimate improves power, dramatically in (a), (b) and (e), and mildly in (c) and (d). In all cases larger L and n improve power further. In (a) and (b), LM^G has worse power than W^A, while this is not necessarily the case in (c), (d) and (e). In (a), (b) and (c) our LM^A has best power which improves with increasing L. In (d), while LM^A still reports the best power results, there is less clear pattern on the best choice of L. In (e), W^A has somewhat better power than LM^A and power increases with L in both statistics. It is notable that the power of LM^G and W^G remain much the same across the five distributions for given n, while the adaptive statistics W^A and LM^A report greatest power in (a), followed by (b). In Table 7, naturally the reported Monte Carlo power is greater than in Table 6. Patterns similar to Table 6 are observed, except that W^A occasionally has slightly better power than LM^A. It is remarkable that the improvements from using adaptive statistics LM^A and W^A are so great that even for modest $\lambda = 0.2$, powers are close to 1 in (a) for $n = 198, 392$.

[Tables 6-7 about here]

Results for the SEM (3.4) are reported in Tables 8-10, with one dimensional regressor X_i generated from a uniform distribution on $[0, 1]$, and $\beta_0 = 1$. β_0 was estimated by the LSE. Table 8 reports Monte Carlo size. For (a) and (c) with $n = 198, 392$, sizes for all four statistics are better than in the pure case of Table 5, with the LM statistics in particular having sizes much closer to the nominal ones. In (b), (d) and (e) there is no such clear pattern. The relative performance of the four statistics remains unchanged from Table 5. In terms of best choices of L, there are changes in that L_1 performs best in (b), $n = 198$, and in (c) larger L_2 and L_3 now produce better size results than L_1, and in (d), $n = 198$, L_3 led to better results than L_1. In Tables 9 and 10, powers are reported, the powers of all statistics under (a) being somewhat smaller than under pure SAR, while in other distributions they are similar to Tables 6 and 7. The relative power performance of the four statistics reported from Tables 6-7 continues to hold in the SEM case.

[Tables 8-10 about here]

6. Empirical Application

In this section, we apply our adaptive estimation and testing procedure to a cross-sectional data of property crime rates in 103 Italian provinces. The data are from Buonanno, Montolio and Vanin (2009), who studied effects of social capital on crime rates. Their data contain (report-rate-adjusted) crime rates (z) for three crimes, Robbery, Thefts and Car thefts, four different measures of social capital (SC), and a set of demographic, socioeconomic and geographical controls (DSG), so that $X = (SC, DSG)$. In order to account for possible spatial spillovers of crime across the provinces, Buonanno *et al.* (2009) fitted the MRSAR with three different choices of weight matrix W, one based on the inverse of road travel distance between the capital cities in each province, one based on the inverse of Euclidean distance between their geographic coordinates, and one based on simple contiguity among provinces. Buonanno *et al.* (2009) obtained bootstrapped regionally clustered standard errors for the coefficient estimates and finds the spatial lag coefficient of the MRSAR to be insignificant in all but one of 12 regressions (and significant only at 10 percent level in that one instance).

We focus on the number of blood donations per 100,000 inhabitants (Blood) as the measure of social capital, since it is the least likely to suffer from endogeneity out of the four measures of Buonanno *et al.* (2009) as pointed out by the authors, others being the numbers of recreational and voluntary associations per 100,000 inhabitants and referenda turnout. Because estimates of the spatial lag coefficient are insignificant for all three choices of W in MRSAR, we drop the spatial lag term and instead fit the SEM of (3.4) with (1.1) and test $H_0 : \lambda_0 = 0$. Table 11 reports the Gaussian and adaptive LM and Wald statistics when using the road-traveling distance W, for which Buonanno *et al.* (2009) reported estimation results in their Tables 3-5. In the adaptive tests, we used (ii) in (5.1) for ϕ and $L = 3, 4$ and 5. For the other two W, the test results are unchanged and not reported here. We reject $H_0 : \lambda_0 = 0$ for Robbery and Car thefts, while for Theft, LM statistics fail to reject H_0 and Wald statistics reject at 10% significance level. Hence, we note the importance of accounting for error spatial dependence in the standard errors of regression coefficient estimates, especially for Car thefts and Robbery.

[Table 11 about here]

Buonanno *et al.* (2009) also acknowledged the need to allow for possible error correlation and obtained bootstrapped regionally clustered standard errors for the coefficient estimates in their MRSAR model. The controls DSG include income (GDP), unemployment rate (Unemployment), education (High School), urbanization rate (Urbanization), share of youth (Youth), length of judicial proceedings (Length), crime-specific clear-up rates (Clear Up), a measure of criminal association (Criminal Networks) and geographic dummies, details of which can be found in the appendix of Buonanno *et al.* (2009).

We fitted the SEM model (3.4) with SAR error correlation, and compare the coefficient estimates and their significance with what Buonanno *et al.* (2009) had obtained in MRSAR with regionally clustered standard errors. In Table 12 we report estimation results for the coefficients and standard errors with the road-traveling distance W, and corresponding estimates and standard errors reported in Tables 3-5 of Buonanno *et al.* (2009) for MRSAR with the same W. Standard errors obtained with the other two W are very similar and do not affect significance. For Theft, again the standard error remain much the same.

[Table 12 about here]

Across the two models, the signs of significant coefficients are the same, although magnitude or significance vary somewhat for Length, Urbanization and Clear Up. For the coefficient of the social capital measure, Blood, which was the main interest of Buonanno *et al.* (2009), the estimates and significance are remarkably stable across the two models. Urbanization and Clear Up are the two most significant controls across all three crime types. Our SEM tends to find more controls significant. For Theft, Youth and High School are additionally significant, while for Robbery, Unemployment is the additionally significant control.

This is natural as the presence the spatial lag term Wz in the MRSAR would have taken on some explanatory power of these controls.

TABLE 1. SAR, Relative Monte Carlo Variance $Var(\hat{\lambda})/Var(\tilde{\lambda})$

	n	96			n	198			n	392		
ϕ	$L\backslash\lambda$	0.2	0.4	0.8	$L\backslash\lambda$	0.2	0.4	0.8	$L\backslash\lambda$	0.2	0.4	0.8
(a)(i)	1	2.153	1.348	1.000	3	1.295	0.809	0.614	4	1.226	0.673	0.644
	3	1.495	0.874	0.602	5	0.792	0.453	0.357	6	0.743	0.373	0.367
	5	0.816	0.486	0.312	7	0.571	0.316	0.228	8	0.476	0.237	0.215
(ii)	1	1.502	0.854	0.387	3	0.292	0.143	0.100	4	0.230	0.106	0.097
	3	0.334	0.209	0.122	5	0.213	0.103	0.061	6	0.170	0.081	0.059
	5	0.263	0.168	0.092	7	0.199	0.098	0.060	8	0.157	0.077	0.055
(b)(i)	1	2.155	1.366	1.000	3	0.541	0.270	0.205	4	0.504	0.256	0.217
	3	0.652	0.376	0.235	5	0.507	0.245	0.182	6	0.448	0.234	0.191
	5	0.629	0.376	0.237	7	0.503	0.261	0.203	8	0.413	0.236	0.203
(ii)	1	1.880	1.354	1.564	3	0.468	0.225	0.152	4	0.403	0.209	0.171
	3	0.545	0.322	0.161	5	0.473	0.229	0.160	6	0.397	0.213	0.173
	5	0.556	0.334	0.177	7	0.467	0.240	0.172	8	0.402	0.216	0.177
(c)(i)	1	2.183	1.310	1.000	3	1.827	1.084	0.882	4	1.577	0.935	0.880
	3	2.061	1.194	0.901	5	1.694	0.997	0.845	6	1.466	0.885	0.826
	5	1.879	1.091	0.845	7	1.740	1.075	0.948	8	1.394	0.899	0.846
(ii)	1	2.066	1.171	0.818	3	1.592	0.917	0.767	4	1.342	0.804	0.756
	3	1.849	1.058	0.747	5	1.593	0.943	0.806	6	1.329	0.825	0.742
	5	1.850	1.092	0.791	7	1.570	0.959	0.826	8	1.291	0.837	0.761
(d)(i)	1	2.268	1.323	1.000	3	1.734	1.100	0.929	4	1.637	0.941	0.920
	3	2.181	1.236	0.953	5	1.702	1.114	0.933	6	1.611	0.946	0.916
	5	2.111	1.259	0.971	7	1.650	1.216	1.010	8	1.573	1.028	0.974
(ii)	1	2.168	1.237	0.920	3	1.696	1.096	0.916	4	1.655	0.953	0.905
	3	2.198	1.226	0.942	5	1.704	1.087	0.961	6	1.609	0.963	0.936
	5	2.157	1.206	0.962	7	1.700	1.132	1.011	8	1.610	1.001	0.967
(e)(i)	1	1.000	1.000	1.000	3	0.595	0.581	0.568	4	0.508	0.449	0.434
	3	0.554	0.622	0.559	5	0.470	0.478	0.453	6	0.430	0.401	0.361
	5	0.552	0.58	0.497	7	0.525	0.543	0.530	8	0.506	0.490	0.405
(ii)	1	0.908	0.945	1.012	3	0.622	0.646	0.610	4	0.535	0.492	0.479
	3	0.633	0.686	0.607	5	0.533	0.533	0.491	6	0.456	0.433	0.399
	5	0.575	0.619	0.535	7	0.504	0.491	0.497	8	0.437	0.407	0.376

