A Remark on the Asymptotic Distribution of the OLS Estimator for a Purely Autoregressive Spatial Model

Mynbaev, Kairat and Ullah, Aman

Economics Department, University of California at Riverside

16 January 2006

A Remark on the Asymptotic Distribution of the OLS Estimator for a Purely Autoregressive Spatial Model

January 16, 2006

Kairat Mynbaev[1], Aman Ullah[2]

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-square 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 methods of moments fail. A remedial iterative procedure using the OLS estimator is proposed.

Keywords: spatial model, OLS estimator, asymptotic distribution, maximum likelihood, method of moments

JEL codes: C13, C21

1 Introduction and Main Statements

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 is the error vector with zero mean. Anselin (1988) classifies (1.1) as a first-order spatial autoregressive model. 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 $V_n = \mu M_n V_n + U_n$.

Earlier development in testing and estimation of spatial autoregressive models has been summarized in Anselin (1988), Cressie (1993) and Anselin and Bera (1998), among others. Some of the recent references are Kelejian and Prucha (1999, 2002), Kelejian et al. (2004) and

[1]Economics Department, KIMEP, Almaty, Kazakhstan

[2]Economics Department, University of California at Riverside, Riverside, California, USA

Lee (2001, 2002, 2003, 2004). It is widely recognized in the spatial econometrics literature that the OLS estimator $\hat{\rho}$ of ρ is in general inconsistent, see, for instance, Section 6.1.1 in Anselin (1988). Lee (2002) considers (1.2) and shows that, depending on the assumptions, the OLS estimator can be consistent or can have an asymptotic bias. In the same paper for model (1.1) he provides two examples of inconsistency. Kelejian and Prucha (1999) and Lee (2001) have obtained asymptotic distributions of the ML estimator and the generalized method of moments estimator.

The research in the above references has been moving towards relaxing the assumptions and increasing generality. Along the way the conditions imposed and the results obtained have become more complex. In this paper we move in the opposite direction: we look for simpler conditions and results amenable to easy interpretation, perhaps by narrowing the area of applicability. The highlights of the outcome are as follows. The asymptotic distribution of the OLS estimator is a ratio of infinite linear combinations of χ-square variables; none of the existing sources captures this effect. No normalization is necessary to achieve convergence in distribution. The denominator of the OLS estimator converges to a non-degenerate random variable.

We assume i.i.d. errors and use only low-level conditions. The class of matrices corresponds to the case when a particular economic unit is influenced by many others, so that the interaction between units is stronger than in other researches.[3] Our formula may not be applicable to another practically interesting situation when an economic agent is influenced by a limited number of others, even as $n \to \infty$. Because of the convergence in probability of the denominator and numerator of the formula, it can be used for approximate calculations by truncating the sums.

Intuitively, our result is explained as follows. Assuming that W_n is symmetric with eigenvalues $\lambda_{n1}, ..., \lambda_{nn}$ and V_n is distributed as $N(0, \sigma^2 I)$, it is easy to calculate the finite-sample deviation from the true value

$$\hat{\rho} - \rho = \frac{\sum_{i=1}^n v_i^2 \frac{\lambda_{ni}}{1-\rho\lambda_{ni}}}{\sum_{i=1}^n v_i^2 \left(\frac{\lambda_{ni}}{1-\rho\lambda_{ni}}\right)^2}.$$

Whether this ratio-of-quadratic-forms structure will be preserved in the limit, depends on the assumptions. If the weights of v_i's tend to zero, normality may appear in the limit. Under our conditions they do not.

Since there is asymptotic bias, we attempt to find alternative estimators. This problem has been tackled by Kelejian and Prucha (1999) and Lee (2001), by using the maximum likelihood method and method of moments. They have provided conditions for asymptotic normality and unbiasedness of these estimators. In particular, Lee (2001) imposes the requirement that a certain limit exist and be different from zero, which assures validity of the ML and MM procedures. However, we prove that under our conditions that limit exists and is zero[4]. Therefore under our assumptions the corresponding identification condition fails and maximum likelihood or method of moments cannot be used. For these reasons we look at the OLS estimator more closely and devise an iterative procedure that can be used for finite samples. Again, the problem turns out to be more difficult than usual. It is not possible to

[3]We thank professors Ingmar Prucha and Lung-fei Lee for this and other comments.

[4]In his conditions, Lee introduces a special parameter h_n designed to accommodate different behaviors at infinity. Under our conditions, the only meaningful choice is $h_n = 1$.

prove convergence of the sequence obtained. We indicate how its oscillating character can be exploited to obtain a convergent sequence.

The methodology we develop is interesting in its own right. In particular, the way the spatial matrices are modeled can be used to study more general models of type (1.2). It is based on the idea of approximating discrete objects (sequences of vectors or matrices) with functions of a continuous argument. Such an approximation allows one to widely use the tools of the theory of functions. We rely on the rendition of this general idea contained in Mynbaev (2001). The class of errors can be widened to include linear processes with short-range dependence, at the expense of significantly lengthening the proof.

One conclusion that can be drawn from our exercise is that when dealing with (1.1) or its generalizations, one has to work with an infinite series of type $\sum_{k=0}^{\infty}\rho^k W_n^{k+1}$ in order to avoid high-level conditions. The existing papers on spatial models do not treat such series. To clarify, consider a simple autoregression $y_t = c_1 + c_2 y_{t-1} + e_t$. In this model, one cannot assume that dependence of y_t on y_{t-1} is essential, while all previous values of y are $o_p(1)$. Dependence on all previous values is a distinct feature of autoregressive models.

We start with describing our assumptions and then state the main results. The proofs are given in Sections 2 and 3.

Assumption 1 (on the error term). For each n, one has $V_n = (v_1, ..., v_n)'$ where $v_1, ..., v_n$ are independent, identically distributed variables with mean zero, variance σ^2 and finite moments up to $\mu_4 = Ev_i^4$.

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 as follows. For an integrable function K, d_nK is an $n \times n$ matrix with elements

$$(d_nK)_{ij} = n\int_{q_{ij}} K(x,y)dxdy, \quad i,j = 1, ..., n,$$

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 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 $\{W_n : n = 1, 2, ...\}$ 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

$$\|W_n - d_nK\|_2 = o\left(\frac{1}{\sqrt{n}}\right). \tag{1.3}$$

It is evident that such classes of matrices exist. For example, one can take any function K and put $W_n = d_nK$, in which case the left side of (1.3) is identically zero. In Section 2 we show that Assumption 2 implies

$$\max_{i,j}|w_{nij}| \longrightarrow 0, \quad \sum_{i,j}|w_{nij}| \longrightarrow \infty,\ n \to \infty. \tag{1.4}$$

The first relation means that economic 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 between the units. Some properties often imposed on W_n in practice can be readily visualized in terms of K . For example, in those applications which treat (1.1) as an equilibrium model, it is customary to require W_n to have zeros on the main diagonal. This corresponds to K vanishing in the neighborhood of the 45-degree line. We would like to stress, however, 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.

Assumption 3 (on the function K). The function K is symmetric and the eigenvalues $\lambda_i,\ i = 1, 2, ...,$ of the integral operator

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

are summable: $\sum_{i\ge1} |\lambda_i| < \infty$.

