Available online at www.sciencedirect.com
ScienceDirect
Journal of Multivariate Analysis 99 (2008) 245–277
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Asymptotic distribution of the OLS estimator for a purely autoregressive spatial model

Kairat T. Mynbaev[a,*,1], Aman Ullah[b]

[a] International School of Economics, Kazakh-British Technical University, Tolebi 59, 050000, Almaty, Kazakhstan
[b] Economics Department, University of California at Riverside, Riverside, California, USA

Received 3 March 2006
Available online 18 April 2007

Abstract

We derive the asymptotics of the OLS estimator for a purely autoregressive spatial model. Only low-level conditions are used. As the sample size increases, the spatial matrix is assumed to approach a square-integrable function on the square $(0, 1)^2$. The asymptotic distribution is a ratio of two infinite linear combinations of χ^2 variables. The formula involves eigenvalues of an integral operator associated with the function approached by the spatial matrices. Under the conditions imposed identification conditions for the maximum likelihood method and method of moments fail. A corrective two-step procedure using the OLS estimator is proposed.

AMS 1991 subject classification: 91B72; 91D25; 62J05

Keywords: Spatial model; OLS estimator; Asymptotic distribution; Maximum likelihood; Method of moments

1. Introduction

We consider the model

$$Y_n = \rho W_n Y_n + V_n, \tag{1.1}$$

where Y_n is the observed $n \times 1$ vector, ρ is the real parameter to be estimated, W_n is a predetermined $n \times n$ matrix, called a spatial matrix, and $V_n = (v_1, \ldots, v_n)'$ is the error vector with zero mean. This model is important in multidimensional signal processing, and is extensively studied in

* Corresponding author.
E-mail addresses: kairat@fulbrightweb.org (K.T. Mynbaev), aman.ullah@ucr.edu (A. Ullah).
[1] Research supported by the Fulbright program and generously hosted by UC—Riverside.

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2007.04.002

econometrics, see Anselin [2] and Anselin and Bera [3] for details. The importance of (1.1) increases with the growth of the number of more complex models in which the error itself is generated by a spatial model, such as

$$Y_n = X_n\beta + \rho W_n Y_n + V_n, \tag{1.2}$$

where $V_n = \mu M_n V_n + U_n$, X_n is a matrix of exogenous regressors, M_n is a spatial matrix, possibly different from W_n, and U_n is a new error vector.

The earlier developments in testing and estimation of spatial autoregressive models have been summarized in Cliff and Ord [6], Anselin [2], Cressie [7] and Anselin and Bera [3], among others. Kelejian and Prucha [12] have considered a generalized method of moments (MM) estimator of the parameter ρ in (1.1). Lee [16] has developed the theory of quasi-maximum likelihood (QML) estimation for the model in (1.1) and then Lee [19] extends this approach of estimation to the mixed model (1.2). Further Lee [17] studies consistency and efficiency of ordinary least squares (OLS) estimation for (1.2). Kelejian and Prucha [11] apply two-stage least squares (2SLS) estimation and Lee [18] improves 2SLS to achieve asymptotic optimality. There have been many other developments, such as [13–15]. They are interesting in their own right but less relevant to our subject here.

The research in the above references has been moving towards relaxing the assumptions on which the asymptotic theory results are based. Along the way the conditions imposed and the results obtained have become more complex, to the point that sometimes it is hard to see whether a given condition can be satisfied or whether two different conditions imposed on the same sequence of matrices are compatible. For example, the theorems on convergence of estimators in distribution have been put in the form of convergence to a normal vector. Convergence conditions are complex by themselves. Expressing the variance matrix of the limiting vector as a limit of a complex combination of two or more sequences of matrices adds to the nontransparency of the result, especially if existence of the limit is a new requirement and not a consequence of previous assumptions. As the reader can see from the above references, the situation is much more complex than with the classroom condition

$\lim X_n' X_n / n$ exists and is not singular

commonly used for the classical model

$$Y_n = X_n\beta + V_n. \tag{1.3}$$

Regarding model (1.1) our concern has been more specific. Assuming that W_n is symmetric with eigenvalues $\lambda_{n1}, \ldots, \lambda_{nn}$ and V_n is distributed as $N(0, \sigma^2 I)$, it is easy to write the deviation of OLS estimator $\hat{\rho}$ from the true value as

$$\hat{\rho} - \rho = \frac{\sum_{i=1}^n v_i^2 \lambda_{ni}/(1-\rho\lambda_{ni})}{\sum_{i=1}^n v_i^2 [\lambda_{ni}/(1-\rho\lambda_{ni})]^2}. \tag{1.4}$$

Kelejian and Prucha [13] and Lee [19] have developed central limit theorems for linear-quadratic forms. However, under their assumptions the quadratic part disappears in the limit. We think it is a good idea to be careful with assumptions and try to preserve this ratio-of-quadratic-forms structure in the limit.

Thus, in this paper our main objective has been to simplify and reduce the number of conditions, avoid assumptions with overlapping responsibilities, reveal quadratic forms in the limiting distribution and derive the characteristics of the limiting distribution from the primary low-level conditions. In order to achieve this we model the spatial matrices using the idea of approximating discrete objects (sequences of vectors or matrices) with functions of a continuous argument. Such an approximation allows one to use more widely the tools of the theory of functions. We rely on the rendition of this general idea contained in Mynbaev [21]. The class of matrices corresponds to the case when a particular economic unit is influenced by many others, so that interaction among the units is stronger than in other settings. Using the low-level conditions we concentrate our attention on the OLS estimator, and have shown its asymptotic distribution in Theorem 1. Under our conditions, the QML and MM estimators studied in [12,16] are inconsistent, and this is presented in Theorem 2. In addition we have proposed a new two-step estimator. It is not consistent in the usual sense but satisfies certain consistency-type requirements. A numerical simulation is presented to analyze the behavior of the OLS and two-step estimators.

The plan of the paper is as follows. Section 2 contains the main assumptions and theorems. The proofs are given in Section 3. It is followed by Section 4 with computer simulation results. Section 5 presents the conclusions.

2. Main statements

First we describe our assumptions and then state the main results.

Assumption 1 (*On the error term*). For each n, the error vector $V_n = (v_1, \dots, v_n)'$ has independent identically distributed components $v_1, \dots, v_n$ with mean zero, variance σ^2 and finite moments up to $\mu_4 = Ev_i^4$.

This is the usual condition adopted by our many predecessors. Generalization to martingale differences is possible at the expense of lengthening the proof. For the next assumption we need some notation. On the set of integrable on the square $(0, 1)^2$ functions we can define a discretization operator d_n. For an integrable function K, $d_n K$ is an $n \times n$ matrix with elements

$$(d_n K)_{ij} = n \int_{q_{ij}} K(x, y)\, dx\, dy, \quad i, j = 1, \dots, n,$$

where

$$q_{ij} = \left\{ (x, y) : \frac{i-1}{n} < x < \frac{i}{n}, \ \frac{j-1}{n} < y < \frac{j}{n} \right\}$$

are small squares that partition $(0, 1)^2$. Elements of a matrix A are denoted by $(A)_{ij}$ or a_{ij} and the Euclidean norm of A is $\|A\|_2 = \left(\sum_{ij} a_{ij}^2 \right)^{1/2}$.

Assumption 2 (*On the spatial matrices*). The sequence of matrices $\{W_n : n = 1, 2, \dots\}$ is such that W_n is of size $n \times n$ and there exists a function K which is square-integrable on $(0, 1)^2$ and satisfies

$$\|W_n - d_n K\|_2 = o\left(\frac{1}{\sqrt{n}} \right). \tag{2.1}$$

Evidently, such classes of matrices exist. For example, one can take any function K and put $W_n = d_n K$, in which case the left-hand side of (2.1) is identically zero. In Section 3 we show that Assumption 2 implies

$$\max_{i,j} |w_{nij}| \longrightarrow 0, \quad \sum_{i,j} |w_{nij}| \longrightarrow \infty, \;\; n \to \infty \tag{2.2}$$

(w_{nij} are the elements of W_n). The first relation means that activities of a given unit have weak influence on the other units, whereas the second can be understood as an increase to infinity in total interaction among the units. We would like to stress that in practice, when only one matrix is available, it can be approximated arbitrarily well, so Assumption 2 is rather a mathematical restriction on the regularity of the behavior at infinity of a sequence of matrices than an economic restriction. Conditions similar to (2.1) with continuous K are used in Tanaka [23, Section 5.6].

Assumption 3 (*On the function K*). The function K is symmetric and the eigenvalues λ_i, $i = 1, 2, \ldots$, of the integral operator

$$(\mathcal{K} f)(x) = \int_0^1 K(x, y) f(y)\, dy$$

are summable: $\sum_{i \geqslant 1} |\lambda_i| < \infty$.

$\mathcal{K}$ is considered an operator in the space $L_2(0, 1)$ of square-integrable functions on (0, 1). Its eigenvalues λ_i and eigenfunctions f_i are listed according to their multiplicity; the system of eigenfunctions is complete and orthonormal in $L_2(0, 1)$. For a symmetric and square-integrable K, its eigenvalues are real and square-summable: $\sum_{i \geqslant 1} \lambda_i^2 < \infty$. The summability condition we require is stronger because

$$\left(\sum_{i \geqslant 1} \lambda_i^2 \right)^{1/2} \leqslant \sum_{i \geqslant 1} |\lambda_i|. \tag{2.3}$$

Necessary and sufficient conditions (in terms of K) for summability of eigenvalues can be found in Gohberg and Kreĭn [8, Theorem 10.1].

The completeness and orthonormality of $\{f_i\}$ imply the decomposition

$$K(x, y) = \sum_{i \geqslant 1} \lambda_i f_i(x) f_i(y) \tag{2.4}$$

and the identity

$$\sum_{i \geqslant 1} \lambda_i^2 = \int_0^1 \int_0^1 K^2(x, y)\, dx\, dy \tag{2.5}$$

which are important for understanding both the result and proof.

Now, denoting $Z_n = W_n Y_n$ the regressor in (1.1), we have the following expression for the OLS estimator $\hat{\rho}$ of ρ:

$$\hat{\rho} = (Z_n' Z_n)^{-1} Z_n' Y_n. \tag{2.6}$$

Next, we denote $S_n = S_n(\rho) = I_n - \rho W_n$ and $G_n = W_n S_n^{-1}$ when S_n^{-1} exists. Further, dlim and plim will denote limits in distribution and probability, respectively. We can now present Theorem 1.

Theorem 1. *Suppose Assumptions* 1–3 *hold.*

(1) *If*

$$|\rho| < 1 \Bigg/ \left(\sum_{i \geqslant 1} \lambda_i^2 \right)^{1/2}, \tag{2.7}$$

then the matrices S_n^{-1} exist for all sufficiently large n and have uniformly bounded $\|\cdot\|_2$-norms and the deviation of the OLS estimate from the true value equals

$$\hat{\rho} - \rho = \frac{V_n' G_n' V_n}{V_n' G_n' G_n V_n}. \tag{2.8}$$

(2) *If*

$$|\rho| < 1 \Bigg/ \sum_{i \geqslant 1} |\lambda_i|, \tag{2.9}$$

then

$$\text{dlim}(\hat{\rho} - \rho) = \frac{\sum_{i \geqslant 1} u_i^2 v(\lambda_i)}{\sum_{i \geqslant 1} u_i^2 v^2(\lambda_i)}, \tag{2.10}$$

where $u_i \in N(0, 1)$ are independent and

$$v(\lambda_i) = \frac{\lambda_i}{1 - \rho\lambda_i}.$$

(3) *Eq.* (2.9) *implies convergence*

$$\text{dlim}\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \in N(0, \mu_4 - \sigma^4), \tag{2.11}$$

where

$$\hat{\sigma}^2 = \frac{1}{n-1}(Y_n - \hat{\rho} W_n Y_n)'(Y_n - \hat{\rho} W_n Y_n)$$

is the OLS estimator of σ^2.

The proof is given in Section 3. We present some remarks based on Theorem 1.

Remarks. (a) In (2.10) both the numerator and denominator are nontrivial random variables, unlike many other econometric problems where the numerator is nontrivial and denominator is constant.

(b) In the ratio at the right of (2.10) the top and bottom converge in L_1 and, consequently, in probability. This fact can be used for approximating the ratio by truncating the sums.

(c) If the numerator in (2.8) or (2.10) has mean zero, it does not necessarily mean that the whole fraction has mean zero (see Lemma 5 in Section 3 regarding (2.8)). The characteristic function of an infinite weighted sum of χ^2-variables has been found by Anderson and Darling [1] (see also [24]). However, as far as we know, similar results for ratios of such sums are not known.

(d) We do not know if the difference $\widehat{\rho}-\rho$ converges in probability. However, if it does, then by (2.10) $\operatorname{plim}(\widehat{\rho}-\rho)$ is a random variable whose mean is not zero in general. In this sense the OLS estimator is inconsistent (this fact has been noted in the literature in other formulations, see, for instance, [2, Section 6.1.1] and Lee [17]).

