Asymptotic Distribution of the OLS Estimator for a Mixed Regressive, Spatial Autoregressive Model

Mynbaev, Kairat

Kazakh-British Technical University

1 August 2006

Asymptotic Distribution of the OLS Estimator for a Mixed Regressive, Spatial Autoregressive Model

Kairat T. Mynbaev*†

August 1, 2006

Abstract

We find the asymptotics of the OLS estimator of the parameters β and ρ in the spatial autoregressive model with exogenous regressors $Y_n = X_n\beta + \rho W_n Y_n + V_n$. Only low-level conditions are imposed. Exogenous regressors may be bounded or growing, like polynomial trends. The assumption on the spatial matrix W_n is appropriate for the situation when each economic agent is influenced by many others. The asymptotics contains both linear and quadratic forms in standard normal variables. The conditions and the format of the result are chosen in a way compatible with known results for the model without lags by Anderson (1971) and for the spatial model without exogenous regressors due to Mynbaev and Ullah (2006).

Keywords: mixed regressive spatial autoregressive model, OLS estimator, asymptotic distribution

JEL codes: C21, C31

1 Introduction

We busy ourselves with estimation of parameters β and ρ in the model

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

where X_n is an $n \times k$ matrix of deterministic exogenous regressors, β is an unknown $k \times 1$ parameter, ρ is an unknown real parameter, the $n \times n$ matrix W_n is given and the elements of $W_n Y_n$ represent spatial lags of the n-dimensional dependent vector Y_n. V_n is an unobservable error vector with zero mean.

The early development of spatial econometrics has been summarized in several textbooks (Paelinck and Klaasen (1979), Anselin (1988), Cressie (1993)) and collections of spatial econometrics papers (Anselin (1992), Anselin and Florax (1995), Anselin and Rey (1997)). The recent years have seen new efforts in establishing asymptotic properties of various estimation techniques. Kelejian and Prucha (1998) provide an asymptotic analysis of the

*Kazakh-British Technical University, Almaty, Kazakhstan

†Research supported by the Fulbright program and generously hosted by the Economics Department, University of California - Riverside. Address for correspondence: Kairat T. Mynbaev, Zhandosov 57, kv. 49, 050035, Almaty, Kazakhstan

This paper has been delivered at Second Italian Congress of Econometrics and Empricial Economics, see http://www.cide.info/conf/papers.php#SpatialEconometrics

instrumental variables estimator. Kelejian and Prucha (1999) consider a generalized moments estimator in the absence of X_n. Lee (2002) investigates the OLS approach. Lee (2003, 2004) studies a two-stage least squares procedure and derives an asymptotic distribution of the quasi-maximum likelihood estimator. It has long been noted that the OLS estimator may be inconsistent (see, e.g., Whittle (1954), Ord (1975), Anselin (1988)). The search for consistent estimators or conditions ensuring consistency of the OLS estimator has been partially the motivation of the recent papers.

The structure of model (1.1) makes the analysis of any estimation procedure very lengthy and sophisticated. Along the way, several complex expressions in terms of X_n, W_n and V_n arise. Limits of those expressions need to be evaluated. The existing papers deal with this problem by imposing the condition that the required limits exist and take the desired values. As the level of complexity rises, it is more and more difficult to see how various conditions relate to one another and if they are compatible at all. Thus, the state of the current research calls for a drastic reduction in the number of conditions, with development of a corresponding analytical method.

One such method for the OLS estimator in the absence of exogenous regressors has been proposed by Mynbaev and Ullah (2006). They have found a closed analytical expression for the asymptotic distribution of the OLS estimator. The asymptotic bias is a ratio of two (in general, infinite) linear combinations of independent χ^2 variables. An attractive feature of their method is a significant reduction in the number of assumptions and possibility to calculate all the required limits. In particular, by verifying the corresponding identification conditions from Lee (2001), they have shown that neither the maximum likelihood nor the method of moments work under their set of conditions.

In this paper we develop further their method to apply to the general case (1.1). While doing this, we keep the number of conditions low and use only low-level assumptions. Note that when there are no spatial lags, the asymptotics of the OLS estimator is expressed in terms of a normal vector. In the other extreme case, when the exogenous regressors are absent, the asymptotic result by Mynbaev and Ullah (2006) involves linear combinations of χ^2-variables. The major challenge is to glue these two kinds of asymptotics together. That is, we want to derive an asymptotics for the OLS estimator $\widehat{\delta}$ of $\delta = (\beta', \rho)'$ which would include both linear and quadratic forms. The fact that finite-sample distributions involve linear-quadratic forms of innovations is well-known; the problem is to carry this structure over to infinity. Kelejian and Prucha (1998, 2001) and Lee (2004) prove central limit theorems for linear-quadratic forms but under their conditions the quadratic part disappears in the limit.

Lee's (2002) paper is the most relevant to ours. The main results are not comparable as Lee studies a different situation when W_n is row-normalized. The methodologies, on the other hand, can be compared and the comparison reveals two important differences. Firstly, we retain in the asymptotics both linear and quadratic forms in standard normal variables, while in Lee (2002) and an earlier paper Kelejian and Prucha (2001) the quadratic part disappears. Secondly, in many cases we are able to verify analogs of Lee's conditions, instead of imposing them as independent assumptions. The most notable examples are Assumption 5 from Lee (2002) and Assumption 9 from Lee (2004).

We are sure that ideas and techniques used in this paper can be successfully applied in areas other than spatial econometrics. Therefore the exposition is not limited to just statements and proofs. In addition to explaining the mathematics, we motivate our choice of conditions. Where appropriate, we compare different approaches.

In Section 2 we discuss the advantages of Anderson's (1971) normalization of the regressor matrix. To the simple facts that Anderson's normalizer is convenient and self-adjusting we add a less simple fact from Mynbaev and Castelar (2001) that it is unique in some sense. Section 3 is an introduction to the L_p-approximability theory developed by Mynbaev (2001). It allows one to avoid high-level conditions when working with deterministic regressors and should be distinguished from the L_p-approximability of stochastic processes defined in Pötscher and Prucha (1991). In Section 4 we review the main ideas and tools used by Mynbaev and Ullah (2006) to the extent necessary to study the general case.

To keep the exposition as much nontechnical as possible we separate the conceptual part of the general case (Section 5) from the proofs (Section 6). The choice of the multiplier for the autoregressive part (the function m_n) and Assumption 5, though may seem simple, is a culmination of the sequence of assumptions. The idea has been borrowed from Mynbaev (2006) who studies a time series autoregressive model with one exogenous regressor.

A limit in distribution is denoted $\xrightarrow{d}$ or dlim. Likewise, symbols $\xrightarrow{p}$ or plim are used interchangeably for limits in probability.

2 The Choice of Conditions Determines the Result You Obtain

We start with the classical model

$$Y_n = X_n\beta + V_n \tag{2.1}$$

where X_n is a deterministic matrix and V_n satisfies

Assumption 1. The components $v_1, ..., v_n$ of V_n are independent identically distributed with mean zero and variance σ^2 and finite moments up to $\mu_4 = Ev_i^4$.

The classical $\sqrt{n}$-normalization arises as follows. From the formula of the OLS estimator

$$\widehat{\beta} = (X_n'X_n)^{-1}X_n'Y_n = \beta + (X_n'X_n)^{-1}X_n'V_n \tag{2.2}$$

it is easy to obtain

$$\sqrt{n}(\widehat{\beta} - \beta) = \left(\frac{X_n'X_n}{n}\right)^{-1}\frac{X_n'V_n}{\sqrt{n}}, \tag{2.3}$$

Then one imposes the condition

$$\text{the limit } \lim_{n\to\infty}\frac{X_n'X_n}{n} = \Omega \text{ exists and is nonsingular} \tag{2.4}$$

and makes additional assumptions about the error to prove that

$$\frac{X_n'V_n}{\sqrt{n}} \text{ converges in distribution to } N(0, \sigma^2\Omega). \tag{2.5}$$

Then (2.3), (2.4) and (2.5) will immediately give convergence in distribution of $\sqrt{n}(\widehat{\beta} - \beta)$.

The approach based on (2.3) is rather restrictive, as we shall see momentarily. Denote $X_{n1}, ..., X_{nk}$ the columns of X_n so that X_n is partitioned as $X_n = (X_{n1}, ..., X_{nk})$ and let $\|x\|_2 = (x'x)^{1/2} = \left(\sum_{j=1}^n x_j^2\right)^{1/2}$ be the Euclidean norm of $x \in R^n$. For the diagonal elements condition (2.4) gives

$$\lim_{n\to\infty}\frac{\|X_{ni}\|_2^2}{n} = \omega_{ii} > 0 \tag{2.6}$$

where ω_{ij} denote elements of Ω. All numbers $\omega_{11},...,\omega_{kk}$ are positive because if, say, $\omega_{ii}=0$, then by the Cauchy-Schwartz inequality

$$|\omega_{ij}|=\lim_{n\to\infty}\frac{|X'_{ni}X_{nj}|}{n}\le\lim_{n\to\infty}\frac{\|X_{ni}\|_2}{\sqrt{n}}\frac{\|X_{nj}\|_2}{\sqrt{n}}=\sqrt{\omega_{ii}\omega_{jj}}=0,\ j=1,...,k,$$

and Ω is singular. (2.6) shows that by requiring (2.4) you force the norms of the columns to grow at the same $\sqrt{n}$-rate:

$$\|X_{ni}\|_2\sim\sqrt{n\omega_{ii}}.$$

This excludes, for example, a polynomial trend $T_n=(1^l,...,n^l)'$ for which $\|T_n\|_2\sim n^{l+1/2}$ (see, for example, Hamilton (1994)).