Table 2. SMA, Relative Monte Carlo Variance $Var(\hat{\lambda})/Var(\tilde{\lambda})$

	n	96			n	198			n	392		
ϕ	$L\backslash\lambda$	0.2	0.4	0.8	$L\backslash\lambda$	0.2	0.4	0.8	$L\backslash\lambda$	0.2	0.4	0.8
(a)(i)	1	1.224	1.011	1.000	3	1.361	1.124	0.635	4	0.800	0.663	0.679
	3	1.310	1.094	0.639	5	0.543	0.432	0.471	6	0.506	0.411	0.436
	5	0.529	0.446	0.412	7	0.409	0.319	0.347	8	0.327	0.294	0.293
(ii)	1	0.584	0.520	0.672	3	0.255	0.194	0.207	4	0.198	0.172	0.167
	3	0.274	0.243	0.213	5	0.183	0.162	0.172	6	0.151	0.136	0.130
	5	0.226	0.204	0.187	7	0.178	0.158	0.163	8	0.147	0.128	0.120
(b)(i)	1	1.273	1.004	1.000	3	1.122	1.045	0.856	4	0.386	0.298	0.324
	3	1.116	1.042	0.852	5	0.348	0.277	0.306	6	0.345	0.277	0.299
	5	0.450	0.345	0.352	7	0.381	0.298	0.317	8	0.332	0.279	0.301
(ii)	1	1.842	1.402	0.969	3	0.320	0.263	0.274	4	0.313	0.255	0.277
	3	0.383	0.294	0.281	5	0.328	0.269	0.275	6	0.312	0.261	0.276
	5	0.404	0.302	0.290	7	0.338	0.273	0.288	8	0.317	0.263	0.274
(c)(i)	1	1.221	1.006	1.000	3	1.461	1.162	0.524	4	0.978	0.899	0.915
	3	1.424	1.138	0.518	5	1.067	0.825	0.896	6	0.929	0.868	0.858
	5	1.155	0.893	0.914	7	1.181	0.919	0.976	8	0.938	0.889	0.874
(ii)	1	1.084	0.866	0.927	3	1.017	0.783	0.805	4	0.873	0.807	0.792
	3	1.107	0.857	0.838	5	1.000	0.821	0.839	6	0.842	0.818	0.813
	5	1.116	0.890	0.878	7	1.023	0.866	0.863	8	0.852	0.831	0.830
(d)(i)	1	1.210	1.024	1.000	3	1.486	1.159	0.517	4	1.064	0.901	0.931
	3	1.426	1.138	0.518	5	1.133	0.939	0.964	6	1.066	0.920	0.942
	5	1.230	1.034	1.001	7	1.208	1.077	1.058	8	1.134	0.998	1.013
(ii)	1	1.139	0.976	0.937	3	1.106	0.915	0.962	4	1.076	0.916	0.945
	3	1.199	0.983	0.964	5	1.144	0.927	0.993	6	1.066	0.943	0.980
	5	1.229	1.023	0.994	7	1.194	1.012	1.051	8	1.090	0.987	1.017
(e)(i)	1	1.174	1.069	0.771	3	1.533	1.365	1.028	4	1.715	1.485	1.080
	3	1.501	1.366	1.116	5	1.581	1.410	1.133	6	1.750	1.514	1.144
	5	1.513	1.371	1.185	7	1.492	1.356	1.125	8	1.607	1.431	1.152
(ii)	1	1.227	1.101	0.800	3	1.488	1.335	0.999	4	1.718	1.466	1.059
	3	1.470	1.329	1.077	5	1.552	1.380	1.101	6	1.718	1.494	1.123
	5	1.496	1.372	1.166	7	1.531	1.395	1.115	8	1.712	1.486	1.153

Table 3. MESS, Relative Monte Carlo Variance $Var(\hat{\lambda})/Var(\tilde{\lambda})$

	n	96			n	198			n	392		
ϕ	$L\backslash\lambda$	1	2	3	$L\backslash\lambda$	1	2	3	$L\backslash\lambda$	1	2	3
(a)(i)	1	1.000	1.000	1.000	3	0.675	0.686	0.676	4	0.690	0.644	0.694
	3	0.666	0.660	0.658	5	0.429	0.404	0.433	6	0.423	0.380	0.431
	5	0.408	0.405	0.383	7	0.307	0.286	0.300	8	0.263	0.247	0.272
(ii)	1	0.522	0.522	0.508	3	0.167	0.142	0.154	4	0.138	0.121	0.129
	3	0.177	0.185	0.174	5	0.118	0.106	0.113	6	0.101	0.092	0.092
	5	0.147	0.148	0.138	7	0.114	0.104	0.111	8	0.097	0.089	0.087
(b)(i)	1	1.000	1.000	1.000	3	0.273	0.268	0.284	4	0.525	0.456	0.525
	3	0.332	0.304	0.320	5	0.257	0.242	0.260	6	0.494	0.438	0.486
	5	0.339	0.308	0.321	7	0.292	0.268	0.276	8	0.535	0.469	0.549
(ii)	1	1.290	1.297	1.298	3	0.227	0.222	0.225	4	0.239	0.221	0.235
	3	0.263	0.248	0.245	5	0.236	0.228	0.231	6	0.239	0.226	0.236
	5	0.287	0.262	0.258	7	0.246	0.241	0.245	8	0.246	0.230	0.238
(c)(i)	1	1.000	1.000	1.000	3	0.924	0.889	0.908	4	0.892	0.894	0.906
	3	0.942	0.919	0.923	5	0.884	0.831	0.886	6	0.839	0.856	0.855
	5	0.935	0.875	0.903	7	1.003	0.923	0.978	8	0.844	0.876	0.874
(ii)	1	0.887	0.857	0.861	3	0.829	0.777	0.808	4	0.778	0.790	0.790
	3	0.873	0.838	0.825	5	0.831	0.801	0.831	6	0.768	0.803	0.791
	5	0.908	0.872	0.857	7	0.862	0.834	0.858	8	0.781	0.819	0.813
(d)(i)	1	1.000	1.000	1.000	3	0.936	0.923	0.947	4	0.915	0.891	0.936
	3	0.966	0.949	0.951	5	0.945	0.946	0.958	6	0.923	0.905	0.939
	5	1.032	0.999	0.997	7	1.069	1.074	1.041	8	1.006	0.982	1.005
(ii)	1	0.938	0.935	0.924	3	0.924	0.920	0.942	4	0.927	0.903	0.934
	3	0.984	0.947	0.959	5	0.957	0.924	0.981	6	0.923	0.924	0.968
	5	1.021	0.970	0.990	7	1.008	0.995	1.038	8	0.952	0.966	1.000
(e)(i)	1	1.000	1.000	1.000	3	0.651	0.629	0.641	4	0.546	0.503	0.519
	3	0.610	0.648	0.634	5	0.512	0.535	0.513	6	0.477	0.452	0.449
	5	0.585	0.605	0.561	7	0.549	0.576	0.564	8	0.547	0.516	0.493
(ii)	1	0.977	0.990	0.989	3	0.670	0.683	0.671	4	0.594	0.557	0.559
	3	0.671	0.706	0.670	5	0.574	0.584	0.557	6	0.505	0.488	0.482
	5	0.614	0.643	0.595	7	0.544	0.548	0.550	8	0.479	0.459	0.458

