ADAPTIVE INFERENCE ON PURE SPATIAL MODELS

JUNGYOON LEE[†] AND PETER M. ROBINSON[‡]

Abstract. We consider adaptive tests and estimates which are asymptotically efficient in the presence of unknown, nonparametric, distributional form in pure spatial models. A novel adaptive Lagrange Multiplier testing procedure for lack of spatial dependence is proposed and extended to linear regression with spatially correlated errors. 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. The paper covers a general class of semiparametric spatial models allowing nonlinearity in the parameters and/or the weight matrix, in addition to unknown distribution.

JEL classifications: C12, C13, C14, C21.
Keywords: Efficient test; adaptive estimation; spatial models.

....

[†]Economics Department, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. Tel:+44 (0)1784 276506, 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

Acknowledgement: We thank Paolo Buonanno for permission to use the data in our empirical example.

1. Introduction

Spatial autoregressive models were introduced by Cliff and Ord (1968), and have since been extensively developed in the econometric literature. In particular, they often model spatially observed explanatory variables in terms of explanatory variables as well as a weight matrix structure that introduces spatial correlation. But as originally introduced they were designed to model spatial correlation, without the presence of explanatory variables, echoing the earlier work of Moran (1950) on testing for spatial correlation. Pure spatial models, in which observations are spatially dependent but not influenced by explanatory variables, are known to lead to rather different statistical properties from models that do include explanatory variables. In particular least squares estimates (LSE) of pure spatial models are inconsistent, and, with instrumental variables being unavailable, the leading alternative, Gaussian pseudo maximum likelihood estimate (PMLE), may converge more slowly than at the parametric rate. Here we consider a quite general class of 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 tests, 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).

While Wald statistics based on our adaptive estimate have greater efficiency compared to those based on less efficient estimates, we also provide adaptive Lagrange Multiplier (LM) tests which have the advantage of being based on the restricted model only. Many authors including Cliff and Ord (1972), Burridge(1980), Kelejian and Prucha (2001), Robinson (2008) and Robinson and Rossi (2014) have considered Gaussian LM tests for lack of spatial dependence in SAR model, extending Moran (1950). Although Gaussian LM tests enjoy the same robustness property as Gaussian PMLE, there is a scope for further efficiency improvement which our adaptive LM tests set to achieve. To enhance the relevance of our methods, we also extend our results to cover testing spatial correlation in error terms of a linear regression model.

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 a $n \times 1$ vector of 1's, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^T$ is a vector of independent identically distributed 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 $w_{ii} \equiv 0$. 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 the general model (1.1). Let $\underline{eig}(W)$ denote the minimum eigenvalue of the matrix W.

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

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

where I is the $n \times n$ identity matrix and $\lambda_0 \in (\underline{eig}(W)^{-1}, 1)$.

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

$$Q(\lambda_0) = (I + \lambda_0 W)^{-1}, \tag{1.3}$$

for $\lambda_0 \in (\underline{eig}(W)^{-1}, 1)$.

(3) MESS (matrix spatial exponential model, see LeSage and Pace (2009)):

$$Q(\lambda_0) = \exp(\lambda_0 W). \tag{1.4}$$

The models (1.2)-(1.4) are sometimes referred to as “pure”models, to reflect the absence of regressors. When ε, and thus y, is Gaussian, the model (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 model under some regularity conditions, the estimate $\tilde{\theta}$ is consistent and asymptotically normal. In fact, these latter properties hold over a much wider class of distributions of the ε_i, for which the estimate $\tilde{\theta}$ is termed a (Gaussian) PMLE. Such robustness is also shared by the LM test for $H_0 : \lambda = 0$ based on the Gaussian likelihood.

However, 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 (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 $\theta_0 = \left(\lambda_0, \mu_0, \sigma_0, \zeta_0^T\right)^T$, and denote by $\theta = \left(\lambda, \mu, \sigma, \zeta^T\right)^T$ any admissible value of θ_0. Write the corresponding log likelihood as

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

where $Q_i^T(\lambda)$ denotes the ith row of $Q(\lambda)$. The MLE $\bar{\theta} = \left(\bar{\lambda}, \bar{\mu}, \bar{\sigma}, \bar{\zeta}^T\right)^T$ 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 $L(\theta)$ with respect to λ evaluated at $\lambda = 0$. Unfortunately 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 the fact that $\mu_0 = 0$ 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 the Gaussian PMLE $\tilde{\lambda}$, using nonparametric (series) estimation of the score function. In a similar vein, a 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 for estimation in a spatial context by Robinson (2010), for the mixed regressive SAR model of order 1:

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

where X is a $n \times k$ matrix of observed regressors 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 behaviours of estimates of λ_0 under the two models can differ, even their convergence rates. Consequently, the feasibility and implementation of such adaptive estimation in the pure spatial models, including pure SAR model, need to be established separately.

The method of estimation we employ is very similar to that of Robinson (2010), but the asymptotic variance matrix of his estimate of $(\lambda_0, \beta_0^T)^T$ corresponds to that found in the classical adaptive estimation literature, whereas that of 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 (typically less) than in the classical outcome. Somewhat unusually in the spatial econometric literature, we cover several possible functional forms by treating (1.1) with $Q(\lambda)$ being a parametric function that can take several forms, such as (1.2)-(1.4).

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 in the previous paragraph. Sections 3 and 4 describe, respectively, our estimate $\hat{\lambda}$ and its asymptotic distribution. The nonparametric estimation of the score function for ε_1 introduced in Section 3 is used in Section 5 to construct an adaptive LM testing procedure for lack of spatial dependence in the model (1.1), and also when (1.1) is used for the unobserved error terms in a linear regression model. Section 6 presents results of a small Monte Carlo study of finite sample performance of our adaptive estimate and LM testing procedure. Both estimation and testing led to substantial efficiency gains compared those based on Gaussian likelihood, while it is notable that LM testing improves significantly in the size performance compared to Wald test statics that accompany substantial undersizing, which was also reported in panel data setting in Robinson and Rossi (2015). Section 7 contains application of our methods to an economic dataset on crime rates across Italian provinces.

2. Block-diagonality of the information matrix

The feasibility of adaptive estimation of λ_0 is shown via establishing the block-diagonality of the information matrix. Denote $M(\lambda) := -dQ(\lambda)/d\lambda$, $M = M(\lambda_0) = (m_{ij})$. 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 $n \to \infty$,*

$$\max_{1\le i\le n} \sum_{j=1}^{n} |m_{ij}| = O(1) \quad \text{and} \quad \max_{1\le j\le n} \sum_{i=1}^{n} |m_{ij}| = O(1).$$

(ii) *For a sequence $h = h_n$ such that $h^{-1} + h/n \to 0$ as $n \to \infty$, $\max_{1\le i,j\le n} |m_{ij}| = O\big(h^{-1}\big)$.*

(iii) *For all sufficiently large n, $Q := Q(\lambda_0)$ is non-singular and Q^{-1} is uniformly bounded in both row and column sums.*

The sequence h is important in the asymptotic analysis, defining the rate of convergence of estimates of λ_0. For SAR model, Assumption 1 is typically assumed with $M = W$ and $Q = (I - \lambda_0 W)$ (see e.g. Assumptions 2-5 of Lee (2004), with $\lambda_0 \in (-1, 1)$). These are in fact sufficient conditions for Assumption 1 to also hold for SMA and MESS models, based on some basic matrix results given in e.g. Lee (2004, p.1918).

Assumption 2. *The limits*

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

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

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

To show the feasibility of adaptive estimation of λ_0, we establish block-diagonality of the information matrix between λ_0 and the other parameters. Introduce:

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

Proposition 1. *Under Assumptions 1-2,* $\Xi := \lim_{n\to\infty} D^{-1}E\big(-\frac{d^2L(\theta_0)}{d\theta d\theta^T}\big)D^{-1}$ *exists and*

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

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 supplementary appendix.

3. Adaptive Estimation

With f, f' respectively denoting the nonparametric density and derivative-of-density of ε_1, the score function of ε_1 is given by $\psi(s) = -f'(s)/f(s)$, when $f(s) \neq 0$. The nonparametric estimate of ψ we use in adaptive estimation is a series one, whose advantages over kernel estimation are discussed in Robinson (2010). To formulate the adaptive estimate, denote first by $\phi_\ell(s)$, $\ell = 1, 2, ...$ a sequence of smooth functions, to be used in series estimation of $\psi(\cdot)$. For an integer $L \geq 1$, where $L = L_n$ will be regarded as increasing with n, define the $L \times 1$ vectors

$$\phi^{(L)}(s) = (\phi_1(s), \cdots, \phi_L(s))^T, \quad \bar{\phi}^{(L)}(s) = \phi^{(L)}(s) - E\left\{\phi^{(L)}(\varepsilon_i)\right\}, \tag{3.1}$$
$$\phi^{'(L)}(s) = (\phi_1^{'}(s), \cdots, \phi_L^{'}(s))^T.$$

L is the number of approximating functions that are used in series estimation of $\psi(\cdot)$. Allowing $L \to \infty$ as $n \to \infty$ enables nonparametric estimation of $\psi(\cdot)$. See Robinson (2010) for discussion on the choice of ϕ and L.

We regard $\psi(s)$ as being approximated by

$$\psi(s, a^{(L)}) = \bar{\phi}^{(L)}(s)^T a^{(L)}, \tag{3.2}$$

for unknown vector $a^{(L)} = \left[E\left\{\bar{\phi}^{(L)}(\varepsilon_i)\bar{\phi}^{(L)}(\varepsilon_i)^T\right\}\right]^{-1} E\left\{\phi^{'(L)}(\varepsilon_i)\right\}$ as explained in Robinson (2010). Given a vector of observable proxies $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_n)^T$, we estimate $a^{(L)}$ by $\tilde{a}^{(L)}$, a sample analogue constructed as follows. For a generic vector $x = (x_1, ..., x_n)^T \in \mathbb{R}^n$, define

$$\tilde{a}^{(L)}(x) = W^{(L)}(x)^{-1}w^{(L)}(x),$$

where

$$W^{(L)}(x) = \frac{1}{n}\sum_{i=1}^n \Phi^{(L)}(x_i)\Phi^{(L)}(x_i)^T, \quad \Phi^{(L)}(x_i) = \phi^{(L)}(x_i) - \frac{1}{n}\sum_{j=1}^n \phi^{(L)}(x_j),$$

and $w^{(L)}(x) := n^{-1}\sum_{i=1}^n \phi^{'(L)}(x_i)$. Likewise, define $\psi^{(L)}\left(x_i; \tilde{a}^{(L)}(x)\right) := \Phi^{(L)}(x_i)^T\tilde{a}^{(L)}(x)$. The estimate $\tilde{\psi}_{iL} := \psi^{(L)}\left(\tilde{\varepsilon}_i; \tilde{a}^{(L)}(\tilde{\varepsilon}_i)\right)$ of $\psi(\varepsilon_i)$ for a given vector $\tilde{\varepsilon}$, will be used to construct the adaptive estimate in (3.5).

The above discussion is based on a given proxy $\tilde{\varepsilon}$ for ε. The specific one we use is constructed as follows. 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

$$\epsilon(\lambda) = (\epsilon_1(\lambda), \cdots, \epsilon_n(\lambda))^T = HQ(\lambda)y, \quad i = 1, \cdots, n. \tag{3.3}$$

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

$$\tilde{\varepsilon} := \frac{\epsilon(\tilde{\lambda})}{\tilde{\sigma}}.$$

For convenience, set $\tilde{\psi}_{iL} := \tilde{\psi}_{iL}(\tilde{\lambda}, \tilde{\sigma})$, where $\tilde{\psi}_{iL}(\lambda, \sigma) := \Phi^L(\epsilon_i(\lambda)/\sigma)^T \tilde{a}^L(\epsilon(\lambda)/\sigma)$.

Introduce the estimate $\tilde{\mathcal{J}}_L := \tilde{\mathcal{J}}_L(\tilde{\lambda}, \tilde{\sigma})$ of the information $\mathcal{J}$, where

$$\tilde{\mathcal{J}}_L(\lambda, \sigma) = \frac{1}{n}\sum_{i=1}^n \tilde{\psi}_{iL}^2(\lambda, \sigma). \tag{3.4}$$

Denote $P(\lambda) := M(\lambda)Q^{-1}(\lambda), P = P(\lambda_0) = (p_{ij})$. 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) $P(\lambda) = -W$.