However, there is a better normalization appeared in Anderson (1971) and Schmidt (1976) who did not compare it to the classical one. Amemiya (1985) does such a comparison, while Mynbaev and Castelar (2001) prove that it is better than *any other* normalizer (see also Mynbaev and Lemos (2004)). Here we repeat the main points because these sources are not easily accessible. Put

$$M_n=\text{diag}[\|X_{n1}\|_2,...,\|X_{nk}\|_2].$$

Instead of (2.3) we now have

$$M_n(\widehat{\beta}-\beta)=M_n(X'_nX_n)^{-1}M_nM_n^{-1}X'_nV_n=(H'_nH_n)^{-1}H'_nV_n \tag{2.7}$$

where

$$H_n=X_nM_n^{-1}=(X_{n1}/\|X_{n1}\|_2,...,X_{nk}/\|X_{nk}\|_2).$$

The conditions

$$\text{the limit }\lim_{n\to\infty}H'_nH_n=\Gamma_1\text{ exists and is nonsingular} \tag{2.8}$$

and

$$H'_nV_n\text{ converges in distribution to }N(0,\sigma^2\Gamma_1) \tag{2.9}$$

replace (2.4) and (2.5), respectively. Since the columns of H are normalized, (2.8) has the advantages that it is more likely to be satisfied than (2.4) and it does not exclude regressors with quickly growing norms. (2.9) is also better since the components of H'_nV_n have constant variances if the error is subject to Assumption 1. (2.3), (2.4), (2.5) or (2.7), (2.8), (2.9) represent the line of reasoning we call a *conventional scheme*.

To cover regressors with norms growing at a rate different from $\sqrt{n}$, you might want to play with different functions of n as a normalizer. For example, in case of the polynomial trend it is common to use $f(n)=n^{l+1/2}$. This is not a good idea, though, because each time you will need to figure out the rate of growth of $\|X_{ni}\|_2$ and the result you obtain will be tied to a function $f(n)$ with a particular behavior at infinity. In fact, you obtain as many "results" as there are functions with different asymptotics at infinity. With the Anderson normalizer you don't have this multitude of results because it is self-regulating: it adjusts itself to regressors instead of separating a narrow class thereof. This is especially important in applications where one usually has just an irregular set of numbers without any analytical pattern.

Speaking of applications, what happens to the usual statistics if you use Anderson's normalizer instead of the classical square root? The analysis in Mynbaev and Castelar (2001) shows that the usual tests of scalar and vector restrictions based on t and F statistics apply. The underlying assumptions and proofs change but the form of the statistics does not. This means everybody can continue using the same statistical software.

As it happens, if in addition to (2.8) one requires the *errors contribution negligibility condition*

$$\lim_{n\to\infty} \max_{i,j} |h_{nij}| = 0 \tag{2.10}$$

or, in terms of the original regressor matrix,

$$\lim_{n\to\infty} \max_{i,j} \frac{x_{nij}}{\|X_{ni}\|_2} = 0$$

and the error satisfies Assumption 1, then (2.9) is true. This is more or less how T.W. Anderson came up with the next theorem (which slightly differs from the original but the idea is the same).

Theorem (Anderson (1971)). If the error satisfies Assumption 1 and the regressors are subject to (2.8) and (2.10), then

$$M_n(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \Gamma_1^{-1}). \tag{2.11}$$

The final and, perhaps, most important point about M_n requires a definition. Let $\overline{M}_n$ be some diagonal $k \times k$ matrix with positive elements on the main diagonal and let $\overline{H}_n = X_n \overline{M}_n^{-1}$. Generalizing upon (2.8) and (2.9), we say that $\overline{M}_n$ is a *Conventional-Scheme-Compliant (CSC) normalizer* if

$$\text{the limit } \lim_{n\to\infty} \overline{H}_n' \overline{H}_n = \overline{\Gamma} \text{ exists and is nonsingular and}$$
$$\overline{H}_n' V_n \xrightarrow{d} N(0, \sigma^2 \overline{\Gamma}) \text{ for all } V_n \text{ satisfying Assumption 1.}$$

A CSC normalizer is not unique (if it exists) because if

$$\Delta_n = \text{diag}[m_{n1}, ..., m_{nk}], \text{ limits } m_k = \lim_{n\to\infty} m_{nk} \text{ exist and are positive,} \tag{2.12}$$

then $\Delta_n \overline{M}_n$ is also a CSC normalizer.

Theorem (Mynbaev and Castelar (2001)). Anderson's normalizer is unique in the class of CSC normalizers up to a factor satisfying (2.12), that is if $\overline{M}_n$ is any other CSC normalizer, then there exists Δ_n such that (2.12) holds and $\overline{M}_n = \Delta_n M_n$.

Summarizing, M_n is universally applicable (if any other CSC normalizer works, then M_n also works), self-adjusting (you don't need to worry about the rates of growth of regressors) and unique (up to an asymptotically constant factor).

3 Want nice sequences of vectors? Look no further than L_2-approximability

For finite n, linear independence of columns of X_n is equivalent to $\det H_n' H_n \neq 0$. By way of generalization, nonsingularity of Γ_1 from (2.8) can be interpreted as an *asymptotic linear independence* condition. The question is: can the word "asymptotic" be removed from this interpretation? Put it differently, are there any vectors for which $\det \Gamma_1 \neq 0$ would mean just linear independence?

The answer is "no" if you try to use columns $H_{n1}, ..., H_{nk}$. They belong to spaces R^n of growing dimension but that is not a problem because we can think of R^n as being embedded into the space l_2 of infinite sequences $x = (x_1, x_2, ...)$ provided with the norm

$$\|x\|_2 = \left(\sum_{i\geq 1} x_i^2\right)^{1/2}.$$

The problem is that, due to (2.10), coordinates of the columns tend to zero and, consequently, the columns do not converge in l_2 either.

The answer may be "yes", if the columns are represented as images of some functions of a continuous argument. This statement will be clear after a couple of definitions.

Consider the space $L_2(0,1)$ of square-integrable on $(0,1)$ functions h provided with the norm

$$\|h\|_2 = \left(\int_0^1 h^2(t)dt\right)^{1/2}.$$

$(x,y)_{L_2}$ is the corresponding scalar product. Let $d_n : L_2(0,1) \to R^n$ be a *discretization operator* defined as follows. For $h \in L_2(0,1)$, $d_n h \in R^n$ is a vector with components

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

where $q_i = \left(\frac{i-1}{n}, \frac{i}{n}\right)$ are small intervals that partition $(0,1)$. Using Hölder's inequality it is easy to check that

$$\|d_n h\|_2 \leq \|h\|_2 \text{ for all } h \text{ and } n \tag{3.1}$$

(the norm at the left is the Euclidean norm in R^n).

One can go back from R^n to $L_2(0,1)$ by way of piece-wise interpolation. If 1_{q_i} denotes the indicator of q_i ($1_{q_i} = 1$ on q_i and $1_{q_i} = 0$ outside q_i), then the *interpolation operator* D_n takes a vector $x \in R^n$ to

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

We shall use the notation

$$(x,y)_{l_2} = \sum_{i\in I} x_i y_i$$

for scalar products of all vectors encountered in this paper; the set of indices I will depend on the context. It is easy to see that D_n preserves scalar products

$$(D_n x, D_n y)_{L_2} = (x,y)_{l_2} \text{ for all } x, y \in R^n \text{ and } n \tag{3.2}$$

and that the product $D_n d_n$ coincides with the *Haar projector* P_n defined by

$$P_n h = n\sum_{i=1}^n \int_{q_i} h(x)dx 1_{q_i}.$$

Its main property is that it approximates the identity operator:

$$\lim_{n\to\infty} \|P_n h - h\|_2 = 0 \text{ for any } h \in L_2(0,1). \tag{3.3}$$

For a fixed $h \in L_2(0,1)$, the sequence $\{d_n h : n = 1, 2, ...\}$ is called *L_2-generated*. L_2-generated sequences have been introduced by Moussatat (1976) and used in some statistical papers (see Milbrodt (1992) and Millar (1982)). Now, if we take two functions $h_1, h_2 \in L_2(0,1)$, then (3.3) and continuity of scalar products imply $(P_n h_1, P_n h_2)_{L_2} \to (h_1, h_2)_{L_2}$. If, further, we put $H_{n1} = d_n h_1$, $H_{n2} = d_n h_2$, then by (3.2) we have

$$H'_{n1} H_{n2} = (d_n h_1, d_n h_2)_{l_2} = (P_n h_1, P_n h_2)_{L_2} \to (h_1, h_2)_{L_2}.$$

This tells us that if columns of H_n are L_2-generated by $h_1, ..., h_k \in L_2(0,1)$ and $h_1, ..., h_k$ are linearly independent, then (2.8) will be true with a nonsingular matrix

$$\Gamma_1 = \begin{pmatrix} (h_1, h_1)_{L_2} & ... & (h_1, h_k)_{L_2} \\ ... & ... & ... \\ (h_k, h_1)_{L_2} & ... & (h_k, h_k)_{L_2} \end{pmatrix} \tag{3.4}$$

which is called a *Gram matrix* of the system $h_1, ..., h_k$. (2.10) will also hold because by Hölder's inequality and absolute continuity of the Lebesgue integral

$$\max_i |(d_n h)_i| = \max_i \left(\int_{q_i} h^2(x) dx \right)^{1/2} \to 0, \ n \to \infty.$$

Thus, (2.11) will be true if instead of requiring (2.8) and (2.10) we just say that the columns of H_n are L_2-generated by linearly independent functions $h_1, ..., h_k$. Much of the finite-dimensional geometric intuition works in $L_2(0,1)$. Linear (in)dependence, orthogonality of vectors, orthoprojectors can be used in full if asymptotic properties like (2.8) are looked at from the point of view of their counterparts in $L_2(0,1)$.

Practitioners may object by saying that requiring the columns of H_n to be *exact* images of some functions under the mapping d_n will void potential applications in econometrics. The next definition from Mynbaev (2001) is a way around this obstacle.

Definition. Let $\{h_n\}$ be some sequence of vectors such that $h_n \in R^n$ for each n. We say that $\{h_n\}$ is *L_2-approximable* if there exists a function $h \in L_2(0,1)$ such that

$$\|h_n - d_n h\|_2 = \left(\sum_{i=1}^{n} (h_{ni} - (d_n h)_i)^2 \right)^{1/2} \to 0, \ n \to \infty.$$

In this case we also say that $\{h_n\}$ is *L_2-close* to h.

This definition introduces some degree of freedom by allowing h_n to deviate from exact images.

Assumption 2. The columns $H_{n1}, ..., H_{nk}$ of H_n are L_2-close to $h_1, ..., h_k \in L_2(0,1)$, respectively.