$\mathcal{K}$ is considered an operator in the space $L_2(0, 1)$ of square-integrable on $(0, 1)$ functions. 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: $\sum_{i\ge1} \lambda_i^2 < \infty$. The summability condition we require is stronger because

$$\left(\sum_{i\ge1} \lambda_i^2\right)^{1/2} \le \sum_{i\ge1} |\lambda_i|. \tag{1.5}$$

Necessary and sufficient conditions for summability of lambdas can be found in Gohberg and Kreĭn (1969).

The decomposition

$$K(x, y) = \sum_{i\ge1} \lambda_i f_i(x)f_i(y) \tag{1.6}$$

and the identity

$$\sum_{i\ge1} \lambda_i^2 = \int_0^1 \int_0^1 K^2(x, y)dxdy \tag{1.7}$$

are important to understand both the result and the proof.

Denoting $Z_n = W_nY_n$ the regressor in (1.1), we have the following expression for the OLS estimator $\hat\rho$ of ρ:

$$\hat\rho = (Z_n'Z_n)^{-1}Z_n'Y_n. \tag{1.8}$$

Put $S_n = I_n - \rho W_n$ and $G_n = W_nS_n^{-1}$ when S_n^{-1} exists.

Theorem 1. Suppose Assumptions 1, 2 and 3 hold.

1) If

$$|\rho| < 1/\left(\sum_{i\ge1} \lambda_i^2\right)^{1/2}, \tag{1.9}$$

then the matrices S_n^{-1} exist for all sufficiently large n and have uniformly bounded $\|\cdot\|_2$-norms and the bias equals

$$\hat\rho - \rho = \frac{V_n'G_n'V_n}{V_n'G_n'G_nV_n}. \tag{1.10}$$

2) If

$$|\rho| < 1/\sum_{i\ge1} |\lambda_i|, \tag{1.11}$$

then

$$\hat{\rho} - \rho \xrightarrow{d} \frac{\sum_{i\geq 1} u_i^2 \nu(\lambda_i)}{\sum_{i\geq 1} u_i^2 \nu^2(\lambda_i)} \tag{1.12}$$

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

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

3) (1.11) implies convergence

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4) \tag{1.13}$$

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.

Remarks. The peculiarity of the fraction in (1.10) is that both the numerator and denominator are non-trivial distributions (both series converging in L_1 and, consequently, in probability), unlike many other econometric problems where the numerator is non-trivial and the denominator is a constant. If the numerator in (1.10) or (1.12) has mean zero, it does not necessarily mean that the whole fraction has mean zero (see Lemma 5 in Section 2 regarding (1.10)). Of course, infinite summation adds a lot of complexity. The characteristic function of an infinite weighted sum of χ-square variables has been found by Anderson and Darling (1952) (see also Varberg (1966)). Because of convergence in L_1 of the numerator and denominator in (1.12) the whole fraction converges in probability. Therefore by truncating the sums an expression for approximate calculation of the fraction can be obtained.

The next issue is to find a better estimator which would converge in some sense to the true parameter. Lee (2001) has proved that under some conditions the Quasi-Maximum Likelihood Estimator (QMLE) for (1.1) is consistent and asymptotically normally distributed. One of key elements in his proof consists in applying White's (1994) identification uniqueness condition. Lee has developed conditions sufficient for local and global identification. Those conditions involve positive numbers h_n which in our case should be chosen to be identically 1 (for the statement of his Theorem 2 to be true). Then the local identification condition takes the form

$$\text{the limit of the sequence } \frac{1}{n}[\mathrm{tr}(G_n' G_n) + \mathrm{tr}(G_n^2) - \frac{2}{n}\mathrm{tr}^2(G_n)] \text{ exists and is positive} \tag{1.14}$$

and the global one looks like this:

$$\begin{gathered}\text{for any } \rho \text{ different from the true value } \rho_0 \text{ the limit} \\ \lim_{n\to\infty} \frac{1}{n}\left(\ln|\sigma_0^2 S_n^{-1} S_n'^{-1}| - \ln|\sigma_n^2(\rho) S_n^{-1}(\rho) S_n'^{-1}(\rho)|\right) \text{ exists and is not zero}\end{gathered} \tag{1.15}$$

where

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

Another avenue to think about is the method of moments considered by Kelejian and Prucha (1999) for the problem under consideration. Lee (2001) has simplified their approach and worked out an identification condition stated in terms of a 2×2 matrix A_n with elements

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

Both versions of the method of moments considered by Lee (2001) require the limit

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}A_n \tag{1.16}$$

to exist and be nonsingular.

Theorem 2. Under the assumptions of Theorem 1 the limits in (1.14), (1.15) and (1.16) are zero.

Without the consistency of QMLE, the derivation of the asymptotic distribution based on the formula

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

does not work. Here $\ln L_n(\rho)$ is the log likelihood function (see (1.18) below) and $\widetilde{\rho}$ lies between $\hat{\rho}_{QMLE}$ and ρ.

The problems we have just described force us to analyse the OLS estimator more closely. The solution we have found has its limitations and advantages. Firstly, we have not been able to prove that the procedure described below gives a consistent estimator. Rather, it is essentially a finite-sample instrument which answers the question: if only one sample is available and the OLS estimator has been obtained, then how that estimator can be improved? Secondly, unlike the ML estimator, it is an iterative procedure with a well-defined initial point: the OLS estimate. Thirdly, its format parallels the asymptotic result in Theorem 1. Namely, from the point of view of (1.12), instead of requiring $\operatorname{plim}\hat{\rho} = \rho$ (consistency) it would be correct to require

$$\operatorname{plim}\hat{\rho} = \rho + \kappa \text{ where } E\kappa = 0 \tag{1.17}$$

and in Theorem 3 we try to satisfy this condition. Finally, the ensuing discussion will be less rigorous than the preceding one in that we shall impose one high-level condition on matrices W_n. It is possible to obtain such a condition as a consequence of a low-level one but we refrain from doing that to maintain transparency. Plausibility of the new assumption will be seen from Lemma 6 of Section 2. Besides, the error term will be assumed to be normal. This will enable us to use exact results about ratios of quadratic forms of normal variables.

The expression for the ML estimator will help the reader to understand the idea behind our construction. The ML estimator has been derived in a more general situation by Ord (1975), among others. In our case the log likelihood function is

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\rho)| - \frac{1}{2\sigma^2}(Y_n - \rho W_nY_n)'(Y_n - \rho W_nY_n) \tag{1.18}$$

where $\theta = (\rho, \sigma^2)$. Using (3.35) we get

$$\begin{aligned}
\frac{\partial \ln L_n(\theta)}{\partial \rho} &= -\mathrm{tr}(W_n S_n^{-1}(\rho)) - \frac{1}{2\sigma^2}(-Y_n' W_n Y_n - Y_n' W_n' Y_n + 2\rho (W_n Y_n)' W_n Y_n) \\
&= -\mathrm{tr}(W_n S_n^{-1}(\rho)) + \frac{1}{\sigma^2}(Y_n' W_n Y_n - \rho (W_n Y_n)' W_n Y_n), \\
\frac{\partial \ln L_n(\theta)}{\partial \sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}(Y_n - \rho W_n Y_n)'(Y_n - \rho W_n Y_n).
\end{aligned}$$

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

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

Of course, these estimators are not feasible as they contain an unknown ρ.