Since the OLS estimator is inconsistent, we now try to find alternative estimators which are consistent. Earlier Kelejian and Prucha [12] and Lee [16] have looked into this issue and suggested QML and MM estimators, proving their consistency and asymptotic normality. We show that, under our set of assumptions and conditions, these estimators are in fact not applicable. This is because, based on White's [25, Chapter 3] identification uniqueness condition, the local and global identification conditions used in their proofs do not hold under our assumptions. Following Lee [16] these identification conditions for the consistency of QML and MM estimators are as given below. [2]

Local identification condition for QML: The limit

$$\lim_{n\to\infty}\frac{1}{n}\left[\operatorname{tr}(G_n'G_n)+\operatorname{tr}(G_n^2)-\frac{2}{n}\operatorname{tr}^2(G_n)\right] \tag{2.12}$$

exists and is positive.

Global identification condition for QML: For any ρ different from the true value ρ_0 the limit

$$\lim_{n\to\infty}\frac{1}{n}\left(\ln|\sigma_0^2S_n^{-1}S_n'^{-1}|-\ln|\sigma_n^2(\rho)S_n^{-1}(\rho)S_n'^{-1}(\rho)|\right) \tag{2.13}$$

exists and is not zero where $S_n=S_n(\rho_0)$ and

$$\sigma_n^2(\rho)=\frac{\sigma_0^2}{n}\operatorname{tr}(S_n'^{-1}S_n'(\rho)S_n(\rho)S_n^{-1}).$$

Identification condition for MM: The limit

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}A_n \tag{2.14}$$

exists and is nonsingular, where the elements of the 2×2 matrix A_n are given by

$$\begin{aligned}
a_{n11}&=2\left[Y_n'W_n'^2W_nY_n-\operatorname{tr}(W_n'W_n)\frac{1}{n}Y_n'W_nY_n\right],\\
a_{n12}&=-Y_n'W_n'^2W_n^2Y_n+\operatorname{tr}(W_n'W_n)\frac{1}{n}Y_n'W_n'W_nY_n,\\
a_{n21}&=Y_n'W_n^2Y_n+Y_n'W_n'W_nY_n,\\
a_{n22}&=-Y_n'W_n'^2W_nY_n.
\end{aligned}$$

We can now present our Theorem 2.

Theorem 2. *Under the assumptions of Theorem* 1 *limits* (2.12)–(2.14) *are zero*.

The proof of Theorem 2 is given in Section 3. The results in Theorem 2 show that, under our assumptions, the identification conditions used in Lee [16] and Kelejian and Prucha [12] for

[2] Lee introduces a special parameter h_n designed to accommodate different asymptotics of matrices at infinity. Under our conditions, the only meaningful choice is $h_n=1$.

proving the consistency of QML and MM estimators do not hold. Thus, these estimators are not consistent, and without consistency the derivation of the asymptotic distribution based on the formula

$$\hat{\rho}_{\mathrm{QML}} - \rho_0 = \left(\frac{\partial^2 \ln L_n(\widetilde{\rho})}{\partial \rho^2} \right)^{-1} \frac{\partial \ln L_n(\rho)}{\partial \rho}$$

does not work. Here $\ln L_n(\rho)$ is the log likelihood function (see (2.16) below) and $\widetilde{\rho}$ lies between $\hat{\rho}_{\mathrm{QML}}$ and ρ_0.

The problems we have described above about MM and QML estimators force us to analyze the OLS estimator more closely. We devise a two-step procedure whose format is dictated by (2.10): instead of requiring plim $\hat{\rho} = \rho$ (consistency) it would be correct to require

$$\operatorname{plim} \hat{\rho} = \rho + \kappa \quad \text{where } E\kappa = 0. \tag{2.15}$$

The assumptions will be more restrictive than in Theorem 1.

The ML estimator expression will help the reader understand the idea behind our construction. The ML estimator has been derived in a more general situation by Ord [22], among others. In our case the log likelihood function is

$$\begin{aligned} \ln L_n(\theta) = &-\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |S_n(\rho)| \\ &-\frac{1}{2\sigma^2} (Y_n - \rho W_n Y_n)'(Y_n - \rho W_n Y_n), \end{aligned} \tag{2.16}$$

where $\theta = (\rho, \sigma^2)$. Denoting A, B square matrices of order n and δ_{ij} the Kronecker symbol ($\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 0$), we have

$$\frac{\partial \ln |A|}{\partial A} = (A')^{-1} = (A^{-1})', \quad \operatorname{tr}(AB) = \sum_{i,j=1}^{n} a_{ij} b_{ji},$$

$$\frac{\partial (S_n(\rho))_{ij}}{\partial \rho} = \frac{\partial}{\partial \rho} (\delta_{ij} - \rho w_{nij}) = -w_{nij}$$

(the first of these equations is true when $|A| = \det A > 0$ and can be found in [20, p. 473]). These equations imply

$$\begin{aligned} \frac{\partial \ln |S_n(\rho)|}{\partial \rho} &= \sum_{i,j=1}^{n} \frac{\partial \ln |S_n(\rho)|}{\partial (S_n(\rho))_{ij}} \frac{\partial (S_n(\rho))_{ij}}{\partial \rho} \\ &= -\sum_{i,j=1}^{n} (S_n^{-1}(\rho))_{ji} (W_n)_{ij} = -\operatorname{tr}[W_n S_n^{-1}(\rho)] \end{aligned} \tag{2.17}$$

for ρ such that the determinant $|S_n(\rho)|$ is positive. Using (2.17) we get

$$\begin{aligned} \frac{\partial \ln L_n(\theta)}{\partial \rho} &= -\operatorname{tr}(W_n S_n^{-1}(\rho)) - \frac{1}{2\sigma^2} (-Y_n' W_n Y_n - Y_n' W_n' Y_n + 2\rho (W_n Y_n)' W_n Y_n) \\ &= -\operatorname{tr}(W_n S_n^{-1}(\rho)) + \frac{1}{\sigma^2} (Y_n' W_n Y_n - \rho (W_n Y_n)' W_n Y_n), \end{aligned}$$

$$\frac{\partial \ln L_n(\theta)}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} (Y_n - \rho W_n Y_n)'(Y_n - \rho W_n Y_n).$$

The first-order conditions for maximization of $\ln L_n(\theta)$ give the estimators

$$\hat{\rho}_{\mathrm{ML}} = \frac{Y_n' W_n Y_n - \hat{\sigma}^2_{\mathrm{ML}} \mathrm{tr}(W_n S_n^{-1}(\rho))}{(W_n Y_n)' W_n Y_n},$$

$$\hat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}(Y_n - \rho W_n Y_n)'(Y_n - \rho W_n Y_n).$$

Of course, these estimators are not feasible as they contain an unknown ρ. However, a modification of $\hat{\rho}_{\mathrm{ML}}$ can be used to correct the OLS estimate.

Modification factor definition: Since the OLS estimator and the formula we suggest below do not change if W_n is replaced by its symmetric derivative $(W_n + W_n')/2$, in Theorem 3 and its proof we assume without loss of generality that W_n is symmetric. Then W_n can be represented as

$$W_n = P_n \,\mathrm{diag}[\lambda_{n1}, \ldots, \lambda_{nn}] P_n', \tag{2.18}$$

where $\lambda_{n1}, \ldots, \lambda_{nn}$ are eigenvalues of W_n and P_n is an orthogonal matrix: $P_n P_n' = I$. Denote

$$\pi_n(t) = \left[\prod_{i=1}^{n} (1 + 2t v^2(\lambda_{ni}))\right]^{1/2},$$

$$c_n = \int_0^\infty \frac{dt}{\pi_n(t)}, \quad c_{ni} = \int_0^\infty \frac{dt}{\pi_n(t)(1 + 2tv^2(\lambda_{ni}))}, \quad i = 1, \ldots, n.$$

These integrals converge if $n > 2$. The modification factor is defined by

$$A_n = P_n \,\mathrm{diag}\left[\frac{c_{n1}}{c_n}, \ldots, \frac{c_{nn}}{c_n}\right] P_n'.$$

This factor is introduced to satisfy property (2.22) below.

Correction term and two-step estimator definition[3]: Estimate ρ and σ^2 by OLS and put

$$\rho_{\mathrm{corr}} = \frac{Y_n' W_n Y_n - \hat{\sigma}^2 \,\mathrm{tr}(A_n W_n S_n^{-1}(\hat{\rho}))}{(W_n Y_n)' W_n Y_n}, \quad \rho_{2S} = (\hat{\rho} + \rho_{\mathrm{corr}})/2.$$

For analytical purposes we rewrite the correction term as

$$\rho_{\mathrm{corr}} = \frac{V_n' S_n'^{-1} G_n V_n - \hat{\sigma}^2 \,\mathrm{tr}(A_n W_n S_n^{-1}(\hat{\rho}))}{V_n' G_n' G_n V_n}. \tag{2.19}$$

Instead of Assumption 1 we make a stronger assumption.

Assumption 1′. V_n is distributed as $N(0, \sigma^2 I_n)$.

Further we make the following assumption:

Assumption 4. The sums $\sum_{i=1}^n |\lambda_{ni}|^p$ are uniformly bounded for some $p < 2$.

[3] The first version of the paper contained a multistep procedure. Computer simulations show that two steps perform just as well.

It can be shown that (2.1) implies

$$\lim_{n\to\infty}\sum_{i=1}^{n}\lambda_{ni}^2=\sum_{i=1}^{\infty}\lambda_i^2 \tag{2.20}$$

(see Lemma 6 in Section 3) and that a condition stronger than (2.1) can be imposed on the sequence $\{W_n\}$ to make sure that Assumption 4 is satisfied. See Gohberg and Kreĭn [8, Chapter III] for more information.

We can now present the following theorem.

Theorem 3. *Suppose Assumptions* 1′, 2–4 *hold. If the true* ρ *satisfies* (2.9), *then there exist random variables* $\kappa_{n1},\ \kappa_{n2},\ \kappa_{n3}$ *and a deterministic function* ψ_n *such that*

$$\rho_{\text{corr}}=\rho+\kappa_{n1}+\kappa_{n2}+\kappa_{n3}\int_{\hat\rho}^{\rho}\psi_n(t)\,dt, \tag{2.21}$$

$$E\kappa_{n1}=0\quad \textit{for all } n,\qquad \operatorname{plim}\kappa_{n2}=0, \tag{2.22}$$

$$\operatorname{dlim}\kappa_{n3}=\frac{1}{\sum_{i\geqslant 1}u_i^2v^2(\lambda_i)}, \tag{2.23}$$

where u_i *are independent standard normal and* κ_{n3} *and* ψ_n *are positive almost everywhere.*

The proof can be found in Section 3. The remarks below explain what this theorem gives.

Remarks. (a) Property (2.22) is in line with (2.15).

(b) Heuristically, the definition of ρ_{2S} can be explained as follows. By the mean value theorem (2.21) and (2.22) imply $\rho_{\text{corr}}\approx\rho+\kappa_{n3}\psi_n(t^*)(\rho-\hat\rho)$ so that the true parameter is a weighted sum of ρ_{corr} and $\hat\rho$:

$$\rho\approx\frac{\rho_{\text{corr}}+\kappa_{n3}\psi_n(t^*)\hat\rho}{1+\kappa_{n3}\psi_n(t^*)}. \tag{2.24}$$

Here t^* is some point between the true value and the OLS estimate. Since the weights are unknown we choose one half for each which seems to work pretty well.

3. Proofs of lemmas and theorems

First we give the main notation used in proving lemmas and theorems. Depending on the context, $\|\cdot\|_2$ may mean any of the norms

$$\|x\|_2=\left(\sum_{i\in I}x_i^2\right)^{1/2},\quad \|f\|_2=\left(\int_0^1 f^2(x)\,dx\right)^{1/2},$$

$$\|K\|_2=\left(\int_0^1\int_0^1 K^2(x,y)\,dx\,dy\right)^{1/2}.$$

Here the set of indices I can be finite or infinite. $(\cdot,\cdot)_{l_2}$ denotes the scalar product associated with the first of these norms and $(\cdot,\cdot)_{L_2}$ stands for the scalar product that generates the last two norms.

Let $(\Omega, \mathfrak{F}, P)$ be a probability space. Among the norms

$$\|X\|_p = \left(\int_\Omega |X(\omega)|^p \, dP(\omega)\right)^{1/p}, \quad 1 \leqslant p < \infty,$$

$\|\cdot\|_1$ and $\|\cdot\|_2$ will be particularly useful.

$c, c_1, c_2, \ldots$ will denote various inconsequential positive constants (which do not depend on the variables of interest). For an $n \times n$ matrix A we find it handy to use the notation

$$N(A) = \left(E\left(V_n' A V_n\right)^2\right)^{1/2}.$$

3.1. *Proofs of lemmas*

Here we are going to present lemmas and their proofs that are necessary to prove the theorems in the later subsections.

Lemma 1. (a) *With any square matrix A such that* $|\rho| \, \|A\|_2 < 1$ *one can associate the matrix*

$$s(A) = \sum_{k=0}^{\infty} \rho^k A^{k+1}.$$

If $|\rho| \, \|W_n\|_2 < 1$, *then* $G_n = s(W_n)$.