We are now ready to define our adaptive estimate of λ_0, based on a preliminary estimate $\tilde{\lambda}$, as follows:

$$\hat{\lambda} = \tilde{\lambda} + \left(\tilde{\mathcal{J}}_L \cdot \operatorname{tr}\left\{P(\tilde{\lambda})P(\tilde{\lambda})^T\right\} + \operatorname{tr}\left\{P(\tilde{\lambda})^2\right\}\right)^{-1}$$
$$\left(\frac{1}{\tilde{\sigma}}(\tilde{\psi}_{1L}, \cdots, \tilde{\psi}_{nL})M(\tilde{\lambda})Hy - \operatorname{tr}\left\{P(\tilde{\lambda})\right\}\right). \tag{3.5}$$

The second term of (3.5) represents a Newton step, based on nonparametric estimate of the score function $\psi(\cdot)$.

4. Asymptotic Normality and Efficiency

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 $s^2 f''(s) \to 0$ as $|s| \to \infty$ and*

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

Assumption 4. *In (3.1) and (3.2), $\phi_\ell(s) = \phi(s)^\ell$, $l = 1, ..., L$, where $\phi(s)$ is a strictly increasing and thrice differentiable function such that for some $\kappa \geq 0$, $K > 0$,*

$$|\phi(s)| \leq 1 + |s|^\kappa, \quad |\phi'(s)| + |\phi''(s)| + |\phi'''(s)| \leq C(1 + |\phi(s)|^K), \quad s \in \mathbb{R}. \tag{4.1}$$

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 of (3.1) satisfy one of the following conditions with κ as in (4.1).*

(i) $\kappa = 0$, $E(\varepsilon_i^4) < \infty$, and for some $A > \eta \max(\varphi, 1)$, $L \leq \log h / 8 \log A, \quad n \to \infty$.

(ii) $\kappa > 0$, for some $\omega > 0$ and $t > 0$, $E\left(e^{t|\varepsilon_i|^{\omega}}\right) < \infty$, and for some $B > 8\kappa \max(1, \frac{1}{\omega})$, $L \log L \leq \log h / B, \quad n \to \infty$.

(iii) $\kappa > 0$, the random variables ε_i's are almost surely bounded, and for some $C > 4\kappa$, $L \log L \leq \log h / C, \quad n \to \infty$.

Assumption 5 is an amended version of Assumption 5 of Robinson (2010). It captures the trade-offs in the choice of series functions and restrictions imposed on the ε_i's, L and h. If bounded ϕ is used and $E\varepsilon_i^4 < \infty$, then Assumption 5 (i) entails a relatively modest upper bound on the rate of growth of 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 choice of series functions to provide an approximation error of $\psi(\cdot)$ (c.f. (3.2)) that decreases at a suitably fast rate 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.

The following theorem states asymptotic normality of the adaptive estimate $\hat{\lambda}$ of (3.5).

Theorem 1. *Let y follow the model (1.1) with $\lambda_0 \in (\underline{eig}(W)^{-1}, 1)$ and Assumptions 1 - 7 be satisfied. Then, as $n \to \infty$,*

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

4.1. **Efficiency comparison of adaptive estimate and PMLE.** In Lee (2004) it was shown that for the pure SAR model

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

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 to that in other settings.

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

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

For some W, it follows that $\mathrm{tr}(G^2) < 0$, $\omega_2 < 0$. However, if all elements of G are non-negative, which is implied if $w_{ij} \geq 0$ 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. 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 time series setting. For example, for the circulant matrix given below with one negative and one positive element in each row, we have $\omega_2 = -(1+\lambda^2)^{-3/2} - \lambda^{-2}((1+\lambda^2)^{-1/2} - 1) < 0$ for $0 < \lambda^2 < (1+\sqrt{5})/2$:

$$W = \frac{1}{2}\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & -1 \\ -1 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \dots & -1 & 0 \end{pmatrix}.$$

When the matrix W above is used in MESS model, $\omega_2 = -0.5$, while in SMA model ω_2 is the same as in SAR model as given above.

5. Testing for lack of spatial dependence

One can construct an "adaptive" LM test 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 test has the advantage of estimating only the restricted model and the statistic is based on the following standardized residuals from the restricted model denoted $\tilde{\varepsilon}_i^{(r)}$:

$$\tilde{\varepsilon}_i^{(r)} = \tilde{\epsilon}_i^{(r)}/\tilde{\sigma}_{(r)}, \text{ where } \quad \tilde{\epsilon}_i^{(r)} = y_i - \bar{y},\ \tilde{\sigma}^2_{(r)} = \tilde{\epsilon}^{(r)T}\tilde{\epsilon}^{(r)}/n. \qquad (5.1)$$

For SAR, the Gaussian LM test statistic LM^G_{SAR} takes the form:

$$LM^G_{SAR} = \frac{\big(\tilde{\varepsilon}^{(r)T} W \tilde{\varepsilon}^{(r)}\big)^2}{\mathrm{tr}(\mathrm{WW^T}) + \mathrm{tr}(\mathrm{W}^2)}. \tag{5.2}$$

Burridge (1980) noted that LM^G_{SAR} is also the Gaussian LM test statistic for SMA model.

Whilst the Gaussian LM test shares the robustness of the Gaussian PMLE, one expects power to improve when using instead a correctly specified error distribution to derive the LM statistic. To build a LM statistic which adapts to an unknown error distribution of nonparametric form, note that under H_0:

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

$$\lim_{n\to\infty} \frac{h}{n} E\Big(\frac{\partial^2 L(\theta_0)}{\partial \lambda^2}\Big|_{H_0}\Big) = \mathcal{J}\omega_1(0) + \omega_2(0),$$

with $\omega_1(0)$ and $\omega_2(0)$ evaluated at $\lambda_0 = 0$.

To build a LM statistic, one needs estimates of $\mathcal{J}$ and $\psi((y_i - \mu_0)/\sigma_0)$. We use the series estimation of score function given in Section 3, with the restricted standardized residual $\tilde{\varepsilon}_i^{(r)}$. Denote $\tilde{\psi}_{iL}^{(r)} := \Phi^L(\tilde{\varepsilon}_i^{(r)})^T \tilde{a}^L(\tilde{\varepsilon}^{(r)})$ and $\tilde{\mathcal{J}}_L^{(r)} = \sum_{i=1}^n (\tilde{\psi}_{iL}^{(r)})^2/n$. Our adaptive LM statistic is given by

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

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

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

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

The adaptive LM test statistic specified for the SAR (and SMA) is

$$LM^A_{SAR} = \frac{\Big(\sum_{i=1}^n W_i^T \tilde{\varepsilon}^{(r)} \cdot \tilde{\psi}_{iL}^{(r)}\Big)^2}{\tilde{\mathcal{J}}_L^{(r)} \mathrm{tr}(\mathrm{WW^T}) + \mathrm{tr}(\mathrm{W}^2)}. \tag{5.3}$$

LM^A_{SAR} is in fact identical to what would have been derived under MESS also.

One could extend our LM testing to a linear regression model with spatially dependent errors

$$z = X\beta + y, \tag{5.4}$$

with $n \times 1$ vector of dependent variables z, $n \times k$ matrix of regressors $X = (X_1, ..., X_n)^T$, and the error terms y following (1.1). The model (5.4) along with (1.1) is called the Spatial Error model (SEM), see e.g. Anselin (1988). Under Assumption 1, the dependence in the error terms y is weak and usual estimates of β such as ordinary LSE are $\sqrt{n}$-consistent. Denoting such estimates $\tilde{\beta}$ and corresponding fitted residuals $\tilde{y} := Z - X\tilde{\beta}$, we replace y with $\tilde{y}$ in getting the proxy to be used in the LM statistic in (5.1): $\tilde{\epsilon}^{(r)} = H\tilde{y}$, with H being redundant if X contains an intercept.

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

Theorem 2.

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

(ii) Let $\tilde{y}$ be fitted residuals from model (5.4), (1.1) with $\lambda_0 \in (\underline{eig}(W)^{-1}, 1)$ and Assumptions 1-8 be satisfied. Then, as $n \to \infty$ under $H_0 : \lambda_0 = 0$, $LM^A \to_d \chi^2(1)$.

6. Monte Carlo Study of Finite Sample Performance

In this section, we report results from a small Monte Carlo study of the 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 estimate $\tilde{\lambda}$ under differing error distributions, sample sizes, and magnitudes of spatial dependence, and then compare size and power performance of our adaptive tests and the existing ones. 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:

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

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 three different values of $\lambda_0 = 0.2, 0.4, 0.8$ for SAR and SMA and $\lambda_0 = 1, 2, 3$ for MESS. As in the Monte Carlo study of Robinson (2010), the following four distributions of ε_i are used with asymptotic relative efficiency (ARE) $(= 2/(\mathcal{J} + 1))$ of $\hat{\lambda}$ to $\tilde{\lambda}$ as reported below. The *ARE* was calculated based on the reported values of $1/\mathcal{J}$ from Robinson (2010).

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

$$f(u) = \frac{0.05}{\sqrt{50\pi}} exp\big(-\frac{u^2}{50}\big) + \frac{0.95}{\sqrt{2\pi}} exp\big(-\frac{u^2}{2}\big), \quad u \in \mathbb{R} \quad ARE = 0.679.$$

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

$$f(u) = \frac{0.5}{\sqrt{2\pi}} exp\big(-\frac{(u-3)^2}{2}\big) + \frac{0.5}{\sqrt{2\pi}} exp\big(-\frac{(u+3)^2}{2}\big), \quad u \in \mathbb{R} \quad ARE = 0.188.$$

(c) Laplace, $f(u) = \exp(-|s|\sqrt{2})\sqrt{2}$, $ARE = 0.666$.

(d) Student t_5, $\varepsilon_i = u\sqrt{3/5}$, where $u \sim t_5$, $ARE = 0.685$.

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

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

6.1. **Efficiency improvement in estimation.** Based on 1000 replications, the Monte Carlo variance of the two estimates of λ_0 were computed in each setting, and their ratios are presented in Tables 1-3 for SAR, SMA and MESS, respectively. A ratio less than 1 indicates efficiency improvement. Substantial improvements are reported in the 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 for $\lambda = 0.2$ (except for SMA with $n = 392$), and for $\lambda = 0.4, 0.8$ 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 most settings, 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 apart from 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 the bias has been adversely affected by the adaptive estimation for the choice (ii) of ϕ. In fact, the relative MSE often exhibited greater improvement than the relative variance, suggesting bias has been also reduced.

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

[Tables 1-4 about here]

6.2. **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 test LM^G_{SAR} of (5.2), the adaptive LM test LM^A_{SAR} of (5.3), and Wald tests based on Gaussian PMLE $\tilde{\lambda}$ and our adaptive estimate $\hat{\lambda}$:

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

For $n = 96, 198, 392$ we report results with three choices of L:, $L_1 = 1, 3, 4$, $L_2 = 3, 5, 6$ and $L_3 = 5, 7, 8$. 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 four distributions. This is in line with what Robinson and Rossi (2015) observed in a panel data setting, notably for normal 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 normal cumulative distribution function (cdf) offer poor approximation for the exact cdf of the Gaussian PMLE derived under normality, even when $\lambda = 0$. 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) and (b). 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 four distributions. Size results based on LM statistics are much more encouraging, with LM^G_{SAR} reporting better size results than W^G and W^A in all four distributions. Our adaptive LM^A_{SAR} improves the size results even further, with the exception of $n = 96$ in (c), 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) and (c). In (a) and (b), size results are best for larger L and for (c) and (d) size often is best with the smaller L, with the exception of (c) $n = 198$. In all cases but one ((d), $n = 392$), our adaptive LM^A_{SAR} generated the best size results out of the four statistics.