L_2-approximable sequences inherit all properties of L_2-generated ones. In particular,

$$H'_{nl} H_{nm} \to (h_l, h_m)_{L_2} \text{ for } 1 \le l, m \le k. \tag{3.5}$$

Assumption 2 is strictly stronger than the combination (2.8) + (2.10). This can be seen from the characterization of L_2-approximability given in Mynbaev (2001). Due to their regularity properties, L_2-approximable sequences are to others as normal errors to more general ones. It is common to require econometric results to be true at least for normal errors. Similarly, when imposing some condition on sequences of vectors, your justification could be: "I have checked that this condition holds for L_2-approximable sequences".

In the rest of this section we state and comment some properties of L_2-approximable sequences. Note that the coordinates of $H_n' V_n$ are of form $\sum_{i=1}^{n} w_{ni} v_n$ where w_{ni} are deterministic weights. The next theorem describes the asymptotic behavior of such sums when the weights are L_2-approximable.

Central Limit Theorem (Mynbaev (2001)). If Assumptions 1 and 2 hold and $h_1, ..., h_k$ are linearly independent, then one has

$$H_n' V_n \xrightarrow{d} N(0, \sigma^2 \Gamma_1), \ \lim_{n\to\infty} \operatorname{var}(H_n' V_n) = \sigma^2 \Gamma_1. \tag{3.6}$$

Theorem 4.1 from Mynbaev (2001) actually covers also weighted sums of linear processes $\sum_{j=-\infty}^{j=\infty} e_{t-j}\psi_j$ with short-range dependence ($\sum_{j=-\infty}^{j=\infty} |\psi_j| < \infty$). The second relation in (3.6) is unusual for CLTs. Mynbaev and Castelar (2001) have shown that sequences obtained by normalizing a polynomial trend (T_n) and logarithmic trend ($L_n = (\ln^k 1, \ldots, \ln^k n)$, k is natural) are L_2-approximable and those obtained from a geometric progression ($G_n = (a^0, a^1, \ldots, a^{n-1})$, a is real) and exponential trend ($E_n = (e^a, \ldots, e^{na})$, a is real) are not. Linear independence of $h_1, ..., h_k$ means that Γ_1 is positive definite. The next corollary shows that this condition can be omitted.

Corollary. Under Assumptions 1 and 2 (3.6) remains true if $h_1, ..., h_k$ are linearly dependent.

Definitions of d_n and D_n easily modify for a two-dimensional case. For an integrable on the square $(0,1)^2$ function W, $d_n W$ is an $n \times n$ matrix with elements

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

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

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

Analogs of (3.2), (3.3), (3.5) are true in the two-dimensional case. A sequence of matrices $\{W_n\}$ such that W_n is of size $n \times n$ for each n is called *L_2-approximable* if there is a function $W \in L_2((0,1)^2)$ satisfying $\|W_n - d_n W\|_2 \to 0, \ n \to \infty$. Some statements in the next section require a stronger

Assumption 3. For the spatial matrices W_n there exists a function $W \in L_2((0,1)^2)$ such that

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

4 Purely Autoregressive Spatial Model

For the case $\rho = 0$ we refer to Anderson's theorem from Section 2. The other extreme case, $\beta = 0$, will be discussed here. To show the intuition behind the main result of this section, we calculate the finite-sample deviation of the OLS estimator from the true parameter under simplified assumptions.

Thus, here we deal with the model

$$Y_n = \rho W_n Y_n + V_n \tag{4.1}$$

and the OLS estimator $\widehat{\rho}$ of ρ. In many applications (1.1) and (4.1) are considered equilibrium models. In the language of the theory of simultaneous equations, a reduced-form equation is the one which does not contain the dependent variable on the right. If ρ is such that the matrix

$$S_n = I_n - \rho W_n$$

is nonsingular, then the reduced form of (4.1) is $Y_n = S_n^{-1} V_n$. Denoting additionally $Z_n = W_n Y_n$ the regressor in (4.1) and

$$G_n = W_n S_n^{-1}$$

we get the formula

$$\widehat{\rho} = (Z_n' Z_n)^{-1} Z_n' Y_n = \rho + \frac{V_n' G_n' V_n}{V_n' G_n' G_n V_n} \tag{4.2}$$

which can be used for analysis.

Obviously, in (4.2) we have a ratio of two quadratic forms in random variables. Without loss of generality we can think of W_n as a symmetric matrix because otherwise it can be replaced by $(W_n + W_n')/2$ without changing the value of (4.2). Then each W_n can be represented as

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

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

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

$$G_n = P_n \mathrm{diag}\left[\frac{\lambda_{n1}}{1 - \rho\lambda_{n1}}, ..., \frac{\lambda_{nn}}{1 - \rho\lambda_{nn}}\right] P_n'.$$

Assume for a moment that V_n is distributed as $N(0, \sigma^2 I)$. Putting

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

and noting that $\widetilde{V}_n = P_n' V_n$ is also distributed as $N(0, \sigma^2 I)$, we have

$$\widehat{\rho} - \rho = \frac{\sum_{i=1}^n \left(\frac{\widetilde{v}_i}{\sigma}\right)^2 \nu(\lambda_{ni})}{\sum_{i=1}^n \left(\frac{\widetilde{v}_i}{\sigma}\right)^2 \nu^2(\lambda_{ni})} \tag{4.3}$$

where both the numerator and denominator are linear combinations of χ^2-variables with one degree of freedom. Whether this ratio-of-quadratic-forms structure will be preserved in the limit depends on assumptions. For example, if you encounter a fraction f_n/g_n, represent the denominator as $g_n = Eg_n(1 + \frac{g_n - Eg_n}{Eg_n})$ and require Eg_n to converge to some non-zero value and $\frac{g_n - Eg_n}{Eg_n}$ to converge in probability to zero, you will get rid of randomness in the denominator after sending $n \to \infty$ (this is what happens in Lee (2002)).

We are taken by our assumptions to another world where the limit of $\widehat{\rho} - \rho$ is a ratio of two *infinite* linear combinations of χ^2-variables. Our choice has nothing to do with value judgments as to which world is better; we just want to stick to low-level assumptions and trace their implications to whatever world they take us.

From (4.3) one can surmise that, as $n \to \infty$, the eigenvalues λ_{ni} may approach some numbers and those numbers should be eigenvalues of something. This idea is formalized in the next assumption. Denote $\mathcal{W}$ the integral operator in $L_2(0,1)$ with the kernel W (see Assumption 3)

$$(\mathcal{W}f)(x) = \int_0^1 W(x,y)f(y)dy, \ f \in L_2(0,1).$$

Assumption 4. W is symmetric, which together with square-integrability of W implies that the eigenvalues $\lambda_i, \ i = 1, 2, ...,$ of $\mathcal{W}$ are real and satisfy $\sum_{i\geq 1} \lambda_i^2 < \infty$. We assume further that the eigenvalues are summable: $\sum_{i\geq 1} |\lambda_i| < \infty$.

Here the eigenvalues λ_i and eigenfunctions f_i of $\mathcal{W}$ are listed according to their multiplicity; the system of eigenfunctions is complete and orthonormal in $L_2(0,1)$. The kernel can be decomposed into the series

$$W(x,y) = \sum_{i\geq 1} \lambda_i f_i(x) f_i(y) \tag{4.4}$$

which converges in $L_2\left((0,1)^2\right)$. This decomposition leads to the identity

$$\int_0^1 \int_0^1 W^2(x,y)dxdy = \sum_{i\geq 1} \lambda_i^2 \tag{4.5}$$

which show that the condition $W \in L_2((0,1)^2)$ is equivalent to the square-summability of eigenvalues. The eigenvalues summability assumption is stronger because

$$\left(\sum_{i\geq 1} \lambda_i^2\right)^{1/2} \leq \sum_{i\geq 1} |\lambda_i|.$$

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

The main statement about asymptotics of (4.2) is next and it will be followed by commentaries and pieces of the proof needed later.

Theorem (Mynbaev and Ullah (2006)). Suppose Assumptions 1, 3 and 4 hold.

1) If

$$|\rho| < 1/\left(\sum_{i\geq 1} \lambda_i^2\right)^{1/2}, \tag{4.6}$$

then the matrices S_n^{-1} exist for all sufficiently large n and have uniformly bounded $\|\cdot\|_2$-norms, so that (4.2) can be used.

2) If

$$|\rho| < 1/\sum_{i\geq 1} |\lambda_i|, \tag{4.7}$$

then

$$\widehat{\rho} - \rho \xrightarrow{d} \frac{\sum_{i\geq 1} u_i^2 \nu(\lambda_i)}{\sum_{i\geq 1} u_i^2 \nu^2(\lambda_i)} \tag{4.8}$$

where u_i are independent standard normal.

3) (4.7) implies convergence

$$\sqrt{n}(\widehat{\sigma}^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4) \tag{4.9}$$

where

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

is the OLS estimator of σ^2.

Commentaries

The statement about uniform boundedness of $\|S_n^{-1}\|_2$ is one of high-level conditions often imposed in the literature.

Because of (4.5), condition (4.6) is the same as $|\rho|\|W\|_2 < 1$. One can show that L_2-approximability contained in Assumption 3 implies

$$\lim_{n\to\infty} \|W_n\|_2 = \lim_{n\to\infty} \|d_n W\|_2 = \|W\|_2 .$$

Therefore (4.6) implies $|\rho|\|W_n\|_2 < 1$ for all large n and S_n^{-1} can be represented as

$$S_n^{-1} = \sum_{l=0}^{\infty} \rho^l W_n^l. \tag{4.10}$$

By analogy with time series autoregressions, one might think that $|\rho| < 1$ is the stability condition. Based on statement 1), we can say that (4.6) is the stability condition under Assumptions 1, 3 and 4. Another deviation from the routine is that in (4.8) no normalization is necessary to achieve convergence in distribution.