Since the OLS estimator and the formula we suggest below do not change if W_n is replaced by its symmetric derivative $(W_n + W_n')/2$, we can assume without loss of generality that all of W_n are symmetric. Then each W_n can be represented as

$$W_n = P_n \mathrm{diag}[\lambda_{n1}, ..., \lambda_{nn}] P_n' \tag{1.19}$$

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

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

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

These integrals converge if $n > 2$. Let

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

Iterative procedure. Estimate ρ and σ^2 by OLS and put

$$\rho_0 = \hat{\rho}, \; \rho_j = \frac{Y_n' W_n Y_n - \hat{\sigma}^2 \mathrm{tr}(A_n W_n S_n^{-1}(\rho_{j-1}))}{(W_n Y_n)' W_n Y_n}, \; j = 1, 2, ...$$

For analytical purposes we rewrite the recurrent formula as

$$\rho_j = \frac{V_n' S_n'^{-1} G_n V_n - \hat{\sigma}^2 \mathrm{tr}(A_n W_n S_n^{-1}(\rho_{j-1}))}{V_n' G_n' G_n V_n}, \; j = 1, 2, ... \tag{1.20}$$

Instead of Assumption 1 we make a stronger

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

The high-level condition we talked about above is

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

It can be shown that (1.3) implies

$$\lim_{n\to\infty} \sum_{i=1}^{n} \lambda_{ni}^2 = \sum_{i=1}^{\infty} \lambda_i^2 \tag{1.21}$$

(see Lemma 6 in Section 2) and that a condition stronger than (1.3) can be imposed on the sequence $\{W_n\}$ to make sure that Assumption 4 is satisfied. See Gohberg and Kreĭn (1969), Chapter III, for more information.

Theorem 3. Suppose Assumptions $1'$, 2, 3 and 4 hold. If the true ρ satisfies (1.11), then there exist random variables $\kappa_{n1},\ \kappa_{n2},\ \kappa_{n3}$ and a deterministic function ψ_n such that

$$\rho_j = \rho + \kappa_{n1} + \kappa_{n2} + \kappa_{n3} \int_{\rho_{j-1}}^{\rho} \psi_n(t)dt, \tag{1.22}$$

$$E\kappa_{n1} = 0 \text{ for all } n,\ \operatorname{plim} \kappa_{n2} = 0, \tag{1.23}$$

$$\operatorname{plim} \kappa_{n3} = \frac{1}{\sum_{i\geq 1} u_i^2 \nu^2(\lambda_i)}, \tag{1.24}$$

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

Property (1.23) is in line with (1.17).

Denote $h_j = \rho_j - \rho_{j-1}$ the step from ρ_{j-1} to ρ_j . Then (1.22) implies

$$h_j = \kappa_{n3} \int_{\rho_{j-1}}^{\rho_{j-2}} \psi_n(t)dt. \tag{1.25}$$

The main point about this formula is that successive steps are made in opposite directions. For example, suppose that the step h_1 is positive ($\rho_1 > \rho_0 = \hat{\rho}$). Because of positivity of κ_{n3} and ψ_n (1.25) gives $h_2 < 0$ ($\rho_2 < \rho_1$). We have not been able to prove convergence of the sequence $\{\rho_j\}$. Even convergence of the steps to zero is not guaranteed. By the mean value theorem $|h_j| = \kappa_{n3}\psi_n(\widetilde{\rho})|h_{j-1}|$ where $\widetilde{\rho}$ lies between ρ_{j-2} and ρ_{j-1}. Since in general $\kappa_{n3}\psi_n(\widetilde{\rho})$ is not less than 1, the steps may not decline. Our suggestion is to apply one of methods of summation of divergent series to the sequence of steps. Such methods have a good feature that if the original series actually converges, its generalized limit ascribed by a summation method gives the same value. The simple average in $\hat{\rho} + 1/n\sum_{j=1}^n h_j$ is a particular case of Cèsaro methods (see Hardy (1949) for details).

2 Auxiliary Statements

Depending on the context, $\|\cdot\|_2$ may mean any of the norms

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

Here the set of indices I can be finite or infinite. $(\cdot,\cdot)_{l_2}$ denotes the scalar product associated with the norm $\|\cdot\|_2$.

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

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

$\|\cdot\|_1$ and $\|\cdot\|_2$ will be particularly useful. A limit in distribution is denoted $\xrightarrow{d}$ or dlim. Likewise, symbols $\xrightarrow{p}$ or plim are used interchangeably for limits in probability.

$c,\ c_1,\ c_2, ...$ 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

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

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 all integer $k \geq 0$

$$\left\|A^{k+1} - B^{k+1}\right\|_2 \leq \|A - B\|_2 \, (k+1) \left(\max\left\{\|A\|_2, \|B\|_2\right\}\right)^k. \tag{2.1}$$

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

$$\|s(A) - s(B)\|_2 \leq \varphi(\rho, A, B) \, \|A - B\|_2 \tag{2.2}$$

where

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

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

$$N(AB) \leq c \, \|A\|_2 \, \|B\|_2. \tag{2.3}$$

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

$$N(A) \leq c\sqrt{n} \, \|A\|_2. \tag{2.4}$$

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

$$N\left(A^{k+1} - B^{k+1}\right) \leq c \, \|A - B\|_2 \, (k+1) \left(\max\left\{\|A\|_2, \|B\|_2\right\}\right)^k. \tag{2.5}$$

Proof. a) follows from the well-known fact that if $\|A\| < 1$ and the norm $\|\cdot\|$ is submultiplicative ($\|AB\| \leq \|A\| \, \|B\|$), then the series $\sum_{k \geq 0} A^k$ 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 $k = 0$, (2.1) is trivial. If $k > 0$, then the identity

$$A^{k+1} - B^{k+1} = A^k (A - B) + A^{k-1}(A - B)B + ... + (A - B)B^k \tag{2.6}$$

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

$$\begin{aligned} \left\|A^{k+1} - B^{k+1}\right\|_2 &\leq \|A\|_2^k \, \|A - B\|_2 + ... + \|A - B\|_2 \, \|B\|_2^k \\ &\leq \|A - B\|_2 \, (k+1) \left(\max\left\{\|A\|_2, \|B\|_2\right\}\right)^k. \end{aligned} \tag{2.7}$$

c) (2.2) follows from (2.1):

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

d) For any square matrix A of order n Lee's (2004) Lemma A.11[5] yields

$$N(A) \le c\left(\|A\|_2 + |\mathrm{tr} A|\right). \tag{2.8}$$

Since $\|AB\|_2 \le \|A\|_2 \|B\|_2$ and $|\mathrm{tr}(AB)| \le \|A\|_2 \|B\|_2$, (2.8) gives (2.3).

e) Because of the growing factor $\sqrt{n}$ in (2.4), it is not a good idea to estimate the left side of (2.5) using (2.1). Instead, we apply identity (2.6) directly (this is why the assumption $k > 0$ is important). By (2.3) and Minkowski's inequality

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

The rest is the same as in (2.7).

We 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 1, possesses the property

$$\|d_n K\|_2 \le \|K\|_2 \text{ for all } K \text{ and } n \tag{2.9}$$

(just apply Hölder's inequality to prove). 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:

$$\|D_n A\|_2 = \|A\|_2 . \tag{2.10}$$

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

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

Its main property is that it approximates the identity operator:

$$\lim_{n\to\infty} \|P_n K - K\|_2 = 0 \text{ for any } K \in L_2\left((0,1)^2\right). \tag{2.11}$$

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

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

and

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

They possess properties similar to (2.9), (2.10) and (2.11).