Table 4. Relative Monte Carlo MSE $MSE(\hat{\lambda})/MSE(\tilde{\lambda})$, $\phi = (ii)$

	n	96			n	198			n	392		
	$L\backslash\lambda$	0.2	0.4	0.8	$L\backslash\lambda$	0.2	0.4	0.8	$L\backslash\lambda$	0.2	0.4	0.8
SAR (a)	1	1.497	0.706	0.317	3	0.267	0.125	0.095	4	0.215	0.094	0.094
	3	0.334	0.176	0.111	5	0.196	0.086	0.052	6	0.156	0.070	0.049
	5	0.260	0.138	0.076	7	0.183	0.082	0.052	8	0.145	0.067	0.047
(b)	1	2.438	1.804	2.371	3	0.449	0.196	0.138	4	0.390	0.185	0.156
	3	0.532	0.285	0.144	5	0.459	0.206	0.148	6	0.384	0.189	0.159
	5	0.540	0.299	0.159	7	0.453	0.224	0.169	8	0.388	0.193	0.164
(c)	1	2.184	1.099	0.742	3	1.585	0.890	0.767	4	1.378	0.796	0.751
	3	1.924	1.013	0.721	5	1.569	0.902	0.776	6	1.348	0.794	0.712
	5	1.904	1.025	0.743	7	1.531	0.914	0.797	8	1.307	0.815	0.735
(d)	1	2.345	1.190	0.902	3	1.771	1.072	0.895	4	1.732	0.930	0.888
	3	2.344	1.180	0.933	5	1.763	1.063	0.943	6	1.678	0.939	0.918
	5	2.274	1.150	0.958	7	1.725	1.096	0.986	8	1.663	0.975	0.949
(e)	1	0.907	0.957	1.029	3	0.589	0.604	0.565	4	0.501	0.460	0.463
	3	0.593	0.632	0.563	5	0.501	0.498	0.457	6	0.427	0.401	0.369
	5	0.532	0.566	0.494	7	0.474	0.460	0.464	8	0.411	0.376	0.345
SMA (a)	1	0.526	0.463	0.634	3	0.271	0.246	0.286	4	0.228	0.220	0.231
	3	0.318	0.300	0.306	5	0.199	0.189	0.206	6	0.173	0.162	0.149
	5	0.261	0.243	0.237	7	0.197	0.188	0.203	8	0.171	0.158	0.150
(b)	1	1.929	1.808	1.934	3	0.337	0.297	0.314	4	0.331	0.277	0.309
	3	0.396	0.359	0.332	5	0.343	0.303	0.316	6	0.330	0.283	0.310
	5	0.414	0.367	0.339	7	0.354	0.309	0.329	8	0.335	0.285	0.309
(c)	1	0.995	0.838	0.854	3	0.957	0.804	0.841	4	0.874	0.839	0.820
	3	1.032	0.876	0.848	5	0.934	0.826	0.843	6	0.835	0.825	0.810
	5	1.043	0.892	0.861	7	0.956	0.868	0.870	8	0.848	0.848	0.833
(d)	1	1.063	0.956	0.934	3	1.052	0.918	0.953	4	1.051	0.910	0.940
	3	1.114	0.969	0.970	5	1.086	0.933	0.990	6	1.043	0.937	0.974
	5	1.146	1.004	1.009	7	1.124	1.006	1.042	8	1.064	0.981	1.012
(e)	1	1.449	0.974	0.658	3	1.501	1.250	0.884	4	1.700	1.391	0.957
	3	1.471	1.190	0.962	5	1.531	1.308	1.008	6	1.689	1.431	1.037
	5	1.466	1.266	1.079	7	1.503	1.339	1.041	8	1.682	1.436	1.079
MESS (a)	1	0.529	0.527	0.519	3	0.155	0.140	0.152	4	0.134	0.119	0.126
	3	0.177	0.177	0.173	5	0.113	0.105	0.116	6	0.101	0.094	0.092
	5	0.149	0.145	0.140	7	0.110	0.104	0.114	8	0.096	0.092	0.087
(b)	1	1.281	1.279	1.286	3	0.222	0.221	0.225	4	0.237	0.216	0.235
	3	0.255	0.255	0.245	5	0.230	0.227	0.230	6	0.237	0.221	0.237
	5	0.278	0.271	0.259	7	0.241	0.242	0.247	8	0.244	0.226	0.241
(c)	1	0.880	0.859	0.848	3	0.791	0.766	0.798	4	0.770	0.789	0.774
	3	0.844	0.821	0.785	5	0.794	0.791	0.822	6	0.761	0.795	0.778
	5	0.881	0.848	0.814	7	0.818	0.815	0.839	8	0.773	0.815	0.795
(d)	1	0.930	0.921	0.927	3	0.919	0.919	0.935	4	0.933	0.896	0.933
	3	0.965	0.930	0.961	5	0.945	0.921	0.971	6	0.925	0.911	0.958
	5	0.998	0.943	0.992	7	0.982	0.978	1.017	8	0.947	0.950	0.988
(e)	1	0.968	0.984	0.980	3	0.655	0.662	0.649	4	0.574	0.535	0.538
	3	0.650	0.680	0.646	5	0.566	0.571	0.547	6	0.494	0.473	0.460
	5	0.592	0.621	0.579	7	0.537	0.538	0.544	8	0.473	0.446	0.437

TABLE 5. Size of test of $H_0 : \lambda = 0$, SAR, $\phi = (ii)$

		$n = 96$			$n = 198$			$n = 392$		
	s	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(a)	LM^G	0.064	0.023	0.006	0.063	0.031	0.01	0.069	0.033	0.007
	$LM^A(L_1)$	0.147	0.062	0.015	0.074	0.044	0.008	0.067	0.035	0.008
	$LM^A(L_2)$	0.083	0.04	0.011	0.094	0.05	0.011	0.082	0.045	0.016
	$LM^A(L_3)$	0.091	0.052	0.012	0.089	0.047	0.013	0.081	0.047	0.012
	W^G	0.03	0.01	0.004	0.026	0.009	0.002	0.024	0.014	0.003
	$W^A(L_1)$	0.037	0.015	0	0.068	0.039	0.009	0.054	0.021	0.005
	$W^A(L_2)$	0.045	0.015	0.003	0.056	0.023	0.008	0.056	0.02	0.002
	$W^A(L_3)$	0.054	0.028	0.004	0.07	0.031	0.008	0.059	0.027	0.002
(b)	LM^G	0.062	0.03	0.012	0.077	0.025	0.009	0.082	0.034	0.009
	$LM^A(L_1)$	0.019	0.01	0.001	0.079	0.04	0.01	0.093	0.049	0.016
	$LM^A(L_2)$	0.078	0.036	0.012	0.084	0.042	0.01	0.09	0.048	0.014
	$LM^A(L_3)$	0.083	0.043	0.011	0.084	0.043	0.014	0.1	0.048	0.012
	W^G	0.031	0.018	0.005	0.025	0.015	0.003	0.033	0.015	0.002
	$W^A(L_1)$	0.017	0.008	0.001	0.048	0.021	0.004	0.068	0.031	0.007
	$W^A(L_2)$	0.047	0.017	0.003	0.049	0.022	0.005	0.063	0.035	0.009
	$W^A(L_3)$	0.059	0.027	0.008	0.052	0.03	0.008	0.068	0.042	0.009
(c)	LM^G	0.057	0.024	0.007	0.053	0.017	0.009	0.064	0.022	0.004
	$LM^A(L_1)$	0.083	0.034	0.014	0.054	0.018	0.009	0.064	0.024	0.004
	$LM^A(L_2)$	0.071	0.035	0.013	0.058	0.024	0.008	0.06	0.02	0.002
	$LM^A(L_3)$	0.073	0.035	0.007	0.06	0.026	0.01	0.047	0.019	0.003
	W^G	0.027	0.014	0.002	0.017	0.011	0.003	0.018	0.006	0
	$W^A(L_1)$	0.036	0.02	0.002	0.032	0.017	0.005	0.029	0.01	0.001
	$W^A(L_2)$	0.045	0.025	0.007	0.036	0.018	0.006	0.031	0.012	0.001
	$W^A(L_3)$	0.048	0.035	0.012	0.055	0.022	0.006	0.031	0.015	0.003
(d)	LM^G	0.063	0.018	0.006	0.068	0.034	0.013	0.08	0.03	0.01
	$LM^A(L_1)$	0.068	0.026	0.008	0.074	0.031	0.009	0.078	0.03	0.007
	$LM^A(L_2)$	0.064	0.024	0.009	0.073	0.028	0.006	0.074	0.031	0.008
	$LM^A(L_3)$	0.064	0.027	0.009	0.068	0.031	0.008	0.07	0.028	0.007
	W^G	0.022	0.012	0.001	0.036	0.018	0.004	0.027	0.017	0.004
	$W^A(L_1)$	0.032	0.014	0.001	0.032	0.015	0.005	0.031	0.016	0.003
	$W^A(L_2)$	0.038	0.023	0.004	0.036	0.021	0.005	0.033	0.017	0.006
	$W^A(L_3)$	0.051	0.029	0.013	0.058	0.03	0.01	0.041	0.021	0.007
(e)	LM^G	0.067	0.031	0.006	0.082	0.037	0.014	0.067	0.024	0.006
	$LM^A(L_1)$	0.069	0.024	0.005	0.083	0.033	0.008	0.069	0.039	0.009
	$LM^A(L_2)$	0.07	0.029	0.015	0.098	0.039	0.007	0.077	0.042	0.009
	$LM^A(L_3)$	0.068	0.028	0.012	0.089	0.041	0.011	0.081	0.04	0.009
	W^G	0.03	0.012	0.001	0.026	0.015	0.002	0.044	0.023	0.007
	$W^A(L_1)$	0.028	0.015	0.002	0.034	0.013	0.003	0.063	0.035	0.01
	$W^A(L_2)$	0.042	0.023	0.006	0.045	0.017	0.007	0.062	0.033	0.012
	$W^A(L_3)$	0.052	0.028	0.011	0.05	0.023	0.005	0.075	0.035	0.013

$(L_1, L_2, L_3) = (1, 3, 5), (3, 5, 7), (4, 6, 8)$ for $n = 96, 198, 392$, respectively.