The region (4.7) is narrower than (4.6). It would be interesting to find out whether (4.8) is true for $1/\sum_{i\geq 1}|\lambda_i| \leq |\rho| < 1/\left(\sum_{i\geq 1}\lambda_i^2\right)^{1/2}$ or, even better, for any $\rho \neq 1/\lambda_i,\ i = 1, 2, ...$

The expected value of the numerator in (4.8) is zero if and only if $\sum_{i\geq 1}\nu(\lambda_i) = 0$. This fact is of little use, however, because the expected value of a fraction is not necessarily proportional to the expectation of the numerator. Characteristic functions of infinite linear combinations of χ^2-variables have been derived by Anderson and Darling (1952). We have not heard of such results for the ratio in (4.8).

(4.9), in particular, means that $\widehat{\sigma}^2$ is a consistent estimator of σ^2, despite the fact that it is based on $\widehat{\rho}$, which, in general, is inconsistent.

Details of the proof

i) In many statistical expressions containing fractions the numerator converges in distribution and the denominator – in probability (this is how the conventional scheme from Section 2 works). Therefore it is possible to use the implication

$$\left.\begin{array}{c} \text{dlim} f_n = f \\ \text{plim} g_n = g,\ g \neq 0 \text{ almost surely} \end{array}\right\} \Longrightarrow \text{dlim}\frac{f_n}{g_n} = \frac{f}{f}.$$

This is not the case with (4.2) where both the numerator and denominator converge just in distribution. To circumvent this problem, one has to prove convergence of the vector (f_n, g_n) to the vector (f, g) in joint distribution and then apply the Continuous Mapping Theorem (CMT) to f_n/g_n. Mynbaev and Ullah (2006) apply this trick to the vector $(V_n'G_n'V_n, V_n'G_n'G_nV_n)$.

ii) Most CLTs are about convergence to normal vectors. If you apply such a CLT, you will get in the limit a normal vector and nothing but. More general CLTs treat convergence to the so-called stable distributions. A linear combination of χ^2-variables is not one of them. If you want to retain χ^2 in the limit, you have to express your process as a continuous function of a linear process and apply a CLT in conjunction with CMT.

iii) L_2-approximability is a device to jump from finite dimensions to infinite dimension. Another such tool is the approximation of (4.4) by its initial segment

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

From an analytical perspective, there is a place in the proof where one must work with finite L.

iv) Denote

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

for any square matrix A such that $|\rho| \|A\|_2 < 1$. Multiplication of (4.10) by W_n gives

$$G_n = s(W_n).$$

Working with infinite series of this type is a must in spatial econometrics if one wants to avoid high-level conditions. We can draw a parallel with 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 $o_p(1)$. One has to unwind the dependence $y_t = c_1 + c_1 c_2 + c_2^2 y_{t-2} + c_2 e_{t-1} + e_t$ and so on to infinity or to the initial point y_0 (in spatial econometrics there is no initial point).

The four ideas we have just explained are embodied in the representation

$$X_n = \alpha_n + \beta_{nL} + \gamma_{nL} + \delta_{nL} \tag{4.11}$$

where

$$X_n = \begin{pmatrix} V_n' G_n' V_n \\ V_n' G_n' G_n V_n \end{pmatrix}$$

(vector composed of numerator and denominator of (4.2)),

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

(intuitively, if W_n is close to $d_n W$, then $G_n' = s(W_n')$ and $G_n' G_n = s(W_n') s(W_n)$ should be close to $s(d_n W)$ and $s^2(d_n W)$, resp.),

$$\beta_{nL} = \begin{pmatrix} V_n' \big(s(d_n W) - s(d_n W_L) \big) V_n \\ V_n' \big(s^2(d_n W) - s^2(d_n W_L) \big) V_n \end{pmatrix}$$

(this is the jump from finite to infinite L),

$$\gamma_{nL} = \begin{pmatrix} V_n' s(d_n W_L) V_n \\ V_n' s^2(d_n W_L) V_n \end{pmatrix} - \delta_{nL}$$

(a small correction needed to obtain a continuous function of an asymptotically normal vector) and

$$\delta_{nL} = \sum_{i=1}^{L} \left(V_n' d_n f_i \right)^2 \nu(\lambda_i) \begin{pmatrix} 1 \\ \nu(\lambda_i) \end{pmatrix}$$

(allows for application of CLT and CMT).

To complete the scheme, Billingsley's (1968) Theorem 4.2 is used to manage the arising double-indexed family of vectors.

Most statements in the rest of this section depend on Assumptions 1, 3, 4 and (4.7).

Symmetry of W implies symmetry of d_nW.

L_2-approximability implies that $s(d_nW)$ is close to $s(W_n)$:

$$\|s(W_n)-s(d_nW)\|_2 \le c\,\|W_n-d_nW\|_2 \text{ for all large } n \tag{4.12}$$

and that $s(d_nW)$ and $G_n = s(W_n)$ have uniformly bounded norms:

$$\sup_{n\ge n_0}\|s(W_n)\|_2<\infty,\ \sup_{n\ge n_0}\|s(d_nW)\|_2<\infty \tag{4.13}$$

where n_0 depends on how close ρ is to $1/\sum_{i\ge 1}|\lambda_i|$.

With the eigenfunctions f_i of $\mathcal{W}$ in mind, for a collection of indices $i=(i_1,...,i_{l+1})$, where all of i_j's are positive integers, denote

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

and

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

Then for all i

$$\lim_{n\to\infty}\mu_{ni}=\mu_{\infty i}. \tag{4.14}$$

This property is a simple consequence of $(d_nf_i,d_nf_j)_{l_2}=(P_nf_i,P_nf_j)_{l_2}\to(f_i,f_j)_{l_2}$ and orthonormality of $\{f_i\}$.

The functions μ_{ni} allow us to write elements of the series $s(d_nW_L)$ and $s^2(d_nW_L)$ in a relatively compact form

$$\begin{aligned}(s(d_nW_L))_{st} &= \sum_{p\ge 0}\rho^p\sum_{i_1,...,i_{p+1}\le L}\prod_{j=1}^{p+1}\lambda_{i_j}\mu_{ni}(d_nf_{i_1})_s(d_nf_{i_{p+1}})_t,\\ (s^2(d_nW_L))_{st} &= \sum_{p\ge 0}\rho^p(p+1)\sum_{i_1,...,i_{p+2}\le L}\prod_{j=1}^{p+2}\lambda_{i_j}\mu_{ni}(d_nf_{i_1})_s(d_nf_{i_{p+2}})_t,\\ &\qquad s,t=1,...,n.\end{aligned} \tag{4.15}$$

For any i,j,n

$$\left(E(V_n'd_nf_iV_n'd_nf_j)^2\right)^{1/2}\le c. \tag{4.16}$$

For $\nu(\lambda_i)$ and $\nu^2(\lambda_i)$ one has expansions

$$\nu(\lambda_i)=\sum_{p\ge 0}\rho^p\lambda_i^{p+1},\ \nu^2(\lambda_i)=\sum_{p\ge 0}\rho^p(p+1)\lambda_i^{p+2}. \tag{4.17}$$

The inequalities

$$\sup_{n,L}\|s(d_nW_L)\|_2<\infty,\ \sup_n\|s(d_nW)-s(d_nW_L)\|_2\le c\sum_{i>L}|\lambda_i|, \tag{4.18}$$

where c does not depend on L, enable us to realize the approximation of $s(d_nW)$ by $s(d_nW_L)$.

Under condition (4.7) one has an equivalence

$$\sum_{i\geq 1}|\lambda_i|<\infty \text{ if and only if } \sum_{i\geq 1}|\nu(\lambda_i)|<\infty. \tag{4.19}$$

The most important elements of the proof are about convergence of variables participating in (4.11):

$$\text{plim}_{n\to\infty}\,\alpha_n=0,\ \ \text{plim}_{n\to\infty}\gamma_{nL}=0 \text{ for any fixed } L, \tag{4.20}$$

$$\begin{gathered}\text{there is a constant } c>0 \text{ such that for any positive } \varepsilon,n,L\\ P\left(|\beta_{nL1}|+|\beta_{nL2}|>\varepsilon\right)\leq \tfrac{c}{\varepsilon^2}\textstyle\sum_{i>L}|\lambda_i|.\end{gathered} \tag{4.21}$$

If we denote

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

where u_i are independent standard normal, then

$$\text{dlim}_{n\to\infty}\,X_n=\Delta_\infty,\ \text{dlim}_{n\to\infty}\,\delta_{nL}=\Delta_L,\ \text{dlim}L\to\infty\,\Delta_L=\Delta_\infty. \tag{4.22}$$

5 General Case: Preliminary Analysis and Main Results

A little calculation will reveal the OLS estimator structure for the main model (1.1). Denoting $\delta=(\beta',\rho)'$ and $Z_n=(X_n,W_nY_n)$ we can rewrite the model as $Y_n=Z_n\delta+V_n$. Until we work out the condition for nonsingularity of $Z_n'Z_n$ it is safer to work with the normal equation $Z_n'Z_n(\widehat{\delta}-\delta)=Z_n'V_n$. Recalling Anderson's normalizer M_n for X_n, let

$$\overline{M}_n=\begin{pmatrix}M_n & 0\\ 0 & m_n\end{pmatrix}$$

be an extended normalizer for Z_n, so that the normalized regressor is

$$\overline{H}_n=Z_n\overline{M}_n^{-1}.$$

Here $m_n>0$ is to be defined later. Then the normal equation becomes

$$\Omega_n\overline{M}_n(\widehat{\delta}-\delta)=\xi_n \text{ where by definition } \Omega_n=\overline{H}_n'\overline{H}_n,\ \xi_n=\overline{H}_n'V_n.$$

Denoting

$$\kappa_n=\frac{1}{m_n}M_n\beta,$$

from the reduced-form equation $Y_n=S_n^{-1}X_n\beta+S_n^{-1}V_n$ we get

$$W_nY_n=G_nX_n\beta+G_nV_n=G_nH_nM_n\beta+G_nV_n=m_nG_nH_n\kappa_n+G_nV_n,$$

which leads to another expression for the normalized regressor

$$\overline{H}_n = (X_n, W_n Y_n) \begin{pmatrix} M_n^{-1} & 0 \\ 0 & m_n^{-1} \end{pmatrix} = \left(H_n, G_n H_n \kappa_n + \frac{1}{m_n} G_n V_n \right).$$

Thus, the right side of the normal equation is

$$\xi_n = \begin{pmatrix} H_n' V_n \\ \kappa_n' H_n' G_n' V_n + \frac{1}{m_n} V_n' G_n' V_n \end{pmatrix} \tag{5.1}$$

and the blocks of the matrix

$$\Omega_n = \begin{pmatrix} \Omega_{n11} & \Omega_{n12} \\ \Omega_{n21} & \Omega_{n22} \end{pmatrix}$$

are

$$\begin{aligned} \Omega_{n11} &= H_n' H_n \\ \Omega_{n12} &= H_n' G_n H_n \kappa_n + \frac{1}{m_n} H_n' G_n V_n \\ \Omega_{n21} &= \Omega_{n12}' \\ \Omega_{n22} &= \kappa_n' H_n' G_n' G_n H_n \kappa_n + \frac{2}{m_n} \kappa_n' H_n' G_n' G_n V_n + \frac{1}{m_n^2} V_n' G_n' G_n V_n. \end{aligned} \tag{5.2}$$

We need to squeeze the most out of the assumptions imposed so far to keep the number of the new ones low. In the next two lemmas we show that all parts of ξ_n and Ω_n not involving m_n and κ_n converge. In the first lemma we consider the nonstochastic terms.