[Table 5 about here]

In Tables 6 and 7, we report Monte Carlo power for nominal sizes $s = 01, 0.05, 0.01$ 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 the power, dramatically in (a) and (b) and mildly in (c) and (d) and in all cases larger L and n improve power further. In (a) and (b), LM^G_{SAR} has worse power than W^A, while this is not necessarily the case in (c) and (d). In (a), (b), (c), our LM^A_{SAR} has best power which improves with increasing L. In (d), while LM^A_{SAR} still reports the best power results, there is less clear pattern on the best choice of L. It is notable that the power of LM^G_{SAR} and W^G remain much the same across the four distributions for given n, while adaptive statistics W^A and LM^A_{SAR} report greatest power in (a) and then in (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 often has slightly better power than LM^A_{SAR}. It is remarkable that the power improvements from using adaptive statistics LM^A_{SAR} 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 (5.4) are reported in Tables 8-10, with one dimensional regressor X_i generated from uniform distribution on $[0, 1]$ and $\beta = 1$. β was estimated by 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) and (d) there is no such clear pattern. Relative performance of the four statistics remain 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. Relative power performance of the four statistics reported from Tables 6-7 continue to hold in the SEM case.

[Tables 8-10 about here]

7. 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) which studies effects of social capital on crime rates. Their data contains (report-rate-adjusted) crime rates (Y) for three crimes, robbery, thefts and car thefts, four different measures of social capital (SC), and a set of demographic, socioeconomics 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) had fitted the mixed regressive SAR model of (1.6) 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 ρ to be insignificant in all but one of of 12 regressions (just at the 10 percent level).

In this paper, 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 social capital measures of Buonanno *et al.* (2009) as pointed out by the authors, other measures being the number of recreational and voluntary associations per 100,000 inhabitants and referenda turnout. As estimates of ρ are insignificant for all three choices of W, we drop the spatial lag and fit the spatial error model below

$$Y = \beta_1 SC + DSG\beta_2 + u, \qquad u = \lambda W u + \varepsilon, \tag{7.1}$$

and test $H_0 : \lambda = 0$. Table 11 reports the Gaussian and adaptive LM and Wald test statistics (L=3,4,5) when using the road-traveling distance weight matrix, for which Buonanno *et al.* reported estimation results in their Tables 3-5. In adaptive tests, we have used for ϕ, the choice (ii) given in the previous section. For the other two choices of weight matrix, the test results are unchanged and not reported here. We reject $H_0 : \lambda = 0$ for robbery and car thefts, while for theft, LM statistics fail to reject H_0 and Wald statistics reject at 10% significance level.

[Table 11 about here]

To account for possible spatial correlation in the error term, Buonanno *et al.* (2009) obtained bootstrapped regionally clustered standard errors for the coefficient estimates in the mixed SAR 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).

In Table 12 we report estimation results for the coefficients and standard errors based on (7.1) with the road-traveling distance weight matrix, and corresponding estimates and standard errors reported in Tables 3-5 of Buonanno *et al.* (2009) for mixed regressive SAR when using the same weight matrix. Standard errors that are obtained for (7.1) with the other two weight matrices are very similar to the ones obtained with the road-traveling distance weight matrix and do not affect significance of any coefficient estimates. For Theft, we have also tried working out the standard errors under $H_0 : \lambda = 0$, and again the standard error remain much the same.

[Table 12 about here]

Across the two models, the signs of coefficients which are significant 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 variables that are most significant controls across all three crime types. The SEM (7.1) tends to find more controls significant. For Theft, Youth and High School are additionally identified as significant, while for Robbery, Unemployment is the additionally significant control. This is natural as the presence the spatial lag term WY in the mixed SAR model 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

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

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

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
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
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

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

		$n = 96$			$n = 198$			$n = 392$		
	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

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

		$n = 96$			$n = 198$			$n = 392$		
	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

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

		$n = 96$			$n = 198$			$n = 392$		
	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

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

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

		$n = 96$			$n = 198$			$n = 392$		
	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

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

		$n = 96$			$n = 198$			$n = 392$		
	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

Table 11. Test statistics for $H_0 : \lambda = 0$

	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 mixed regressive SAR and SEM

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

Y is logarithm crime rates per 1,000 inhabitants. Results for mixed SAR model are taken from Tables 3-5 of Buonanno *et al.* (2009). Standard errors in parenthesis.
*Significant at 10%, **Significant at 5%, ***Significant at 1%.

8. Appendix: Proofs

The sequence h introduced in Assumption 1 plays important role in asymptotic analysis. Combining (ii) and (iii) of Assumption 1 leads to (cf. Lee (2004, p. 1918)):

$$\max_{1\le i,j\le n} |p_{ij}| = O(h^{-1}). \tag{8.1}$$

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

$$\max_{1\le i\le n} \sum_{j=1}^{n} |p_{ij}| = O(1) \quad \text{and} \quad \max_{1\le j\le n} \sum_{i=1}^{n} |p_{ij}| = O(1), \text{ as } \quad n\to\infty. \tag{8.2}$$

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. For brevity, we will denote $M = M(\lambda_0), Q = Q(\lambda_0), P = P(\lambda_0)$. 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 $Q(\lambda)$. Firstly, notice 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

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

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. For brevity, we denote $\psi_i = \psi(\varepsilon_i)$. Introduce notations $N(\lambda) := dM(\lambda)/d\lambda$ and $R(\lambda) := N(\lambda)Q^{-1}(\lambda)$ with $N = N(\lambda_0)$ and $R = R(\lambda_0)$.
$(1,1)^{th}$ *element of* Ξ. We observe first that

$$\frac{\partial \operatorname{tr}(P)}{\partial\lambda} = \operatorname{tr}\Big(M\frac{\partial Q^{-1}}{\partial\lambda} + \frac{\partial M}{\partial\lambda}Q^{-1}\Big) = \operatorname{tr}\Big(-MQ^{-1}\frac{\partial Q}{\partial\lambda}Q^{-1} + \frac{\partial M}{\partial\lambda}Q^{-1}\Big)$$
$$= \operatorname{tr}\Big(P^2 + NQ^{-1}\Big) = \operatorname{tr}\Big(P^2 + R\Big); \tag{8.3}$$

$$\begin{aligned}
&\frac{\partial}{\partial \lambda_0}\Big\{\sum_{i=1}^n \frac{M_i^T(y-\mu_0 1_n)}{\sigma_0}\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big)\Big\}\\
&= \sum_{i=1}^n P_i^T\varepsilon\Big(\psi^2(Q_i^T(y-\mu_0 1_n)/\sigma_0) - \frac{f''(Q_i^T(y-\mu_0 1_n)/\sigma_0)}{f(Q_i^T(y-\mu_0 1_n)/\sigma_0)}\Big)\frac{\partial}{\partial\lambda}\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\\
&\qquad + \sum_{i=1}^n N_i^T\frac{(y-\mu_0 1_n)}{\sigma_0^2}\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big)\\
&= -\sum_{i=1}^n P_i^T\varepsilon\Big(\psi^2(Q_i^T(y-\mu_0 1_n)/\sigma_0) - \frac{f''(Q_i^T(y-\mu_0 1_n)/\sigma_0)}{f(Q_i^T(y-\mu_0 1_n)/\sigma_0)}\Big)P_i^T\varepsilon\\
&\qquad + \sum_{i=1}^n R_i^T\varepsilon\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big)\\
&\qquad\qquad = -\sum_{i=1}^n (P_i^T\varepsilon)^2\cdot\Big(\psi_i^2 - \frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big) + (R_i^T\varepsilon)\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big). \qquad (8.4)
\end{aligned}$$

Then, in the last line of (8.4), the expectation of the first term is

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

since $p_{ii} = O(1/h)$ uniformly in i, see (8.1). Next, taking the expectation of the second product of (8.4) and noting $E\big(f''(\varepsilon_i)/f(\varepsilon_i)\big) = 0$,

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

since under Assumption 3, $E(\varepsilon_1^2 f''(\varepsilon_1)/f(\varepsilon_1)) = 2$. Now, noting that $E(\varepsilon_i\psi_i) = 1$, the expectation of the third term of (8.4) is $\sum_{i=1}^n R_{ii}E(\varepsilon_i\psi_i) = \mathrm{tr}(R)$, which cancels out the same term in (8.3).

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

$$\frac{-\partial^2}{\partial\mu\partial\lambda}L(\theta_0) = \sum_{i=1}^{n}\frac{M_i^T 1_n}{\sigma_0}\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big)$$

$$+\sum_{i=1}^{n}\frac{M_i^T(y-\mu_0 1_n)}{\sigma_0}\Big(\psi^2(Q_i^T(y-\mu_0 1_n)/\sigma_0)+\frac{f''(Q_i^T(y-\mu_0 1_n)/\sigma_0)}{f(Q_i^T(y-\mu_0 1_n)/\sigma_0)}\Big)\frac{Q_i^T 1_n}{\sigma_0}$$

$$= \sum_{i=1}^{n}\frac{M_i^T 1_n}{\sigma_0}\psi_i + \sum_{i=1}^{n}P_i^T\varepsilon\Big(\psi_i^2+\frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)\frac{Q_i^T 1_n}{\sigma_0}.$$

Taking expectations, and noting (8.1),

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

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

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

$$\frac{-\partial^2}{\partial\sigma^2\partial\lambda}L(\theta_0) = \sum_{i=1}^{n}\frac{M_i^T(y-\mu_0 1_n)}{2\sigma_0^3}\psi\Big(\frac{Q_i^T(y-\mu_0 1_n)}{\sigma_0}\Big)$$

$$+\sum_{i=1}^{n}\frac{Q_i^T(y-\mu_0 1_n)}{2\sigma_0^3}\Big(\psi^2(Q_i^T(y-\mu_0 1_n)/\sigma_0)+\frac{f''(Q_i^T(y-\mu_0 1_n)/\sigma_0)}{f(Q_i^T(y-\mu_0 1_n)/\sigma_0)}\Big)\frac{M_i^T(y-\mu_0 1_n)}{\sigma_0}$$

$$= \sum_{i=1}^{n}\frac{P_i^T\varepsilon}{2\sigma_0^2}\psi(\varepsilon_i) + \sum_{i=1}^{n}\frac{\varepsilon_i}{2\sigma_0^2}\Big(\psi^2(\varepsilon_i)+\frac{f''(\varepsilon_i)}{f(\varepsilon_i)}\Big)P_i^T\varepsilon.$$

Taking expectations yields

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

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

$(1,4)^{th}$ *element.* In deriving the $(1,4)^{th}$ element of Ξ, the following result is used repeatedly.

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

The second order cross derivative of $L(\theta_0)$ with respect to λ and ζ is

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

Taking expectations yields

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

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

Proof of Theorem 1. Define $\tilde{\epsilon}' := -M(\tilde{\lambda})Hy$. The estimate $\hat{\lambda}$ can be written alternatively as follows. Denote

$$\begin{aligned} r_L(\lambda,\sigma) &:= -\sum_{i=1}^n \tilde{\psi}_{iL}(\lambda,\sigma)\frac{\epsilon_i'}{\sigma} - \operatorname{tr}\{P(\lambda)\} \\ &= \frac{1}{\sigma}\Big(\tilde{\psi}_{1L}(\lambda,\sigma),\cdots,\tilde{\psi}_{nL}(\lambda,\sigma)\Big)M(\lambda)Hy - \operatorname{tr}\{P(\lambda)\}. \end{aligned} \tag{8.5}$$

Then, $\hat{\lambda}$ of (3.5) can be written as

$$\hat{\lambda}-\lambda_0 = (\tilde{\lambda}-\lambda_0) + \Big(\tilde{\mathcal{J}}_L\cdot\operatorname{tr}\Big\{P(\tilde{\lambda})P(\tilde{\lambda})^T\Big\} + \operatorname{tr}\Big\{P(\tilde{\lambda})^2\Big\}\Big)^{-1} r_L(\tilde{\lambda},\tilde{\sigma}). \tag{8.6}$$

Set $\tilde{\omega}_1 = (h/n)\operatorname{tr}\Big(P(\tilde{\lambda})P(\tilde{\lambda})^T\Big)$, $\quad \tilde{\omega}_2 = (h/n)\operatorname{tr}\Big(P(\tilde{\lambda})^2\Big)$.