[5] See his supplement available at http://economics.sbs.ohio-state.edu/lee/

Lemma 2. a) Assumptions 2 and 3 imply

$$\lim_{n\to\infty}\|W_n\|_2=\lim_{n\to\infty}\|d_nK\|_2=\|K\|_2 \tag{2.12}$$

and

$$\|W_n'-d_nK\|_2=o\left(\frac{1}{\sqrt{n}}\right). \tag{2.13}$$

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

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

and

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

Then for all i

$$\lim_{n\to\infty}\mu_{ni}=\mu_{\infty i}. \tag{2.14}$$

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

d) If $\|W_n-d_nK\|_2\to 0$, then (1.4) is true. A similar property holds in the one-dimensional case.

Proof. a) Continuity of norms and (2.11) yield $\|P_nK\|_2\to\|K\|_2$. $\|d_nK\|_2\to\|K\|_2$ follows because by (2.10) $\|d_nK\|_2=\|D_nd_nK\|_2=\|P_nK\|_2$. To prove the other equation in (2.12) note that by (2.10), (1.3) and (2.11)

$$\|D_nW_n-K\|_2\leq\|D_nW_n-P_nK\|_2+\|P_nK-K\|_2=\|W_n-d_nK\|_2+\|P_nK-K\|_2\to 0.$$

Therefore $\|W_n\|_2=\|D_nW_n\|_2\to\|K\|_2$.

To prove (2.13), observe that $(x,y)\in q_{ij}$ if and only if $(y,x)\in q_{ji}$ and, therefore, for a symmetric K, d_nK is also symmetric. Thus,

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

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

$$(D_nx,D_ny)_{l_2}=(x,y)_{l_2},\ \ x,y\in\mathbb{R}^n.$$

Using this fact, continuity of scalar products, and (2.11) we see that

$$(d_nf_i,d_nf_j)_{l_2}=(P_nf_i,P_nf_j)_{l_2}\longrightarrow(f_i,f_j)_{l_2}=\begin{cases}1, & i=j,\\ 0, & i\neq j.\end{cases} \tag{2.15}$$

Turning to (2.14), if $k>0$ and among $i_1,...,i_{k+1}$ there are at least two different indices, then at least two adjacent ones must be unequal. Hence, (2.14) is a direct consequence of (2.15).

c) obtains by calculation.

d) First note that

$$\max_{i,j}|w_{nij}| \le \|W_n - d_nK\|_2 + \max_{i,j}|(d_nK)_{ij}|$$

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

$$|(d_nK)_{ij}| = n\left|\int_{q_{ij}} K(x,y)dxdy\right| \le \left(\int_{q_{ij}} K^2(x,y)dxdy\right)^{1/2} \longrightarrow 0,\ n \to \infty$$

uniformly in i,j. This proves the first of the limit relations in (1.4). By (2.12) for some $c>0$ we have $c \le \|W_n\|_2^2 \le \|W_n\|_\infty \|W_n\|_1$ which implies $\|W_n\|_1 \ge c/\|W_n\|_\infty \to \infty$.

For natural n, L consider the random vector

$$U_{nL} = \begin{pmatrix} \sum_{s=1}^n (d_nf_1)_s v_s \\ \dots \\ \sum_{s=1}^n (d_nf_L)_s v_s \end{pmatrix} = \begin{pmatrix} V_n' d_n f_1 \\ \dots \\ V_n' d_n f_L \end{pmatrix}.$$

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

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

where u_i are independent standard normal.

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

a) For any fixed L

$$\operatorname*{dlim}_{n\to\infty} \delta_{nL} = \Delta_L, \tag{2.16}$$

$$\lim_{n\to\infty} E\delta_{nL} = E\Delta_L = \sigma^2 \sum_{i=1}^L \nu(\lambda_i) \begin{pmatrix} 1 \\ \nu(\lambda_i) \end{pmatrix}, \tag{2.17}$$

$$\lim_{n\to\infty} \operatorname{var}(\delta_{nL}) = \operatorname{var}(\Delta_L) = 2\sigma^4 \sum_{i=1}^L \nu^2(\lambda_i) \begin{pmatrix} 1 & \nu(\lambda_i) \\ \nu(\lambda_i) & \nu^2(\lambda_i) \end{pmatrix}. \tag{2.18}$$

b) If

$$\sum_{i\ge 1} |\nu(\lambda_i)| < \infty, \tag{2.19}$$

then

$$\Delta_L \xrightarrow{L_1(\Omega)} \Delta_\infty \text{ as } L \to \infty \tag{2.20}$$

and

$$\lim_{L\to\infty} \operatorname{var}(\Delta_L) = \operatorname{var}(\Delta_\infty) = 2\sigma^4 \sum_{i=1}^\infty \nu^2(\lambda_i) \begin{pmatrix} 1 & \nu(\lambda_i) \\ \nu(\lambda_i) & \nu^2(\lambda_i) \end{pmatrix}. \tag{2.21}$$

Proof. a) The central limit theorem from Mynbaev (2001) states that under the conditions of the lemma for any L

$$U_{nL} \xrightarrow{d} N(0, \sigma^2 I_L), \quad \operatorname{var}(U_{nL}) \longrightarrow \sigma^2 I_L \text{ as } n \to \infty. \tag{2.22}$$

The vector δ_{nL} is a continuous function of U_{nL}. Since $U_{nLi}^2 \xrightarrow{d} \sigma^2 u_i^2$, $n \to \infty$, (2.16) is true. The second relation in (2.22) implies (2.17):

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

To prove (2.18), we start with

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

Here

$$\begin{aligned} EU_{nLi}^2 U_{nLj}^2 &= E\left(\sum_{s=1}^{n} (d_n f_i)_s v_s \right)^2 \left(\sum_{p=1}^{n} (d_n f_j)_p v_p \right)^2 \\ &= \sum_{s,t,p,q=1}^{n} (d_n f_i)_s (d_n f_i)_t (d_n f_j)_p (d_n f_j)_q E v_s v_t v_p v_q. \end{aligned}$$

From Assumption 1 it follows that

$$E v_s v_t v_p v_q = \begin{cases} \sigma^4, & \text{if } (s = t) \neq (p = q) \text{ or } (s = p) \neq (t = q) \text{ or } (s = q) \neq (t = p), \\ \mu_4, & \text{if } s = t = p = q, \\ 0, & \text{in all other cases.} \end{cases}$$

Hence,

$$\begin{aligned} EU_{nLi}^2 U_{nLj}^2 &= \sigma^4 \left[\sum_{s=1}^{n} (d_n f_i)_s^2 \sum_{p=1}^{n} (d_n f_j)_p^2 + 2 \sum_{s=1}^{n} (d_n f_i)_s (d_n f_j)_s \sum_{p=1}^{n} (d_n f_i)_p (d_n f_j)_p \right] \\ &\quad + \mu_4 \sum_{s=1}^{n} (d_n f_i)_s^2 (d_n f_j)_s^2 \\ &= \sigma^4 \left[\|d_n f_i\|_2^2 \, \|d_n f_j\|_2^2 + 2 (d_n f_i, d_n f_j)_{l_2}^2 \right] + \mu_4 \sum_{s=1}^{n} (d_n f_i)_s^2 (d_n f_j)_s^2. \end{aligned}$$

By Lemma 2d) and (2.15)

$$\|d_n f_i\|_2 \longrightarrow 1, \ (d_n f_i, d_n f_j)_{l_2} \longrightarrow \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad \max_s |(d_n f_i)_s| \longrightarrow 0,$$

so that

$$\sum_{s=1}^{n} (d_n f_i)_s^2 (d_n f_j)_s^2 \leq \max_s (d_n f_i)_s^2 \, \|d_n f_j\|_2^2 \longrightarrow 0$$

and

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

These equations together with the formula for $\mathrm{var}(\delta_{nL})$ above prove that the left and right members of (2.18) are equal.