Using the system of eigenfunctions $\{f_i\}$ and remembering that summability of eigenvalues λ_i implies summability of $\nu(\lambda_i)$ (see (4.19)), define an operator $\mathcal{A}$ in $L_2(0,1)$ by

$$h = \sum_{i \geq 1} (h, f_i)_{L_2} f_i \Longrightarrow \mathcal{A} h = \sum_{i \geq 1} \nu(\lambda_i)(h, f_i)_{L_2} f_i.$$

The Parseval-type identities are true:

$$(\mathcal{A}^j h_1, h_2)_{L_2} = (h_1, \mathcal{A}^j h_2)_{L_2} = \sum_{i \geq 1} \nu^j(\lambda_i)(h_1, f_i)_{L_2}(h_2, f_i)_{L_2}, \ j = 1, 2.$$

Lemma 1. If Assumptions 2, 3, 4 and (4.7) are satisfied, then
1) $\lim_{n\to\infty} H_n' G_n H_n = \lim_{n\to\infty} H_n' G_n' H_n = \Gamma_2$ where

$$\Gamma_2 = \begin{pmatrix} (\mathcal{A} h_1, h_2)_{L_2} & ... & (\mathcal{A} h_1, h_k)_{L_2} \\ ... & ... & ... \\ (\mathcal{A} h_k, h_1)_{L_2} & ... & (\mathcal{A} h_k, h_k)_{L_2} \end{pmatrix}. \tag{5.3}$$

2) $\lim_{n\to\infty} H_n' G_n' G_n H_n = \Gamma_3$ where

$$\Gamma_3 = \begin{pmatrix} (\mathcal{A}^2 h_1, h_2)_{L_2} & ... & (\mathcal{A}^2 h_1, h_k)_{L_2} \\ ... & ... & ... \\ (\mathcal{A}^2 h_k, h_1)_{L_2} & ... & (\mathcal{A}^2 h_k, h_k)_{L_2} \end{pmatrix}. \tag{5.4}$$

In the next vector we collect all random objects (vectors and real variables) from Ω_n and ξ_n which do not depend on m_n and κ_n:

$$X_n = \begin{pmatrix} X_{n1} \\ X_{n2} \\ X_{n3} \\ X_{n4} \\ X_{n5} \end{pmatrix} = \begin{pmatrix} H_n'V_n \\ H_n'G_nV_n \\ H_n'G_n'G_nV_n \\ V_n'G_n'V_n \\ V_n'G_n'G_nV_n \end{pmatrix} \tag{5.5}$$

($H_n'G_n'V_n$ is not included because it has the same limit in distribution as $H_n'G_nV_n$; the ordering of components of X_n does not matter). Denote $h = (h_1, ..., h_k)'$, with a natural implication that $(f_i, h)_{L_2} = ((f_i, h_1)_{L_2}, ..., (f_i, h_k)_{L_2})'$. Following the scheme outlined in Section 4, we represent X_n as (4.11). The main part at the right of (4.11) is

$$\delta_{nL} = \begin{pmatrix} \delta_{nL1} \\ \delta_{nL2} \\ \delta_{nL3} \\ \delta_{nL4} \\ \delta_{nL5} \end{pmatrix} = \begin{pmatrix} H_n'V_n \\ \sum_{i=1}^L \nu(\lambda_i)(f_i, h)_{L_2}U_{nL,k+i} \\ \sum_{i=1}^L \nu^2(\lambda_i)(f_i, h)_{L_2}U_{nL,k+i} \\ \sum_{i=1}^L \nu(\lambda_i)U_{nL,k+i}^2 \\ \sum_{i=1}^L \nu^2(\lambda_i)U_{nL,k+i}^2 \end{pmatrix}$$

where U_{nL} is a random vector with $k + L$ real components

$$U_{nL} = \begin{pmatrix} H_{n1}'V_n \\ ... \\ H_{nk}'V_n \\ (d_nf_1)'V_n \\ ... \\ (d_nf_L)'V_n \end{pmatrix}.$$

The other terms of (4.11) are defined in Section 6.

Lemma 2. 1) Let Assumptions 1, 2, 3 hold and let $\sum_{i\geq 1} |\nu(\lambda_i)| < \infty$. Put

$$\Delta_L = \begin{pmatrix} \Delta_{L1} \\ \Delta_{L2} \\ \Delta_{L3} \\ \Delta_{L4} \\ \Delta_{L5} \end{pmatrix} = \sigma \begin{pmatrix} \sum_{i=1}^\infty (f_i, h)_{L_2}u_i \\ \sum_{i=1}^L \nu(\lambda_i)(f_i, h)_{L_2}u_i \\ \sum_{i=1}^L \nu^2(\lambda_i)(f_i, h)_{L_2}u_i \\ \sigma \sum_{i=1}^L \nu(\lambda_i)u_i^2 \\ \sigma \sum_{i=1}^L \nu^2(\lambda_i)u_i^2 \end{pmatrix}, \ 1 \leq L \leq \infty, \tag{5.6}$$

where $u_1, u_2, ...$ are independent standard normal. Then

$$\text{dlim}_{n\to\infty}\, \delta_{nL} = \Delta_L \text{ for all } L < \infty, \tag{5.7}$$

$$\text{plim}_{L\to\infty}\, \Delta_L = \Delta_\infty. \tag{5.8}$$

2) Under Assumptions 1, 2, 3, 4 and (4.7) one has

$$\text{dlim}_{n\to\infty} X_n = \Delta_\infty. \tag{5.9}$$

From (5.1), (5.2) and Lemmas 1 and 2 we see that we are only lacking information about κ_n and m_n. Comparison with a similar situation in Mynbaev (2006) shows that

$$m_n = \max\{\|X_{n1}\|_2|\beta_1|, ..., \|X_{nk}\|_2|\beta_k|, 1\}$$

is the right choice. Note that always $m_n \geq 1$ and $|\kappa_{ni}| \leq 1$. This definition and the next assumption are critical to the whole paper.

Assumption 5. The limits

$$m_\infty = \lim_{n\to\infty} m_n \in [1,\infty] \text{ and } \kappa_{\infty i} = \lim_{n\to\infty} \frac{\|X_{ni}\|_2\beta_i}{m_n} \in [-1,1]$$

exist.

The next lemma partially answers the question of what this assumption means in terms of regressors and β.

Lemma 3. Under Assumption 5 the following is true:

a) If $\beta_i = 0$, then X_{ni} is arbitrary.

b) Let $\beta_i \neq 0$. Then

b$_1$) $\kappa_{\infty i} = 0$ is equivalent to $\|X_{ni}\|_2 = o(m_n)$.

b$_2$) $\kappa_{\infty i} \neq 0$ is equivalent to $\|X_{ni}\|_2/m_n \to c_i > 0$.

c) Conditions

$$\max_i |\kappa_{\infty i}| < 1 \text{ and } m_\infty > 1 \tag{5.10}$$

are mutually exclusive.

d) $\kappa_\infty = 0$ if and only if either

i) $\beta = 0$

or

ii) $\beta \neq 0$ and $\lim \|X_{ni}\|_2 = 0$ for any i such that $\beta_i \neq 0$.

In either case $m_n = 1$ for all large n and $m_\infty = 1$.

e) If $m_\infty = \infty$, then $\kappa_\infty \neq 0$.

Definition. When $m_\infty = \infty$, we say that *exogenous regressors dominate*. In this case Lemma 2, Lemma 3e), (5.1) and (5.2) show that

$$\xi_n = \begin{pmatrix} H_n' V_n \\ \kappa_n' H_n' G_n' V_n + o_p(1) \end{pmatrix}, \ \Omega_n = \begin{pmatrix} H_n' H_n & H_n' G_n H_n \kappa_n + o_p(1) \\ \kappa_n' H_n' G_n' H_n + o_p(1) & \kappa_n' H_n' G_n' G_n H_n \kappa_n + o_p(1) \end{pmatrix}$$

where $\kappa_n \to \kappa_\infty \neq 0$. The quadratic part in ξ_n and Ω_n disappears. If $\kappa_\infty = 0$, we say that the *autoregressive part dominates*. In this case by Lemma 1 and Lemma 3d)

$$\xi_n = \begin{pmatrix} \xi_{n1} \\ \xi_{n2} \end{pmatrix} = \begin{pmatrix} H_n' V_n \\ V_n' G_n' V_n + o_p(1) \end{pmatrix}, \ \Omega_n = \begin{pmatrix} H_n' H_n & H_n' G_n V_n + o(1) \\ V_n' G_n' H_n + o(1) & V_n' G_n' G_n V_n + o_p(1) \end{pmatrix}$$

and the linear part in ξ_{n2} and Ω_{n22} asymptotically vanishes.