By the mean value theorem applied to $r_L(\tilde{\lambda},\tilde{\sigma})$ in (8.5),

$$r_L(\hat{\lambda},\hat{\sigma}) = r_L(\lambda_0,\sigma_0) + \bar{s}_{1L}(\hat{\sigma}-\sigma_0) + \bar{s}_{2L}(\hat{\lambda}-\lambda_0),$$

where $\bar{s}_{1L} = (\partial/\partial\sigma)r_L(\bar{\lambda},\bar{\sigma})$ and $\bar{s}_{2L} = (\partial/\partial\lambda)r_L(\bar{\lambda},\bar{\sigma})$ are the first derivatives of r_L at some $(\bar{\lambda},\bar{\sigma})$ such that $|\bar{\lambda}-\lambda_0|\le|\tilde{\lambda}-\lambda_0|$ and $|\bar{\sigma}-\sigma_0|\le|\tilde{\sigma}-\sigma_0|$. Thus,

$$\begin{aligned} \hat{\lambda}-\lambda_0 &= (\tilde{\lambda}-\lambda_0)\left[1+\Big\{\tilde{\mathcal{J}}\tilde{\omega}_1+\tilde{\omega}_2\Big\}^{-1}\frac{h}{n}\cdot\bar{s}_{1L}\right] \\ &\quad + \Big\{\tilde{\mathcal{J}}\tilde{\omega}_1+\tilde{\omega}_2\Big\}^{-1}\frac{h}{n}\left[\bar{s}_{2L}(\tilde{\sigma}-\sigma_0)+r_L(\lambda_0,\sigma_0)\right]. \end{aligned} \tag{8.7}$$

Let $\mathcal{N} = \Big(\lambda,\sigma : |\lambda-\lambda_0|\le\sqrt{h/n}, |\sigma-\sigma_0|\le\sqrt{1/n}\Big)$ be a neighborhood of (λ_0,σ_0), which takes into account the different rates of convergence for the two parameters λ and σ.

As in Robinson (2010), the proof of consistency and asymptotic normality of the adaptive estimates $(\hat{\lambda},\hat{\sigma})$ consist of showing

$$\sqrt{\frac{h}{n}}r_L(\lambda_0,\sigma_0)\to_d N(0,\mathcal{J}\omega_1+\omega_2), \tag{8.8}$$

in addition to

$$\tilde{\omega}_1 \to_p \omega_1, \quad \tilde{\omega}_2 \to_p \omega_2, \tag{8.9}$$

$$\frac{h}{n} \cdot s_{1L}(\lambda_0, \sigma_0) \to_p -\mathcal{J}\omega_1 - \omega_2, \quad \frac{\sqrt{h}}{n} s_{2L}(\lambda_0, \sigma_0) \to_p 0, \tag{8.10}$$

$$\tilde{\mathcal{J}}_L(\lambda_0, \sigma_0) \to_p \mathcal{J}, \quad \sup_{\mathcal{N}} |\tilde{\mathcal{J}}_L(\lambda, \sigma) - \tilde{\mathcal{J}}_L(\lambda_0, \sigma_0)| = o_p(1), \tag{8.11}$$

$$\sup_{\mathcal{N}} |s_{iL}(\lambda, \sigma) - s_{iL}(\lambda_0, \sigma_0)| = o_p\left(\frac{n}{h}\right), \quad i = 1, 2. \tag{8.12}$$

Proof of (8.8). We verify (8.8), by establishing

$$\sqrt{\frac{h}{n}} \frac{\partial L(\theta_0)}{\partial \lambda} \to_d N(0, \mathcal{J}\omega_1 + \omega_2), \tag{8.13}$$

$$r_L(\lambda_0, \sigma_0) - \frac{\partial L(\theta_0)}{\partial \lambda} = o_p(\sqrt{n/h}). \tag{8.14}$$

To prove (8.13), write

$$\frac{\partial L(\theta_0)}{\partial \lambda} = \sum_{i=1}^{n} \frac{M_i^T(y - \mu_0 \ell)}{\sigma_0} \psi_i - tr(P) = (\psi_1, \cdots, \psi_n) P\varepsilon - tr(P) = \sum_{i=1}^{n} \eta_i,$$

as the sum of martingale differences $\eta_i := (\varepsilon_i \psi_i - 1) p_{ii} + \varepsilon_i \sum_{j<i} \psi_j p_{ij} + \psi_i \sum_{j<i} \varepsilon_j p_{ji}$, which satisfy $E(\eta_i | \mathcal{F}_{i-1}) = 0$, $\mathcal{F}_i = \sigma(\varepsilon_j, j \le i)$. Therefore, we establish (8.13) by verifying the following sufficient conditions of central limit theorem for martingale differences, see Hall and Heyde (1980):

$$\frac{h}{n} \sum_{i=1} E(\eta_i^2 | \mathcal{F}_{i-1}) \to_p \mathcal{J}\omega_1 + \omega_2, \tag{8.15}$$

$$\left(\frac{h}{n}\right)^{2+\delta} \sum_{i=1}^{n} E|\eta_i|^{2+\delta} \to 0. \tag{8.16}$$

Proof of (8.16). Firstly, noting $E(\varepsilon_i \psi_i) = 1, E(\varepsilon_i^2) = 1$ and $E(\psi_i^2) = \mathcal{J}$ and using *i.i.d.* property of $\{\varepsilon_i\}$, it follows that

$$\begin{aligned} E(\eta_i^2) &= p_{ii}^2[E(\varepsilon_i^2 \psi_i^2) - 1] + \mathcal{J} \sum_{1 \le j < i} p_{ij}^2 + \mathcal{J} \sum_{1 \le j < i} p_{ji}^2 + 2 \sum_{1 \le j < i} p_{ij} p_{ji}, \\ \sum_{i=1}^{n} E(\eta_i^2) &= \sum_{i=1}^{n} p_{ii}^2 [E(\varepsilon_i^2 \psi_i^2) - 2 - \mathcal{J}] + \mathcal{J} \sum_{i,j=1}^{n} p_{ij}^2 + \sum_{i,j=1}^{n} p_{ij} p_{ji} \\ &= O(1) \sum_{i=1}^{n} p_{ii}^2 + \mathcal{J} \mathrm{tr}(PP^T) + \mathrm{tr}(P^2). \end{aligned}$$

Therefore, by Assumption 2,

$$\frac{h}{n} \sum_{i=1}^{n} E(\eta_i^2) \to \mathcal{J}\omega_1 + \omega_2, \tag{8.17}$$

since $n^{-1}h\sum_{i=1}^n p_{ii}^2 = O\big(n^{-1}h \times h^{-2}n\big) = O\big(h^{-1}\big) = o(1)$. Direct calculation, noting that $E\varepsilon_i^2\psi_i = 0$ under Assumption 3, gives

$$\begin{aligned} E(\eta_i^2|\mathcal{F}_{i-1}) - E(\eta_i^2) &= \sum_{j,j'<i:j\neq j'} \psi_j\psi_{j'}p_{ij}p_{ij'} + \sum_{j<i} p_{ij}^2(\psi_j^2 - \mathcal{J}) \\ &+\mathcal{J}\sum_{j,j'<i:j\neq j'} \varepsilon_j\varepsilon_{j'}p_{ji}p_{j'i} + \mathcal{J}\sum_{j<i} p_{ji}^2(\varepsilon_j^2-1) + 2p_{ii}E(\psi_i^2\varepsilon_i)\sum_{j<i} p_{ji}\varepsilon_j \\ &+2\sum_{j<i}\sum_{j'<i:j\neq j'} \psi_j\varepsilon_{j'}p_{ij}p_{j'i} + 2\sum_{j<i}(\psi_j\varepsilon_j - 1)p_{ij}p_{ji} =: m_{1i} + \cdots + m_{7i}. \end{aligned}$$

In view of (8.17), to prove (8.16), it suffices to show that

$$\frac{h}{n}\sum_{i=1}^n \big[E(\eta_i^2|\mathcal{F}_{i-1}) - E(\eta_i^2)\big] = \frac{h}{n}\sum_{i=1}^n (m_{1i} + \cdots + m_{7i}) = o_p(1),$$

which is verified once we establish

$$E\big[\big(\frac{n}{h}\sum_{i=1}^n m_{di}\big)^2\big] = o\big(1\big), \quad \text{for} \quad d = 1, \cdots, 7. \tag{8.18}$$

We first verify (8.18) for d=1.

$$\begin{aligned} &E\big[\big(\sum_{i=1}^n m_{1i}\big)^2\big] = E\big[\big(\sum_{i=1}^n \sum_{j,j'<i:j\neq j'} \psi_j\psi_{j'}p_{ij}p_{ij'}\big)^2\big] \\ &\leq 2\sum_{i,i'=1}^n \sum_{j<i}\sum_{k<i'} |p_{ij}p_{i'j}p_{ik}p_{i'k}||E(\psi_j^2)E(\psi_k^2)| \leq C\sum_{i,i',j,k=1}^n |p_{ij}p_{i'j}p_{ik}p_{i'k}|. \end{aligned}$$

Thus from (8.2),

$$\begin{aligned} &E\big[\big(\sum_{i=1}^n m_{1i}\big)^2\big] \leq Ch^{-1}\sum_{i,i',j,k=1}^n |p_{ij}p_{i'j}p_{i'k}| \\ &\leq Ch^{-1}\big(\sum_{i=1}^n 1\big)\cdot \max_{i'}\sum_{k=1}^n |p_{i'k}| \max_j \sum_{i'=1}^n |p_{i'j}| \max_i \sum_{j=1}^n |p_{ij}| \leq C\Big(\frac{n}{h}\Big) = o\Big(\frac{n^2}{h^2}\Big). \end{aligned}$$

Verification of (8.18) for $d = 3, 6$ follows similar steps as in the proof for $d = 1$. To establish (8.18) for $d = 2$, recall that $\psi_i^2 - \mathcal{J} = \psi_i^2 - E\psi_i^2$ is an $i.i.d.$ sequence.

Thus,

$$
\begin{aligned}
E\big[\big(\sum_{i=1}^{n} m_{2i}\big)^2\big] &= E\big[\big(\sum_{i=1}^{n}\sum_{j<i} p_{ij}^2(\psi_j^2-\mathcal{J})\big)^2\big] \le E\big((\psi_1^2-E\psi_i^2)^2\big)\sum_{i,i',j=1}^{n} p_{ij}^2 p_{i'j}^2 \\
&= C\sum_{j=1}^{n}\Big(\sum_{i=1}^{n} p_{ij}^2\Big)^2 \le C\sum_{j=1}^{n}\Big(\max_i |p_{ij}|\sum_{i=1}^{n}|p_{ij}|\Big)^2 \\
&\le \big(\max_{i,j}|p_{ij}|\big)^2\sum_{j=1}^{n}\max_j\sum_{i=1}^{n}|p_{ij}|\sum_{i'=1}^{n}|p_{i'j}| \le Ch^{-2}nO(1)O(1) = o\Big(\frac{n^2}{h^2}\Big).
\end{aligned}
$$

Verifications of (8.18) for $d = 4, 5, 7$ follows similar steps.

Proof of (8.16). Since $|a+b|^{2+\delta} \le C(|a|^{2+\delta} + |b|^{2+\delta})$,

$$
\begin{aligned}
\sum_{i=1}^{n} E|\eta_i|^{2+\delta} &\le C\Big(\sum_{i=1}^{n}|p_{ii}|^{2+\delta}E|\varepsilon_i\psi_i|^{2+\delta} + \sum_{i=1}^{n}E|\varepsilon_i|^{2+\delta}E|\sum_{j<i}p_{ij}\psi_j|^{2+\delta} \\
&\qquad + \sum_{i=1}^{n}E|\psi_i|^{2+\delta}E|\sum_{j<i}p_{ji}\varepsilon_j|^{2+\delta}\Big) \\
&\le C\Big(\sum_{i=1}^{n}|p_{ii}|^{2+\delta} + \sum_{j<i}E|p_{ij}\psi_j|^{2+\delta} + \sum_{j<i}E|p_{ji}\varepsilon_j|^{2+\delta}\Big) =: C(c_{1n}+c_{2n}+c_{3n}).
\end{aligned}
$$

To prove (8.16), we need to verify that $c_{dn} = o\big((n/h)^{2+\delta}\big)$ for $d = 1, 2, 3$. Firstly, for $d = 1$, using $|p_{ii}| = O(1/h)$, $c_{1n} = O\big(h^{-2-\delta}n\big) = o\big((n/h)^{2+\delta}\big)$. For $d = 2$, by the Rosenthal inequality,

$$
\sum_{i=1}^{n} E|\sum_{j<i}p_{ij}\psi_j|^{2+\delta} \le C\sum_{i=1}^{n}\max\Big(\sum_{j=1}^{n}E|p_{ij}\psi_j|^{2+\delta}, \big(\sum_{j=1}^{n}E(p_{ij}^2\psi_j^2)\big)^{(2+\delta)/2}\Big),
$$

where for any $i = 1, \cdots, n$, by Assumption 3, (8.1) and (8.2)

$$
\begin{aligned}
\sum_{j=1}^{n}E|p_{ij}\psi_j|^{2+\delta} &= C\sum_{j=1}^{n}|p_{ij}|^{2+\delta} \le C\max_j|p_{ij}|^{1+\delta}\sum_{j=1}^{n}|p_{ij}| = O\big(\frac{1}{h^{1+\delta}}\big), \\
\Big(\sum_{j=1}^{n}E(p_{ij}^2\psi_j^2)\Big)^{(2+\delta)/2} &= C\Big(\sum_{j=1}^{n}p_{ij}^2\Big)^{(2+\delta)/2} \le C\Big(\max_j|p_{ij}|\sum_{j=1}^{n}|p_{ij}|\Big)^{(2+\delta)/2} = O\big(\frac{1}{h^{1+\delta/2}}\big).
\end{aligned}
$$

Therefore, $c_{2n} = O\big(h^{-1-\delta/2}n\big) = o\big((n/h)^{2+\delta}\big)$. Proof of $c_{3n} = o\big((n/h)^{2+\delta}\big)$ follows similar steps.