(b) *For square matrices* A, B *and any integer* $k \geqslant 0$

$$\left\| A^{k+1} - B^{k+1} \right\|_2 \leqslant \|A - B\|_2 \, (k+1) \left(\max\{\|A\|_2, \|B\|_2\}\right)^k. \tag{3.1}$$

(c) *For square matrices* A, B *such that* $|\rho| \max\{\|A\|_2, \|B\|_2\} < 1$ *one has*

$$\|s(A) - s(B)\|_2 \leqslant \varphi(\rho, A, B) \, \|A - B\|_2, \tag{3.2}$$

where

$$\varphi(\rho, A, B) \equiv \sum_{k \geqslant 0} (k+1) \left(|\rho| \max\{\|A\|_2, \|B\|_2\}\right)^k < \infty.$$

(d) *If* V_n *satisfies Assumption* 1 *and* A, B *are square matrices of order n, then*

$$N(AB) \leqslant c \, \|A\|_2 \, \|B\|_2. \tag{3.3}$$

In particular, by choosing $B = I$ *we get*

$$N(A) \leqslant c\sqrt{n} \, \|A\|_2. \tag{3.4}$$

(e) *Under the same conditions as in* (d) *for all* $k > 0$

$$N\left(A^{k+1} - B^{k+1}\right) \leqslant c \, \|A - B\|_2 \, (k+1) \left(\max\{\|A\|_2, \|B\|_2\}\right)^k. \tag{3.5}$$

Proof. (a) Follows from the well-known fact that if $\|A\| < 1$ and the norm $\|\cdot\|$ is submultiplicative ($\|AB\| \leqslant \|A\| \, \|B\|$), then the series $\sum_{k \geqslant 0} A^k$ converges and represents $(I - A)^{-1}$. We apply this fact to S_n^{-1} and multiply it by W_n to obtain G_n.

(b) For $k = 0$, (3.1) is trivial. If $k > 0$, then the identity

$$A^{k+1} - B^{k+1} = A^k(A - B) + A^{k-1}(A - B)B + \cdots + (A - B)B^k \tag{3.6}$$

and submultiplicativity of the norm $\|\cdot\|_2$ gives the desired result:

$$\begin{aligned}\left\|A^{k+1} - B^{k+1}\right\|_2 &\leqslant \|A\|_2^k \|A - B\|_2 + \cdots + \|A - B\|_2 \|B\|_2^k \\ &\leqslant \|A - B\|_2 (k + 1) (\max\{\|A\|_2, \|B\|_2\})^k .\end{aligned} \tag{3.7}$$

(c) Eq. (3.2) follows from (3.1):

$$\|s(A) - s(B)\|_2 \leqslant \sum_{k \geqslant 0} |\rho|^k \left\|A^{k+1} - B^{k+1}\right\|_2 \leqslant \varphi(\rho, A, B) \|A - B\|_2 .$$

(d) For any square matrix A of order n Lee's [19, Lemma A.11][4] yields

$$N(A) \leqslant c (\|A\|_2 + |\mathrm{tr}\, A|) . \tag{3.8}$$

Since $\|AB\|_2 \leqslant \|A\|_2 \|B\|_2$ and $|\mathrm{tr}(AB)| \leqslant \|A\|_2 \|B\|_2$, (3.8) gives (3.3).

(e) Because of the growing factor $\sqrt{n}$ in (3.4), it is not a good idea to estimate the left-hand side of (3.5) using (3.1). Instead, we apply identity (3.6) directly (this is why the assumption $k > 0$ is important). By (3.3) and Minkowski's inequality

$$\begin{aligned}N\left(A^{k+1} - B^{k+1}\right) &\leqslant N\left(A^k(A - B)\right) + \cdots + N((A - B)B^k) \\ &\leqslant c\left(\|A\|_2^k \|A - B\|_2 + \cdots + \|A - B\|_2 \|B\|_2^k\right).\end{aligned}$$

The rest is the same as in (3.7). □

The further proofs use several operators which relate functions of discrete and continuous arguments to one another. One of them, the discretization operator d_n defined in Section 2, possesses the property

$$\|d_n K\|_2 \leqslant \|K\|_2 \quad \text{for all } K \text{ and } n \tag{3.9}$$

(apply Hölder's inequality to prove it). The interpolation operator D_n takes a square matrix A of order n to a piece-wise constant function on $(0, 1)^2$ according to

$$D_n A = n \sum_{i,j=1}^{n} a_{ij} 1_{q_{ij}},$$

where 1_S stands for the indicator of a set S: $1_S(x) = 1$, if $x \in S$, and $1_S(x) = 0$, if $x \notin S$. D_n preserves norms:

$$\|D_n A\|_2 = \|A\|_2 . \tag{3.10}$$

The product $D_n d_n$ coincides with the Haar projector P_n defined by

$$P_n K = n^2 \sum_{i,j=1}^{n} \int_{q_{ij}} K(x, y)\, dx\, dy 1_{q_{ij}} .$$

[4] See his supplement http://economics.sbs.ohio-state.edu/lee/wp/sar-qml-r-appen-04feb.pdf

Its main property is that it approximates the identity operator:

$$\lim_{n\to\infty} \|P_n K - K\|_2 = 0 \quad \text{for any } K \in L_2\left((0,1)^2\right). \tag{3.11}$$

Denote $q_i = \left\{x \in \mathbb{R} : \frac{i-1}{n} < x < \frac{i}{n}\right\}$, $i = 1, \ldots, n$. One-dimensional analogs of d_n and D_n are defined, respectively, by

$$(d_n f)_i = \sqrt{n} \int_{q_i} f(x)\, dx, \quad i = 1, \ldots, n, \ f \in L_2(0,1),$$

and

$$D_n x = \sqrt{n} \sum_{i=1}^{n} x_i 1_{q_i}, \quad x \in \mathbb{R}^n.$$

They possess properties similar to (3.9), (3.10) and (3.11).

Now we can proceed with the next lemma.

Lemma 2. (a) *Assumption* 2 *and symmetry of K imply*

$$\lim_{n\to\infty} \|W_n\|_2 = \lim_{n\to\infty} \|d_n K\|_2 = \|K\|_2 \tag{3.12}$$

and

$$\left\| W_n' - d_n K \right\|_2 = o\left(\frac{1}{\sqrt{n}}\right). \tag{3.13}$$

(b) *Consider any orthonormal system* $\{f_i : i \geqslant 1\}$ *in* $L_2(0,1)$. *For a collection of indices* $i = (i_1, \ldots, i_{k+1})$, *where all of* i_j*'s are positive integers, denote*

$$\mu_{ni} = \begin{cases} (d_n f_{i_1}, d_n f_{i_2})_{l_2} (d_n f_{i_2}, d_n f_{i_3})_{l_2} \ldots (d_n f_{i_k}, d_n f_{i_{k+1}})_{l_2} & \text{if } k > 0, \\ 1 & \text{if } k = 0, \end{cases}$$

and

$$\mu_{\infty i} = \begin{cases} 1 & (i_1 = i_2 = \cdots = i_{k+1} \text{ and } k > 0) \text{ or } (k = 0), \\ 0 & \text{otherwise}. \end{cases}$$

Then for all i

$$\lim_{n\to\infty} \mu_{ni} = \mu_{\infty i}. \tag{3.14}$$

(c) *Denote the two-dimensional discretization operator by* d_n^2 *and its one-dimensional counterpart by* d_n^1. *If* $F(x, y) = G(x)H(y)$, *then*

$$\left(d_n^2 F\right)_{st} = \left(d_n^1 G\right)_s \left(d_n^1 H\right)_t, \quad s, t = 1, \ldots, n.$$

(d) *If* $\|W_n - d_n K\|_2 \to 0$, *then* (2.2) *is true. A similar property holds in the one-dimensional case.*

Proof. (a) Continuity of norms and (3.11) yield $\|P_n K\|_2 \to \|K\|_2$. $\|d_n K\|_2 \to \|K\|_2$ follows because by (3.10) $\|d_n K\|_2 = \|D_n d_n K\|_2 = \|P_n K\|_2$. To prove the other equation in (3.12) note that by (3.10), (2.1) and (3.11)

$$\begin{aligned}\|D_n W_n - K\|_2 &\leqslant \|D_n W_n - P_n K\|_2 + \|P_n K - K\|_2 \\ &= \|W_n - d_n K\|_2 + \|P_n K - K\|_2 \to 0.\end{aligned}$$

Therefore, $\|W_n\|_2 = \|D_n W_n\|_2 \to \|K\|_2$.

To prove (3.13), observe that $(x, y) \in q_{ij}$ if and only if $(y, x) \in q_{ji}$ and, therefore, for a symmetric K, $d_n K$ is also symmetric. Thus,

$$\|W_n' - d_n K\|_2 = \|(W_n - d_n K)'\|_2 = \|W_n - d_n K\|_2 = o\left(\frac{1}{\sqrt{n}}\right).$$

(b) It is easy to check that D_n preserves not only norms but also scalar products. For example, in the one-dimensional case that we need right now

$$(D_n x, D_n y)_{L_2} = (x, y)_{l_2}, \quad x, y \in \mathbb{R}^n.$$

Using this fact, continuity of scalar products, and (3.11) we see that

$$(d_n f_i, d_n f_j)_{l_2} = (P_n f_i, P_n f_j)_{L_2} \longrightarrow (f_i, f_j)_{L_2} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{3.15}$$

Turning to (3.14), if $k > 0$ and among $i_1, \ldots, i_{k+1}$ there are at least two different indices, then at least two adjacent ones must be unequal. Hence, (3.14) is a direct consequence of (3.15).

(c) This is straightforward to show.

(d) First note that

$$\max_{i,j} |w_{nij}| \leqslant \|W_n - d_n K\|_2 + \max_{i,j} |(d_n K)_{ij}|$$

and then that by Hölder's inequality and absolute continuity of the Lebesgue integral

$$|(d_n K)_{ij}| = n \left| \int_{q_{ij}} K(x, y)\, dx\, dy \right| \leqslant \left(\int_{q_{ij}} K^2(x, y)\, dx\, dy \right)^{1/2} \longrightarrow 0, \quad n \to \infty$$

uniformly in i, j. This proves the first of the limit relations in (2.2). By (3.12) for some $c > 0$ we have $c \leqslant \|W_n\|_2^2 \leqslant \|W_n\|_\infty \|W_n\|_1$ which implies $\|W_n\|_1 \geqslant c/\|W_n\|_\infty \to \infty$. $\square$

For natural n, L consider the random vector

$$U_{nL} = \begin{pmatrix} \sum_{s=1}^n (d_n f_1)_s v_s \\ \ldots \\ \sum_{s=1}^n (d_n f_L)_s v_s \end{pmatrix} = \begin{pmatrix} V_n' d_n f_1 \\ \ldots \\ V_n' d_n f_L \end{pmatrix}.$$

We need the following two-dimensional function of U_{nL}:

$$\delta_{nL} = \sum_{i=1}^L U_{nLi}^2 \nu(\lambda_i) \begin{pmatrix} 1 \\ \nu(\lambda_i) \end{pmatrix}.$$

The limiting behavior of δ_{nL} is described in terms of the vectors

$$\Delta_L = \sigma^2 \sum_{i=1}^{L} u_i^2 v(\lambda_i) \begin{pmatrix} 1 \\ v(\lambda_i) \end{pmatrix}, \quad \Delta_\infty = \sigma^2 \sum_{i=1}^{\infty} u_i^2 v(\lambda_i) \begin{pmatrix} 1 \\ v(\lambda_i) \end{pmatrix},$$

where u_i are independent standard normal.