Theorem 1. Under Assumptions 1 through 5 and $|\rho| < 1/\sum_{i\geq 1} |\lambda_i|$ one has

$$\mathrm{dlim}_{n\to\infty}\Omega_n = \Omega_\infty, \ \mathrm{dlim}_{n\to\infty}\xi_n = \xi_\infty$$

where

$$\xi_\infty = \begin{pmatrix} \Delta_{\infty 1} \\ \kappa_\infty' \Delta_{\infty 2} + \frac{1}{m_\infty}\Delta_{\infty 4} \end{pmatrix},$$

$$\Omega_\infty = \begin{pmatrix} \Gamma_1 & \Gamma_2 \kappa_\infty + \frac{1}{m_\infty}\Delta_{\infty 2} \\ \kappa_\infty' \Gamma_2 + \frac{1}{m_\infty}\Delta_{\infty 2}' & \kappa_\infty' \Gamma_3 \kappa_\infty + \frac{2}{m_\infty}\kappa_\infty' \Delta_{\infty 3} + \frac{1}{m_\infty^2}\Delta_{\infty 5} \end{pmatrix}.$$

In particular, if the exogenous regressors dominate, then

$$\xi_\infty = \begin{pmatrix} \Delta_{\infty 1} \\ \kappa_\infty' \Delta_{\infty 2} \end{pmatrix}, \ \Omega_\infty = \begin{pmatrix} \Gamma_1 & \Gamma_2 \kappa_\infty \\ \kappa_\infty' \Gamma_2 & \kappa_\infty' \Gamma_3 \kappa_\infty \end{pmatrix}, \ \kappa_\infty \neq 0,$$

and if the autoregressive part dominates, then

$$\xi_\infty = \begin{pmatrix} \Delta_{\infty 1} \\ \Delta_{\infty 4} \end{pmatrix}, \ \Omega_\infty = \begin{pmatrix} \Gamma_1 & \Delta_{\infty 2} \\ \Delta'_{\infty 2} & \Delta_{\infty 5} \end{pmatrix}.$$

The proof follows from Lemmas 1 and 2 and comparison of (5.1), (5.2), (5.5) and (5.6). Γ_1, Γ_2, Γ_3 have been defined in (3.4), (5.3) and (5.4), resp. All components of Δ_∞ are obtained from (5.6) with $L = \infty$.

Without loss of generality we can suppose that $h_1, ..., h_k$ are linearly independent and $|\Gamma_1| \neq 0$.

According to the standard results about partitioned matrices one has $|\Omega_\infty| = |\Gamma_1|\pi$ where

$$\pi = \Omega_{\infty 22} - \Omega'_{\infty 12}\Gamma_1^{-1}\Omega_{\infty 12}$$

is different from zero if and only if Ω_∞ is nonsingular; the inverse is

$$\Omega_\infty^{-1} = \begin{pmatrix} \Gamma_1^{-1} + \frac{1}{\pi}EE' & -\frac{1}{\pi}E \\ -\frac{1}{\pi}E' & \frac{1}{\pi} \end{pmatrix}$$

where $E = \Gamma_1^{-1}\Omega_{\infty 12}$. In the next theorem we make one step further by revealing the geometric nature of π in case of dominating exogenous regressors and by showing that $E|\Omega_\infty| \neq 0$ in the general case. Note that $\mathcal{W}(I-\rho\mathcal{W})^{-1}$ is an infinite-dimensional version of $W_n(I-\rho W_n)^{-1}$.

Theorem 2. Let conditions of Theorem 1 be satisfied and suppose that $|\Gamma_1| \neq 0$.

a) If the exogenous regressors dominate, then Ω_∞ is nonsingular if and only if the vector $\mathcal{W}(I - \rho\mathcal{W})^{-1}\kappa'_\infty h$ is linearly independent of $h_1, ..., h_k$. Besides,

$$\pi = \text{dist}^2(\mathcal{W}(I - \rho\mathcal{W})^{-1}\kappa'_\infty h, \mathcal{H})$$

where $\mathcal{H}$ is the linear span of $h_1, ..., h_k$.

b) In the general case, if the vector $\mathcal{W}(I-\rho\mathcal{W})^{-1}\kappa'_\infty h$ is linearly independent of $h_1, ..., h_k$, then

$$E|\Omega_\infty| \geq |\Gamma_1|\text{dist}^2(\mathcal{W}(I - \rho\mathcal{W})^{-1}\kappa'_\infty h, \mathcal{H}) > 0.$$

6 Proofs

Proof of Corollary from Section 3. Suppose Γ_1 is singular. If necessary, we can renumber $h_1, ..., h_k$ in such a way that $h_1, ..., h_l$ will be linearly independent and $h_{l+1}, ..., h_k$ will be their linear combinations:

$$h_j = \sum_{i=1}^{l} c_{ij}h_i, \ j = l + 1, ..., k. \tag{6.1}$$

By the Cramér-Wold theorem, to prove convergence of H'_nV_n in joint distribution, it suffices to prove convergence in distribution of

$$\sum_{j=1}^{l} x_jH'_{nj}V_n + \sum_{j=l+1}^{k} x_jH'_{nj}V_n$$

for all $x \in R^k$. But because of (6.1) this is the same as

$$\begin{gathered}\sum_{j=1}^k x_j(H_{nj} - d_nh_j)'V_n + \sum_{j=1}^l x_j(d_nh_j)'V_n + \sum_{j=l+1}^k x_j \sum_{i=1}^l c_{ij}(d_nh_i)'V_n \\ = \sum_{j=1}^k x_j(H_{nj} - d_nh_j)'V_n + \sum_{j=1}^l \left(x_j + \sum_{i=l+1}^k c_{ji}x_i\right)(d_nh_j)'V_n.\end{gathered}$$

The second sum converges in distribution by Mynbaev CLT. The first sum converges in $L_2(\Omega)$ to zero because by Assumptions 1 and 2

$$\begin{aligned}\|(H_{nj}-d_nh_j)'V_n\|^2_{L_2(\Omega)} &= E[(H_{nj}-d_nh_j)'V_nV_n'(H_{nj}-d_nh_j)] \\ &= \sigma^2\left\|H_{nj}-d_nh_j\right\|_2^2\to 0,\ n\to\infty,\ j=1,...,k.\end{aligned}$$

Proof of Lemma 1. 1) The elements of the matrix $H_n'G_nH_n$ are $H_{nl}'G_nH_{nm}$, $1\le l,m\le k$. For any l,m

$$H_{nl}'G_nH_{nm}=H_{nl}'(s(W_n)-s(d_nW))H_{nm}+H_{nl}'s(d_nW)H_{nm}.$$

Here the first term tends to zero by (3.5) and (4.12):

$$|H_{nl}'(s(W_n)-s(d_nW))H_{nm}|\le c\|H_{nl}\|_2\left\|W_n-d_nW\right\|_2\|H_{nm}\|_2\to 0.$$

For the second term (4.15) gives

$$H_{nl}'s(d_nW)H_{nm}=\sum_{p\ge 0}\rho^p\sum_{i_1,...,i_{p+1}=1}^{\infty}\prod_{j=1}^{p+1}\lambda_{i_j}\mu_{ni}(d_nf_{i_1},H_{nl})_{l_2}(d_nf_{i_{p+1}},H_{nm})_{l_2}.$$

The series converge uniformly because

$$|H_{nl}'s(d_nW)H_{nm}|\le c\sum_{p\ge 0}|\rho|^p\sum_{i_1,...,i_{p+1}=1}^{\infty}|\lambda_{i_1}...\lambda_{i_{p+1}}|=c\sum_{p\ge 0}\left(|\rho|\sum_{i=1}^{\infty}|\lambda_i|\right)^p\sum_{i=1}^{\infty}|\lambda_i|<\infty.$$

Besides, by (4.14) and (3.5) we have element-wise convergence, so

$$\begin{aligned}H_{nl}'s(d_nW)H_{nm} &\to \sum_{p\ge 0}\rho^p\sum_{i_1,...,i_{p+1}=1}^{\infty}\prod_{j=1}^{p+1}\lambda_{i_j}\mu_{\infty i}(f_{i_1},h_l)_{L_2}(f_{i_{p+1}},h_m)_{L_2} \\ &= \sum_{p\ge 0}\rho^p\sum_{i=1}^{\infty}\lambda_i^{p+1}(f_i,h_l)_{L_2}(f_i,h_m)_{L_2} \\ &= \sum_{i=1}^{\infty}\nu(\lambda_i)(f_i,h_l)_{L_2}(f_i,h_m)_{L_2}=(\mathcal{A}h_l,h_m)_{L_2}.\end{aligned}$$

We have taken into account (4.17) and the fact that $\mu_{\infty i}$ vanishes outside the line $i_1=...=i_{p+1}$.

2) As above, we note that $H_n'G_n'G_nH_n$ has $H_{nl}'G_n'G_nH_{nm}$ as its elements and

$$H_{nl}'G_n'G_nH_{nm}=H_{nl}'(G_n'G_n-s^2(d_nW))H_{nm}+H_{nl}'s^2(d_nW)H_{nm}.$$

The first term is estimated using (3.5), (4.12) and (4.13):

$$\begin{aligned}&|H_{nl}'(G_n'G_n-s^2(d_nW))H_{nm}| \\ &\le \|H_{nl}\|_2(\|G_n'-s(d_nW)\|_2\left\|G_n\right\|_2+\left\|s(d_nW)\right\|_2\left\|G_n-s(d_nW)\|_2\right)\|H_{nm}\|_2 \\ &\le c\left\|W_n-d_nW\right\|_2\to 0.\end{aligned}$$

By (4.15) the second term rewrites as

$$H'_{nl}s^2(d_nW)H_{nm}=\sum_{p\geq 0}\rho^p(p+1)\sum_{i_1,\dots,i_{p+2}=1}^{\infty}\prod_{j=1}^{p+2}\lambda_{i_j}\mu_{ni}(d_nf_{i_1},H_{nl})_{l_2}(d_nf_{i_{p+2}},H_{nm})_{l_2}$$

with the series converging uniformly. After letting $n\to\infty$ and applying (4.14), (3.5) and (4.17) we obtain

$$\begin{aligned}H'_{nl}s^2(d_nW)H_{nm} &\to \sum_{p\geq 0}\rho^p(p+1)\sum_{i_1,\dots,i_{p+2}=1}^{\infty}\prod_{j=1}^{p+2}\lambda_{i_j}\mu_{\infty i}(f_{i_1},h_l)_{L_2}(f_{i_{p+2}},h_m)_{L_2}\\ &= \sum_{i=1}^{\infty}\left(\sum_{p\geq 0}\rho^p(p+1)\lambda_i^{p+2}\right)(f_i,h_l)_{L_2}(f_i,h_m)_{L_2}\\ &= \sum_{i=1}^{\infty}\nu^2(\lambda_i)(f_i,h_l)_{L_2}(f_i,h_m)_{L_2}=(\mathcal{A}^2h_l,h_m)_{L_2}.\end{aligned}$$