Proof of (8.14). Let, for the brevity, r_L and $\tilde{\psi}_{iL0}$ denote quantities evaluated at the true parameter values θ_0. Noting $MHy = MH(Q^{-1}\sigma_0\varepsilon+\mu_0 1_n) = \sigma_0 MHQ^{-1}\varepsilon$

since $H1_n = 0$, we can write

$$
\begin{aligned}
r_L - \frac{\partial L(\theta_0)}{\partial \lambda} &= \left(\tilde{\psi}_{1L0}, \cdots, \tilde{\psi}_{nL0}\right) MHQ^{-1}\varepsilon - \operatorname{tr}\{P\} - (\psi_1, ..., \psi_n)\, MQ^{-1}\varepsilon + \operatorname{tr}(P) \\
&= \left(\tilde{\psi}_{1L0}, \cdots, \tilde{\psi}_{nL0}\right) HMQ^{-1}\varepsilon + \left(\tilde{\psi}_{1L0}, \cdots, \tilde{\psi}_{nL0}\right) \frac{1}{n} 1_n 1_n^T MQ^{-1}\varepsilon \\
&\quad - \frac{1}{n}\left(\tilde{\psi}_{1L0}, \cdots, \tilde{\psi}_{nL0}\right) M 1_n 1_n^T Q^{-1}\varepsilon - (\psi_1, ..., \psi_n)\, MQ^{-1}\varepsilon \\
&= \left(\tilde{\psi}_{1L0} - \psi_1, \cdots, \tilde{\psi}_{nL0} - \psi_n\right) HP\varepsilon - (\psi_1, ..., \psi_n)\,(H - I)P\varepsilon \\
&\quad - \left(\tilde{\psi}_{1L0}, \cdots, \tilde{\psi}_{nL0}\right)(H - I)P\varepsilon + \left(\tilde{\psi}_{1L0}, \cdots, \tilde{\psi}_{nL0}\right) M(H - I)Q^{-1}\varepsilon \\
&= q_{n,1} - q_{n,2} - q_{n,3} + q_{n,4}.
\end{aligned}
$$

It remains to show $q_{n,i} = o_p(\sqrt{n/h})$, $i = 1, 2, 3, 4$. First consider $i = 2$. Denote $P = (p_{ij})$. Then,

$$
\begin{aligned}
q_{n,2} &= (\psi_1, ..., \psi_n)\,(H - I)P\varepsilon = -\frac{1}{n}(\psi_1, ..., \psi_n)\,(1_n 1_n^T)P\varepsilon \\
&= -n^{-1}\Big[\sum_{j=1}^n \varepsilon_j \big(\sum_{i=1}^n p_{ij}\big)\Big] \cdot \sum_{m=1}^n \psi_m,
\end{aligned}
$$

where $\sum_{m=1}^n \psi_m = \sum_{m=1}^n \psi(\varepsilon_m) = O_p(\sqrt{n})$ due to ε_j's being *i.i.d.* By Assumption 1, $\max_{1\le j\le n} \big(\sum_{i=1}^n |p_{ij}|\big) < C$ uniformly over j. Then, $n^{-1}E\big(\sum_{j=1}^n \varepsilon_j \big(\sum_{i=1}^n p_{ij}\big)\big) = 0$, and

$$
\operatorname{Var}\big(n^{-1}\sum_{j=1}^n \varepsilon_j \big(\sum_{i=1}^n p_{ij}\big)\big) = n^{-2}\big(\sum_{j=1}^n \big(\sum_{i=1}^n p_{ij}\big)^2\big) = O\left(n^{-1}\right).
$$

Hence, $q_{n,2} = O_p(n^{-1/2})O_p(n^{1/2}) = O_p(1) = o_p(\sqrt{n/h})$, because $n/h \to \infty$.

Next, we show $q_{n,1} = o_p(\sqrt{n/h})$. In the following quantities introduced below, the triangular array structure is present but the n-subscript is suppressed. Let

$$
t_{ij} := \ell_j^T P^T \ell_i - \ell_j^T P^T 1_n/n = p_{ij} - \frac{1}{n}\sum_{m=1}^n p_{mj}, \quad \chi_i := \varepsilon^T P^T(\ell_i - n^{-1}1_n) = \sum_{j=1}^n \varepsilon_j t_{ij},
$$

where ℓ_i stands for the i^{th} column of I and the equality $\sum_{i=1}^n \chi_i = 0$ holds, arising from $\sum_{i=1}^n t_{ij} = 0$ for $j = 1, ..., n$. As pointed out in Robinson (2010, p. 18), Assumption 1 implies $|t_{ij}| = O(1/h)$ uniformly over i and j, following from $\max_{1\le i,j\le n} |p_{ij}| = O(1/h)$. Let $a_l := \sum_{i=1}^n b_{li}\chi_i$, $l = 2, 3, 4$. Write

$$
\begin{aligned}
\tilde{\psi}^{(L)}(\lambda_0, \sigma_0) - \psi(\varepsilon_i) &= \left[\bar{\psi}^{(L)}\left(\varepsilon_i; a^{(L)}\right) - \psi(\varepsilon_i)\right] + \left[\psi^{(L)}\left(\varepsilon_i; \tilde{a}^{(L)}(\varepsilon)\right) - \bar{\psi}^{(L)}\left(\varepsilon_i; a^{(L)}\right)\right] \\
&+ \left[\tilde{\psi}^{(L)}(\lambda_0, \sigma_0) - \psi^{(L)}\left(\varepsilon_i; \tilde{a}^{(L)}(\varepsilon)\right)\right] =: c_{2i} + c_{3i} + c_{4i}. \qquad (8.19)
\end{aligned}
$$

We can rewrite

$$q_{n,1} = \left(\tilde{\psi}_{1L0} - \psi_1, \cdots, \tilde{\psi}_{nL0} - \psi_n\right) HP\varepsilon = \sum_{i=1}^{n} c_{2i}\chi_i + \sum_{i=1}^{n} c_{3i}\chi_i + \sum_{i=1}^{n} c_{4i}\chi_i := a_2 + a_3 + a_4. \quad (8.20)$$

To prove $q_{n,1} = o_p(\sqrt{n/h})$, we show that

$$a_\ell = o_p(\sqrt{n/h}), \quad \ell = 2, 3, 4. \quad (8.21)$$

Proof of (8.21) for $i = 2$. It requires the projection error, arising from projecting the score function onto the space spanned by the functionals of our series estimation, to be of small enough order, as required in Assumption 6.

Write down a_2 as in (A.27) of Robinson (2010):

$$a_2 = \sum_{i=1}^{n} c_{2i}\varepsilon_i t_{ii} + \sum_{i,j=1:j\neq i}^{n} c_{2i}\varepsilon_i t_{ij}, \quad (8.22)$$

recalling $c_{2i} = \bar{\psi}^{(L)}\left(\varepsilon_i; a^{(L)}\right) - \psi(\varepsilon_i)$. Then,

$$E\Big|\sum_{i=1}^{n} c_{2i}\varepsilon_i t_{ii}\Big| \leq \left\{E\left(c_{2i}^2\right)\right\}^{\frac{1}{2}} \sum_{i=1}^{n} |t_{ii}| = o\Big(\sqrt{\frac{h}{n}}\Big) \cdot O\Big(\frac{n}{h}\Big) = o\left(\sqrt{\frac{n}{h}}\right),$$

by Assumptions 1 and 6. The second term of (8.22) has zero mean and

$$\begin{aligned} \mathrm{Var}\Big(\sum_{i,j=1:j\neq i}^{n} c_{2i}\varepsilon_i t_{ij}\Big) &= E\Big[\big(\sum_{i<j} c_{2i}\varepsilon_i t_{ij}\big) + \big(\sum_{j\leq i} c_{2i}\varepsilon_i t_{ij}\big)\Big]^2 \\ &\leq 2E\Big[\big(\sum_{i<j} c_{2i}\varepsilon_i t_{ij}\big)^2\Big] + 2E\Big[\big(\sum_{j\leq i} c_{2i}\varepsilon_i t_{ij}\big)^2\Big]. \quad (8.23) \end{aligned}$$

The first expectation can be bounded by

$$\begin{aligned} &2\sum_{i<j}\sum_{i'<j'} \big|E[c_{2i}\varepsilon_j c_{2i'}\varepsilon_{j'}] t_{ij} t_{i'j'}\big| \\ &\quad \leq 2\sum_{i<j} E(c_{2i}^2)E(\varepsilon_i^2)|t_{ij}t_{i'j'}| = 2\sum_{i<j} O_p\big(\frac{h}{n}\big) O_p\big(\frac{1}{h}\big) O_p\big(\frac{1}{h}\big) = o_p\big(\frac{n}{h}\big), \end{aligned}$$

using independence of ε_j's, the bound $Ec_{i2}^2 = o_p(h/n)$ from Assumption 6 and $t_{ij} = O(1/h)$.

The same bound holds for the second term in (8.23) which yields $a_2 = o_p(\sqrt{n/h})$.

To prove (8.21) for $i = 3, 4$, we shall use the following notation. Let

$$\begin{aligned} \pi_L &:= (\log L)\eta^{2L} 1(\varphi < 1) + (L \log L)\eta^{2L} 1(\varphi = 1) + (\log L)(\eta\varphi)^{2L} 1(\varphi > 1) \\ &\leq L(\log L)A^{2L}, \quad (8.24) \end{aligned}$$

with $A = \eta \max(\varphi, 1)$. Note that $A > 1$.

Set

$$\begin{aligned}\rho_{uL} &= CL, \quad \text{if } u=0,\\ &= (CL)^{uL/\omega}, \quad \text{if } u>0 \text{ and Assumption 5(ii) holds},\\ &= C^L, \quad \text{if } u>0 \text{ and Assumption 5(iii) holds.}\end{aligned}$$

Proof of (8.21) for $i=3$. Proof is based on an extensive use of Assumption 5. Equations (A.31)-(A.39) of pages 19-20 of Robinson (2010) yield the upper bound on the stochastic order of a_3:

$$\begin{aligned}a_3 &= O_p\left(\frac{\sqrt{n}}{h}L^{3/2}\rho_{2\kappa L}\rho_{4\kappa L}^{\frac{1}{2}}\pi_L^2\right)\\ &= O_p\left(\sqrt{\frac{n}{h}}\frac{H_3}{\sqrt{h}}\right), \quad H_3 := L^{3/2}\rho_{2\kappa L}\rho_{4\kappa L}^{\frac{1}{2}}\pi_L^2. \qquad (8.25)\end{aligned}$$

To prove (8.21) for $i=3$, it remains to show

$$H_3 = o(\sqrt{h}). \qquad (8.26)$$

Case 1. Let Assumption 5 (i) hold. Then, $\rho_{2\kappa L} = \rho_{4\kappa L} = CL$ and $H_3 = C^{3/2}L^3\pi_L^2$. Notice that for any $p>0$ and $\varepsilon>0$,

$$L^p = o\big((1+\varepsilon)^L\big). \qquad (8.27)$$

Hence, as $L\to\infty$,

$$\pi_L^2 = o\big((1+\varepsilon)^L A^{4L}\big), \quad \forall\varepsilon>0. \qquad (8.28)$$

Combining (8.27) and (8.28), we obtain $H_3 = o\big([(1+\varepsilon)A]^{4L}\big), \quad \forall\varepsilon>0$.