Table 6. Power of test of $H_0 : \lambda = 0$ when $\lambda_0 = 0.1$, SAR, $\phi = (ii)$

		$n = 96$			$n = 198$			$n = 392$		
	s	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(a)	LM^G	0.116	0.067	0.025	0.115	0.068	0.033	0.163	0.103	0.05
	$LM^A(L_1)$	0.208	0.149	0.09	0.431	0.351	0.2	0.527	0.429	0.288
	$LM^A(L_2)$	0.341	0.265	0.141	0.563	0.472	0.33	0.677	0.594	0.42
	$LM^A(L_3)$	0.44	0.363	0.242	0.602	0.5	0.35	0.7	0.61	0.442
	W^G	0.082	0.044	0.014	0.092	0.059	0.026	0.116	0.065	0.025
	$W^A(L_1)$	0.137	0.093	0.033	0.376	0.298	0.14	0.475	0.374	0.206
	$W^A(L_2)$	0.291	0.207	0.098	0.518	0.411	0.249	0.642	0.533	0.328
	$W^A(L_3)$	0.39	0.299	0.169	0.561	0.458	0.271	0.669	0.561	0.337
(b)	LM^G	0.119	0.067	0.029	0.141	0.092	0.049	0.161	0.104	0.045
	$LM^A(L_1)$	0.044	0.026	0.008	0.308	0.234	0.124	0.388	0.306	0.16
	$LM^A(L_2)$	0.218	0.159	0.073	0.306	0.233	0.132	0.378	0.304	0.167
	$LM^A(L_3)$	0.209	0.15	0.069	0.296	0.232	0.118	0.381	0.302	0.165
	W^G	0.075	0.046	0.014	0.098	0.067	0.021	0.113	0.068	0.022
	$W^A(L_1)$	0.035	0.021	0.007	0.267	0.195	0.082	0.343	0.253	0.115
	$W^A(L_2)$	0.191	0.123	0.048	0.281	0.199	0.089	0.345	0.265	0.124
	$W^A(L_3)$	0.197	0.13	0.053	0.301	0.225	0.101	0.372	0.281	0.132
(c)	LM^G	0.11	0.071	0.032	0.156	0.099	0.05	0.162	0.099	0.04
	$LM^A(L_1)$	0.148	0.083	0.041	0.162	0.103	0.061	0.156	0.108	0.058
	$LM^A(L_2)$	0.138	0.094	0.041	0.171	0.105	0.061	0.164	0.119	0.062
	$LM^A(L_3)$	0.142	0.088	0.047	0.17	0.116	0.058	0.167	0.114	0.062
	W^G	0.077	0.046	0.012	0.103	0.067	0.026	0.111	0.061	0.017
	$W^A(L_1)$	0.095	0.059	0.018	0.12	0.079	0.029	0.125	0.086	0.036
	$W^A(L_2)$	0.105	0.067	0.023	0.148	0.104	0.047	0.14	0.1	0.049
	$W^A(L_3)$	0.131	0.085	0.04	0.149	0.114	0.051	0.156	0.107	0.054
(d)	LM^G	0.115	0.066	0.029	0.155	0.098	0.046	0.161	0.105	0.044
	$LM^A(L_1)$	0.118	0.079	0.041	0.158	0.102	0.051	0.173	0.11	0.056
	$LM^A(L_2)$	0.12	0.078	0.035	0.142	0.097	0.043	0.177	0.116	0.051
	$LM^A(L_3)$	0.114	0.071	0.037	0.134	0.091	0.046	0.173	0.112	0.051
	W^G	0.072	0.046	0.013	0.099	0.068	0.022	0.115	0.07	0.024
	$W^A(L_1)$	0.086	0.053	0.026	0.115	0.082	0.021	0.143	0.083	0.027
	$W^A(L_2)$	0.093	0.062	0.029	0.122	0.084	0.035	0.148	0.095	0.039
	$W^A(L_3)$	0.116	0.073	0.036	0.138	0.106	0.056	0.171	0.11	0.044
(e)	LM^G	0.116	0.068	0.034	0.136	0.097	0.047	0.157	0.112	0.056
	$LM^A(L_1)$	0.117	0.069	0.027	0.181	0.121	0.059	0.21	0.152	0.081
	$LM^A(L_2)$	0.153	0.101	0.05	0.194	0.133	0.061	0.223	0.159	0.083
	$LM^A(L_3)$	0.153	0.108	0.047	0.199	0.141	0.066	0.22	0.16	0.088
	W^G	0.088	0.056	0.017	0.111	0.074	0.021	0.107	0.069	0.024
	$W^A(L_1)$	0.087	0.046	0.014	0.133	0.085	0.029	0.2	0.124	0.045
	$W^A(L_2)$	0.144	0.1	0.04	0.179	0.122	0.048	0.21	0.155	0.066
	$W^A(L_3)$	0.172	0.127	0.059	0.207	0.15	0.073	0.232	0.176	0.088

$(L_1, L_2, L_3) = (1, 3, 5), (3, 5, 7), (4, 6, 8)$ for $n = 96, 198, 392$, respectively.

TABLE 7. Power of test of $H_0 : \lambda = 0$ when $\lambda_0 = 0.2$, SAR, $\phi = (ii)$

		$n = 96$			$n = 198$			$n = 392$		
	s	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(a)	LM^G	0.266	0.187	0.1	0.294	0.229	0.132	0.321	0.265	0.174
	$LM^A(L_1)$	0.441	0.376	0.279	0.871	0.825	0.705	0.936	0.908	0.822
	$LM^A(L_2)$	0.759	0.696	0.55	0.934	0.901	0.816	0.969	0.959	0.917
	$LM^A(L_3)$	0.849	0.797	0.691	0.943	0.921	0.848	0.98	0.966	0.928
	W^G	0.206	0.131	0.062	0.259	0.174	0.081	0.277	0.188	0.086
	$W^A(L_1)$	0.399	0.318	0.182	0.852	0.791	0.66	0.937	0.909	0.818
	$W^A(L_2)$	0.753	0.674	0.489	0.927	0.9	0.833	0.979	0.968	0.915
	$W^A(L_3)$	0.866	0.803	0.671	0.941	0.916	0.852	0.988	0.973	0.935
(b)	LM^G	0.263	0.209	0.125	0.297	0.238	0.146	0.348	0.269	0.17
	$LM^A(L_1)$	0.132	0.087	0.031	0.684	0.62	0.493	0.796	0.741	0.605
	$LM^A(L_2)$	0.59	0.512	0.384	0.693	0.616	0.502	0.791	0.742	0.607
	$LM^A(L_3)$	0.584	0.514	0.383	0.689	0.609	0.491	0.797	0.746	0.594
	W^G	0.22	0.161	0.074	0.242	0.173	0.078	0.283	0.214	0.106
	$W^A(L_1)$	0.108	0.06	0.022	0.675	0.596	0.438	0.795	0.718	0.552
	$W^A(L_2)$	0.569	0.492	0.329	0.691	0.606	0.452	0.795	0.725	0.555
	$W^A(L_3)$	0.593	0.516	0.36	0.696	0.616	0.476	0.807	0.745	0.567
(c)	LM^G	0.229	0.17	0.092	0.282	0.216	0.132	0.357	0.283	0.162
	$LM^A(L_1)$	0.276	0.211	0.13	0.351	0.284	0.167	0.426	0.344	0.226
	$LM^A(L_2)$	0.282	0.218	0.134	0.364	0.282	0.172	0.416	0.348	0.238
	$LM^A(L_3)$	0.29	0.219	0.132	0.356	0.274	0.165	0.426	0.339	0.236
	W^G	0.179	0.121	0.052	0.233	0.172	0.075	0.292	0.218	0.107
	$W^A(L_1)$	0.218	0.156	0.076	0.291	0.216	0.12	0.365	0.28	0.161
	$W^A(L_2)$	0.248	0.181	0.097	0.321	0.242	0.129	0.388	0.299	0.182
	$W^A(L_3)$	0.269	0.207	0.12	0.346	0.274	0.157	0.402	0.333	0.202
(d)	LM^G	0.235	0.18	0.095	0.303	0.236	0.139	0.362	0.294	0.183
	$LM^A(L_1)$	0.256	0.196	0.113	0.33	0.252	0.162	0.38	0.316	0.209
	$LM^A(L_2)$	0.252	0.189	0.11	0.321	0.247	0.147	0.373	0.312	0.206
	$LM^A(L_3)$	0.234	0.185	0.111	0.309	0.243	0.142	0.364	0.306	0.204
	W^G	0.193	0.13	0.058	0.249	0.181	0.088	0.312	0.231	0.115
	$W^A(L_1)$	0.197	0.139	0.058	0.287	0.207	0.116	0.343	0.271	0.141
	$W^A(L_2)$	0.214	0.156	0.074	0.294	0.215	0.125	0.355	0.287	0.158
	$W^A(L_3)$	0.231	0.179	0.093	0.325	0.25	0.142	0.367	0.294	0.179
(e)	LM^G	0.235	0.178	0.098	0.274	0.195	0.113	0.342	0.268	0.157
	$LM^A(L_1)$	0.247	0.19	0.103	0.372	0.306	0.183	0.497	0.404	0.283
	$LM^A(L_2)$	0.31	0.253	0.154	0.406	0.333	0.212	0.533	0.46	0.315
	$LM^A(L_3)$	0.32	0.26	0.156	0.422	0.343	0.215	0.545	0.468	0.333
	W^G	0.181	0.125	0.049	0.254	0.179	0.089	0.284	0.205	0.107
	$W^A(L_1)$	0.186	0.121	0.047	0.374	0.292	0.162	0.462	0.355	0.207
	$W^A(L_2)$	0.279	0.208	0.109	0.432	0.339	0.203	0.53	0.43	0.264
	$W^A(L_3)$	0.35	0.27	0.168	0.499	0.412	0.259	0.576	0.497	0.321