Proof of Lemma 2, part 1). By Corollary from Section 3 U_{nL} converges in distribution to a normal vector with zero mean and variance-covariance matrix equal to σ^2 times the Gram matrix of the system $h_1,\dots,h_k,f_1,\dots,f_L$. Putting $F_L=(f_1,\dots,f_L)$ and using the usual vector operations we can write that matrix in the form

$$\sigma^2\begin{pmatrix}(h,h')_{L_2} & (h,F'_L)_{L_2}\\ (F_L,h')_{L_2} & (F_L,F'_L)_{L_2}\end{pmatrix}=\begin{pmatrix}\sum_{i=1}^{\infty}(f_i,h)_{L_2}(f_i,h')_{L_2} & (f_1,h)_{L_2} & \dots & (f_L,h)_{L_2}\\ (f_1,h')_{L_2} & 1 & & 0\\ \dots & & \ddots & \\ (f_L,h')_{L_2} & 0 & & 1\end{pmatrix}.$$

If we take a sequence of independent standard normal variables $u_1,u_2,\dots$ and put

$$U_L=\sigma\begin{pmatrix}\sum_{i=1}^{\infty}(f_i,h)_{L_2}u_i\\ u_1\\ \dots\\ u_L\end{pmatrix},$$

it will have the required mean and variance. Hence, $U_{nL}\xrightarrow{d}U_L,\ n\to\infty$. δ_{nL}, being a continuous function of U_{nL}, converges in distribution to the same function of U_L. To obtain (5.7), it suffices to substitute for H'_nV_n and $U_{nL,k+i}$ their limits in distribution.

The component $\sum_{i=1}^{\infty}(f_i,h)_{L_2}u_i$ converges in $L_2(\Omega)$, all others converge, as $L\to\infty$, in $L_1(\Omega)$, due to summability of $\nu(\lambda_i)$ and $\nu^2(\lambda_i)$. This proves (5.8).

Part 2). We have given the required definitions and estimates for the last two components of X_n in Section 4. Therefore here we consider only the first three components of alphas, betas and gammas. Thus, the missing elements of (4.11) are

$$\alpha_n=\begin{pmatrix}\alpha_{n1}\\ \alpha_{n2}\\ \alpha_{n3}\end{pmatrix}=\begin{pmatrix}0\\ H'_n\big(G_n-s(d_nW)\big)V_n\\ H'_n\big(G'_nG_n-s^2(d_nW)\big)V_n\end{pmatrix}$$

(the first block of α_n is zero because Mynbaev CLT is directly applicable to H'_nV_n),

$$\beta_{nL}=\begin{pmatrix}\beta_{nL1}\\ \beta_{nL2}\\ \beta_{nL3}\end{pmatrix}=\begin{pmatrix}0\\ H'_n\big(s(d_nW)-s(d_nW_L)\big)V_n\\ H'_n\big(s^2(d_nW)-s^2(d_nW_L)\big)V_n\end{pmatrix},$$

$$\gamma_{nL}=\begin{pmatrix}\gamma_{nL1}\\ \gamma_{nL2}\\ \gamma_{nL3}\end{pmatrix}=\begin{pmatrix}0\\ H_n's(d_nW_L)V_n\\ H_n's^2(d_nW_L)V_n\end{pmatrix}-\begin{pmatrix}0\\ \delta_{nL2}\\ \delta_{nL3}\end{pmatrix}.$$

Our task is to show that the alphas, betas and gammas are asymptotically negligible.

For any $n\times n$ matrix A_n we can write by Assumption 1 and boundedness of $\|H_{nl}\|_2$

$$\begin{aligned}(E\|H_n'A_nV_n\|_2^2)^{1/2} &= \left[E\left(\sum_{l=1}^k H_{nl}'A_nV_nV_n'A_n'H_{nl}\right)\right]^{1/2}\\ &= \sigma\left(\sum_{l=1}^k H_{nl}'A_nA_n'H_{nl}\right)^{1/2}\\ &\le \sigma\left(\sum_{l=1}^k \|H_{nl}\|_2^2\|A_n\|_2^2\right)^{1/2}\le c\|A_n\|_2.\end{aligned}\tag{6.2}$$

One can use this fact to prove that

$$E\|\alpha_{ni}\|_2^2\to 0,\ n\to\infty,\ i=2,3.\tag{6.3}$$

Indeed, for $i=2$ it suffices to use (4.12), whereas for $i=3$ by (4.12), (4.13) and symmetry of d_nW

$$\begin{aligned}&(E\|\alpha_{n3}\|_2^2)^{1/2}\le c\|G_n'G_n-s^2(d_nW)\|_2\\ &\le c[\|s(W_n')-s(d_nW)\|_2\,\|s(W_n)\|_2+\|s(d_nW)\|_2\,\|s(W_n)-s(d_nW)\|_2]\\ &\le c_1\,\|W_n-d_nW\|_2\to 0.\end{aligned}$$

We claim that

$$\left(E\|\beta_{nLi}\|_2^2\right)^{1/2}\le c\sum_{i>L}|\lambda_i|,\ i=2,3,\tag{6.4}$$

where c does not depend on n,L. The inequality for $i=2$ is an immediate consequence of (6.2) and the second bound in (4.18). For $i=3$ we use (6.2), (4.13) and (4.18):

$$\begin{aligned}\left(E\|\beta_{nL3}\|_2^2\right)^{1/2} &\le c\|s^2(d_nW)-s^2(d_nW_L)\|_2\\ &\le c(\|s(d_nW)\|_2+\|s(d_nW_L)\|_2)\,\|s(d_nW)-s(d_nW_L)\|_2\le c_1\sum_{i>L}|\lambda_i|.\end{aligned}$$

Now we prove that for any $\varepsilon,L>0$ there exists $n_0=n_0(\varepsilon,L)$ such that

$$E|(\gamma_{nLj})_l|\le c\varepsilon,\ n\ge n_0,\ l=1,...,k,\ j=2,3.\tag{6.5}$$

By the first equation in (4.15) and using the vector U_{nL} we have for $l=1,...,k$

$$H_{nl}'s(d_nW_L)V_L=\sum_{p\ge 0}\rho^p\sum_{i_1,...,i_{p+1}\le L}\prod_{j=1}^{p+1}\lambda_{i_j}\mu_{ni}(d_nf_{i_1},H_{nl})_{l_2}U_{nL,k+i_{p+1}}.$$

On the other hand, using the first equation in (4.17) and the definition of $\mu_{\infty i}$ the lth coordinate of δ_{nL2} can be rearranged like this:

$$\begin{aligned}(\delta_{nL2})_l &= \sum_{i=1}^L\sum_{p\ge 0}\rho^p\lambda_i^{p+1}(f_i,h_l)_{L_2}U_{nL,k+i}\\ &= \sum_{p\ge 0}\rho^p\sum_{i_1,...,i_{p+1}\le L}\prod_{j=1}^{p+1}\lambda_{i_j}\mu_{\infty i}(f_{i_1},h_l)_{L_2}U_{nL,k+i_{p+1}}.\end{aligned}$$

The last two equations give the next expression for the lth component of γ_{nL2}:

$$(\gamma_{nL2})_l = \sum_{p\geq 0} \rho^p \sum_{i_1,...,i_{p+1}\leq L} \prod_{j=1}^{p+1} \lambda_{i_j} [\mu_{ni}(d_n f_{i_1}, H_{nl})_{l_2} - \mu_{\infty i}(f_{i_1}, h_l)_{L2}] U_{nL,k+i_{p+1}}.$$

Applying continuity (3.5) and (4.14) we can say that for any $\varepsilon, L > 0$ there exists $n_0 = n_0(\varepsilon, L)$ such that

$$|\mu_{ni}(d_n f_{i_1}, H_{nl})_{l_2} - \mu_{\infty i}(f_{i_1}, h_l)_{L_2}| < \varepsilon, \ n \geq n_0,$$

for all i which enter $(\gamma_{nL2})_l$. Besides, by (4.16)

$$E|U_{nL,k+i_{p+1}}| \leq (E(V_n' d_n f_{i_{p+1}} V_n' d_n f_{i_{p+1}}))^{1/2} \leq c.$$

Hence, the estimate in (6.5) for $j = 2$ follows. The proof for $j = 3$ goes in a similar fashion, except that when dealing with $s^2(d_n W_L)$ one has to use the second equations in (4.15) and (4.17), in place of the first ones.

After these preparatory steps we can conclude the proof of Lemma 2. Due to (4.19), we can apply part 1) of Lemma 2. (4.20) and (6.3) show that $\mathrm{plim}_{n\to\infty} \alpha_n = 0$. From (4.20) and (6.5) we see that $\mathrm{plim}_{n\to\infty} \gamma_{nL} = 0$ for any fixed L. Because of (6.4) and the Chebyshev inequality

$$P\left(\|\beta_{nL2}\|_2 + \|\beta_{nL3}\|_2 > \varepsilon\right) \leq \frac{c}{\varepsilon^2} \sum_{i>L} |\lambda_i|.$$

This bound, (4.21) and (4.11) imply

$$\limsup_{n\to\infty} P(\|X_n - \delta_{nL}\|_2 > \varepsilon) \leq \frac{c}{\varepsilon^2} \sum_{i>L} |\lambda_i|.$$

By Billinsgley's (1968) Theorem 4.2 then the statement of part 2) of Lemma 2 follows.