Thus, to prove that $H_3 = o(\sqrt{h})$, it suffices to show that

$$\begin{aligned}&[(1+\varepsilon)A]^{4L} \le \sqrt{h}, \quad \text{i.e.} \qquad (8.29)\\ &4L\log[(1+\varepsilon)A] \le (1/2)\log h, \quad \text{or} \quad L \le \frac{\log h}{8\log[(1+\varepsilon)A]},\end{aligned}$$

which is valid for small $\varepsilon\ge 0$ by Assumption 5 (i).

Case 2. Let Assumption 5 (ii) hold. Then, $\rho_{aL} = (CL)^{\frac{aL}{\omega}}$ and

$$H_3 = L^{\frac{3}{2}}\rho_{2\kappa L}\rho_{4\kappa L}^{\frac{1}{2}}\pi_L^2 = L^{\frac{3}{2}}C^{\frac{4\kappa L}{\omega}}L^{\frac{4\kappa L}{\omega}}\pi_L^2.$$

Observe that for any $C>0$, $p>0$, $a>0$ and $\varepsilon>0$,

$$L^p = o(L^{\varepsilon L}), \quad C^{aL} = o(L^{\varepsilon L}). \qquad (8.30)$$

Hence by (8.24),

$$\pi_L^2 = o(L^{\varepsilon L}), \quad \forall\varepsilon>0, \qquad (8.31)$$

and $H_3 = o\big(L^{L(\frac{4\kappa}{\omega}+\varepsilon)}\big)$, $\forall\varepsilon>0$. Thus, $H_3 = o(\sqrt{h})$ holds if

$$\begin{aligned}&L^{L(\frac{4\kappa}{\omega}+\varepsilon)} \le \sqrt{h}, \quad \text{i.e.} \qquad (8.32)\\ &\big(\frac{4\kappa}{\omega}+\varepsilon\big)L\log L \le \frac{1}{2}\log h, \quad \text{or} \quad L\log L \le \frac{\log h}{2\big(\frac{4\kappa}{\omega}+\varepsilon\big)},\end{aligned}$$

which is valid for small $\varepsilon > 0$ by Assumption 5 (ii).

Case 3. Let Assumption 5(iii) hold. Then, $\rho_{aL} = C^L$, and $H_3 = L^{\frac{3}{2}}\rho_{2\kappa L}\rho_{4\kappa L}^{\frac{1}{2}}\pi_L^2 = L^{\frac{3}{2}}C^{\frac{3L}{2}}\pi_L^2$. Then by (8.30) and (8.31), $H_3 = o(L^{\varepsilon L}), \quad \forall\varepsilon > 0$. Thus, $H_3 = o(\sqrt{h})$, if

$$L^{\varepsilon L} \leq \sqrt{h}, \quad \text{i.e.} \tag{8.33}$$
$$\varepsilon L \log L \leq \frac{1}{2}\log h, \quad \text{or} \quad L\log L \leq \frac{1}{2\varepsilon}\log h,$$

which is valid for sufficiently small $\varepsilon > 0$ by Assumption 5 (ii).

Now, we prove (8.21) for $i = 4$. Following (A.45)-(A.56) of Robinson (2010), we obtain the following upper bound

$$a_4 = O_p\big(\frac{\sqrt{n}}{h}H_4\big),$$
$$H_4 := \rho_{2\kappa L}\pi_L \times \left\{C^{\kappa L}L^{\frac{7}{2}} + \rho_{2\kappa L}\pi_L L^2 + \rho_{2\kappa L}\pi_L(CL)^{4\kappa L+3}n^{-\frac{1}{2}}\log n\right\}.$$

It remains to show that

$$H_4 = o_p(\sqrt{h}). \tag{8.34}$$

Case 1. Under Assumption 5 (i), $\rho_{2\kappa L} = CL$, and

$$H_4 = \pi_L L^{\frac{9}{2}} + \pi_L^2 L^4 + \pi_L^2 L^5 n^{-\frac{1}{2}}\log n.$$

By (8.27) and (8.28),

$$H_4 = o\big([(1+\varepsilon)A]^{2L} + [(1+\varepsilon)A]^{4L}(1 + n^{-1/2}\log n)\big) = o\big([(1+\varepsilon)A]^{4L}\big).$$

Hence $H_4 = o(\sqrt{h})$, if $[(1+\varepsilon)A]^{4L} \leq \sqrt{h}$, which is true for small $\varepsilon > 0$ as shown in (8.29).

Case 2. Let Assumption 5 (ii) hold. Then $\rho_{aL} = (CL)^{\frac{aL}{\omega}}$ and

$$H_4 = C^{\kappa L}(CL)^{2\kappa L/\omega}L^{7/2}\pi_L + (CL)^{4\kappa L/\omega}L^2\pi_L^2 + \frac{(CL)^{4\kappa L(1+1/\omega)}L^3\pi_L^2}{\sqrt{n}/\log n}.$$

By (8.30) and (8.31),

$$\begin{aligned} H_4 &= o\big(L^{(\frac{2\kappa}{\omega}+\varepsilon)L} + L^{(\frac{4\kappa}{\omega}+\varepsilon)L} + \frac{L^{(4\kappa(1+1/\omega)+\varepsilon)L}}{\sqrt{n}/\log n}\big) \\ &= o\Big(L^{(\frac{4\kappa}{\omega}+\varepsilon)L}(1 + \frac{L^{4\kappa L}}{\sqrt{n}/\log n})\Big). \end{aligned}$$

By (8.32), $L^{(\frac{4\kappa}{\omega}+\varepsilon)L} \leq \sqrt{h}$, if $\varepsilon > 0$ is small. Next, for any $\delta > 0$, $\sqrt{n}/\log n \geq n^{\frac{1}{2}-\delta} \geq h^{\frac{1}{2}-\delta}$. Hence by the same arguments as in proving (8.29), we obtain that

$$\frac{L^{4\kappa L}}{n^{1/2}/\log n} \leq \frac{L^{4\kappa L}}{n^{\frac{1}{2}-\delta}} \leq 1, \tag{8.35}$$

if $L\log L \leq (\frac{1}{2}-\delta)\log h/4\kappa$ which holds for small δ. Hence $H = o(\sqrt{h})$, and (8.34) holds.

Case 3. Under Assumption 5(iii), $\rho_{aL} = C^L$ and

$$H_4 = C^{(\kappa+1)L} L^{7/2} \pi_L + C^{2L} L^2 \pi_L^2 + C^{(4\kappa+2)L+3} L^{4\kappa L+3} \pi_L^2 n^{-\frac{1}{2}} \log n.$$

By (8.30) and (8.31), $H_4 = o\big(L^{\varepsilon L} + \frac{L^{(4\kappa+\varepsilon)L}}{\sqrt{n}/\log n}\big)$. By (8.33), $L^{\varepsilon L} \leq \sqrt{h}$. Hence, to prove that $H_4 = o(\sqrt{h})$, it remains to show that

$$\frac{L^{(4\kappa+\varepsilon)L}}{\sqrt{n}/\log n} \leq \frac{L^{(4\kappa+\varepsilon)L}}{\sqrt{h}/\log h} \leq \sqrt{h},$$

where the first inequality holds because $h \leq n$. For that we shall verify that for small $\delta > 0$, $L^{(4\kappa+\varepsilon)L} \leq h^{1-\delta}$, i.e.

$$(4\kappa + \varepsilon) L \log L \leq (1-\delta) \log h, \quad \text{or} \quad L \log L \leq \frac{(1-\delta)}{(4\kappa+\delta)} \log h,$$

which follows from Assumption 5 (iii) when δ and ε are small enough.

Now we show $q_{n,3} = o_p(\sqrt{n/h})$. Note

$$\begin{aligned} q_{n,3} &= q_{n,2} - \left(\tilde{\psi}_{1L0} - \psi_1, \cdots, \tilde{\psi}_{nL0} - \psi_n\right) \frac{1}{n} 1_n 1_n^T P \varepsilon \\ &= q_{n,2} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\tilde{\psi}_{iL0} - \psi_i\right) \frac{1}{\sqrt{n}} \sum_{i,j=1}^{n} p_{ij} \varepsilon_j. \end{aligned}$$

From (8.2) implied by Assumption 1,

$$Var\big(\frac{1}{\sqrt{n}} \sum_{i,j=1}^{n} p_{ij} \varepsilon_j\big) = \frac{1}{n} \sum_{i=1}^{n} \big(\sum_{j=1}^{n} p_{ij}\big)^2 = O(1).$$

Therefore, $q_{n,3} = o_p(\sqrt{n/h})$ if $\sum_{i=1}^n \left(\tilde{\psi}_{iL0} - \psi_i\right)/\sqrt{n} = o(\sqrt{n/h})$, which in turn follows if for all $1 \leq i \leq n$,

$$|\tilde{\psi}_{iL0} - \psi_i| = o_p(1/\sqrt{h}). \tag{8.36}$$

Proof of (8.36). Recall

$$\begin{aligned} \tilde{\psi}^{(L)}(\lambda_0, \sigma_0) - \psi(\varepsilon_i) \;=\; & \left[\bar{\psi}^{(L)}\left(\varepsilon_i; a^{(L)}\right) - \psi(\varepsilon_i)\right] + \left[\psi^{(L)}\left(\varepsilon_i; \tilde{a}^{(L)}(\varepsilon)\right) - \bar{\psi}^{(L)}\left(\varepsilon_i; a^{(L)}\right)\right] \\ + \; & \left[\tilde{\psi}^{(L)}(\lambda_0, \sigma_0) - \psi^{(L)}\left(\varepsilon_i; \tilde{a}^{(L)}(\varepsilon)\right)\right] = c_{2i} + c_{3i} + c_{4i}, \end{aligned}$$

and we will verify $c_{di} = o_p(\sqrt{1/h})$ for $d = 2, 3, 4$. From Assumption 6, $c_{2i} = o_p(\sqrt{h/n}) = o_p(\sqrt{1/h})$.

We have

$$c_{3i} \;=\; \bar{\phi}^{(L)}(\varepsilon_i)^T (\tilde{a}^{(L)}(\varepsilon) - a^{(L)}) + (\Phi^{(L)}(\varepsilon_i) - \bar{\phi}^{(L)}(\varepsilon_i))^T \tilde{a}^{(L)}(\varepsilon) = o_p(H_3 + H_4) = o_p(\sqrt{1/h}),$$

from Lemma 9 and 10 of Robinson (2005).

Similarly, Lemma 10 and 19 of Robinson (2005) together with

$$\|\Phi^{(L)}(\epsilon_i/\sigma_0) - \Phi^{(L)}(\varepsilon_i)\| = O_p\big(\frac{1}{\sqrt{n}} C^L L^{3/2} \rho_{2\kappa L}^{1/2}\big), \tag{8.37}$$

imply

$$\begin{aligned} c_{4i} &= \Phi^{(L)}(\varepsilon_i)^T(\tilde{a}^{(L)}(\epsilon/\sigma_0) - \tilde{a}^{(L)}(\varepsilon)) + (\Phi^{(L)}(\epsilon_i/\sigma_0) - \Phi^{(L)}(\varepsilon_i))^T\tilde{a}^{(L)}(\epsilon/\sigma_0) \\ &= o_p(H_3 + H_4) = o_p(\sqrt{1/h}), \end{aligned}$$

where $\epsilon = HQy$.