$(L_1, L_2, L_3) = (1, 3, 5), (3, 5, 7), (4, 6, 8)$ for $n = 96, 198, 392$, respectively.

TABLE 8. Size of test of $H_0 : \lambda = 0$, SEM, $\phi = (ii)$

		$n=96$			$n=198$			$n=392$		
	s	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(a)	LM^G	0.062	0.024	0.007	0.09	0.041	0.011	0.086	0.038	0.009
	$LM^A(L_1)$	0.142	0.065	0.019	0.102	0.063	0.019	0.085	0.045	0.008
	$LM^A(L_2)$	0.099	0.045	0.015	0.104	0.055	0.016	0.098	0.048	0.011
	$LM^A(L_3)$	0.086	0.041	0.015	0.1	0.055	0.017	0.094	0.044	0.011
	W^G	0.027	0.011	0.001	0.037	0.016	0.003	0.034	0.019	0.003
	$W^A(L_1)$	0.033	0.015	0.002	0.063	0.038	0.006	0.056	0.028	0.003
	$W^A(L_2)$	0.046	0.023	0.009	0.057	0.031	0.007	0.054	0.025	0.002
	$W^A(L_3)$	0.043	0.024	0.007	0.062	0.037	0.009	0.065	0.029	0.003
(b)	LM^G	0.079	0.038	0.016	0.082	0.036	0.012	0.071	0.032	0.005
	$LM^A(L_1)$	0.02	0.014	0.005	0.088	0.039	0.005	0.081	0.038	0.007
	$LM^A(L_2)$	0.082	0.041	0.01	0.084	0.039	0.006	0.08	0.043	0.007
	$LM^A(L_3)$	0.083	0.044	0.008	0.085	0.037	0.01	0.087	0.038	0.01
	W^G	0.043	0.017	0.007	0.031	0.015	0.005	0.029	0.012	0.003
	$W^A(L_1)$	0.018	0.009	0.005	0.04	0.015	0.002	0.045	0.02	0.003
	$W^A(L_2)$	0.053	0.023	0.007	0.05	0.018	0.003	0.044	0.019	0.004
	$W^A(L_3)$	0.057	0.032	0.01	0.05	0.023	0.004	0.05	0.021	0.005
(c)	LM^G	0.056	0.023	0.004	0.069	0.024	0.01	0.077	0.033	0.013
	$LM^A(L_1)$	0.073	0.031	0.012	0.057	0.021	0.007	0.09	0.041	0.011
	$LM^A(L_2)$	0.061	0.035	0.01	0.074	0.028	0.006	0.087	0.046	0.011
	$LM^A(L_3)$	0.075	0.037	0.01	0.073	0.025	0.006	0.092	0.047	0.012
	W^G	0.027	0.012	0.001	0.025	0.015	0.004	0.03	0.018	0.005
	$W^A(L_1)$	0.032	0.016	0.001	0.021	0.01	0.005	0.046	0.024	0.005
	$W^A(L_2)$	0.042	0.025	0.005	0.036	0.018	0.005	0.049	0.033	0.009
	$W^A(L_3)$	0.052	0.031	0.011	0.046	0.021	0.008	0.056	0.036	0.011
(d)	LM^G	0.054	0.023	0.011	0.056	0.017	0.007	0.072	0.027	0.005
	$LM^A(L_1)$	0.069	0.021	0.01	0.066	0.025	0.012	0.084	0.026	0.007
	$LM^A(L_2)$	0.062	0.026	0.011	0.062	0.027	0.01	0.079	0.026	0.005
	$LM^A(L_3)$	0.064	0.026	0.011	0.068	0.03	0.01	0.077	0.027	0.006
	W^G	0.027	0.011	0.006	0.017	0.01	0.001	0.02	0.011	0.003
	$W^A(L_1)$	0.025	0.012	0.005	0.033	0.021	0.007	0.03	0.012	0.004
	$W^A(L_2)$	0.028	0.018	0.007	0.041	0.024	0.007	0.037	0.019	0.005
	$W^A(L_3)$	0.041	0.026	0.011	0.054	0.033	0.01	0.042	0.027	0.005
(e)	LM^G	0.056	0.024	0.009	0.081	0.032	0.009	0.072	0.031	0.009
	$LM^A(L_1)$	0.054	0.022	0.009	0.076	0.024	0.011	0.064	0.027	0.011
	$LM^A(L_2)$	0.073	0.04	0.009	0.073	0.027	0.002	0.067	0.029	0.008
	$LM^A(L_3)$	0.069	0.038	0.009	0.064	0.024	0.003	0.073	0.03	0.006
	W^G	0.03	0.016	0.003	0.041	0.021	0.003	0.023	0.013	0.001
	$W^A(L_1)$	0.031	0.014	0.002	0.042	0.021	0.005	0.047	0.02	0.003
	$W^A(L_2)$	0.04	0.024	0.01	0.048	0.025	0.01	0.056	0.027	0.005
	$W^A(L_3)$	0.052	0.029	0.012	0.061	0.036	0.013	0.075	0.036	0.009

$(L_1, L_2, L_3) = (1,3,5), (3,5,7), (4,6,8)$ for $n = 96, 198, 392$, respectively.

Table 9. Power of test of $H_0 : \lambda = 0$ when $\lambda_0 = 0.1$, SEM, $\phi = (ii)$