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

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

b) Inequality (1.5) applied to $\{\nu(\lambda_i)\}$ and condition (2.19) show that both components of Δ_L converge to those of Δ_∞ in $L_1(\Omega)$. (2.21) is proved similarly to (2.18).

Lemma 4. Suppose that for each L, $\delta_{nL} \xrightarrow{d} \Delta_L$ as $n \to \infty$ and that $\Delta_L \xrightarrow{d} \Delta_\infty$ 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 $X_n \xrightarrow{d} \Delta_\infty$ as $n \to \infty$.

This is just Theorem 4.2 from Billingsley (1968) with the notation adapted to ours.

Lemma 5. One has

$$0 < c_{ni} \le c_n < \infty,\ i = 1, ..., n, \tag{2.23}$$

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

$$E\left(\frac{\sum_{i=1}^n (c_n u_i^2 - \sigma^2 c_{ni})\nu(\lambda_{ni})}{\sum_{i=1}^n u_i^2\nu^2(\lambda_{ni})}\right) = 0. \tag{2.24}$$

Proof. (2.23) is obvious ($c_n < \infty$ because $n > 2$). Hoque (1985) has proved that if S and B are symmetric matrices, B is positive definite and $u \sim N(0, \Omega)$, then

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

In our case

$$S = \mathrm{diag}[\nu(\lambda_{n1}), ..., \nu(\lambda_{nn})],\ B = \mathrm{diag}[\nu^2(\lambda_{n1}), ..., \nu^2(\lambda_{nn})],$$

$$\Omega = \sigma^2 I,\ I + 2t\Omega B = \mathrm{diag}[1 + 2t\sigma^2\nu^2(\lambda_{n1}), ..., 1 + 2t\sigma^2\nu^2(\lambda_{nn})],$$

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

Then

$$E\left(\frac{\sum_{i=1}^n u_i^2\nu(\lambda_{ni})}{\sum_{i=1}^n u_i^2\nu^2(\lambda_{ni})}\right) = \int_0^\infty \sum_{i=1}^n \frac{\sigma^2\nu(\lambda_{ni})}{1+2t\sigma^2\nu^2(\lambda_{ni})}\frac{dt}{\pi_n(\sigma^2 t)} = \sum_{i=1}^n c_{ni}\nu(\lambda_{ni}). \tag{2.25}$$

On the other hand, formula (10) from Jones (1986) yields

$$E\left(\frac{\sigma^2}{\sum_{i=1}^n u_i^2\nu^2(\lambda_{ni})}\right)=\int_0^\infty \frac{dt}{\pi_n(t)}=c_n. \tag{2.26}$$

Combining (2.25) and (2.26) we get

$$E\left(\frac{\sum_{i=1}^n(c_nu_i^2-\sigma^2c_{ni})\nu(\lambda_{ni})}{\sum_{i=1}^n u_i^2\nu^2(\lambda_{ni})}\right)=c_nE\left(\frac{\sum_{i=1}^n u_i^2\nu(\lambda_{ni})}{\sum_{i=1}^n u_i^2\nu^2(\lambda_{ni})}\right)$$

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

Lemma 6. (1.3) implies (1.21).

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 $z\in R^n$ put

$$D_n^2W_n=n\sum_{i,j=1}^n w_{nij}1_{q_{ij}},\ D_n^1z=\sqrt{n}\sum_{i=1}^n z_i1_{q_i}.$$

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

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

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

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

so that

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

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 $x\in q_i$. $\mathcal{W}_nf=\lambda f$ implies

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

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

The statement we have just proved is sufficient for our purposes because the sums in (1.21) are not affected by zero eigenvalues. In the second part of the proof we need some facts from Gohberg and Kreĭn (1969). 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 $||A||_{\sigma_p} = \left(\sum_{i=1}^{\infty} s_j^p(\mathcal{K})\right)^{1/p}$, $1 \le p < \infty$, is a norm (p. 92).

These facts and (1.3) give

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

3 Proofs of Main Results

Proof of Theorem 1

1) Due to identity (1.7), condition (1.9) is equivalent to

$$|\rho| \|K\|_2 < 1. \tag{3.1}$$

Hence, $|\rho| \|K\|_2 \le 1 - 2\varepsilon$ for some sufficiently small $\varepsilon > 0$ and then (2.12) shows that there exists $n_0 = n_0(\varepsilon)$ such that

$$\sup_{n \ge n_0} |\rho| \|W_n\|_2 \le 1 - \varepsilon. \tag{3.2}$$

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

$$\|G_n\|_2 \le \sum_{k \ge 0} |\rho|^k \, \|W_n\|_2^{k+1} = \frac{\|W_n\|_2}{1 - |\rho| \, \|W_n\|_2} \le c \text{ for all } n \ge n_0. \tag{3.3}$$

The reduced form $Y_n = S_n^{-1} V_n$ of the basic model (1.1) and (1.8) lead to (1.10) in the usual way:

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

2) Here is the plan of the proof. The numerator and denominator of (1.10) 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 (1.6):

$$K_L(x, y) = \sum_{i=1}^{L} \lambda_i f_i(x) f_i(y). \tag{3.4}$$

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

The scheme we have just explained is realized through the representation

$$X_n = \alpha_n + \beta_{nL} + \gamma_{nL} + \delta_{nL} \tag{3.5}$$

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.

Bounding α_n. We evaluate coordinates of the alphas, betas, and gammas separately. Using (2.4) for $k = 0$ and (2.5) for positive k, we have

$$\begin{aligned} \|\alpha_{n1}\|_2 &= N\left(\sum_{k\geq 0} \rho^k \big((W_n')^{k+1} - (d_n K)^{k+1}\big)\right) \leq N\left(W_n' - d_n K\right) \\ &+ \sum_{k>0} |\rho|^k N\left((W_n')^{k+1} - (d_n K)^{k+1}\right) \leq c\sqrt{n}\left\|W_n' - d_n K\right\|_2 \\ &+ c\left\|W_n' - d_n K\right\|_2 \sum_{k>0} (k+1)\left(|\rho| \max\left\{\|W_n'\|_2, \|d_n K\|_2\right\}\right)^k. \end{aligned} \tag{3.6}$$

Because of (1.5), assumption (1.11) implies (1.9) and, consequently, (3.1). Hence, in the way we derived (3.2) we can now derive

$$\sup_{n\geq n_0} |\rho| \max\left\{\|W_n'\|_2, \|d_n K\|_2\right\} \leq 1 - \varepsilon. \tag{3.7}$$

This allows us to continue (3.6) using (2.13)

$$\|\alpha_{n1}\|_2 \leq c\sqrt{n}\left\|W_n' - d_n K\right\|_2 = o(1). \tag{3.8}$$

Repeating the argument which led us to (3.3) we can assert that for the ε from (3.7) there exists $n_0 = n_0(\varepsilon)$ such that

$$\sup_{n\geq n_0} \|G_n\|_2 < \infty, \ \sup_{n\geq n_0} \|s(d_n K)\|_2 < \infty. \tag{3.9}$$