Lemma 3. *Let V_n satisfy Assumption* 1 *and suppose that* $\{f_i : i = 1, 2, \ldots\}$ *is any orthonormal system in* $L_2(0, 1)$. *Then*

(a) *For any fixed L*

$$\operatorname{dlim}_{n\to\infty} \delta_{nL} = \Delta_L, \tag{3.16}$$

$$\lim_{n\to\infty} E\delta_{nL} = E\Delta_L = \sigma^2 \sum_{i=1}^{L} v(\lambda_i) \begin{pmatrix} 1 \\ v(\lambda_i) \end{pmatrix}, \tag{3.17}$$

$$\lim_{n\to\infty} \operatorname{var}(\delta_{nL}) = \operatorname{var}(\Delta_L) = 2\sigma^4 \sum_{i=1}^{L} v^2(\lambda_i) \begin{pmatrix} 1 & v(\lambda_i) \\ v(\lambda_i) & v^2(\lambda_i) \end{pmatrix}. \tag{3.18}$$

(b) *If*

$$\sum_{i\geqslant 1} |v(\lambda_i)| < \infty, \tag{3.19}$$

then in the sense of $L_1(\Omega)$

$$\lim_{n\to\infty} \Delta_L = \Delta_\infty \tag{3.20}$$

and

$$\lim_{L\to\infty} \operatorname{var}(\Delta_L) = \operatorname{var}(\Delta_\infty) = 2\sigma^4 \sum_{i=1}^{\infty} v^2(\lambda_i) \begin{pmatrix} 1 & v(\lambda_i) \\ v(\lambda_i) & v^2(\lambda_i) \end{pmatrix}. \tag{3.21}$$

Proof. (a) The central limit theorem from Mynbaev [21] states that under the conditions of the lemma for any L

$$\operatorname{dlim} U_{nL} \in N(0, \sigma^2 I_L), \quad \operatorname{var}(U_{nL}) \longrightarrow \sigma^2 I_L \text{ as } n \to \infty. \tag{3.22}$$

The vector δ_{nL} is a continuous function of U_{nL}. Since $\operatorname{dlim} U_{nLi}^2 = \sigma^2 u_i^2$, as $n \to \infty$, (3.16) is true. The second relation in (3.22) implies (3.17):

$$E\delta_{nL} = \sum_{i=1}^{L} v(\lambda_i) \begin{pmatrix} 1 \\ v(\lambda_i) \end{pmatrix} EU_{nLi}^2 \longrightarrow E\Delta_L.$$

To prove (3.18), we start with

$$\operatorname{var}(\delta_{nL}) = E\delta_{nL}\delta_{nL}' - E\delta_{nL}E\delta_{nL}' = \sum_{i,j=1}^{L} \left(EU_{nLi}^2 U_{nLj}^2 - EU_{nLi}^2 EU_{nLj}^2\right)$$
$$\times v(\lambda_i)v(\lambda_j) \begin{pmatrix} 1 & v(\lambda_i) \\ v(\lambda_j) & v(\lambda_i)v(\lambda_j) \end{pmatrix}.$$

Here

$$EU_{nLi}^2U_{nLj}^2 = E\left(\sum_{s=1}^n (d_nf_i)_s v_s\right)^2\left(\sum_{p=1}^n (d_nf_j)_p v_p\right)^2$$
$$= \sum_{s,t,p,q=1}^n (d_nf_i)_s(d_nf_i)_t(d_nf_j)_p(d_nf_j)_q Ev_sv_tv_pv_q.$$

From Assumption 1 it follows that

$$Ev_sv_tv_pv_q = \begin{cases} \sigma^4 & \text{if } [(s=t)\neq(p=q)] \text{ or } [(s=p)\neq(t=q)] \\ & \text{or } [(s=q)\neq(t=p)], \\ \mu_4 & \text{if } s=t=p=q, \\ 0 & \text{in all other cases.}\end{cases}$$

Hence,

$$EU_{nLi}^2U_{nLj}^2 = \sigma^4\left[\sum_{s=1}^n (d_nf_i)_s^2\sum_{p=1}^n (d_nf_j)_p^2 + 2\sum_{s=1}^n (d_nf_i)_s(d_nf_j)_s\sum_{p=1}^n (d_nf_i)_p(d_nf_j)_p\right]$$
$$+\mu_4\sum_{s=1}^n (d_nf_i)_s^2(d_nf_j)_s^2$$
$$= \sigma^4\left[\|d_nf_i\|_2^2\left\|d_nf_j\right\|_2^2 + 2(d_nf_i, d_nf_j)_{l_2}^2\right] + \mu_4\sum_{s=1}^n (d_nf_i)_s^2(d_nf_j)_s^2.$$

By Lemma 2(d) and (3.15)

$$\max_s |(d_nf_i)_s| \to 0, \quad \|d_nf_i\|_2 \to 1, \quad (d_nf_i, d_nf_j)_{l_2} \to \delta_{ij} = \begin{cases} 1, & i=j, \\ 0, & i\neq j,\end{cases}$$

so that

$$\sum_{s=1}^n (d_nf_i)_s^2(d_nf_j)_s^2 \leqslant \max_s (d_nf_i)_s^2\left\|d_nf_j\right\|_2^2 \to 0$$

and

$$EU_{nLi}^2U_{nLj}^2 \to \sigma^4(1+2\delta_{ij}), \quad EU_{nLi}^2EU_{nLj}^2 \to \sigma^4 \text{ for all } i, j.$$

These equations together with the formula for $\mathrm{var}(\delta_{nL})$ above prove that the left and right members of (3.18) are equal.

Standard normal variables satisfy $\mu_4 = 3\sigma^4 = 3$, so

$$\mathrm{var}(\Delta_L) = E\Delta_L\Delta_L' - E\Delta_LE\Delta_L'$$
$$= \sigma^4\sum_{i,j=1}^L \left(Eu_i^2u_j^2 - 1\right)v(\lambda_i)v(\lambda_j)\begin{pmatrix} 1 & v(\lambda_i) \\ v(\lambda_j) & v(\lambda_i)v(\lambda_j)\end{pmatrix}$$
$$= \sigma^4\sum_{i=1}^L (3-1)\,v^2(\lambda_i)\begin{pmatrix} 1 & v(\lambda_i) \\ v(\lambda_i) & v(\lambda_i)v(\lambda_i)\end{pmatrix}$$
$$= 2\sigma^4\sum_{i=1}^L v^2(\lambda_i)\begin{pmatrix} 1 & v(\lambda_i) \\ v(\lambda_i) & v^2(\lambda_i)\end{pmatrix}.$$

(b) Inequality (2.3) applied to $\{v(\lambda_i)\}$ and condition (3.19) show that both components of Δ_L converge to those of Δ_∞ in $L_1(\Omega)$. Eq. (3.21) is proved similarly to (3.18). □

Lemma 4. *Suppose that for each* L, $\operatorname{dlim} \delta_{nL} = \Delta_L$ *as* $n \to \infty$ *and that* $\operatorname{dlim} \Delta_L = \Delta_\infty$ *as* $L \to \infty$. *Suppose further that*

$$\lim_{L\to\infty} \limsup_{n\to\infty} P(|X_{n1} - \delta_{nL1}| + |X_{n2} - \delta_{nL2}| > \varepsilon) = 0$$

for each positive ε. *Then* $\operatorname{dlim} X_n = \Delta_\infty$ *as* $n \to \infty$.

This is just Theorem 4.2 from Billingsley [4] with the notation adapted to ours.

Lemma 5. *One has*

$$0 < c_{ni} \leqslant c_n < \infty, \quad i = 1, \dots, n, \tag{3.23}$$

and for $u \sim N(0, \sigma^2 I)$

$$E\left(\frac{\sum_{i=1}^n (c_n u_i^2 - \sigma^2 c_{ni}) v(\lambda_{ni})}{\sum_{i=1}^n u_i^2 v^2(\lambda_{ni})}\right) = 0. \tag{3.24}$$

Proof. Eq. (3.23) is obvious ($c_n < \infty$ because $n > 2$). Hoque [9] has proved that if S and B are symmetric matrices, B is positive definite and $u \sim N(0, \Omega)$, then

$$E\left(\frac{u'Su}{u'Bu}\right) = \int_0^\infty |I + 2t\Omega B|^{-1/2} \operatorname{tr}[(I + 2t\Omega B)^{-1}\Omega S]\, dt.$$

We apply this result to

$$S = \operatorname{diag}[v(\lambda_{n1}), \dots, v(\lambda_{nn})], \quad B = \operatorname{diag}[v^2(\lambda_{n1}), \dots, v^2(\lambda_{nn})],$$

$$\Omega = \sigma^2 I, \quad I + 2t\Omega B = \operatorname{diag}[1 + 2t\sigma^2 v^2(\lambda_{n1}), \dots, 1 + 2t\sigma^2 v^2(\lambda_{nn})],$$

$$(I + 2t\Omega B)^{-1}\Omega S = \operatorname{diag}\left[\frac{\sigma^2 v(\lambda_{n1})}{1 + 2t\sigma^2 v^2(\lambda_{n1})}, \dots, \frac{\sigma^2 v(\lambda_{nn})}{1 + 2t\sigma^2 v^2(\lambda_{nn})}\right].$$

Then

$$E\left(\frac{\sum_{i=1}^n u_i^2 v(\lambda_{ni})}{\sum_{i=1}^n u_i^2 v^2(\lambda_{ni})}\right) = \int_0^\infty \sum_{i=1}^n \frac{\sigma^2 v(\lambda_{ni})}{1 + 2t\sigma^2 v^2(\lambda_{ni})} \frac{dt}{\pi_n(\sigma^2 t)} = \sum_{i=1}^n c_{ni} v(\lambda_{ni}). \tag{3.25}$$

On the other hand, formula (10) from Jones [10] yields

$$E\left(\frac{\sigma^2}{\sum_{i=1}^n u_i^2 v^2(\lambda_{ni})}\right) = \int_0^\infty \frac{dt}{\pi_n(t)} = c_n. \tag{3.26}$$

Combining (3.25) and (3.26) we get

$$E\left(\frac{\sum_{i=1}^{n}(c_n u_i^2-\sigma^2 c_{ni})v(\lambda_{ni})}{\sum_{i=1}^{n}u_i^2 v^2(\lambda_{ni})}\right)=c_n E\left(\frac{\sum_{i=1}^{n}u_i^2 v(\lambda_{ni})}{\sum_{i=1}^{n}u_i^2 v^2(\lambda_{ni})}\right)-\sum_{i=1}^{n}c_{ni}v(\lambda_{ni})E\left(\frac{\sigma^2}{\sum_{i=1}^{n}u_i^2 v^2(\lambda_{ni})}\right)=c_n\sum_{i=1}^{n}c_{ni}v(\lambda_{ni})-c_n\sum_{i=1}^{n}c_{ni}v(\lambda_{ni})=0. \qquad \square$$

Lemma 6. *Eq.* (2.1) *implies* (2.20).

Proof. To avoid ambiguity, we restate the definitions of interpolation operators given earlier, in the form we need now: for an $n\times n$ matrix W_n and $z\in\mathbb{R}^n$ put

$$D_n^2 W_n = n\sum_{i,j=1}^{n} w_{nij}1_{q_{ij}},\qquad D_n^1 z=\sqrt{n}\sum_{i=1}^{n} z_i 1_{q_i}.$$

Denote $\mathcal{W}_n$ the integral operator

$$(\mathcal{W}_n f)(x)=\int_0^1 (D_n^2 W_n)(x,y)f(y)\,dy.$$

The first part of the proof consists in showing that there is a one-to-one correspondence between the set of nonzero eigenvalues of W_n and a similar set of $\mathcal{W}_n$. Let $W_n z=\lambda z$ with some $\lambda\neq 0$ and $z\neq 0$. Put $f=D_n^1 z$. If $x\in[0,1]$, we can assume that $x\in q_i$ for some *i* (thereby neglecting a finite number of points). Then

$$(D_n^2 W_n)(x,y)=n\sum_j w_{nij}1_{q_{ij}}(x,y),\qquad f(x)=\sqrt{n}z_i,$$

so that

$$(\mathcal{W}_n f)(x)=\sum_j\int_{q_j} n\sum_j w_{nij}1_{q_{ij}}\sqrt{n}z_j\,dy=\sum_j w_{nij}\sqrt{n}z_j=\lambda f(x).$$

Since *f* is nontrivial, λ is an eigenvalue of $\mathcal{W}_n$ (in this part of the proof the assumption $\lambda\neq 0$ is not necessary). Conversely, let $\lambda\neq 0$ be an eigenvalue of $\mathcal{W}_n$. Suppose $x\in q_i$. $\mathcal{W}_n f=\lambda f$ implies

$$n\sum_j w_{nij}\int_{q_j} f(y)\,dy=\lambda f(x).$$

Since the left-hand side is constant and $\lambda\neq 0$, *f* is constant on q_i: $f(x)=z_i$. Hence, the last equation yields $\sum_j w_{nij}z_j=\lambda z_i,\ i=1,\dots,n$, or $W_n z=\lambda z$. z is nontrivial because otherwise *f* is trivial.

The statement we have just proved is sufficient for our purposes because the sums in (2.20) are not affected by zero eigenvalues. In the second part of the proof we need some facts from Gohberg and Kreĭn [8]. *s*-numbers of an operator *A* in a Hilbert space *H* are defined as eigenvalues of the operator $(A'A)^{1/2}$: $s_j(A) = \lambda_j((A'A)^{1/2})$. The facts we need are:

(1) For self-adjoint operators $s_j(A) = |\lambda_j(A)|$ (p. 27).

(2) For an integral operator $\mathcal{K}$ with a square-integrable kernel K one has $\|K\|_2 = \left(\sum_{i=1}^{\infty} s_j^2(\mathcal{K})\right)^{1/2}$ (pp. 108–109).

(3) The expression $\|A\|_{\sigma_p} = \left(\sum_{i=1}^{\infty} s_i^p(\mathcal{K})\right)^{1/p}$, $1 \leqslant p < \infty$, is a norm (p. 92).