		$n = 96$			$n = 198$			$n = 392$		
	s	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(a)	LM^G	0.107	0.061	0.03	0.116	0.077	0.025	0.158	0.117	0.055
	$LM^A(L_1)$	0.221	0.159	0.074	0.413	0.334	0.192	0.517	0.414	0.274
	$LM^A(L_2)$	0.339	0.261	0.145	0.527	0.434	0.287	0.638	0.552	0.395
	$LM^A(L_3)$	0.429	0.351	0.22	0.549	0.456	0.307	0.655	0.576	0.422
	W^G	0.065	0.039	0.01	0.076	0.041	0.011	0.119	0.078	0.026
	$W^A(L_1)$	0.135	0.075	0.029	0.363	0.274	0.142	0.48	0.38	0.195
	$W^A(L_2)$	0.265	0.195	0.096	0.485	0.387	0.225	0.614	0.514	0.307
	$W^A(L_3)$	0.362	0.279	0.159	0.528	0.412	0.251	0.635	0.543	0.347
(b)	LM^G	0.116	0.068	0.025	0.13	0.089	0.044	0.163	0.109	0.046
	$LM^A(L_1)$	0.042	0.023	0.008	0.292	0.212	0.125	0.36	0.281	0.164
	$LM^A(L_2)$	0.224	0.158	0.07	0.293	0.202	0.122	0.353	0.286	0.167
	$LM^A(L_3)$	0.214	0.158	0.077	0.288	0.211	0.115	0.351	0.273	0.156
	W^G	0.076	0.045	0.016	0.093	0.061	0.018	0.114	0.075	0.019
	$W^A(L_1)$	0.033	0.017	0.005	0.258	0.18	0.077	0.315	0.235	0.103
	$W^A(L_2)$	0.182	0.127	0.048	0.261	0.177	0.077	0.315	0.237	0.104
	$W^A(L_3)$	0.209	0.142	0.06	0.281	0.204	0.089	0.335	0.234	0.114
(c)	LM^G	0.106	0.062	0.034	0.154	0.096	0.04	0.18	0.117	0.049
	$LM^A(L_1)$	0.13	0.082	0.04	0.17	0.116	0.052	0.185	0.132	0.06
	$LM^A(L_2)$	0.132	0.083	0.038	0.169	0.119	0.056	0.191	0.142	0.063
	$LM^A(L_3)$	0.146	0.088	0.038	0.162	0.114	0.046	0.182	0.132	0.064
	W^G	0.07	0.048	0.019	0.102	0.062	0.021	0.131	0.077	0.022
	$W^A(L_1)$	0.09	0.054	0.02	0.125	0.087	0.029	0.148	0.091	0.038
	$W^A(L_2)$	0.106	0.059	0.027	0.136	0.092	0.038	0.153	0.103	0.041
	$W^A(L_3)$	0.124	0.077	0.039	0.157	0.096	0.049	0.164	0.11	0.044
(d)	LM^G	0.118	0.078	0.032	0.157	0.115	0.058	0.163	0.115	0.052
	$LM^A(L_1)$	0.122	0.077	0.037	0.164	0.112	0.065	0.171	0.113	0.055
	$LM^A(L_2)$	0.114	0.079	0.038	0.162	0.104	0.063	0.172	0.117	0.049
	$LM^A(L_3)$	0.12	0.081	0.032	0.166	0.105	0.058	0.175	0.112	0.05
	W^G	0.081	0.049	0.015	0.121	0.078	0.037	0.128	0.083	0.022
	$W^A(L_1)$	0.08	0.055	0.019	0.134	0.086	0.039	0.133	0.085	0.03
	$W^A(L_2)$	0.096	0.062	0.024	0.151	0.093	0.045	0.034	0.141	0.09
	$W^A(L_3)$	0.116	0.082	0.03	0.166	0.117	0.063	0.164	0.107	0.043
(e)	LM^G	0.11	0.07	0.026	0.149	0.103	0.042	0.155	0.113	0.052
	$LM^A(L_1)$	0.104	0.065	0.029	0.182	0.127	0.061	0.192	0.144	0.074
	$LM^A(L_2)$	0.139	0.098	0.047	0.203	0.146	0.065	0.212	0.145	0.088
	$LM^A(L_3)$	0.154	0.112	0.048	0.216	0.146	0.066	0.221	0.164	0.083
	W^G	0.074	0.043	0.016	0.104	0.065	0.023	0.115	0.078	0.028
	$W^A(L_1)$	0.079	0.039	0.013	0.144	0.094	0.043	0.18	0.122	0.054
	$W^A(L_2)$	0.131	0.09	0.033	0.167	0.121	0.049	0.217	0.147	0.063
	$W^A(L_3)$	0.164	0.116	0.053	0.206	0.151	0.069	0.246	0.162	0.083

$(L_1, L_2, L_3) = (1, 3, 5), (3, 5, 7), (4, 6, 8)$ for $n = 96, 198, 392$, respectively.

Table 10. Power of test of $H_0 : \lambda = 0$ when $\lambda_0 = 0.2$, SEM, $\phi = (ii)$

		$n = 96$			$n = 198$			$n = 392$		
	s	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
(a)	LM^G	0.243	0.183	0.104	0.2720	0.2160	0.1280	0.33	0.248	0.156
	$LM^A(L_1)$	0.418	0.354	0.252	0.8400	0.7900	0.687	0.91	0.869	0.773
	$LM^A(L_2)$	0.705	0.639	0.507	0.91	0.871	0.79	0.969	0.933	0.878
	$LM^A(L_3)$	0.791	0.733	0.636	0.929	0.892	0.812	0.97	0.95	0.889
	W^G	0.202	0.145	0.059	0.227	0.156	0.069	0.266	0.194	0.094
	$W^A(L_1)$	0.374	0.291	0.168	0.851	0.79	0.654	0.924	0.887	0.788
	$W^A(L_2)$	0.697	0.625	0.464	0.922	0.892	0.814	0.972	0.955	0.899
	$W^A(L_3)$	0.802	0.741	0.623	0.941	0.917	0.833	0.976	0.963	0.916
(b)	LM^G	0.236	0.175	0.097	0.332	0.26	0.162	0.31	0.253	0.162
	$LM^A(L_1)$	0.103	0.071	0.025	0.725	0.662	0.536	0.788	0.709	0.59
	$LM^A(L_2)$	0.573	0.505	0.372	0.728	0.665	0.536	0.782	0.714	0.585
	$LM^A(L_3)$	0.563	0.488	0.366	0.718	0.654	0.512	0.783	0.72	0.589
	W^G	0.193	0.125	0.061	0.284	0.206	0.092	0.266	0.196	0.103
	$W^A(L_1)$	0.083	0.045	0.018	0.711	0.634	0.477	0.775	0.711	0.554
	$W^A(L_2)$	0.565	0.477	0.333	0.732	0.643	0.482	0.778	0.705	0.553
	$W^A(L_3)$	0.57	0.493	0.345	0.738	0.661	0.507	0.795	0.726	0.567
(c)	LM^G	0.229	0.175	0.104	0.293	0.232	0.143	0.357	0.288	0.179
	$LM^A(L_1)$	0.27	0.206	0.14	0.351	0.29	0.17	0.417	0.341	0.225
	$LM^A(L_2)$	0.267	0.201	0.132	0.362	0.296	0.182	0.424	0.344	0.235
	$LM^A(L_3)$	0.278	0.211	0.133	0.359	0.279	0.166	0.416	0.33	0.231
	W^G	0.19	0.133	0.062	0.246	0.179	0.087	0.301	0.233	0.119
	$W^A(L_1)$	0.213	0.161	0.083	0.311	0.229	0.11	0.367	0.28	0.154
	$W^A(L_2)$	0.223	0.171	0.091	0.329	0.26	0.143	0.391	0.302	0.175
	$W^A(L_3)$	0.268	0.209	0.111	0.352	0.273	0.167	0.403	0.319	0.185
(d)	LM^G	0.262	0.198	0.123	0.307	0.232	0.138	0.338	0.269	0.165
	$LM^A(L_1)$	0.278	0.224	0.143	0.335	0.255	0.16	0.367	0.283	0.185
	$LM^A(L_2)$	0.278	0.208	0.136	0.324	0.25	0.153	0.349	0.283	0.18
	$LM^A(L_3)$	0.267	0.205	0.124	0.308	0.234	0.146	0.344	0.277	0.174
	W^G	0.216	0.147	0.086	0.247	0.173	0.081	0.281	0.202	0.101
	$W^A(L_1)$	0.229	0.164	0.102	0.285	0.21	0.115	0.322	0.23	0.127
	$W^A(L_2)$	0.245	0.17	0.109	0.299	0.234	0.124	0.329	0.247	0.136
	$W^A(L_3)$	0.268	0.211	0.122	0.312	0.252	0.146	0.339	0.266	0.144
(e)	LM^G	0.217	0.155	0.088	0.329	0.239	0.137	0.353	0.28	0.186
	$LM^A(L_1)$	0.223	0.162	0.084	0.418	0.335	0.215	0.495	0.423	0.274
	$LM^A(L_2)$	0.302	0.238	0.142	0.45	0.38	0.242	0.534	0.461	0.318
	$LM^A(L_3)$	0.328	0.25	0.149	0.463	0.39	0.26	0.554	0.474	0.341
	W^G	0.188	0.132	0.06	0.214	0.139	0.071	0.286	0.208	0.099
	$W^A(L_1)$	0.202	0.136	0.066	0.365	0.276	0.153	0.465	0.376	0.205
	$W^A(L_2)$	0.286	0.225	0.106	0.434	0.341	0.197	0.529	0.434	0.28
	$W^A(L_3)$	0.342	0.271	0.17	0.487	0.406	0.242	0.57	0.473	0.318

$(L_1, L_2, L_3) = (1, 3, 5), (3, 5, 7), (4, 6, 8)$ for $n = 96, 198, 392$, respectively.