Proof of Lemma 3. a) is obvious.

b) If $\beta_i \neq 0$. then $\|X_{ni}\|_2 = \kappa_{ni} m_n / \beta_i$. This equation implies b$_1$) and b$_2$).

c) Suppose that (5.10) is true and denote $\varepsilon = 1 - \max_i |\kappa_{\infty i}|$. Then for all large n $m_n = \max\{\|X_{n1}\|_2 |\beta_1|, ..., \|X_{nk}\|_2 |\beta_k|\}$ and $|\kappa_{ni}| = \|X_{ni}\|_2 |\beta_i| / m_n \leq 1 - \varepsilon/2$. This leads to a contradiction: $m_n \leq (1 - \varepsilon/2) m_n$.

d) Let $\kappa_\infty = 0$. If $\beta = 0$, there is nothing to prove. If $\beta \neq 0$, then consider any i such that $\beta_i \neq 0$. By b$_1$) for any such i we have $\|X_{ni}\|_2 = o(m_n)$. This is possible only if $m_n = 1$ for all large n and $\|X_{ni}\|_2 \to 0$. Conversely, if i) is true, then trivially $\kappa_\infty = 0$. If ii) is true, then $m_n = 1$ for all large n and $\kappa_{ni} = \|X_{ni}\|_2 \beta_i \to 0$ for any i such that $\beta_i \neq 0$. Hence, $\kappa_\infty = 0$.

e) If $m_\infty = \infty$, then by c) $\max_i |\kappa_{\infty i}| = 1$ and $\kappa_\infty \neq 0$.

Proof of Theorem 2. a) From Theorem 1 $\pi = \kappa_\infty' \Gamma_3 \kappa_\infty - \kappa_\infty' \Gamma_2 \Gamma_1^{-1} \Gamma_2 \kappa_\infty$. Consider an operator $\mathcal{B} : L_2(0,1) \to l_2$ defined by

$$\mathcal{B} f = ((f, f_1)_{L_2}, (f, f_2)_{L_2}, ...).$$

$\mathcal{B}$ is linear and norm-preserving:

$$\|\mathcal{B} f\|_2 = \left(\sum_{i\geq 1} (f, f_i)_{L_2}^2\right)^{1/2} = \|f\|_2.$$

Therefore $\mathcal{B}h_1, ..., \mathcal{B}h_k$ are linearly independent in l_2. The matrix $G = (\mathcal{B}h_1, ..., \mathcal{B}h_k)$ with infinite square-summable columns can be manipulated as a finite-dimensional matrix. Let $A = \text{diag}[\nu(\lambda_1), \nu(\lambda_2), ...]$ be a diagonal operator in l_2. Then

$$\mathcal{BA}f = \mathcal{B}\left(\sum_{i\geq 1}\nu(\lambda_i)(f, f_i)_{L_2}f_i\right) = (\nu(\lambda_1)(f, f_1)_{L_2}, \nu(\lambda_2)(f, f_2)_{L_2}, ...) = A\mathcal{B}f,$$

that is $\mathcal{BA} = A\mathcal{B}$.

It is easy to see that

$$\Gamma_1 = G'G,\ \Gamma_2 = G'AG,\ \Gamma_3 = G'A^2G$$

and that $P = G(G'G)^{-1}G'$ and $Q = I - P$ are orthoprojectors: $P^2 = P = P'$, $Q^2 = Q = Q'$. Therefore

$$\begin{aligned}\pi &= \kappa'_\infty G'A^2G\kappa_\infty - \kappa'_\infty G'AG(G'G)^{-1}G'AG\kappa_\infty \\ &= \kappa'_\infty G'A(I - G(G'G)^{-1}G')AG\kappa_\infty = \kappa'_\infty G'AQ'QAG\kappa_\infty = \|QAG\kappa_\infty\|_2^2.\end{aligned}$$

Q projects onto the subspace orthogonal to the image $\text{Im}(P)$ and $\|Qx\|_2$ is the distance from x to $\text{Im}(P)$. Thus,

$$\pi = \text{dist}^2(AG\kappa_\infty, \text{Im}(P)).$$

$\text{Im}(P)$ coincides with $\text{Im}(\mathcal{B})$: for any $x \in l_2$ we have $y = (G'G)^{-1}G'x \in R^k$ and

$$Pf = G(G'G)^{-1}G'x = \sum_{l=1}^{k} y_l\mathcal{B}h_l = \mathcal{B}\sum_{l=1}^{k} y_l h_l.$$

From the functional calculus $\mathcal{A} = \mathcal{W}(I - \rho\mathcal{W})^{-1}$ and

$$\begin{aligned}AG\kappa_\infty &= A\sum_{l=1}^{k}\kappa_{\infty l}\mathcal{B}h_l = \sum_{l=1}^{k}\kappa_{\infty l}\mathcal{BA}h_l = \mathcal{B}\left(\sum_{l=1}^{k}\kappa_{\infty l}\mathcal{A}h_l\right) \\ &= \mathcal{B}(\mathcal{W}(I - \rho\mathcal{W})^{-1}\kappa'_\infty h).\end{aligned}$$

Since $\mathcal{B}$ is norm-preserving, we get

$$\begin{aligned}\pi &= \text{dist}^2(\mathcal{B}(\mathcal{W}(I - \rho\mathcal{W})^{-1}\kappa'_\infty h), \text{Im}(\mathcal{B})) \\ &= \text{dist}^2(\mathcal{W}(I - \rho\mathcal{W})^{-1}\kappa'_\infty h, \text{Im}(\mathcal{H})).\end{aligned}$$

b) In the general case

$$\begin{aligned}\pi &= \kappa'_\infty\Gamma_3\kappa_\infty - \kappa'_\infty\Gamma_2\Gamma_1^{-1}\Gamma_2\kappa_\infty + \frac{2}{m_\infty}\kappa'_\infty\Delta_{\infty 3} + \frac{1}{m_\infty^2}\Delta_{\infty 5} \\ &\quad - \frac{2}{m_\infty}\kappa'_\infty\Gamma_2\Gamma_1^{-1}\Delta_{\infty 2} - \frac{1}{m_\infty^2}\Delta'_{\infty 2}\Gamma_1^{-1}\Delta_{\infty 2}.\end{aligned}$$

Since $\Delta_{\infty 2}$ and $\Delta_{\infty 3}$ are linear in normal variables, we have

$$E\pi = \text{dist}^2(\mathcal{W}(I - \rho\mathcal{W})^{-1}\kappa'_\infty h, \mathcal{H}) + \frac{1}{m_\infty^2}E(\Delta_{\infty 5} - \Delta'_{\infty 2}\Gamma_1^{-1}\Delta_{\infty 2}).$$

As $\Delta_{\infty 2}$ and $\Delta_{\infty 5}$ converge in $L_2(\Omega)$ and $L_1(\Omega)$, respectively, we can write

$$E(\Delta_{\infty 5} - \Delta'_{\infty 2}\Gamma_1^{-1}\Delta_{\infty 2}) = \sigma^2 \lim_{L\to\infty} E\zeta_L \text{ where } \zeta_L = \Delta_{L_5} - \Delta'_{L_2}\Gamma_1^{-1}\Delta_{L_2}.$$

Let $\overline{u}_L = (u_1, ..., u_L, 0, ...)$. Then $\Delta_{L_5} = \overline{u}'_L A^2 \overline{u}_L - \overline{u}'_L AG(G'G)^{-1}G'A\overline{u}_L = \|QA\overline{u}_L\|_2^2 \geq 0$. This proves the statement.

Bibliography

Amemiya, T. (1985) Advanced Econometrics. Blackwell, Oxford, UK.

Anderson, T. W. (1971) The Statistical Analysis of Time Series. Wiley, New York.

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

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

Anselin, L. (1992) Space and Applied Econometrics. Anselin, ed. Special Issue, Regional Science and Urban Economics 22.

Anselin, L. and R. Florax (1995) New Directions in Spatial Econometrics. Springer-Verlag, Berlin.

Anselin, L. and S. Rey (1997) Spatial Econometrics. Anselin, L. and S. Rey, ed. Special Issue, International Regional Science Review 20.

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

Cressie, N. (1993) Statistics for Spatial Data. Wiley, New York.

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

Hamilton, J. D. (1994) Time Series Analysis. Princeton University Press, Princeton, New Jersey.

Kelejian, H.H. and I.R. Prucha (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99-121.

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

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

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.

Lee, L.F. (2002) Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252-277.

Lee, L.F. (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, No. 4, 307-335.

Lee, L.F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, No. 6, 1899-1925.

Milbrodt, H. (1992) Testing stationarity in the mean of autoregressive processes with a nonparametric regression trend. Ann. Statist. 20, 1426-1440.

Millar, P. W. (1982) Optimal estimation of a general regression function, Ann. Statist. 10, 717-740.

Moussatat, M. W. (1976) On the Asymptotic Theory of Statistical Experiments and Some of Its Applications. Ph.D. dissertation, Univ. of California, Berkeley.

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.

Mynbaev, K.T. (2006) Asymptotic properties of OLS estimates in autoregressions with bounded or slowly growing deterministic trends. Communications in Statistics 35, No. 3, 499-520.

Mynbaev, K.T. and A. Lemos (2004) Econometrics (in Portuguese). Fundação Getúlio Vargas, Brazil.

Mynbaev, K.T. and I. Castelar (2001) The Strengths and Weaknesses of L_2-approximable Regressors. Two Essays on Econometrics. Expressão Gráfica, Fortaleza, v.1.

Mynbaev, K.T. and Ullah, A. (2006) A remark on the asymptotic distribution of the OLS estimator for a purely autoregressive spatial model. The North American Summer Meetings of the Econometric Society, Minneapolis, MN, June 22 - June 25.

Pötscher, B.M. and I.R. Prucha. (1991) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, Part I: Consistency and approximation concepts. Econometric Reviews **10**, 125-216

Ord, J.K. (1975) Estimation methods for models of spatial interaction. Journal of American Statistical Association 70, 120-126.

Paelinck, J. and L. Klaassen (1979) Spatial Econometrics. Saxon House, Farnborough.

Schmidt, P. (1976) Econometrics. Marcel Dekker, New York and Basel.

Varberg, D.E. (1966) Convergence of quadratic forms in independent random variables. Ann. Math. Stat. 37, 567-576.

Whittle, P. (1954) On stationary processes in the plane. Biometrika 41, 434-449.