These facts and (2.1) give

$$\left|\left(\sum_{i=1}^{n} \lambda_{ni}^2\right)^{1/2} - \left(\sum_{i=1}^{\infty} \lambda_i^2\right)^{1/2}\right| = \left|\|\mathcal{W}_n\|_{\sigma_2} - \|\mathcal{K}\|_{\sigma_2}\right| \leqslant \|\mathcal{W}_n - \mathcal{K}\|_{\sigma_2}$$
$$= \|D_n^2 W_n - K\|_2 \to 0.$$

The last line follows from (2.1) and (3.11). □

3.2. *Proof of Theorem 1*

Part (1): Due to identity (2.5), condition (2.7) is equivalent to

$$|\rho|\|K\|_2 < 1. \tag{3.27}$$

Hence, $|\rho|\|K\|_2 \leqslant 1 - 2\varepsilon$ for some sufficiently small $\varepsilon > 0$ and then (3.12) shows that there exists $n_0 = n_0(\varepsilon)$ such that

$$\sup_{n \geqslant n_0} |\rho|\|W_n\|_2 \leqslant 1 - \varepsilon. \tag{3.28}$$

By Lemma 1(a) $G_n = s(W_n)$ exists and, moreover,

$$\|G_n\|_2 \leqslant \sum_{k \geqslant 0} |\rho|^k \|W_n\|_2^{k+1} = \frac{\|W_n\|_2}{1 - |\rho| \|W_n\|_2} \leqslant c \quad \text{for all } n \geqslant n_0. \tag{3.29}$$

The reduced form $Y_n = S_n^{-1} V_n$ of the basic model (1.1) and (2.6) lead to (2.8) in the usual way:

$$\hat{\rho} = (Z_n' Z_n)^{-1} Z_n' (\rho Z_n + V_n) = \rho + (V_n' G_n' G_n V_n)^{-1} V_n' G_n' V_n. \qquad \square$$

Part (2): Here is the plan of the proof. The numerator and denominator of (2.8) will be considered coordinates of a new random vector X_n. X_n will be approximated by another vector with $s(d_n K)$ instead of $G_n = s(W_n)$. That second vector, in turn, will be approximated by yet another vector with $s(d_n K_L)$ where K_L is an initial segment of (2.4):

$$K_L(x, y) = \sum_{i=1}^{L} \lambda_i f_i(x) f_i(y). \tag{3.30}$$

To this last vector we shall be able to apply Lemma 3. Billingsley's Lemma 4 will help us handle a double-indexed family of vectors that occurs in the course of the proof.

The scheme we have just explained is realized through the representation

$$X_n = \alpha_n + \beta_{nL} + \gamma_{nL} + \delta_{nL}, \tag{3.31}$$

where

$$X_n = \begin{pmatrix} V_n' G_n' V_n \\ V_n' G_n' G_n V_n \end{pmatrix}, \quad \alpha_n = \begin{pmatrix} V_n'\big(G_n' - s(d_n K)\big)V_n \\ V_n'\big(G_n' G_n - s^2(d_n K)\big)V_n \end{pmatrix},$$

$$\beta_{nL} = \begin{pmatrix} V_n'\big(s(d_n K) - s(d_n K_L)\big)V_n \\ V_n'\big(s^2(d_n K) - s^2(d_n K_L)\big)V_n \end{pmatrix}, \quad \gamma_{nL} = \begin{pmatrix} V_n' s(d_n K_L) V_n \\ V_n' s^2(d_n K_L) V_n \end{pmatrix} - \delta_{nL},$$

δ_{nL} has been defined before Lemma 3. Our goal is to show that α_n, β_{nL} and γ_{nL} are negligible in some sense and therefore δ_{nL} represents the main part of X_n. We evaluate coordinates of the alphas, betas, and gammas separately.

Bounding α_n: Using (3.4) for $k = 0$ and (3.5) for positive k, we have

$$\begin{aligned} \|\alpha_{n1}\|_2 &= N\left(\sum_{k\geqslant 0} \rho^k \big((W_n')^{k+1} - (d_n K)^{k+1}\big)\right) \\ &\leqslant N\left(W_n' - d_n K\right) + \sum_{k>0} |\rho|^k N\left((W_n')^{k+1} - (d_n K)^{k+1}\right) \\ &\leqslant c\left\|W_n' - d_n K\right\|_2 \left[\sqrt{n} + \sum_{k>0}(k+1)\left(|\rho| \max\left\{\left\|W_n'\right\|_2, \|d_n K\|_2\right\}\right)^k\right]. \end{aligned} \tag{3.32}$$

Because of (2.3), assumption (2.9) implies (2.7) and, consequently, (3.27). Hence, in the way we have derived (3.28) we can now derive

$$\sup_{n\geqslant n_0} |\rho| \max\left\{\left\|W_n'\right\|_2, \|d_n K\|_2\right\} \leqslant 1 - \varepsilon. \tag{3.33}$$

This allows us to continue (3.32) using (3.13)

$$\|\alpha_{n1}\|_2 \leqslant c\sqrt{n}\left\|W_n' - d_n K\right\|_2 = o(1). \tag{3.34}$$

Repeating the argument which led us to (3.29) we can assert that for the ε from (3.33) there exists $n_0 = n_0(\varepsilon)$ such that

$$\sup_{n\geqslant n_0} \|G_n\|_2 < \infty, \quad \sup_{n\geqslant n_0} \|s(d_n K)\|_2 < \infty. \tag{3.35}$$

By (3.2)

$$\left\|G_n' - s(d_n K)\right\|_2 = \|G_n - s(d_n K)\|_2 \leqslant c\|W_n - d_n K\|_2, \tag{3.36}$$

where we have used the symmetry of $s(d_n K)$ (see the proof of Lemma 2(a)) and the fact that $\varphi(\rho, W_n, d_n K) < \infty$ because of (3.33). Now we may use (3.3), (3.35) and (3.36) to obtain

$$\begin{aligned} \|\alpha_{n2}\|_2 &\leqslant N\left(\big(G_n' - s(d_n K)\big)G_n\right) + N\left(s(d_n K)\big(G_n - s(d_n K)\big)\right) \\ &\leqslant c(\|G_n' - s(d_n K)\|_2\|G_n\|_2 + \|s(d_n K)\|_2\|G_n - s(d_n K)\|_2) \\ &\leqslant c_1\|W_n - d_n K\|_2. \end{aligned} \tag{3.37}$$

Bounding β_{nL}: For any $1 \leqslant L < M \leqslant \infty$ we can write by Lemma 2(c)

$$\left(d_n\left(\sum_{i=L}^{M} \lambda_i f_i(x) f_i(y)\right)\right)_{st} = \sum_{i=L}^{M} \lambda_i (d_n f_i)_s (d_n f_i)_t, \quad s, t = 1, \ldots, n \tag{3.38}$$

(here the d_n at the left is two-dimensional and at the right one-dimensional). Since for any n, i, j by the Cauchy–Schwartz inequality and (3.9)

$$|(d_n f_i, d_n f_j)_{l_2}| \leqslant \|d_n f_i\|_2 \left\|d_n f_j\right\|_2 \leqslant \|f_i\|_2 \left\|f_j\right\|_2 = 1, \tag{3.39}$$

we deduce from (3.38)

$$\begin{aligned}\left\|d_n\left(\sum_{i=L}^{M}\lambda_i f_i(x) f_i(y)\right)\right\|_2^2 &= \sum_{s,t=1}^{n}\sum_{i,j=L}^{M}\lambda_i\lambda_j (d_n f_i)_s (d_n f_i)_t (d_n f_j)_s (d_n f_j)_t \\ &= \sum_{i,j=L}^{M}\lambda_i\lambda_j (d_n f_i, d_n f_j)_{l_2}^2 \leqslant \left(\sum_{i=L}^{M}|\lambda_i|\right)^2.\end{aligned}$$

This bound along with decompositions (2.4) and (3.30) of K and K_L produces three particular cases:

$$\|d_n K\|_2 \leqslant \sum_{i\geqslant 1}|\lambda_i|, \quad \|d_n K_L\|_2 \leqslant \sum_{i\geqslant L}|\lambda_i|, \quad \|d_n K - d_n K_L\|_2 \leqslant \sum_{i>L}|\lambda_i|. \tag{3.40}$$

The last bound will be used for estimating the terms in β_{nL} with $k > 0$. For $k = 0$, (3.8), (3.38) and (3.40) give the inequality

$$N(d_n K - d_n K_L) \leqslant c\left(\|d_n K - d_n K_L\|_2 + \left|\sum_{i>L}\lambda_i \|d_n f_i\|_2^2\right|\right) \leqslant c_1 \sum_{i>L}|\lambda_i|. \tag{3.41}$$

Overall, utilizing (3.5), (3.40) and (3.41) we can bound the first component of β_{nL} as follows:

$$\begin{aligned}\left\|\beta_{nL1}\right\|_2 &\leqslant N(d_n K - d_n K_L) + \sum_{k>0}|\rho|^k N\left((d_n K)^{k+1} - (d_n K_L)^{k+1}\right) \\ &\leqslant c_1 \sum_{i>L}|\lambda_i|\left[1 + \sum_{k>0}(k+1)\left(|\rho|\sum_{i\geqslant 1}|\lambda_i|\right)^k\right] \leqslant c_2 \sum_{i>L}|\lambda_i|.\end{aligned} \tag{3.42}$$

It is important that c_2 here does not depend on n.

Eq. (3.40) trivially leads to the bound

$$\max\left\{\|s(d_n K)\|_2, \|s(d_n K_L)\|_2\right\} \leqslant \sum_{k\geqslant 0}|\rho|^k \left(\sum_{i\geqslant 1}|\lambda_i|\right)^{k+1} \leqslant c, \tag{3.43}$$

which is uniform in n and L, while (3.2) and (3.40) guarantee that

$$\|s(d_n K) - s(d_n K_L)\|_2 \leqslant c\|d_n K - d_n K_L\|_2 \leqslant c\sum_{i>L}|\lambda_i|, \tag{3.44}$$

where

$$c = \varphi(\rho, d_n K, d_n K_L) \leqslant \sum_{k\geqslant 0}(k+1)\left(|\rho|\sum_{i\geqslant 1}|\lambda_i|\right)^k < \infty.$$

It follows from (3.3), (3.43) and (3.44) that

$$\begin{aligned}\|\beta_{nL2}\|_2 &\leqslant N\left(\left(s(d_nK)-s(d_nK_L)\right)s(d_nK)\right)+N\left(s(d_nK_L)\left(s(d_nK)-s(d_nK_L)\right)\right)\\ &\leqslant c\|s(d_nK)-s(d_nK_L)\|_2(\|s(d_nK)\|_2+\|s(d_nK_L)\|_2)\\ &\leqslant c_1\sum_{i>L}|\lambda_i|. \end{aligned} \tag{3.45}$$

Estimating γ_{nL}: Using formula (3.38) it is easy to show by induction that (see Lemma 2(b) for the notation μ_{ni})

$$(d_nK_L)^{k+1}_{st}=\sum_{i_1,\ldots,i_{k+1}\leqslant L}\prod_{j=1}^{k+1}\lambda_{i_j}\mu_{ni}(d_nf_{i_1})_s(d_nf_{i_{k+1}})_t. \tag{3.46}$$

Hence, in terms of the vector U_{nL} used in Lemma 3

$$\begin{aligned}V_n's(d_nK_L)V_n &= \sum_{s,t=1}^{n}\sum_{k\geqslant 0}\rho^k(d_nK_L)^{k+1}_{st}v_sv_t\\ &=\sum_{k\geqslant 0}\rho^k\sum_{i_1,\ldots,i_{k+1}\leqslant L}\prod_{j=1}^{k+1}\lambda_{i_j}\mu_{ni}U_{nLi_1}U_{nLi_{k+1}}.\end{aligned}$$

We need to express δ_{nL1} in similar terms. Replacing $1/(1-\rho\lambda_i)$ by $\sum_{k\geqslant 0}(\rho\lambda_i)^k$ gives

$$\delta_{nL1}=\sum_{i=1}^{L}U^2_{nLi}\sum_{k\geqslant 0}\rho^k\lambda_i{}^{k+1}=\sum_{k\geqslant 0}\rho^k\sum_{i=1}^{L}\lambda_i{}^{k+1}U^2_{nLi}.$$

Since $\mu_{\infty i}$ vanishes for i with different components, this is the same as

$$\delta_{nL1}=\sum_{k\geqslant 0}\rho^k\sum_{i_1,\ldots,i_{k+1}\leqslant L}\prod_{j=1}^{k+1}\lambda_{i_j}\mu_{\infty i}U_{nLi_1}U_{nLi_{k+1}}.$$

The result is the representation

$$\gamma_{nL1}=\sum_{k\geqslant 0}\rho^k\sum_{i_1,\ldots,i_{k+1}\leqslant L}\prod_{j=1}^{k+1}\lambda_{i_j}(\mu_{ni}-\mu_{\infty i})U_{nLi_1}U_{nLi_{k+1}}, \tag{3.47}$$

which can be used for bounding.