TABLE 11. Test statistics for $H_0 : \lambda = 0$ in SAR disturbances in SEM

	LM^G	$LM^A(3)$	$LM^A(4)$	$LM^A(5)$	W^G	$W^A(3)$	$W^A(4)$	$W^A(5)$
Thefts	1.052	0.736	0.691	0.78	1.593	1.672*	1.675*	1.676*
Car thefts	5.578**	5.155**	5.178**	3.225*	3.829***	3.859***	3.965***	3.967***
Robbery	3.917**	4.217**	4.346**	3.078*	3.197***	3.158***	3.27***	3.54***

*Significant at 10%, **Significant at 5%, ***Significant at 1%.

TABLE 12. Estimates of coefficients in MRSAR and SEM

Y	Theft		Car theft		Robbery	
model	MRSAR	SEM	MRSAR	SEM	MRSAR	SEM
Blood	−0.006***	−0.006***	−0.007**	−0.007**	−0.007**	−0.006**
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Criminal Networks	0.248*	0.249**	0.68**	0.667***	0.869***	0.802***
	(0.134)	(0.118)	(0.324)	(0.174)	(0.228)	(0.212)
Length	0.066**	0.048*	0.11***	0.086**	0.091	0.073
	(0.03)	(0.027)	(0.036)	(0.039)	(0.062)	(0.047)
Youth	−0.109	−0.153***	0.019	−0.001	0.145	0.04
	(0.093)	(0.059)	(0.136)	(0.089)	(0.122)	(0.108)
High School	0.026	0.032**	−0.007	−0.005	−0.074*	−0.081***
	(0.018)	(0.015)	(0.028)	(0.022)	(0.04)	(0.026)
Unemployment	0.001	0.004	−0.016	−0.007	−0.061	−0.05**
	(0.024)	(0.012)	(0.028)	(0.018)	(0.042)	(0.021)
GDP	−0.002	−0.01	0.02	0.047	0.056	0.058
	(0.031)	(0.028)	(0.049)	(0.042)	(0.055)	(0.05)
Urbanization	0.006**	0.005***	0.009***	0.006**	0.011***	0.01***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Clear Up	−0.048***	−0.078***	−0.078***	−0.094***	−0.012**	−0.028***
	(0.012)	(0.018)	(0.015)	(0.014)	(0.006)	(0.006)

Y is logarithm of crime rates per 1,000 inhabitants. Standard errors in parenthesis.
*Significant at 10%, **Significant at 5%, ***Significant at 1%.
Results for MRSAR are taken from Tables 3-5 of Buonanno *et al.* (2009), with bootstrapped regionally clustered standard errors. The reported standard errors for LSE in SEM with SAR error are based on:
$\sqrt{n}(\tilde{\beta} - \beta_0) \to_d N(0, V)$ where $V := \operatorname*{plim}_{n\to\infty} \sigma^2 (\frac{X'X}{n})^{-1} \frac{X'Q^{-1}Q^{-1\prime}X}{n} (\frac{X'X}{n})^{-1}$,
since Assumption 1 implies spatial dependence in disturbance term is weak.

7. Appendix: Proof of Proposition 1.

We derive the elements of the first row of Ξ which suffices for block-diagonality. Other terms' derivations can be found in the supplementary appendix. From (1.5),

$$L(\theta) = \sum_{i=1}^{n} \log f\Big(\frac{Q_i^T(\lambda)(y-\mu 1_n)}{\sigma};\zeta\Big) + \log\det\{Q(\lambda)\} - \frac{n}{2}\log\sigma^2,$$

where $Q_i^T(\lambda)$ denotes the i-th row of $Q(\lambda)$. From (1.1),

$$\varepsilon = (\varepsilon_1,\cdots,\varepsilon_n)^T = \frac{Q(y-\mu_0 1_n)}{\sigma_0}, \quad \varepsilon_i = \frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}, \quad i=1,\cdots,n.$$

The first derivative of $L(\theta)$ w.r.t λ at $\theta_0 = (\lambda_0,\mu_0,\sigma_0^2,\zeta_0)^T$ is given by

$$\frac{\partial L(\theta_0)}{\partial\lambda} = \sum_{i=1}^{n}\frac{M_i^T(y-\mu_0 1_n)}{\sigma_0}\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big) - \mathrm{tr}(P).$$

The following facts are used below:

$$\frac{\partial\psi(s)}{\partial s} = \frac{(f'(s))^2 - f''(s)f(s)}{f^2(s)} = \psi^2(s) - \frac{f''(s)}{f(s)},$$
$$\frac{\partial}{\partial\lambda}\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0} = \frac{-M_i^T(y-\mu_0 1_n)}{\sigma_0} = -M_i^T Q^{-1}\varepsilon = -P_i^T\varepsilon,$$
$$\frac{\partial}{\partial\mu}\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0} = \frac{-Q_i^T 1_n}{\sigma_0}, \quad \frac{\partial}{\partial\sigma^2}\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0} = \frac{-Q_i^T(y-\mu_0 1_n)}{2\sigma_0^3} = \frac{-\varepsilon_i}{2\sigma_0^2},$$

where P_i^T denotes the ith row of P. Next, we derive the elements in the first row of Ξ. Denote $\psi_i = \psi(\varepsilon_i)$, $N(\lambda) := dM(\lambda)/d\lambda$ and $R(\lambda) := N(\lambda)Q^{-1}(\lambda)$ with $N = N(\lambda_0)$ and $R = R(\lambda_0)$.

Combining (ii) and (iii) of Assumption 1 leads to (cf. Lee (2004, p. 1918)):

$$\max_{1\le i,j\le n}|p_{ij}| = O(h^{-1}). \tag{7.1}$$

Assumption 1 also implies that for all sufficiently large n, P is uniformly bounded in both row and column sums:

$$\max_{1\le i\le n}\sum_{j=1}^{n}|p_{ij}|=O(1)\quad\text{and}\quad\max_{1\le j\le n}\sum_{i=1}^{n}|p_{ij}|=O(1),\ \text{as}\quad n\to\infty. \tag{7.2}$$

$(1,1)^{th}$ *element of* Ξ. We observe first that

$$\begin{aligned}\frac{\partial \mathrm{tr}(P)}{\partial\lambda}&=\mathrm{tr}\Big(M\frac{\partial Q^{-1}}{\partial\lambda}+\frac{\partial M}{\partial\lambda}Q^{-1}\Big)=\mathrm{tr}\Big(-MQ^{-1}\frac{\partial Q}{\partial\lambda}Q^{-1}+\frac{\partial M}{\partial\lambda}Q^{-1}\Big)\\&=\mathrm{tr}\Big(P^2+NQ^{-1}\Big)=\mathrm{tr}\Big(P^2+R\Big).\end{aligned} \tag{7.3}$$

Writing $\varepsilon_i=Q_i^T(y-\mu_0 1_n)/\sigma_0$,

$$\begin{aligned}&\frac{\partial}{\partial\lambda_0}\Big\{\sum_{i=1}^{n}\frac{M_i^T(y-\mu_0 1_n)}{\sigma_0}\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big)\Big\}\\&=\sum_{i=1}^{n}P_i^T\varepsilon\Big(\psi_i^2-\frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)\frac{\partial}{\partial\lambda}\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}+\sum_{i=1}^{n}N_i^T\frac{(y-\mu_0 1_n)}{\sigma_0^2}\psi_i\\&=-\sum_{i=1}^{n}\big\{(P_i^T\varepsilon)^2\cdot\Big(\psi_i^2-\frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)+(R_i^T\varepsilon)\psi_i\big\}.\end{aligned} \tag{7.4}$$

Now,

$$\begin{aligned}-\sum_{i=1}^{n}E[(P_i^T\varepsilon)^2\psi_i^2]&=-\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}^2E[\varepsilon_j^2\psi_i^2]\\&=-E(\psi_1^2)E(\varepsilon_1^2)\cdot\sum_{i=1}^{n}\sum_{j=1}^{n}p_{ij}^2+\big(E(\psi_1^2)E(\varepsilon_1^2)-E(\varepsilon_1^2\psi_1^2)\big)\cdot\sum_{i=1}^{n}p_{ii}^2\\&=-\mathcal{J}\cdot\mathrm{tr}(PP^T)+O\Big(n/h^2\Big),\end{aligned}$$

since $p_{ii}=O(1/h)$ uniformly in i, see (7.1). Noting $E\big(f''(\varepsilon_i)/f(\varepsilon_i)\big)=0$,

$$\sum_{i=1}^{n}E\Big((P_i^T\varepsilon)^2\frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)=E(\varepsilon_1^2f''(\varepsilon_1)/f(\varepsilon_1))\cdot\sum_{i=1}^{n}p_{ii}^2=2\mathrm{tr}(P^2)=O\big(\frac{n}{h^2}\big),$$

since under Assumption 3, $E(\varepsilon_1^2 f''(\varepsilon_1)/f(\varepsilon_1)) = 2$. Noting that $E(\varepsilon_i\psi_i) = 1$, $\sum_{i=1}^n R_{ii}E(\varepsilon_i\psi_i) = \mathrm{tr}(R)$, which cancels out the same term in (7.3).