By (2.2)

$$\|G_n' - s(d_n K)\|_2 = \|G_n - s(d_n K)\|_2 \leq c\left\|W_n - d_n K\right\|_2 \tag{3.10}$$

where we have used the symmetry of $s(d_n K)$ (see the proof of Lemma 2a)) and the fact that $\varphi(\rho, W_n, d_n K) < \infty$ because of (3.7). Now we may use (2.3), (3.9) and (3.10) to obtain

$$\begin{aligned} \|\alpha_{n2}\|_2 &\leq N\left(\big(G_n' - s(d_n K)\big)G_n\right) + N\left(s(d_n K)\big(G_n - s(d_n K)\big)\right) \\ &\leq c(||G_n' - s(d_n K)||_2 ||G_n||_2 + ||s(d_n K)||_2 ||G_n - s(d_n K)||_2) \\ &\leq c_1 \left\|W_n - d_n K\right\|_2. \end{aligned} \tag{3.11}$$

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

$$\left(d_n\left(\sum_{i=L}^{M} \lambda_i f_i(x) f_i(y)\right)\right)_{st} = \sum_{i=L}^{M} \lambda_i (d_n f_i)_s (d_n f_i)_t, \quad s, t = 1, ..., n \tag{3.12}$$

(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 (2.9)

$$|(d_n f_i, d_n f_j)_{l_2}| \leq \|d_n f_i\|_2 \|d_n f_j\|_2 \leq \|f_i\|_2 \|f_j\|_2 = 1, \tag{3.13}$$

we deduce from (3.12)

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

This bound along with decompositions (1.6) and (3.4) of K and K_L produces three particular cases:

$$\|d_n K\|_2 \leq \sum_{i \geq 1} |\lambda_i|, \ \|d_n K_L\|_2 \leq \sum_{i \geq L} |\lambda_i|, \ \|d_n K - d_n K_L\|_2 \leq \sum_{i > L} |\lambda_i|. \tag{3.14}$$

The last bound will be used for estimating the terms in β_{nL} with $k > 0$. For $k = 0$, (2.8), (3.12) and (3.14) give the inequality

$$N(d_n K - d_n K_L) \leq c \left(\|d_n K - d_n K_L\|_2 + \left| \sum_{i > L} \lambda_i \|d_n f_i\|_2^2 \right| \right) \leq c_1 \sum_{i > L} |\lambda_i|. \tag{3.15}$$

Overall, utilizing (2.5), (3.14) and (3.15) we can bound the first component of β_{nL} as follows

$$\begin{aligned} \|\beta_{nL1}\|_2 &\leq N\left(d_n K - d_n K_L\right) + \sum_{k>0} |\rho|^k N\left((d_n K)^{k+1} - (d_n K_L)^{k+1} \right) \\ &\leq c_1 \sum_{i > L} |\lambda_i| + c \sum_{i > L} |\lambda_i| \sum_{k > 0} (k+1) \left(|\rho| \sum_{i \geq 1} |\lambda_i| \right)^k \leq c_2 \sum_{i > L} |\lambda_i|. \end{aligned} \tag{3.16}$$

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

(3.14) trivially leads to the bound

$$\max\left\{ \|s(d_n K)\|_2 , \|s(d_n K_L)\|_2 \right\} \leq \sum_{k \geq 0} |\rho|^k \left(\sum_{i \geq 1} |\lambda_i| \right)^{k+1} \leq c \tag{3.17}$$

which is uniform in n and L, while (2.2) and (3.14) guarantee that

$$||s(d_n K) - s(d_n K_L)||_2 \leq c||d_n K - d_n K_L||_2 \leq c \sum_{i > L} |\lambda_i| \tag{3.18}$$

where

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

It follows from (2.3), (3.17) and (3.18) that

$$\begin{aligned}
\|\beta_{nL2}\|_2 &\leq N\left(\big(s(d_nK)-s(d_nK_L)\big)s(d_nK)\right) \\
&\quad +N\left(s(d_nK_L)\big(s(d_nK)-s(d_nK_L)\big)\right) \\
&\leq c||s(d_nK)-s(d_nK_L)||_2(||s(d_nK)||_2+||s(d_nK_L)||_2)\leq c_1\sum_{i>L}|\lambda_i|. \quad (3.19)
\end{aligned}$$

Estimating γ_{nL}. Using formula (3.12) it is easy to show by induction that (see Lemma 2b) for the notation μ_{ni})

$$(d_nK_L)^{k+1}_{st}=\sum_{i_1,\dots,i_{k+1}\leq L}\prod_{j=1}^{k+1}\lambda_{i_j}\mu_{ni}(d_nf_{i_1})_s(d_nf_{i_{k+1}})_t. \quad (3.20)$$

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

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

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

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

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

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

The result is the representation

$$\gamma_{nL1}=\sum_{k\geq 0}\rho^k\sum_{i_1,\dots,i_{k+1}\leq L}\prod_{j=1}^{k+1}\lambda_{i_j}(\mu_{ni}-\mu_{\infty i})U_{nLi_1}U_{nLi_{k+1}} \quad (3.21)$$

which can be used for bounding.

By the Hölder inequality, (2.8) and (3.13) for any i,j

$$\begin{aligned}
E\left|U_{nLi}U_{nLj}\right| &\leq \left[E\left(V_n'd_nf_iV_n'd_nf_j\right)^2\right]^{1/2}=N(d_nf_id_nf_j') \\
&\leq c\left[\left(\sum_{s,t=1}^{n}(d_nf_i)^2_s(d_nf_j)^2_t\right)^{1/2}+\left|\sum_{s=1}^{n}(d_nf_i)_s(d_nf_j)_s\right|\right] \\
&= c\left[\|d_nf_i\|_2\|d_nf_j\|_2+|(d_nf_i,d_nf_j)_{l_2}|\right]\leq c_1. \quad (3.22)
\end{aligned}$$

According to (2.14), for any positive (small) ε and (large) L we can choose $n_0=n_0(\varepsilon,L)$ so large that

$$|\mu_{ni}-\mu_{\infty i}|\leq\varepsilon \text{ for all } n\geq n_0 \text{ and } i_1,...,i_{k+1}\leq L. \quad (3.23)$$

Finally, we conclude from (3.21), (3.22) and (3.23) that for all $n \geq n_0$

$$E|\gamma_{nL1}| \leq c_1\varepsilon \sum_{k\geq 0} |\rho|^k \sum_{i_1,\ldots,i_{k+1}\leq L} \prod_{j=1}^{k+1} |\lambda_{i_j}| \leq c_1\varepsilon \sum_{k\geq 0} |\rho|^k \left(\sum_{i\geq 1} |\lambda_i| \right)^{k+1} = c_2\varepsilon. \tag{3.24}$$

For numbers or square matrices a one has the identity

$$\left(\sum_{k\geq 0} a^k \right)^2 = \sum_{k,\, l\geq 0} a^{k+l} = \sum_{m\geq 0} a^m (m+1) \tag{3.25}$$