By the Hölder inequality, (3.8) and (3.39) for any i, j

$$\begin{aligned}E\left|U_{nLi}U_{nLj}\right| &\leqslant \left[E\left(V_n'd_nf_iV_n'd_nf_j\right)^2\right]^{1/2}=N(d_nf_id_nf_j')\\ &\leqslant c\left[\left(\sum_{s,t=1}^{n}(d_nf_i)_s^2(d_nf_j)_t^2\right)^{1/2}+\left|\sum_{s=1}^{n}(d_nf_i)_s(d_nf_j)_s\right|\right]\\ &= c\left[\|d_nf_i\|_2\left\|d_nf_j\right\|_2+|(d_nf_i,d_nf_j)_{l_2}|\right]\leqslant c_1.\end{aligned} \tag{3.48}$$

According to (3.14), for any positive (small) ε and (large) L we can choose $n_0=n_0(\varepsilon,L)$ so large that

$$|\mu_{ni}-\mu_{\infty i}|\leqslant\varepsilon\quad\text{for all } n\geqslant n_0 \text{ and } i_1,\ldots,i_{k+1}\leqslant L. \tag{3.49}$$

Finally, we conclude from (3.47)–(3.49) that for all $n \geqslant n_0$

$$E|\gamma_{nL1}| \leqslant c_1\varepsilon \sum_{k\geqslant 0} |\rho|^k \sum_{i_1,\ldots,i_{k+1}\leqslant L} \prod_{j=1}^{k+1} |\lambda_{i_j}| \leqslant c_1\varepsilon \sum_{k\geqslant 0} |\rho|^k \left(\sum_{i\geqslant 1} |\lambda_i|\right)^{k+1} = c_2\varepsilon. \quad (3.50)$$

For numbers or square matrices a one has the identity

$$\left(\sum_{k\geqslant 0} a^k\right)^2 = \sum_{k,\,l\geqslant 0} a^{k+l} = \sum_{m\geqslant 0} a^m(m+1) \quad (3.51)$$

because there are $(m+1)$ pairs (k, l) such that $k + l = m$. If one chooses $a = \rho d_n K_L$ here and then applies (3.46), one gets

$$\begin{aligned} V_n' s^2(d_n K_L) V_n &= V_n' \left(\sum_{k\geqslant 0} (\rho d_n K_L)^k\right)^2 (d_n K_L)^2 V_n \\ &= V_n' \sum_{m\geqslant 0} \rho^m (m+1)(d_n K_L)^{m+2} V_n \\ &= \sum_{m\geqslant 0} \rho^m (m+1) \sum_{s,t=1}^{n} (d_n K_L)_{st}^{m+2} v_s v_t \\ &= \sum_{m\geqslant 0} \rho^m (m+1) \sum_{i_1,\ldots,i_{m+2}\leqslant L} \prod_{j=1}^{m+2} \lambda_{i_j} \mu_{ni} U_{nLi_1} U_{nLi_{m+2}}. \end{aligned} \quad (3.52)$$

Application of (3.51) also provides another expression for

$$\begin{aligned} \delta_{nL2} &= \sum_{i=1}^{L} U_{nLi}^2 \lambda_i^2 \left(\sum_{k\geqslant 0} (\rho\lambda_i)^k\right)^2 = \sum_{i=1}^{L} U_{nLi}^2 \lambda_i^2 \sum_{m\geqslant 0} (\rho\lambda_i)^m (m+1) \\ &= \sum_{m\geqslant 0} \rho^m (m+1) \sum_{i=1}^{L} U_{nLi}^2 \lambda_i^{m+2}. \end{aligned}$$

Since $\mu_{\infty i} = 0$ if among the indices $i_1, \ldots, i_{m+2}$ there are at least two different ones, δ_{nL2} equals

$$\delta_{nL2} = \sum_{m\geqslant 0} \rho^m (m+1) \sum_{i_1,\ldots,i_{m+2}\leqslant L} \prod_{j=1}^{m+2} \lambda_{i_j} \mu_{\infty i} U_{nLi_1} U_{nLi_{m+2}}.$$

Therefore, taking into account also (3.52), we can rewrite γ_{nL2} as

$$\begin{aligned} \gamma_{nL2} &= V_n' s^2(d_n K_L) V_n - \delta_{nL2} \\ &= \sum_{m\geqslant 0} \rho^m (m+1) \sum_{i_1,\ldots,i_{m+2}\leqslant L} \prod_{j=1}^{m+2} \lambda_{i_j} (\mu_{ni} - \mu_{\infty i}) U_{nLi_1} U_{nLi_{m+2}}. \end{aligned}$$

As above, application of (3.48) and (3.49) leads to an analog of (3.50): for any positive ε, L there is $n_0 = n_0(\varepsilon, L)$ such that

$$E|\gamma_{nL2}| \leqslant c_1\varepsilon \sum_{m\geqslant 0} |\rho|^m (m+1) \sum_{i_1,\dots,i_{m+2}\leqslant L} \prod_{j=1}^{m+2} |\lambda_{i_j}| \leqslant c_2\varepsilon \tag{3.53}$$

for all $n \geqslant n_0$.

Proving (2.10): Under condition (2.9) we have

$$0 < c_1 = 1 - |\rho| \sum_{i\geqslant 1} |\lambda_i| \leqslant 1 - |\rho\lambda_i| \leqslant |1 - \rho\lambda_i|$$
$$\leqslant 1 + |\rho\lambda_i| \leqslant 1 + |\rho| \sum_{i\geqslant 1} |\lambda_i| = c_2 < \infty \quad \text{all } i$$

so that

$$\frac{|\lambda_i|}{c_2} \leqslant |v(\lambda_i)| \leqslant \frac{|\lambda_i|}{c_1} \quad \text{all } i, \tag{3.54}$$

where c_1 and c_2 depend on ρ. Hence, the condition $\sum_{i\geqslant 1} |\lambda_i| < \infty$ is equivalent to (3.19), and we can use (3.16) and (3.20).

Eqs. (3.34) and (3.37) show that $\operatorname{plim} \alpha_n = 0$. From (3.42) and (3.45) we have by the Chebyshev inequality

$$P\left(|\beta_{nL1}| + |\beta_{nL2}| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2} \left\| |\beta_{nL1}| + |\beta_{nL2}| \right\|_2 \leqslant \frac{c}{\varepsilon^2} \sum_{i>L} |\lambda_i|,$$

where c does not depend on n. From (3.50) and (3.53) we conclude that for any fixed L $\operatorname{plim}_{n\to\infty} \gamma_{nL} = 0$. Thus, (3.31) implies

$$\limsup_{n\to\infty} P(|X_{n1} - \delta_{nL1}| + |X_{n2} - \delta_{nL2}| > \varepsilon) \leqslant \frac{c}{\varepsilon^2} \sum_{i>L} |\lambda_i|.$$

All conditions of Lemma 4 are satisfied and, consequently,

$$\operatorname*{dlim}_{n\to\infty} X_n = \Delta_\infty.$$

By the continuous mapping theorem (Theorem 5.1 from Billingsley [4]) it follows that

$$\operatorname{dlim}(\hat\rho - \rho) = \operatorname{dlim} \frac{X_{n1}}{X_{n2}} = \frac{\Delta_{\infty 1}}{\Delta_{\infty 2}},$$

which is (2.10). Theorem 5.1 is applicable because $\Delta_{\infty 2} > 0$ almost surely. □

Part (3): *Proving* (2.11). In the definition of $\hat\sigma^2$ we may as well put n instead of $n-1$. Substituting $S_n(\hat\rho)S_n^{-1} = I - (\hat\rho - \rho)G_n$ we have

$$\begin{aligned}\sqrt{n}(\hat\sigma^2 - \sigma^2) &= \sqrt{n}\frac{V_n' S_n'^{-1} S_n'(\hat\rho) S_n(\hat\rho) S_n^{-1} V_n}{n} - \sqrt{n}\sigma^2 \\ &= \sqrt{n}\frac{V_n'V_n - n\sigma^2}{n} + 2\frac{\rho - \hat\rho}{n^\varepsilon}\frac{V_n' G_n' V_n}{n^{1/2-\varepsilon}} + \frac{(\rho - \hat\rho)^2}{n^\varepsilon}\frac{V_n' G_n' G_n V_n}{n^{1/2-\varepsilon}} \\ &= \frac{\sum (v_i^2 - \sigma^2)}{\sqrt{n}} + 2\frac{\rho - \hat\rho}{n^\varepsilon}\frac{X_{n1}}{n^{1/2-\varepsilon}} + \frac{(\rho - \hat\rho)^2}{n^\varepsilon}\frac{X_{n2}}{n^{1/2-\varepsilon}}.\end{aligned} \tag{3.55}$$

Here $\varepsilon \in (0, \frac{1}{2})$ is arbitrary. From the proof of Part (2) we know that X_{n1}, X_{n2}, $\rho - \hat{\rho}$ and $(\rho - \hat{\rho})^2$ converge in distribution. Therefore, the second and third terms in the last line are $o_p(1)$. The first term is known to converge to $N(0, \mu_4 - \sigma^4)$ in distribution. □

3.3. Proof of Theorem 2

Proving that limit (2.12) *is zero*: The next equation is quite similar to the passage from (3.32) to (3.34):

$$
\begin{aligned}
|\mathrm{tr}(G_n) - \mathrm{tr}(s(d_n K))| &= |\mathrm{tr}(s(W_n) - s(d_n K))| \leqslant |\mathrm{tr}\,(W_n - d_n K)\,| \\
&\quad + \sum_{k>0} |\rho|^k \left|\mathrm{tr}\left({W_n}^{k+1} - (d_n K)^{k+1}\right)\right| \\
&\leqslant \|W_n - d_n K\|_2 \left[\sqrt{n}\right. \\
&\quad \left. + \sum_{k>0} (k+1)\,(|\rho| \max\{\|W_n\|_2\,, \|d_n K\|_2\})^k\right] \\
&= o(1).
\end{aligned}
$$

Using (3.46) and (3.14) we see that

$$
\begin{aligned}
\mathrm{tr}(s(d_n K)) &= \sum_{k \geqslant 0} \rho^k \,\mathrm{tr}((d_n K)^{k+1}) \\
&= \sum_{k \geqslant 0} \rho^k \sum_{i_1,\ldots,i_{k+1}=1}^{\infty} \prod_{j=1}^{k+1} \lambda_{i_j} \mu_{ni}(d_n f_{i_1}, d_n f_{i_{k+1}})_{l_2} \\
&\to \sum_{k \geqslant 0} \rho^k \sum_{i_1,\ldots,i_{k+1}=1}^{\infty} \prod_{j=1}^{k+1} \lambda_{i_j} \mu_{\infty i}(f_{i_1}, f_{i_{k+1}})_{L_2} \\
&= \sum_{k \geqslant 0} \rho^k \sum_{i \geqslant 1} \lambda_i^{k+1} = \sum_{i \geqslant 1} v(\lambda_i).
\end{aligned}
$$

Sending $n \to \infty$ here is possible because under condition (2.9) the series converge uniformly. The conclusion is that

$$
\lim_{n \to \infty} \mathrm{tr}(G_n) = \sum_{i \geqslant 1} v(\lambda_i), \tag{3.56}
$$

where the series at the right converges because of (3.54).

Reviewing the argument that took us from (3.35) to (3.37) we see that

$$
\begin{aligned}
|\mathrm{tr}(G_n' G_n) - \mathrm{tr}(s^2(d_n K))| &\leqslant \left|\mathrm{tr}\left(\left(G_n' - s(d_n K)\right) G_n\right)\right| + \left|\mathrm{tr}\left(s(d_n K)\left(G_n - s(d_n K)\right)\right)\right| \\
&\leqslant \|G_n' - s(d_n K)\|_2 \|G_n\|_2 \\
&\quad + \|G_n - s(d_n K)\|_2 \|s(d_n K)\|_2 \to 0.
\end{aligned}
$$

Arguing along the lines following (3.26) we have

$$\begin{aligned}\operatorname{tr}(s^2(d_nK)) &= \operatorname{tr}\left[\left(\sum_{k\geqslant 0}(\rho d_nK)^k\right)^2(d_nK)^2\right]\\ &= \operatorname{tr}\left(\sum_{m\geqslant 0}\rho^m(m+1)(d_nK)^{m+2}\right)\\ &= \sum_{m\geqslant 0}\rho^m(m+1)\sum_{i_1,\ldots,i_{m+2}=1}^{\infty}\prod_{j=1}^{m+2}\lambda_{i_j}\mu_{ni}(d_nf_{i_1},d_nf_{i_{m+2}})_{l_2}.\end{aligned}$$

The last expression tends to

$$\begin{aligned}\sum_{m\geqslant 0}\rho^m(m+1)\sum_{i\geqslant 1}\lambda_i^{m+2} &= \sum_{i\geqslant 1}\lambda_i^2\sum_{m\geqslant 0}(\rho\lambda_i)^m(m+1)\\ &= \sum_{i\geqslant 1}\lambda_i^2\left(\sum_{k\geqslant 0}(\rho\lambda_i)^k\right)^2 = \sum_{i\geqslant 1}v^2(\lambda_i).\end{aligned}$$

Thus,

$$\lim_{n\to\infty}\operatorname{tr}(G_n'G_n) = \sum_{i\geqslant 1}v^2(\lambda_i). \tag{3.57}$$

In this proof we can replace G_n' by G_n. Then instead of (3.57) we have

$$\lim_{n\to\infty}\operatorname{tr}(G_n^2) = \sum_{i\geqslant 1}v^2(\lambda_i). \tag{3.58}$$