Therefore, the $(1,1)^{th}$ element of Ξ is given by

$$\lim_{n\to\infty}\frac{h}{n}E\big(-\frac{d^2L(\theta_0)}{d\lambda^2}\big) = \lim_{n\to\infty}\frac{h}{n}\Big(\mathcal{J}\mathrm{tr}(PP^T) + \mathrm{tr}(P^2)\Big) = \mathcal{J}\omega_1 + \omega_2.$$

$(1,2)^{th}$ *element.* One has

$$\begin{aligned}\frac{-\partial^2}{\partial\mu\partial\lambda}L(\theta_0) &= \sum_{i=1}^n \frac{M_i^T 1_n}{\sigma_0}\psi_i + \sum_{i=1}^n \frac{M_i^T(y-\mu_0 1_n)}{\sigma_0}\Big(\psi_i^2 + \frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)\frac{Q_i^T 1_n}{\sigma_0} \\ &= \sum_{i=1}^n \frac{M_i^T 1_n}{\sigma_0}\psi_i + \sum_{i=1}^n P_i^T\varepsilon\Big(\psi_i^2 + \frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)\frac{Q_i^T 1_n}{\sigma_0}.\end{aligned}$$

Noting (7.1),

$$E\Big(\frac{-\partial^2}{\partial\mu\partial\lambda}L(\theta_0)\Big) = \frac{1}{\sigma_0}\sum_{i=1}^n p_{ii}(Q_i^T 1_n)\Big(E(\varepsilon_i\psi_i^2) + E\big(\varepsilon_i\frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\big)\Big) = O\Big(\sum_{i=1}^n |Q_i^T 1_n||p_{ii}|\Big) = O\big(\frac{n}{h}\big),$$

so the $(1,2)^{th}$ element of Ξ is $O(h^{-1}n)\times n^{-1}\sqrt{h} = O(h^{-1/2}) = o(1)$.

$(1,3)^{th}$ *element.* One has

$$\begin{aligned}\frac{-\partial^2}{\partial\sigma^2\partial\lambda}L(\theta_0) &= \sum_{i=1}^n \frac{M_i^T(y-\mu_0 1_n)}{2\sigma_0^3}\psi_i + \sum_{i=1}^n \frac{Q_i^T(y-\mu_0 1_n)}{2\sigma_0^3}\Big(\psi_i^2 + \frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)\frac{M_i^T(y-\mu_0 1_n)}{\sigma_0} \\ &= \sum_{i=1}^n \frac{P_i^T\varepsilon}{2\sigma_0^2}\psi(\varepsilon_i) + \sum_{i=1}^n \frac{\varepsilon_i}{2\sigma_0^2}\Big(\psi_i^2 + \frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)P_i^T\varepsilon,\end{aligned}$$

so

$$E\Big(\frac{-\partial^2}{\partial\sigma^2\partial\lambda}L(\theta_0)\Big) = \frac{1}{2\sigma_0^2}\sum_{i=1}^n p_{ii}\Big(E\big(\varepsilon_i\psi_i\big) + E\big(\varepsilon_i^2\psi_i^2\big) + 2\Big) = O\Big(\sum_{i=1}^n |p_{ii}|\Big) = O\Big(\frac{n}{h}\Big),$$

and thus, the $(1,3)^{th}$ element of Ξ is $O(h^{-1}n)\times n^{-1}\sqrt{h} = O(h^{-1/2}) = o(1)$.

$(1,4)^{th}$ *element.* We use

$$\begin{aligned}\frac{\partial\psi(\varepsilon_i;\zeta_0)}{\partial\zeta} &= -\frac{f(\varepsilon_i;\zeta_0)\frac{\partial^2}{\partial\varepsilon_i d\zeta}f(\varepsilon_i;\zeta_0) - \frac{\partial}{\partial\varepsilon_i}f(\varepsilon_i;\zeta_0)\frac{\partial}{\partial\zeta}f(\varepsilon_i;\zeta_0)}{f^2(\varepsilon_i;\zeta_0)} \\ &= -\big(\frac{d^2 f(\varepsilon_i;\zeta_0)}{d\varepsilon_i d\zeta}\big)f^{-1}(\varepsilon_i;\zeta_0) + \chi_i\psi_i,\end{aligned}$$

to obtain

$$-\frac{\partial^2}{\partial\lambda\partial\zeta}L(\theta_0) = -\sum_{i=1}^n P_i^T\varepsilon\frac{\partial\psi(\varepsilon_i;\zeta_0)}{\partial\zeta} = \sum_{i=1}^n P_i^T\varepsilon\big(\frac{\partial^2 f(\varepsilon_i;\zeta_0)}{\partial\varepsilon_i\partial\zeta}\big)f^{-1}(\varepsilon_i;\zeta_0) - P_i^T\varepsilon\chi_i\psi_i,$$

and thus

$$\begin{aligned}E\Big(-\frac{\partial^2}{\partial\mu\partial\zeta}L(\theta_0)\Big) &= \sum_{i=1}^n p_{ii}E\Big(\varepsilon_i\big[\frac{\partial^2 f(\varepsilon_i;\zeta_0)}{\partial\varepsilon_i\partial\zeta}\big]f^{-1}(\varepsilon_i;\zeta_0)\Big) - \sum_{i=1}^n p_{ii}E\Big(\varepsilon_i\chi_i\psi_i\Big) \\ &= O(1)\sum_{i=1}^n p_{ii} = O\Big(\frac{n}{h}\Big).\end{aligned}$$

Therefore, the (1,4)th element of Ξ is $O(h^{-1}n)\times n^{-1}\sqrt{h} = O(h^{-1/2}) = o(1)$. ■

Acknowledgement: We are grateful for the comments of three referees which have led to improvements in the paper. We also thank Paolo Buonanno for permission to use the data in Section 6.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers.

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International regional science review*, **26**, 153-166.

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundation and Applications to Regional Analysis.* Springer-Verlag, Berlin.

Baltagi, B. H. and Yang, Z. (2013). Standardized LM tests for spatial error dependence in linear or panel regressions. *Econometrics Journal* , **16**, 103-134.

Beran, R. (1976). Adaptive estimates for autoregressive processes. *Annals of the Institute of Statistical Mathematics* **26**, 77-89.

Buonanno, P., Montolio, D. and Vanin, P. (2009). Does social capital reduce crime? *Journal of Law and Economics*, **52**, 145-170.

Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society, Series B*, **42**, 107-8.

Cliff, A. and Ord, J. K. (1968). The problem of spatial autocorrelation. *Joint Discussion Paper, University of Bristol: Department of Economics*, **26**, Department of Geography, Series A, 15.

Cliff, A. and Ord, J. K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, **4**, 267-84.

Kelejian, H.H. and Prucha, I.R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, **40**, 509-533.

Kelejian, H. H. and Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, **104**, 219-257.

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

Kyriacou, M., Phillips, P.C.B and Rossi, F. (2019). Continuously updated indirect inference in spatial autoregressions with unobserved heterogeneity. Working paper.

Lee, L.F. (2004). Asymptotic distributions of quasi-maximum likelihood estimates for spatial autoregressive models. *Econometrica*, **72**, 1899-1925.

Lin, X. and Lee, L. F.(2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics*. **159**, 303-319.

Liu, X., Lee, L. F and Bollinger, C. (2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics*, **159**, 303-319.

Lesage, J. P. and Pace, R. K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, **140**, 190-214.

Moran, P.A.P. (1950). A test for the serial dependence of residuals. *Biometrika*, **37**, 178-81.

Newey, W. (1988). Adaptive Estimation of regression models via moment restrictions. *Journal of Econometrics*, **38**, 301-339.

Robinson, P.M. (1987). Asymptotically efficient estimation in the Presence of heteroskedasticity of unknown form *Econometrica*, **55**, 875-891.

Robinson, P.M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series. *Annals of Statistics*, **33**, 1800-1842.

Robinson, P.M. (2008). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics*, **147**, 5-16.

Robinson, P.M. (2010). Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics*, **157**, 6-17.

Robinson, P. M. and Rossi, F. (2014). Refined tests for spatial correlation, *Econometric Theory*, **6**, 1-32.

Robinson, P. M. and Rossi, F. (2015). Refinements in maximum likelihood inference on spatial autocorrelation in panel data, *Journal of Econometrics* , **189** 447-456.

Yang, Z. (2015). LM tests of spatial dependence based on bootstrap critical values. *Journal of Econometrics* , **185** 33-59.