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 (3.20), one gets

$$\begin{aligned} V_n' s^2(d_n K_L) V_n &= V_n' \left(\sum_{k\geq 0} (\rho d_n K_L)^k \right)^2 (d_n K_L)^2 V_n = V_n' \sum_{m\geq 0} \rho^m (m+1)(d_n K_L)^{m+2} V_n \\ &= \sum_{m\geq 0} \rho^m (m+1) \sum_{s,t=1}^{n} (d_n K_L)^{m+2}_{st} v_s v_t \\ &= \sum_{m\geq 0} \rho^m (m+1) \sum_{i_1,\ldots,i_{m+2}\leq L} \prod_{j=1}^{m+2} \lambda_{i_j} \mu_{ni} U_{nLi_1} U_{nLi_{m+2}}. \end{aligned} \tag{3.26}$$

Application of (3.25) also provides another expression for

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

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

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

Therefore, taking into account also (3.26), we can rewrite γ_{nL2} as

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

As above, application of (3.22) and (3.23) leads to an analog of (3.24): for any positive ε, L there is $n_0 = n_0(\varepsilon, L)$ such that

$$E|\gamma_{nL2}| \leq c_1\varepsilon \sum_{m\geq 0} |\rho|^m (m+1) \sum_{i_1,\ldots,i_{m+2}\leq L} \prod_{j=1}^{m+2} |\lambda_{i_j}| \leq c_2\varepsilon \tag{3.27}$$

for all $n \geq n_0$.

Proving (1.12). Under condition (1.11) we have

$$\begin{aligned} 0 < c_1 &= 1 - |\rho| \sum_{i \geq 1} |\lambda_i| \leq 1 - |\rho \lambda_i| \leq |1 - \rho \lambda_i| \\ &\leq 1 + |\rho \lambda_i| \leq 1 + |\rho| \sum_{i \geq 1} |\lambda_i| = c_2 < \infty, \text{ all } i \end{aligned}$$

so that

$$\frac{|\lambda_i|}{c_2} \leq |\nu(\lambda_i)| \leq \frac{|\lambda_i|}{c_1}, \text{ all } i \tag{3.28}$$

where c_1 and c_2 depend on ρ. Hence, the condition $\sum_{i \geq 1} |\lambda_i| < \infty$ is equivalent to (2.19), and we can use (2.16) and (2.20).

(3.8) and (3.11) show that $\operatorname{plim} \alpha_n = 0$. From (3.16) and (3.19) we have by the Chebyshev inequality

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

where c does not depend on n. From (3.24) and (3.27) we conclude that for any fixed L $\operatorname{plim}_{n \to \infty} \gamma_{nL} = 0$. Thus, (3.5) implies

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

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 (1968)) it follows that

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

which is (1.11). Theorem 5.1 is applicable because $\Delta_{\infty 2} > 0$ almost surely.

3) *Proving (1.13).* 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

$$\begin{aligned} \sqrt{n}(\hat{\sigma}^2 - \sigma^2) &= \sqrt{n} \frac{V_n' S_n'^{-1} S_n'(\hat{\rho}) S_n(\hat{\rho}) S_n^{-1} V_n}{n} - \sqrt{n} \sigma^2 \\ &= \sqrt{n} \frac{V_n' V_n - n\sigma^2}{n} + 2 \frac{\rho - \hat{\rho}}{n^\varepsilon} \frac{V_n' G_n' V_n}{n^{1/2 - \varepsilon}} + \frac{(\rho - \hat{\rho})^2}{n^\varepsilon} \frac{V_n' G_n' G_n V_n}{n^{1/2 - \varepsilon}} \\ &= \frac{\sum (v_i^2 - \sigma^2)}{\sqrt{n}} + 2 \frac{\rho - \hat{\rho}}{n^\varepsilon} \frac{X_{n1}}{n^{1/2 - \varepsilon}} + \frac{(\rho - \hat{\rho})^2}{n^\varepsilon} \frac{X_{n2}}{n^{1/2 - \varepsilon}}. \end{aligned} \tag{3.29}$$

Here $\varepsilon \in (0, 1/2)$ is arbitrary. From the proof of Theorem 1 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.

Proof of Theorem 2

Proving that the limit in (1.14) is zero. The next equation is quite similar to the passage from (3.6) to (3.8):

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

Using (3.20) and (2.14) we see that

$$\begin{aligned}\mathrm{tr}(s(d_nK)) &= \sum_{k\ge 0} \rho^k \mathrm{tr}((d_nK)^{k+1}) \\ &= \sum_{k\ge 0} \rho^k \sum_{i_1,\ldots,i_{k+1}=1}^{\infty} \prod_{j=1}^{k+1} \lambda_{i_j} \mu_{ni}(d_nf_{i_1}, d_nf_{i_{k+1}})_{l_2} \\ &\longrightarrow \sum_{k\ge 0} \rho^k \sum_{i_1,\ldots,i_{k+1}=1}^{\infty} \prod_{j=1}^{k+1} \lambda_{i_j} \mu_{\infty i}(f_{i_1}, f_{i_{k+1}})_{l_2} \\ &= \sum_{k\ge 0} \rho^k \sum_{i\ge 1} \lambda_i^{k+1} = \sum_{i\ge 1} \nu(\lambda_i).\end{aligned}$$

Sending $n \to \infty$ here is possible because under condition (1.11) the series converge uniformly. The conclusion is that

$$\lim_{n\to\infty} \mathrm{tr}(G_n) = \sum_{i\ge 1} \nu(\lambda_i) \tag{3.30}$$

where the series at the right converges because of (3.28).

Reviewing the argument that took us from (3.9) to (3.11) we see that

$$\begin{aligned}|\mathrm{tr}(G_n'G_n) - \mathrm{tr}(s^2(d_nK))| &\le \left|\mathrm{tr}\left(\left(G_n' - s(d_nK)\right)G_n\right)\right| + \left|\mathrm{tr}\left(\left(G_n - s(d_nK)\right)G_n\right)\right| \\ &\le ||G_n' - s(d_nK)||_2 ||G_n||_2 + ||G_n - s(d_nK)||_2 ||s(d_nK)||_2 \longrightarrow 0.\end{aligned}$$

Arguing along the lines following (3.26) we have

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

The last expression tends to

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

Thus,

$$\lim_{n\to\infty} \operatorname{tr}(G_n' G_n) = \sum_{i\geq 1} \nu^2(\lambda_i). \tag{3.31}$$

In this proof we can replace G_n' by G_n. Then instead of (3.31) we have

$$\lim_{n\to\infty} \operatorname{tr}(G_n^2) = \sum_{i\geq 1} \nu^2(\lambda_i). \tag{3.32}$$

(3.30), (3.31) and (3.32) show that the limit in (1.14) is zero.

Proving that the limit in (1.15) is zero. In accordance with the ML methodology, here we redenote the true value by ρ_0 and use ρ for points close to ρ_0. The transformation in the next equation is analogous to that in (3.29)

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

It is clear from (3.30) and (3.31) that

$$\lim \sigma_n^2(\rho) = \sigma_0^2 \text{ for any } \rho. \tag{3.33}$$

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

$$\ln|\sigma_0^2 S_n^{-1} S_n'^{-1}| - \ln|\sigma_n^2(\rho) S_n^{-1}(\rho) S_n'^{-1}(\rho)| = \ln(\sigma_0^2/\sigma_n^2(\rho)) + 2(\ln|S_n(\rho)| - \ln|S_n|). \tag{3.34}$$

The formula (see Horn and Johnson (1985))

$$\frac{\partial \ln|S_n(\rho)|}{\partial \rho} = -\operatorname{tr}(W_n S_n(\rho)) \tag{3.35}$$

implies (cf. Gohberg and Kreǐn (1969), p.158)

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

where we have denoted $s(t, W_n) = \sum_{k=0}^{\infty} t^k W_n^{k+1}$. Here we are assuming that $|\rho_0| < 1/\sum_{i\geq 1}|\lambda_i|$ 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 (3.30) one can show that

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

uniformly in t from the neghborhood indicated above. Therefore

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

This relation, (3.33) and (3.34) show that the limit in (1.15) is zero for ρ close to ρ_0.