Eqs. (3.56)–(3.58) show that limit (2.12) is zero. □

Proving that limit (2.13) *is zero*: In accordance with the ML methodology, here we recall the true value is ρ_0 and use ρ for points close to ρ_0. The transformation in the next equation is analogous to that in (3.55):

$$\begin{aligned}\sigma_n^2(\rho) &= \frac{\sigma_0^2}{n}\operatorname{tr}\left[(I-(\rho-\rho_0)G_n)'(I-(\rho-\rho_0)G_n)\right]\\ &= \frac{\sigma_0^2}{n}\operatorname{tr}\left[I-2(\rho-\rho_0)G_n+(\rho-\rho_0)^2G_n'G_n\right]\\ &= \sigma_0^2\left[1-2(\rho-\rho_0)\frac{\operatorname{tr}(G_n)}{n}+(\rho-\rho_0)^2\frac{\operatorname{tr}(G_n'G_n)}{n}\right].\end{aligned}$$

It is clear from (3.56) and (3.57) that

$$\lim \sigma_n^2(\rho) = \sigma_0^2 \quad \text{for any } \rho \text{ satisfying (2.9).} \tag{3.59}$$

Using properties of logs, determinants and the fact that $S_n(\rho)$, S_n and their inverses commute with each other (as functions of the same matrix W_n) we have

$$\begin{aligned}&\ln|\sigma_0^2S_n^{-1}S_n'^{-1}| - \ln|\sigma_n^2(\rho)S_n^{-1}(\rho)S_n'^{-1}(\rho)|\\ &\quad = \ln(\sigma_0^2/\sigma_n^2(\rho)) + 2(\ln|S_n(\rho)| - \ln|S_n|).\end{aligned} \tag{3.60}$$

Formula (2.17) implies (cf. Gohberg and Kreĭn [8, p. 158])

$$\ln|S_n(\rho)| - \ln|S_n| = -\int_{\rho_0}^{\rho} \mathrm{tr}(W_n S_n^{-1}(t))\,dt = -\int_{\rho_0}^{\rho} \mathrm{tr}(s(t, W_n))\,dt,$$

where we have denoted $s(t, W_n) = \sum_{k=0}^{\infty} t^k W_n^{k+1}$. Here we are assuming that $|\rho_0| < 1/\sum_{i\geqslant 1}|\lambda_i|$ and ρ is in a small neighborhood of ρ_0 so that $s(t, W_n)$ converges uniformly on the segment connecting ρ_0 and ρ. Similarly to (3.56) one can show that

$$\lim_{n\to\infty} \mathrm{tr}(s(t, W_n)) = \lim_{n\to\infty} \mathrm{tr}(s(t, d_n K)) = \sum_{i\geqslant 1} \frac{\lambda_i}{1 - t\lambda_i}$$

uniformly in t from the neighborhood indicated above. Therefore,

$$\lim_{n\to\infty} (\ln|S_n(\rho)| - \ln|S_n|) = -\int_{\rho_0}^{\rho} \sum_{i\geqslant 1} \frac{\lambda_i}{1 - t\lambda_i}\,dt.$$

This relation, (3.59) and (3.60) show that the limit in (2.13) is zero for ρ close to ρ_0. □

Proving that limit (2.14) *is zero*: The desired result will follow if we show that $L_2(\Omega)$-norms of all elements of A_n are uniformly bounded. To this end, the reader can consult (3.3), (3.12) and statement (1) of Theorem 1 and verify that

$$\left(E(Y_n' W_n'^2 W_n Y_n)^2\right)^{1/2} = N(S_n'^{-1} W_n'^2 W_n S_n^{-1}) \leqslant c_1 \left\|S_n^{-1}\right\|_2^2 \|W_n\|_2^3 \leqslant c_2,$$

$$|\mathrm{tr}(W_n' W_n)| \leqslant \|W_n\|_2^2 \leqslant c,$$

$$\left(E(Y_n' W_n Y_n)^2\right)^{1/2} = N(S_n'^{-1} W_n S_n^{-1}) \leqslant c_1 \left\|S_n^{-1}\right\|_2^2 \|W_n\|_2 \leqslant c_2,$$

$$\left(E(Y_n' W_n'^2 W_n^2 Y_n)^2\right)^{1/2} = N(S_n'^{-1} W_n'^2 W_n^2 S_n^{-1}) \leqslant c_1 \left\|S_n^{-1}\right\|_2^2 \|W_n\|_2^4 \leqslant c_2,$$

$$\left(E(Y_n' W_n' W_n Y_n)^2\right)^{1/2} = N(S_n'^{-1} W_n' W_n S_n^{-1}) \leqslant c_1 \left\|S_n^{-1}\right\|_2^2 \|W_n\|_2^2 \leqslant c_2,$$

$$\left(E(Y_n' W_n^2 Y_n)^2\right)^{1/2} = N(S_n'^{-1} W_n^2 S_n^{-1}) \leqslant c_1 \left\|S_n^{-1}\right\|_2^2 \|W_n\|_2^2 \leqslant c_2. \qquad \square$$

3.4. Proof of Theorem 3

Deriving (2.21): Denoting

$$v(\rho, \lambda_{ni}) = \frac{\lambda_{ni}}{1 - \rho\lambda_{ni}}, \quad i = 1, \ldots, n,$$

and using (2.18), for the matrices involved in (2.19) we have representations

$$S_n(\hat\rho) = P_n\,\mathrm{diag}[1 - \hat\rho\lambda_{n1}, \ldots, 1 - \hat\rho\lambda_{nn}]P_n',$$

$$G_n = P_n\,\mathrm{diag}[v(\lambda_{n1}), \ldots, v(\lambda_{nn})]P_n',$$

$$\mathrm{tr}(A_n W_n S_n^{-1}(\hat\rho)) = \frac{1}{c_n}\sum_{i=1}^{n} c_{ni}\, v(\hat\rho, \lambda_{ni}).$$

It is easy to see that the vector $\widetilde{V}_n = P_n' V_n$ is distributed as $N(0, \sigma^2 I)$. Eq. (2.19) becomes

$$\rho_{\text{corr}} = \frac{\sum_{i=1}^n \widetilde{v}_i^2 v(\lambda_{ni})/(1-\rho\lambda_{ni}) - \hat{\sigma}^2/c_n \sum_{i=1}^n c_{ni} v(\hat{\rho}, \lambda_{ni})}{\sum_{i=1}^n \widetilde{v}_i^2 v^2(\lambda_{ni})}.$$

The numerator can be rearranged as follows:

$$\begin{aligned}
&\sum_{i=1}^n \widetilde{v}_i^2 \frac{v(\lambda_{ni})}{1-\rho\lambda_{ni}} - \frac{\hat{\sigma}^2}{c_n}\sum_{i=1}^n c_{ni} v(\hat{\rho}, \lambda_{ni}) \\
&\quad = \sum_{i=1}^n \widetilde{v}_i^2 \left(\frac{v(\lambda_{ni})}{1-\rho\lambda_{ni}} - v(\lambda_{ni}) \right) + \sum_{i=1}^n \left(\widetilde{v}_i^2 - \frac{\sigma^2 c_{ni}}{c_n} \right) v(\lambda_{ni}) + \frac{\sigma^2 - \hat{\sigma}^2}{c_n} \sum_{i=1}^n c_{ni} v(\lambda_{ni}) \\
&\qquad + \frac{\hat{\sigma}^2}{c_n} \sum_{i=1}^n c_{ni}(v(\lambda_{ni}) - v(\hat{\rho}, \lambda_{ni})).
\end{aligned}$$

Hence, if we denote

$$\kappa_{n0} = \sum_{i=1}^n \widetilde{v}_i^2 v^2(\lambda_{ni}), \quad \kappa_{n1} = \frac{1}{\kappa_{n0}} \sum_{i=1}^n \left(\widetilde{v}_i^2 - \frac{\sigma^2 c_{ni}}{c_n} \right) v(\lambda_{ni}),$$

$$\kappa_{n2} = \frac{\sigma^2 - \hat{\sigma}^2}{\kappa_{n0} c_n} \sum_{i=1}^n c_{ni} v(\lambda_{ni}), \quad \kappa_{n3} = \frac{\hat{\sigma}^2}{\kappa_{n0}},$$

then ρ_{corr} becomes

$$\rho_{\text{corr}} = \rho + \kappa_{n1} + \kappa_{n2} + \kappa_{n3} \sum_{i=1}^n \frac{c_{ni}}{c_n} (v(\lambda_{ni}) - v(\hat{\rho}, \lambda_{ni})).$$

If we also take into account that

$$\begin{aligned}
v(\lambda_{ni}) - v(\hat{\rho}, \lambda_{ni}) &= v(\rho, \lambda_{ni}) - v(\hat{\rho}, \lambda_{ni}) \\
&= \int_{\hat{\rho}}^{\rho} \frac{\partial v(t, \lambda_{ni})}{\partial t}\, dt = \int_{\hat{\rho}}^{\rho} v^2(t, \lambda_{ni})\, dt
\end{aligned}$$

and denote

$$\psi_n(t) = \sum_{i=1}^n \frac{c_{ni}}{c_n} v^2(t, \lambda_{ni}),$$

then ρ_{corr} can be re-written as (2.21).

Final touches: The validity of the first equation in (2.22) follows from (3.24):

$$E\kappa_{n1} = \frac{1}{c_n} E \left(\frac{\sum_{i=1}^n (c_n \widetilde{v}_i{}^2 - \sigma^2 c_{ni}) v(\lambda_{ni})}{\sum_{i=1}^n \widetilde{v}_i{}^2 v^2(\lambda_{ni})} \right) = 0.$$

We claim that (see Lemma 3)

$$\operatorname*{dlim}_{n\to\infty} \kappa_{n0} = \Delta_{\infty 2} = \sigma^2 \sum_{i=1}^{\infty} u_i^2 v^2(\lambda_i). \tag{3.61}$$

This is so because $\kappa_{n0} = V_n' G_n' G_n V_n = X_{n2}$.

Eq. (3.23) and Assumption 4 imply by Hölder's inequality

$$\left|\frac{1}{n^{1/q}}\sum_{i=1}^{n}\frac{c_{ni}}{c_n}v(\lambda_{ni})\right| \leqslant \frac{1}{n^{1/q}}\left(\sum_{i=1}^{n}|v(\lambda_{ni})|^p\right)^{1/p} n^{1/q} \leqslant c. \tag{3.62}$$

Hence, factorizing κ_{n2} as

$$\kappa_{n2} = \frac{1}{n^{1/2-1/q}}\left[\sqrt{n}(\sigma^2-\hat{\sigma}^2)\right]\left[\frac{1}{\kappa_{n0}}\right]\left[\frac{1}{n^{1/q}}\sum_{i=1}^{n}\frac{c_{ni}}{c_n}v(\lambda_{ni})\right]$$

we see that by (2.11), (3.61) and (3.62) the factors in all brackets are $O_p(1)$, so that $\kappa_{n2} = o_p(1)$. We have proved the second relation in (2.22).

Eq. (2.23) is a consequence of (3.61) and consistency of $\hat{\sigma}^2$.

Nonnegativity of κ_{n3} and ψ_n are obvious. □

4. Computer simulations

Following Lee [19] we consider the Case [5] framework with r districts and m farmers in each district. Denote $l_m = (1, \ldots, 1)'$ (m unities) and $B_m = (l_m l_m' - I_m)/(m-1)$. The Case spatial matrix equals $W_n = I_r \otimes B_m$. It is of order $n = rm$. With

$$q_{uv}^{(n)} = \left\{(s,t) : \frac{u-1}{n} < s < \frac{u}{n},\ \frac{v-1}{n} < t < \frac{v}{n}\right\}, \quad 1 \leqslant u, v \leqslant n,$$

let

$$Q = \bigcup_{u=1}^{r} q_{uu}^{(r)}, \quad K = r1_Q.$$

The purpose of the next lemma is to provide the ground for application of Theorem 1 and, in particular, to show an example of matrices which satisfy Assumption 2.

Lemma 7. (a) *W_n has r eigenvalues equal to* 1 *and* $(r-1)m$ *eigenvalues equal to* $1/(1-m)$.

(b) *For any fixed r, the sequence* $\{W_n : m = 1, 2, \ldots\}$ *is L_2-close to K and*

$$\|W_n - d_n K\|_2 = O\left(\frac{1}{\sqrt{n}}\right).$$

(c) *The sequence of matrices* $\widetilde{W}_n = I_r \otimes (l_m l_m')/(m-1)$, $m = 1, 2, \ldots$, *satisfies* (2.1).

Proof. (a) From $(l_m l_m')l_m = m l_m$ we see that $\lambda_1 = m$ is an eigenvalue and $e_1 = l_m$ is the corresponding eigenvector of the matrix $l_m l_m'$. Denote X_m the $(m-1)$-dimensional subspace of $\mathbb{R}^m$ of vectors orthogonal to e_1:

$$X_m = \{x \in \mathbb{R}^m : l_m' x = x_1 + \cdots + x_m = 0\}.$$

For any $x \in X_m$, $l_m l_m' x = 0$. Selecting in X_m a set $e_2, \ldots, e_m$ of pairwise orthogonal vectors we see that they are eigenvectors corresponding to eigenvalues $\lambda_2 = \cdots = \lambda_m = 0$. Since the system $e_1, \ldots, e_m$ is complete in $\mathbb{R}^m$, we have found all eigenvalues of $l_m l_m'$.