To see (8.37), note that using mean value theorem with ϵ_i^* denoting some point that lies between ϵ_i/σ_0 and ε_i and recalling $\epsilon_i/\sigma_0 = \varepsilon_i - \bar{\varepsilon}$, we have that for each $\ell = 1, ..., L$, $\phi_\ell(\epsilon_i/\sigma_0) - \phi_\ell(\varepsilon_i) = -\bar{\varepsilon}\phi'_\ell(\varepsilon_i) + \frac{1}{2}\bar{\varepsilon}^2\phi''_\ell(\epsilon_i^*)$. Hence,

$$|\phi_\ell(\epsilon_i/\sigma_0) - \phi_\ell(\varepsilon_i)| \leq |\bar{\varepsilon}\phi'_\ell(\varepsilon_i)| + |\bar{\varepsilon}^2\phi''_\ell(\epsilon_i^*)|$$

Using arguments deployed in (A.46)-(A.52) of Robinson (2010)

$$E|\bar{\varepsilon}\phi'_\ell(\epsilon_i)| = O(\frac{1}{\sqrt{n}}\ell\mu_{2\kappa(\ell+K)}^{1/2}), \quad E|\bar{\varepsilon}^2\phi''_\ell(\epsilon_i^*)| = O\big(\frac{1}{n}C^{\kappa\ell}\ell^2(1 + \mu_{2\kappa(\ell-1+2K)}^{1/2})\big).$$

Hence due to Lemma 9 of Robinson (2005),

$$\|\Phi^{(L)}(\epsilon_i/\sigma_0) - \Phi^{(L)}(\varepsilon_i)\| = O_p(\frac{1}{\sqrt{n}}L^{3/2}\rho_{2\kappa L}^{1/2} + \frac{1}{n}C^{\kappa L}L^{5/2}\rho_{2\kappa L}^{1/2}) = O_p(\frac{1}{\sqrt{n}}C^L L^{3/2}\rho_{2\kappa L}^{1/2}).$$

Finally, we verify

$$q_{n,4} = (\psi_1, \cdots, \psi_n)\, M\frac{1}{n}1_n 1_n^T Q^{-1}\varepsilon + \Big(\tilde{\psi}_{1L0} - \psi_1, \cdots, \tilde{\psi}_{nL0} - \psi_n\Big)\, M\frac{1}{n}1_n 1_n^T Q^{-1}\varepsilon = o_p(\sqrt{n/h}).$$

First, we have

$$(\psi_1, \cdots, \psi_n)\, M\frac{1}{n}1_n 1_n^T Q^{-1}\varepsilon = \frac{1}{\sqrt{n}}\,(\psi_1, \cdots, \psi_n)\, M 1_n \frac{1}{\sqrt{n}}1_n^T Q^{-1}\varepsilon = O(1) = o(\sqrt{n/h})$$

since $Var\big(\,(\psi_1, ..., \psi_n)\, M1_n/\sqrt{n}\big) = n^{-1}\mathcal{J}\sum_{i=1}^n\big(\sum_{j=1}^n m_{ij}\big)^2 = O(1)$, and $Var\big(1_n^T Q^{-1}\varepsilon/\sqrt{n}\big) = n^{-1}\sum_{i=1}^n\big(\sum_{j=1}^n Q^{ij}\big)^2 = O(1)$. Now,

$$\Big(\tilde{\psi}_{1L0} - \psi_1, \cdots, \tilde{\psi}_{nL0} - \psi_n\Big)\, M1_n 1_n^T Q^{-1}\varepsilon = \frac{1}{\sqrt{n}}\Big(\tilde{\psi}_{1L0} - \psi_1, \cdots, \tilde{\psi}_{nL0} - \psi_n\Big)\, M1_n\frac{1}{\sqrt{n}}1_n^T Q^{-1}\varepsilon,$$

and we have from above $1_n^T Q^{-1}\varepsilon/\sqrt{n} = O_p(1)$. Therefore, we need to establish

$$\frac{1}{\sqrt{n}}\Big(\tilde{\psi}_{1L0} - \psi_1, \cdots, \tilde{\psi}_{nL0} - \psi_n\Big)\, M1_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(\tilde{\psi}_{iL0} - \psi_i\Big)\,\big(\sum_{j=1}^n m_{ij}\big) = o(\sqrt{n/h}).$$

Since $\max_{1\leq i\leq n}\big(\sum_{j=1}^n |m_{ij}|\big) = O(1)$,

$$\frac{\sqrt{h}}{n}\sum_{i=1}^n\Big(\tilde{\psi}_{iL0} - \psi_i\Big)\,\big(\sum_{j=1}^n m_{ij}\big) \leq \max_{1\leq i\leq n}\big(\sum_{j=1}^n |m_{ij}|\big)\frac{\sqrt{h}}{n}\sum_{i=1}^n|\tilde{\psi}_{iL0} - \psi_i| = o_p(1),$$

which follows from (8.36) shown above.

This completes the proof of (8.8), which is by far the most difficult and distinctive part of the Theorem's proof. We omit the proofs of (8.9)-(8.12), which follow standard arguments. ■

Proof of Theorem 2. Theorem 2 (i) follows rather straightforwardly from the Proof of Theorem 1, applying (8.8)-(8.12) to the case $\lambda = 0$. Here we prove Theorem 2 (ii), focusing on additional considerations arising from the use of fitted residuals $\tilde{y}$ in place of y.

In the Spatial Error model considered in Theorem 2 (ii), denote $\tilde{\epsilon}'^{(r)} = -M(0)H\tilde{y}$ and $\tilde{\psi}_{iL}^{(r)}(\sigma) := \Phi^L(\tilde{\epsilon}_i^{(r)}/\sigma)^T \tilde{a}^L(\tilde{\epsilon}^{(r)}/\sigma)$. Let

$$
\begin{aligned}
r_{L,SEM}(0,\sigma) &:= -\sum_{i=1}^{n} \tilde{\psi}_{iL}^{(r)}(\sigma)\frac{\tilde{\epsilon}_i'^{(r)}}{\sigma} - \operatorname{tr}\{P(0)\} \\
&= \frac{1}{\sigma}\left(\tilde{\psi}_{1L}^{(r)}(\sigma), \cdots, \tilde{\psi}_{nL}^{(r)}(\sigma)\right) M(0)H(y + X(\beta - \tilde{\beta})) - \operatorname{tr}\{P(0)\}.
\end{aligned}
$$

One can verify $\sqrt{\frac{h}{n}} r_{L,SEM}(0,\sigma_0) \to_d N(0, \mathcal{J}\omega_1(0) + \omega_2(0))$ by establishing

$$
\sqrt{\frac{h}{n}} \frac{\partial L(\theta_0)}{\partial \lambda}\Big|_{H_0} \to_d N(0, \mathcal{J}\omega_1(0) + \omega_2(0)), \tag{8.38}
$$

$$
r_{L,SEM}(0,\sigma_0) - \frac{\partial L(\theta_0)}{\partial \lambda}\Big|_{H_0} = o_p(\sqrt{n/h}). \tag{8.39}
$$

(8.38) follows on from (8.13).

Proof of (8.39). Let $r_{L,SEM}$, $\tilde{\psi}_{iL0}^{(r)}$, M and P denote quantities evaluated at true parameter $\theta_0 = (0, \mu_0, \sigma_0)^T$ under H_0. Noting $Q(0) = I_n$,

$$
\begin{aligned}
& r_{L,SEM} - \frac{\partial L(\theta_0)}{\partial \lambda}\Big|_{H_0} \\
&= \left(\tilde{\psi}_{1L0}^{(r)}, \cdots, \tilde{\psi}_{nL0}^{(r)}\right) MH(\varepsilon + \frac{1}{\sigma_0}X(\beta - \tilde{\beta})) - \operatorname{tr}\{P\} - (\psi_1, ..., \psi_n) M\varepsilon + \operatorname{tr}\{P\} \\
&\quad = \left(\tilde{\psi}_{1L0}^{(r)} - \psi_1, \cdots, \tilde{\psi}_{nL0}^{(r)} - \psi_n\right) HM\varepsilon - (\psi_1, ..., \psi_n)(H - I)M\varepsilon \\
&\qquad - \left(\tilde{\psi}_{1L0}^{(r)}, \cdots, \tilde{\psi}_{nL0}^{(r)}\right)(H - I)M\varepsilon + \left(\tilde{\psi}_{1L0}^{(r)}, \cdots, \tilde{\psi}_{nL0}^{(r)}\right) M(H - I)\varepsilon \\
&+ \frac{1}{\sigma_0}\left(\tilde{\psi}_{1L0}^{(r)} - \psi_1, \cdots, \tilde{\psi}_{nL0}^{(r)} - \psi_n\right) MHX(\beta - \tilde{\beta}) + \frac{1}{\sigma_0}(\psi_1, ..., \psi_n) MHX(\beta - \tilde{\beta}) \\
&\quad = q_{n,1} - q_{n,2} - q_{n,3} + q_{n,4} + q_{n,5} + q_{n,6}.
\end{aligned}
$$

We need to show $q_{n,i} = o_p(\sqrt{n/h})$, $i = 1, ..., 6$. We closely follow the proof of (8.14) given in the proof of Theorem 1. For $i = 2$, the proof given in that of (8.14) applies without any change.

Introduce notations $d_i = \tilde{\epsilon}_i/\sigma_0 - \varepsilon_i = -\bar{\varepsilon} + (X_i - \bar{X})^T(\beta - \tilde{\beta})/\sigma_0$ and $\delta_{i\ell} = \phi_\ell(\tilde{\epsilon}_i/\sigma_0) - \phi_\ell(\varepsilon_i)$. Denote $r_i = (X_i - \bar{X})^T(\beta - \tilde{\beta})/\sigma_0$. By mean value theorem

$\delta_{\ell i} = (-\bar{\varepsilon} + r_i)\phi'_\ell(\varepsilon_i) + \frac{1}{2}(\bar{\varepsilon}^2 - 2\bar{\varepsilon}r_i + r_i^2)\phi''_\ell(\varepsilon_i^*)$ for $|\varepsilon_i^* - \varepsilon_i| \le |d_i|$ and terms that contain r_i are new to the SEM model case.

For $i = 1$, the term c_{4i} in (8.19) is affected by the presence of $HX(\beta - \tilde{\beta})$ in the argument of $\tilde{\psi}_{iL0}^{(r)}$, which is in the term a_4 of (8.20). a_4 can be decomposed as

$$a_4 = \{\tilde{a}^{(L)}(\frac{\tilde{\epsilon}}{\sigma_0}) - \tilde{a}^{(L)}(\varepsilon)\}\sum_{i=1}^n \Phi^{(L)}(\varepsilon_i)\chi_{in} + \tilde{a}^{(L)}(\frac{\tilde{\epsilon}}{\sigma_0})\sum_{i=1}^n\{\Phi^{(L)}(\frac{\tilde{\epsilon}_i}{\sigma_0}) - \Phi^{(L)}(\varepsilon_i)\}\chi_{in} \tag{8.40}$$

Analysing the first term is unaffected and follows the same steps as in the proof of Theorem 1 since Lemma 19 of Robinson (2005) continues to apply. For the second term in (8.40), we need to show that

$$\tilde{a}^{(L)}(\frac{\tilde{\epsilon}}{\sigma_0})\big[\sum_{i=1}^n\{\Phi^{(L)}(\frac{\tilde{\epsilon}_i}{\sigma_0}) - \Phi^{(L)}(\varepsilon_i)\}\chi_{in}\big] = o_p(\sqrt{n/h}). \tag{8.41}$$

From Lemma 10, 19 of Robinson (2005) and Assumption 5, $\|\tilde{a}^{(L)}(\tilde{\epsilon}/\sigma_0)\| = O_p\big(L\rho_{2\kappa L}^{1/2}\pi_L\big)$. The sum in the square bracket in (8.41) has norm bounded by

$$\|\sum_{i=1}^n\{\Phi^{(L)}(\frac{\tilde{\epsilon}_i}{\sigma_0}) - \Phi^{(L)}(\varepsilon_i)\}\chi_{in}\| \le \Big(\sum_{\ell=1}^L\big(\sum_{i=1}^n \chi_{in}\delta_{\ell i}\big)^2\Big)^{1/2}.$$

We analyse for $\ell = 1, ..., L$

$$\big(\sum_{i=1}^n \chi_{in}\delta_{\ell i}\big)^2 = \Big(\sum_{i=1}^n \chi_{in}\big([-\bar{\varepsilon}\phi'_\ell(\varepsilon_i) + \frac{1}{2}\bar{\varepsilon}^2\phi''_\ell(\varepsilon_i^*)] + [r_i\phi'_\ell(\varepsilon_i) + \frac{1}{2}(-2\bar{\varepsilon}r_i + r_i^2)\phi''_\ell(\varepsilon_i^*)]\big)\Big)^2$$