Proving that limit (1.16) 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 (2.3), (2.12) and statement 1) of Theorem 1 and verify that

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

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

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

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

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

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

Proof of Theorem 3

Deriving (1.22). Denoting

$$\nu(\rho,\lambda_{ni}) = \frac{\lambda_{ni}}{1-\rho\lambda_{ni}},\ i=1,...,n,$$

and using (1.19), for the matrices involved in (1.20) we have representations

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

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

$$\mathrm{tr}(A_nW_nS_n^{-1}(\rho_{j-1})) = \frac{1}{c_n}\sum_{i=1}^n c_{ni}\nu(\rho_{j-1},\lambda_{ni}).$$

It is easy to see that the vector $\widetilde{V}_n = P_n'V_n$ is distributed as $N(0,\sigma^2I)$. (1.20) becomes

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

The numerator can be rearranged as follows:

$$\sum_{i=1}^n\widetilde{v}_i^2\frac{\nu(\lambda_{ni})}{1-\rho\lambda_{ni}} - \frac{\hat{\sigma}^2}{c_n}\sum_{i=1}^n c_{ni}\nu(\rho_{j-1},\lambda_{ni}) = \sum_{i=1}^n\widetilde{v}_i^2\left(\frac{\nu(\lambda_{ni})}{1-\rho\lambda_{ni}} - \nu(\lambda_{ni})\right)$$

$$+\sum_{i=1}^n\left(\widetilde{v}_i^2 - \frac{\sigma^2c_{ni}}{c_n}\right)\nu(\lambda_{ni}) + \frac{\sigma^2-\hat{\sigma}^2}{c_n}\sum_{i=1}^n c_{ni}\nu(\lambda_{ni}) + \frac{\hat{\sigma}^2}{c_n}\sum_{i=1}^n c_{ni}(\nu(\lambda_{ni}) - \nu(\rho_{j-1},\lambda_{ni})).$$

Hence, if we denote

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

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

then ρ_j becomes

$$\rho_j = \rho + \kappa_{n1} + \kappa_{n2} + \kappa_{n3} \sum_{i=1}^{n} \frac{c_{ni}}{c_n} (\nu(\lambda_{ni}) - \nu(\rho_{j-1}, \lambda_{ni})).$$

If we also take into account that

$$\nu(\lambda_{ni}) - \nu(\rho_{j-1}, \lambda_{ni}) = \nu(\rho, \lambda_{ni}) - \nu(\rho_{j-1}, \lambda_{ni}) = \int_{\rho_{j-1}}^{\rho} \frac{\partial \nu(t, \lambda_{ni})}{\partial t} dt = \int_{\rho_{j-1}}^{\rho} \nu^2(t, \lambda_{ni}) dt$$

and denote

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

then ρ_j rewrites as (1.22).

Final touches. The validity of the first equation in (1.23) follows from (2.24):

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

We claim that (see Lemma 3)

$$\operatorname*{dlim}_{n\to\infty} \kappa_{n0} = \Delta_{\infty 2} = \sigma^2 \sum_{i=1}^{\infty} u_i^2 \nu^2(\lambda_i). \tag{3.36}$$

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

(2.23) and Assumption 4 imply by Hölder's inequality

$$\left| \frac{1}{n^{1/q}} \sum_{i=1}^{n} \frac{c_{ni}}{c_n} \nu(\lambda_{ni}) \right| \le \frac{1}{n^{1/q}} \left(\sum_{i=1}^{n} |\nu(\lambda_{ni})|^p \right)^{1/p} n^{1/q} \le c. \tag{3.37}$$

Hence, factorizing κ_{n2} as

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

we see that by (1.13), (3.36) and (3.37) the factors in all brackets are $O_p(1)$, so that $\kappa_{n2} = o_p(1)$. We have proved the second relation in (1.23).

(1.24) is a consequence of (3.36) and consistency of $\hat{\sigma}^2$.

Nonnegativity of κ_{n3} and ψ_n are obvious.

References

[1] Anderson, T.W. and D. A. Darling (1952) Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Annals of Math. Statist. 23, 193-212.

[2] Anselin, L. (1988) Spatial Econometrics: Methods and Models. Boston, Kluwer Academic Publishers.

[3] Anselin, L. and A.K. Bera (1998) Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, in: Handbook of Applied Economics Statistics ed. by A. Ullah and D.E.A. Giles. New York, Marcel Dekker.

[4] Billingsley, P. (1968) Convergence of Probability Measures. John Wiley & Sons, New York.

[5] Cliff, A., Ord, J. (1981) Spatial Processes, Models and Applications. London, Pion.

[6] Cressie, N.A.C. (1993) Statistics of Spatial Data. New York, Wiley.

[7] Gohberg, I.C. and M.G. Kreĭn (1969) Introduction to the Theory of Linear Nonselfadjoint Operators. American Mathematical Society, Providence, Rhode Island.

[8] Hardy, G.H. (1956) Divergent Series. Oxford, Clarendon Press.

[9] Hoque, A. The Exact Moments of Forecast Error in the General Dynamic Model. Sankhya: The Indian Journal of Satistics, 47, Series B, Pt. 1, 128-143.

[10] Horn, R., and C. Johnson (1985) Matrix Analysis. New York, Cambridge University Press.

[11] Jones, M.C. (1986) Expressions for Inverse Moments of Positive Quadratic Forms in Normal Variables, Australian Journal of Statistics, 28(2), 242-250.

[12] Kelejian, H.H. and I.R. Prucha (1999) A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. International Economic Review 40, 509-533.

[13] Kelejian, H.H. and I.R. Prucha (2002) 2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights, Regional Science and Urban Economics, 32, 691-707.

[14] Kelejian, H.H., I.R. Prucha and Y. Yuzefovich (2004) Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results. Forthcoming in J. LeSage and R.K. Pace, Advances in Econometrics, New York: Elsevier.

[15] Lee, L.F. (2001) Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models I: Spatial Autoregressive Processes. Manuscript, Department of Economics, The Ohio State University, Columbus, Ohio.

[16] Lee, L.F. (2002) Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models. Econometric Theory 18, 252-277.

[17] Lee, L.F. (2003) Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. Econometric Reviews 22, No. 4, 307-335.

[18] Lee, L.F. (2004) Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica 72, No. 6, 1899-1925.

[19] Mynbaev, K.T. (2001) L_p-approximable Sequences of Vectors and Limit Distribution of Quadratic Forms of Random Variables. Advances in Applied Mathematics 26, 302-329.

[20] Ord, J.K. (1975) Estimation Methods for Models of Spatial Interaction. Journal of American Statistical Association, 70, 120-126.

[21] Varberg, D.E. (1966) Convergence of Quadratic Forms in Independent Random Variables. Ann. Math. Stat. 37, 567-576.

[22] White, H. (1994) Estimation, Inference and Specification Analysis. New York, Cambridge University Press.