The eigenvalues of B_m then are $\lambda_1 = 1$ and $\lambda_2 = \cdots = \lambda_m = 1/(1-m)$. Since I_r has r eigenvalues equal to 1, the statement follows from the properties of Kronecker products [20, p. 464].

(b) Consider the terms in

$$\|W_n - d_n K\|_2^2 = \sum_{i,j=1}^{n} (w_{nij} - (d_n K)_{ij})^2.$$

Let

$$b_{uv} = \{(i,j) : (u-1)m + 1 \leqslant i \leqslant um, (v-1)m + 1 \leqslant j \leqslant vm\}, \quad 1 \leqslant u, v \leqslant r$$

be the batches of indices corresponding to blocks of W_n of size $m \times m$. The diagonal blocks are all B_m and the others are null matrices.

(b1) Let $(i,j) \in b_{uu}$. From $1 \leqslant i - (u-1)m \leqslant m,\ 1 \leqslant j - (u-1)m \leqslant m$ we see that

$$w_{nij} = \begin{cases} 1/(m-1) & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

On the other hand,

$$q_{ij}^{(n)} \subset q_{uu}^{(r)} \subset Q, \quad (d_n K)_{ij} = n \int_{q_{ij}^{(n)}} r\, dx\, dy = \frac{1}{m}.$$

(b2) Let $(i,j) \in b_{uv}$ with $u \neq v$. Then

$$w_{nij} = 0.$$

Since $q_{ij}^{(n)} \subset \left((0,1)^2 \backslash Q\right)$, we have

$$(d_n K)_{ij} = 0.$$

The equations we have derived imply

$$\sum_{i,j=1}^{n} (w_{nij} - (d_n K)_{ij})^2 = \sum_{i=1}^{n} \frac{1}{m^2} + \sum_{u=1}^{r} \sum_{(i,j)\in b_{uu},\ i\neq j} \left(\frac{1}{m-1} - \frac{1}{m}\right)^2$$
$$= \frac{r}{m} + \frac{1}{m^2(m-1)^2} \sum_{u=1}^{r} (m^2 - m) = O\left(\frac{1}{n}\right).$$

This proves the required bound.

After what we have done statement (c) is obvious. □

Lee studied the performance of the QML estimator for r ranging from 30 to 120 and m from 3 to 100. Our values for r, m are roughly the same. The simulations consist of two parts based on Theorems 1 and 3.

The first part of computer exercises is related to Theorem 1. In a finite-sample framework, there is no sequence of spatial matrices and one never knows the function K which approximates that sequence. Applying the interpolation operator to W_n, one can define K and consider it as the function, which approximates the given and all subsequent (unknown) spatial matrices. With this definition, W_n for the given sample becomes an exact image of K under discretization. As

Table 1
Simulations for Theorem 1

Values of m, r	True ρ	OLS estimator mean (and st. dev.)	"Eigenvalue" formula mean (and st. dev.)
(a) $m = 10$,	(a1) -0.105	-0.2445 (0.1648)	-0.2350 (0.1712)
$r = 100$	(a2) 0.1	0.1789 (0.1085)	0.1781 (0.1189)
	(a3) 0.95	0.9976 (0.0003)	0.9976 (0.0003)
(b) $m = 100$,	(b1) -0.095	-0.4972 (0.8410)	-0.4503 (0.7753)
$r = 10$	(b2) 0.1	0.0165 (0.5512)	0.0212 (0.5461)
	(b3) 0.95	0.9969 (0.0016)	0.9969 (0.0016)
(c) $m = 50$,	(c1) -0.015	-0.0707 (0.2175)	-0.0685 (0.2193)
$r = 50$	(c2) 0.1	0.1728 (0.1662)	0.1605 (0.1669)
	(c3) 0.95	1.0129 (0.1694)	1.0015 (0.0005)

Lemma 6 shows, nonzero eigenvalues of W_n and $\mathcal{K}$ will coincide. Simulation of the asymptotic result of Theorem 1 becomes, effectively, a comparison of simulation results for the finite-sample deviation from the true value (2.8) and its eigenvalue representation (1.4) in case of a symmetric W_n. In this sense simulation of Theorem 1 is trivial. However, it can be useful if evidence is sought against the null hypothesis of normal asymptotics.

We consider the OLS estimator and find its empirical distribution function with 1000 repetitions. As Lemma 7 shows, W_n has a large number of equal negative eigenvalues, denoted by $\lambda_{\min}$, and a small number of equal positive eigenvalues, denoted by $\lambda_{\max}$. Theorem 1 guarantees convergence of $\hat{\rho}$ for ρ in a small neighborhood of 0. The combinations of r and m considered are:

(a) $m = 10, r = 100$ ($|\rho| < 0.0047$);
(b) $m = 100, r = 10$ ($|\rho| < 0.0524$); and
(c) $m = 50, r = 50$ ($|\rho| < 0.01$)

(the intervals in parentheses are (2.9) for which convergence in distribution is affirmed in Theorem 1). For each of the cases (a)–(c) we take three different values of ρ: one in a small neighborhood of 0, another close to $\lambda_{\min}$ and the third one close to $\lambda_{\max}$. Thus, for Theorem 1 we do 9 simulations and for each of them:

(i) Test for normality the distributions of the OLS estimator and its "eigenvalue" counterpart.
(ii) Find sample means and standard deviations of the OLS estimator $\hat{\rho}$ and its expression in terms of eigenvalues.

Table 1 shows that in many cases bias is large and comparable in absolute value with the parameter being estimated. This should not come as a surprise because a ratio of quadratic forms in general does not have mean zero. The main calculations have been made in GAUSS and the empirical distributions have been fed to MINITAB to test for normality. In all cases the null that the distribution is normal is rejected (the p-value of Anderson–Darling statistic is less than 0.005 in all cases).

The second part of computer simulations is related to Theorem 3. The two-step procedure of Theorem 3 is computationally intensive. GAUSS' internal code for calculating integrals is unreliable and we had to use MathCad to find the coefficients c_n and c_{ni}. For moderate values of n (10 and 100) one has to take values from $a = 100$ to 1000 to approximate improper integrals over the half-line by integrals over $[0, a]$. For $n = 1000$ the function $1/\pi_n$ declines very quickly and it

Table 2
Simulations for Theorem 3

Values of m, r	True ρ	Two-step estimator mean (and st. dev.)
(a) $m = 10$,	(a1) -0.105	-0.0950 (0.0051)
$r = 100$	(a2) 0.1	0.1017 (0.0027)
	(a3) 0.95	0.5070 (3.2e−008)
(b) $m = 100$,	(b1) -0.095	-0.1884 (0.1918)
$r = 10$	(b2) 0.1	0.0888 (0.0586)
	(b3) 0.95	0.9467 (1.3e−006)
(c) $m = 40$,	(c1) -0.015	-0.0721 (0.0525)
$r = 40$	(c2) 0.1	0.1556 (0.0483)
	(c3) 0.95	0.9974 (3.7e−007)

Table 3
Comparison of percentage errors

Values of m, r	True ρ	OLS error (%)	"Eigenvalue" formula error (%)	Two-step estimator error (%)
(a) $m = 10$,	(a1) -0.105	132.86	123.81	9.52
$r = 100$	(a2) 0.1	78.90	78.10	1.70
	(a3) 0.95	5.01	5.01	46.63
(b) $m = 100$,	(b1) -0.095	423.37	374.00	98.32
$r = 10$	(b2) 0.1	83.50	78.80	11.20
	(b3) 0.95	4.94	4.94	0.35
(c) $m = 40$,	**(c1) $-$0.015**	371.33	356.67	380.67
$r = 40$	(c2) 0.1	72.80	60.50	55.60
	(c3) 0.95	6.62	5.42	4.99

is sufficient to take $a = 10$. With c_n and c_{ni} at hand we used again GAUSS to realize the two-step procedure. In cases (a) and (b) it took about half an hour on a computer with a processor speed 2.4 MHz to simulate 100 procedures and the total time for each of the six subcases was about 50 min. Therefore, we did not attempt to simulate 1000 times and in case (c) the combination $m = 50,\ r = 50$ has been replaced with $m = 40,\ r = 40$ (the interval (2.9) being $|\rho| < 0.0125$). The results are presented in Table 2.

Table 2 shows an improvement in estimation due to the two-step procedure. Not always the standard deviations are small relative to the parameter values and estimates.

As one can see from Table 3, for ρ close to zero the two-step procedure improves the OLS estimator in all cases. For ρ close to one of the eigenvalues of W_n the evidence is mixed: in two cases (shown in bold) the error has increased.

5. Conclusions

This paper develops the asymptotic theory of the OLS estimator, proves its inconsistency and provides the asymptotic distribution for the autoregressive spatial model. We caution about choosing the conditions to prove asymptotic results. If conditions contradict one another, the results can be formally correct but valid for a void set of objects. Under the restrictive conditions we consider it is shown that QML, and MM estimators, known to be consistent in the literature, are in fact inconsistent. A new two-step estimator based on the OLS estimator is proposed.

The conditions and method contained here allow us to take the ratio-of-quadratic-forms structure of the OLS estimator to the limit. The simulation exercises reject the null hypothesis that the asymptotic distribution is normal and show that the suggested two-step procedure improves the OLS estimator when the true parameter is sufficiently close to zero.

Our results also raise more questions and they will be the subject of a future study. Under our conditions, what happens outside the interval in which convergence in distribution has been established? Does the method work for a mixed spatial model? Our condition on the spatial matrices does not cover situations with uniformly limited interaction of economic agents. Will the existing results about normal asymptotics in such situations be sustained under more transparent conditions?

Acknowledgment

We thank Professors Ingmar Prucha and LungFei Lee for valuable comments on the first version of this manuscript. The remarks by two anonymous referees have been instrumental in improving the exposition and adding Monte Carlo simulations.

References

[1] T.W. Anderson, D.A. Darling, Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes, Ann. Math. Statist. 23 (1952) 193–212.
[2] L. Anselin, Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, Boston, 1988.
[3] L. Anselin, A.K. Bera, Spatial dependence in linear regression models with an introduction to spatial econometrics, in: A. Ullah, D.E.A. Giles (Eds.), Handbook of Applied Economics Statistics, Marcel Dekker, New York, 1998.
[4] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[5] A.C. Case, Spatial Patterns in Household Demand, Econometrica 59 (1991) 953–965.
[6] A. Cliff, J. Ord, Spatial Processes, Models and Applications, Pion, London, 1981.
[7] N.A.C. Cressie, Statistics of Spatial Data, Wiley, New York, 1993.
[8] I.C. Gohberg, M.G. Kreĭn, Introduction to the Theory of Linear Nonselfadjoint Operators, American Mathematical Society, Providence, RI, 1969.
[9] A. Hoque, The exact moments of forecast error in the general dynamic model, Sankhya Ser. B (Part 1) 47 (1985) 128–143.
[10] M.C. Jones, Expressions for inverse moments of positive quadratic forms in normal variables, Austral. J. Statist. 28 (1986) 242–250.
[11] H.H. Kelejian, I.R. Prucha, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, J. Real Estate Finance Econom. 17 (1998) 99–121.
[12] H.H. Kelejian, I.R. Prucha, A generalized moments estimator for the autoregressive parameter in a spatial model, Internat. Econom. Rev. 40 (1999) 509–533.
[13] H.H. Kelejian, I.R. Prucha, On the asymptotic distribution of the Moran I test statistic with applications, J. Econom. 104 (2001) 219–257.
[14] H.H. Kelejian, I.R. Prucha, 2SLS and OLS in a spatial autoregressive model with equal spatial weights, Regional Sci. Urban Econom. 32 (2002) 691–707.
[15] H.H. Kelejian, I.R. Prucha, Y. Yuzefovich, Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results, in: J. LeSage, R.K. Pace (Eds.), Advances in Econometrics, Elsevier, New York, 2004, pp. 163–198.
[16] L.F. Lee, Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models I: Spatial Autoregressive Processes, Manuscript, Department of Economics, The Ohio State University, Columbus, OH, 2001.
[17] L.F. Lee, Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models, Econom. Theory 18 (2002) 252–277.

[18] L.F. Lee, Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, Econom. Rev. 22 (2003) 307–335.
[19] L.F. Lee, Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, Econometrica 72 (2004) 1899–1925.
[20] H. Lütkepohl, Introduction to Multiple Time Series Analysis, Springer, New York, 1991.
[21] K.T. Mynbaev, L_p-approximable sequences of vectors and limit distribution of quadratic forms of random variables, Adv. Appl. Math. 26 (2001) 302–329.
[22] J.K. Ord, Estimation methods for models of spatial interaction, J. Amer. Statist. Assoc. 70 (1975) 120–126.
[23] K. Tanaka, Time Series Analysis: Nonstationary and Noninvertible Distribution Theory, Wiley, New York, 1996.
[24] D.E. Varberg, Convergence of quadratic forms in independent random variables, Ann. Math. Statist. 37 (1966) 567–576.
[25] H. White, Estimation, Inference and Specification Analysis, Cambridge University Press, New York, 1994.