$\big(\sum_i \chi_{in}(-\bar{\varepsilon}\phi'_\ell(\varepsilon_i) + \bar{\varepsilon}^2\phi''_\ell(\varepsilon_i^*)/2)\big)^2$ was the term considered in Theorem 1. In proof of Theorem 1, we followed Robinson's (2010) steps (A.46)-(A.54) which did not involve any r_i terms and all steps there continue to apply here for corresponding terms. Hence, we only need to focus on terms that contain r_i, i.e. latter quadratic term:

$$\Big(\sum_{i=1}^n \chi_{in}(r_i\phi'_\ell(\varepsilon_i) + \frac{1}{2}(-2\bar{\varepsilon}r_i + r_i^2)\phi''_\ell(\varepsilon_i^*))\Big)^2.$$

In the term above, two square terms are dominant so we focus on

$$\big(\sum_{i=1}^n \chi_{in}r_i\phi'_\ell(\varepsilon_i)\big)^2 + \big(\sum_{i=1}^n \chi_{in}(-2\bar{\varepsilon}r_i + r_i^2)\phi''_\ell(\varepsilon_i^*)\big)^2. \tag{8.42}$$

For the first term of (8.42), by Cauchy-Schwarz inequality

$$\big(\frac{1}{\sigma}\sum_{i=1}^n \chi_{in}\phi'_\ell(\varepsilon_i)(X_i - \bar{X})^T(\beta - \tilde{\beta})\big)^2 \le \frac{1}{\sigma^2}\|\beta - \tilde{\beta}\|^2\|\sum_{i=1}^n \chi_{in}\phi'_\ell(\varepsilon_i)(X_i - \bar{X})\|^2$$

We know $\|\beta-\tilde{\beta}\|^2=O_p(n^{-1})$. By Assumption 8

$$\|\sum_{i=1}^n \chi_{in}\phi'_\ell(\varepsilon_i)(X_i-\bar{X})\|^2=O_p\Big(\mathrm{tr}\big(\sum_{i,i'=1}^n E(\chi_{in}\chi_{i'n}\phi'_\ell(\varepsilon_i)\phi'_\ell(\varepsilon'_i))E((X_i-\bar{X})(X_{i'}-\bar{X})^T)\Big).$$

For $d=1,...,k$:

$$\sum_{i,i'=1}^n E(\chi_{in}\chi_{i'n}\phi'_\ell(\varepsilon_i)\phi'_\ell(\varepsilon_i))E((X_{id}-\bar{X}_d)(X_{i'd}-\bar{X}_d)) \tag{8.43}$$

$$=-\frac{Var(X_{1d})}{n}\sum_{i,i'=1}^n E(\chi_{in}\chi_{i'n}\phi'_\ell(\varepsilon_i)\phi'_\ell(\varepsilon_i))+\frac{n+2}{n}Var(X_{1d})\sum_{i=1}^n E(\chi_{in}^2\phi'_\ell(\varepsilon_i)^2)$$

noting $E((X_{1d}-\bar{X}_d)(X_{2d}-\bar{X}_d))=-Var(X_{1d})/n$. By Assumption 8,

$$\sum_{i,i'=1}^n E(\chi_{in}\chi_{i'n}\phi'_\ell(\varepsilon_i)\phi'_\ell(\varepsilon_i))=\sum_{i,i'=1}^n\sum_{j,j'=1}^n t_{ij}t_{i'j'}E(\varepsilon_j\varepsilon'_j\phi'_\ell(\varepsilon_i)\phi'_\ell(\varepsilon'_i))$$

$$=E(\varepsilon_1^2)E^2(\phi'_\ell(\varepsilon_1))\sum_{i,i',j=1}^n t_{ij}t_{i'j}+E^2(\varepsilon_1\phi'_\ell(\varepsilon_1))\sum_{i,i'=1}^n(t_{ii}t_{i'i'}+t_{ii'}t_{i'i})$$

$$+\big(E(\varepsilon_1^2)E(\phi'_\ell(\varepsilon_1)^2)-E(\varepsilon_1^2)E^2(\phi'_\ell(\varepsilon_1))\big)\sum_{i,i'=1}^n t_{ii'}^2$$

$$+\big(E(\varepsilon_1^2\phi'_\ell(\varepsilon_1))E(\phi'_\ell(\varepsilon_1))-E(\varepsilon_1^2)E^2(\phi'_\ell(\varepsilon_1)\big)\sum_{i,i'=1}^n(t_{ii}t_{i'i}+t_{ii'}t_{i'i'})$$

$$+\big(E(\varepsilon_1^2\phi'_\ell(\varepsilon_1)^2)-E(\varepsilon_1^2)E^2(\phi'_\ell(\varepsilon_1)-2E^2(\varepsilon_1\phi'_\ell(\varepsilon_1))\big)\sum_{i=1}^n t_{ii}^2=O(\frac{n^2}{h^2}\ell^2\mu_{2\kappa(\ell+K)})$$

noting that the second term on the right hand side is the dominant term, since

$$\sum_{i,j=1}^n t_{ij}^2=O\Big(\frac{n}{h}\Big),\quad \sum_{i,i'=1}^n|t_{ii}t_{i'i'}|=O(\frac{n^2}{h^2}),\quad \sum_{i,i',j=1}^n|t_{ij}t_{i'j}|=O(n),$$

$$\sum_{i,i'=1}^n|t_{ii'}t_{i'i}|=O(\frac{n}{h}),\quad \sum_{i,i'=1}^n|t_{ii}t_{i'i}|=O(\frac{n}{h}),\quad \sum_{i=1}^n|t_{ii}^2|=O(\frac{n}{h^2}).$$

For the second term in (8.44), we get

$$\frac{n+2}{n}Var(X_{1d})\sum_{i=1}^n E(\chi_{in}^2\phi'_\ell(\varepsilon_i)^2)=O(\frac{n}{h}\ell^2\mu_{4\kappa(\ell+K)}^{1/2})$$

using Cauchy-Schwarz inequality and

$$\max_i E(\chi_i^4) = 3\max_i \sum_{j=1}^n \sum_{k=1}^n t_{ij}^2 t_{ik}^2 = 3\max_i \big(\sum_{j=1}^n t_{ij}^2\big)^2 \le 3\big(\max_{i,j}|t_{ij}|\sum_{j=1}^n |t_{ij}|\big)^2 = O\big(\frac{1}{h^2}\big). \quad (8.44)$$

In the second term of (8.42), the dominant term is given by

$$\big(\sum_{i=1}^n \chi_{in} r_i^2 \phi_\ell''(\varepsilon_i^*)\big)^2 = \big(\frac{1}{\sigma_0}\sum_{i=1}^n \chi_{in}\phi_\ell''(\varepsilon_i^*)\big((X_i-\bar X)^T(\tilde\beta-\beta_0)\big)^2\big)^2$$
$$\le \frac{1}{\sigma_0^2}\sum_{i=1}^n \chi_{in}^2 \phi_\ell''(\varepsilon_i^*)^2 \sum_{i=1}^n \big((X_i-\bar X)^T(\tilde\beta-\beta_0)\big)^4 = O_p\big(h^{-1}\mu_{4\kappa(\ell-1+2K)}^{1/2}\big),$$

using Cauchy-Schwarz inequality. From Assumption 8

$$\sum_{i=1}^n \big((X_i-\bar X)^T(\tilde\beta-\beta_0)\big)^4 \le \sum_{i=1}^n \|X_i-\bar X\|^4\|\tilde\beta-\beta_0\|^4 = O_p(1/n),$$

and steps similar to (A.48) of Robinson (2010) and (8.44) lead to

$$\big|\sum_{i=1}^n \chi_{in}^2\phi_\ell''(\varepsilon_i^*)^2\big| \le C^{2\kappa\ell+2}\ell^4\sum_{i=1}^n \big\{1+|\varepsilon_i|^{2\kappa(\ell-1+2K)}+|\bar\varepsilon|^{2\kappa(\ell-1+2K)}\big\}\chi_{in}^2.$$

By Cauchy-Schwarz inequality

$$E(|\varepsilon_i|^{2\kappa(\ell-1+2K)}\chi_{in}^2) \le Ch^{-1}\mu_{4\kappa(\ell-1+2K)}^{1/2}.$$

Hence for $\ell = 1, ..., L$,

$$\big(\sum_{i=1}^n \chi_{in}\delta_{\ell i}\big)^2 = O_p\big(\frac{n}{h^2}\ell^2\mu_{2\kappa(\ell+K)} + \frac{1}{h}C^{2\kappa\ell+2}\ell^4\mu_{4\kappa(\ell-1+2K)}^{1/2}\big).$$

Applying Lemma 9 of Robinson (2005), we get

$$\sum_{\ell=1}^L\big(\sum_{i=1}^n \chi_{in}\delta_{\ell i}\big)^2 = O_p\big(\frac{n}{h^2}L^2\rho_{2\kappa L} + \frac{1}{h}C^{2\kappa L+2}L^{9/2}\rho_{4\kappa L}^{1/2}\big).$$

To verify (8.41), it remains to show

$$L^2\rho_{2\kappa L}\pi_L^2\big(\frac{n}{h^2}L^2\rho_{2\kappa L} + \frac{1}{h}C^{2\kappa L+2}L^{9/2}\rho_{4\kappa L}^{1/2}\big) = o(n/h),$$

which directly follow from $H_4 = o(\sqrt{h})$ on p. 31.

For $i = 3, 4$, we need to verify that for all $1 \le i \le n$,

$$|\tilde\psi_{iL0}^{(r)} - \psi_i| = o_p(1/\sqrt{h}). \quad (8.45)$$

Compared to the proof of (8.36) given earlier, the only term that is different here is c_{4i} of (8.19) while Lemma 10, 19 of Robinson (2005) continue to hold.

Noting $|\sigma_0\bar{\varepsilon}+(X_i-\bar{X})^T(\beta-\tilde{\beta})| = O_p(n^{-1/2})$ from Assumption 8, and following same steps as in the proof of (8.37), we can show

$$\|\Phi^{(L)}(\tilde{\epsilon}_i/\sigma_0) - \Phi^{(L)}(\varepsilon_i)\| = O_p(\frac{1}{\sqrt{n}}C^L L^{3/2}\rho_{2\kappa L}^{1/2}),$$

from which (8.45) follows on based on the same steps given in (8.36).

For $i = 5$, due to Assumption 8, it suffices to verify

$$\| \left(\tilde{\psi}_{1L0}^{(r)} - \psi_1, \cdots, \tilde{\psi}_{nL0}^{(r)} - \psi_n\right) MHX\| \leq \sum_{i=1}^{n} |\tilde{\psi}_{iL0}^{(r)} - \psi_i| \sum_{j=1}^{n} |m_{ij}| \|X_j - \bar{X}\| = o_p(n/\sqrt{h}),$$

which follows from Assumptions 1, 8 and (8.45).

For $i = 6$, we shall show

$$E\| (\psi_1, \cdots, \psi_n) MHX\|^2 = \mathrm{tr}\Big(E\big(\sum_{i,j,i',j'=1}^{n} \psi_i\psi_{i'}m_{ij}m_{i'j'}(X_j - \bar{X})(X_{j'} - \bar{X})^T\big)\Big) = o(n^2/h).$$

Noting $E(\psi_i) = 0$ and Assumption 8, we have for all $1 \leq d \leq k$:

$$\begin{aligned}
&\sum_{i,j,i',j'=1}^{n} E\big(\psi_i\psi_{i'}m_{ij}m_{i'j'}(X_{jd} - \bar{X}_d)(X_{j'd} - \bar{X}_d)\big) \\
&= \sum_{i=1}^{n} E(\psi_i^2)E\big(\sum_{j=1}^{n} m_{ij}(X_{jd} - \bar{X}_d)\sum_{j'=1}^{n} m_{ij'}(X_{j'd} - \bar{X}_d)\big) \\
&= \mathcal{J}\frac{(n+2)Var(X_{1d})}{n}\sum_{i,j=1}^{n} m_{ij}^2 - \mathcal{J}\frac{Var(X_{1d})}{n}\sum_{i,j,j'=1}^{n} m_{ij}m_{ij'} = o(n^2/h)
\end{aligned}$$

since from Assumption 1

$$\sum_{i,j=1}^{n} m_{ij}^2 = O(n/h) = o(n^2/h), \qquad \sum_{i,j,j'=1}^{n} |m_{ij}m_{ij'}| = O(n) = o(n^2/h).$$

■

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.

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.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

Kelejian, H. H. and Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, **104**, 219-57.

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

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

Lesage, J. P. and Pace, R. K. (2002). 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.

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.