OLS Estimator for a Mixed Regressive, Spatial Autoregressive Model: Extended Version

Mynbaev, Kairat

Kazakh-British Technical University

10 May 2009

OLS Estimator for a Mixed Regressive, Spatial Autoregressive Model: Extended Version

Kairat T. Mynbaev

kairat_mynbayev@yahoo.com

International School of Economics, Kazakh-British Technical University, Almaty, Kazakhstan

Abstract

We find the asymptotic distribution of the OLS estimator of the parameters β and ρ in the mixed spatial model with exogenous regressors $Y_n = X_n\beta + \rho W_n Y_n + V_n$. The exogenous regressors may be bounded or growing, like polynomial trends. The assumption about the spatial matrix W_n is appropriate for the situation when each economic agent is influenced by many others. The error term is a short-memory linear process. The key finding is that in general the asymptotic distribution contains both linear and quadratic forms in standard normal variables and is not normal.

Keywords: L_p-approximability, mixed spatial model, OLS asymptotics

1 INTRODUCTION

This paper is an extended version of [16]. In addition to the main statements and proofs, it contains an example, report on computer simulations and GAUSS code.

We study the Ordinary Least Squares (OLS) estimator of parameters β and ρ in the model

$$Y_n = X_n\beta + \rho W_n Y_n + V_n \tag{1}$$

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 $n \times n$ spatial 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.

Model (1), known as a *mixed spatial model*, has been used in situations where possible dependence across spatial units is an issue: in urban, real estate, regional, agricultural and other areas of economics, as well as geostatistics. A good general account of the theory and applications can be found in Paelinck & Klaasen [1], Cliff & Ord [2], Anselin [3], Cressie [4], and Anselin & Florax [5].

A range of estimation techniques for this model has been investigated in the literature: the Maximum Likelihood (ML) and Quasi-Maximum Likelihood (QML), the Method of Moments (MM) and generalized MM, the Least Squares (LS) and Two-Stage Least Squares (2SLS) and the Instrumental Variables (IV) estimator, see Ord [6], Kelejian & Prucha [7; 8], Smirnov & Anselin [9], and Lee [10; 11; 12; 13]. Despite the conceptual and technical differences in approaches, all those authors have been looking for a normal asymptotics.

Recently, Mynbaev & Ullah [14] in a paper devoted to the OLS estimator $\widehat{\rho}$ for the *purely spatial model*

$$Y_n = \rho W_n Y_n + V_n \tag{2}$$

have established two results that are out of line. Firstly, in their Theorem 1 they have proved that $\widehat{\rho} - \rho$ converges in distribution to a ratio of two infinite linear combinations of χ^2-variables:

$$\widehat{\rho} - \rho \xrightarrow{d} \sum_{i \geq 1} u_i^2 \nu_i \left(\sum_{i \geq 1} u_i^2 \nu_i^2 \right)^{-1} \tag{3}$$

where $\{\nu_i : i \in \mathbb{N}\}$ is a summable sequence of real numbers and $\{u_i : i \in \mathbb{N}\}$ is a sequence of independent standard normal variables. The ratio in (3) is not a normal variable and, in general, does not belong to any class of standard tabulated distributions. Secondly, they have shown that QML and MM estimators as developed earlier by Kelejian & Prucha [8] and Lee [10] are not applicable under the conditions of their Theorem 1.

These results raise legitimate concerns as to what may happen in case of a more general mixed model (1). Do linear combinations of χ^2-variables appear in the asymptotic distribution of the OLS estimator $\widehat{\theta}$ of the parameter $\theta = (\beta', \rho)'$? We answer this question affirmatively, and this answer has profound implications for hypothesis testing for the least squares estimation in spatial econometrics. Based on the available theoretical papers, a practitioner assumes that the asymptotic distribution is normal and selects the test statistics accordingly. However, if the asymptotics is not normal and, more generally, is not of any standard type, then all testing procedures should be revisited.

The possibility of a non-normal asymptotics leads to an important issue of the choice of the theoretical assumptions and format of the asymptotic expression. Suppose that one researcher claims that under Set 1 of assumptions the OLS estimator is asymptotically normal, while another researcher develops Set 2 of conditions under which the OLS asymptotics is not normal. Since assumptions in the asymptotic theory usually involve infinite sequences of matrices, the practitioner, based on his/her data, would not be able to choose between Sets 1 and 2. Thus, the assumptions should be sufficiently general and the asymptotic expression should be flexible enough to include, as particular cases, both normal and non-normal asymptotics. Our result satisfies this requirement. Under the same set of conditions, the asymptotic distribution may include or exclude quadratic forms in standard normal variables, due to built-in automatic switches, which reflect the behavior at infinity of the exogenous regressors and spatial matrices.

The method used here relies on the L_p-approximability theory of deterministic regressors developed in Mynbaev [15] (which should be distinguished from the L_p-approximability of stochastic processes defined in Pötscher & Prucha [17]). Under the L_p-approximability assumption, the exogenous regressors and spatial matrices are close to some functions of a continuous argument. The asymptotic distribution is characterized in terms of those functions. This fact allows us to perform what we call *analysis at infinity*: the limits of the elements of the normal equation can be analyzed further to formulate a precise (necessary and sufficient) condition for multicollinearity (or absence thereof). Such a condition is stated using a special function that can be termed a *multicollinearity detector*. A new phenomenon, perhaps specific to spatial econometrics, is that the multicollinearity detector may be in terms of a random function even though the regressors and spatial matrices are deterministic.

The plan of the paper is as follows. Section 2 contains the main statements (Theorems 1 through 4). In Theorems 1 and 2 we improve the method of Mynbaev and Ullah [14] who consider model (2) with independent identically distributed (i.i.d.) errors. In Theorem 1 here we generalize their Theorem 1 on the purely spatial model by allowing the errors to be linear processes of i.i.d. innovations. In Theorem 2 we show that in their Theorem 3 about a new two-step estimator one condition is superfluous. In Theorem 3 we present the asymptotics of the elements of the normal equation for the mixed spatial model. Theorem 4 contains statements regarding the multicollinearity detector and convergence of the OLS estimator in absence of multicollinearity, as well as two particular cases illustrating how good (that is, normal) or bad (non-normal) the asymptotics may be. An example in the end of Section 2 illustrates the theory. All proofs are given in Section 3. Section 4 contains the concluding remarks. The GAUSS code is given in the Appendix.

2 MAIN RESULTS

2.1 Notation and L_2-approximability

A limit in distribution is denoted $\xrightarrow{d}$ or dlim. Likewise, symbols $\xrightarrow{p}$ or plim are used interchangeably for limits in probability. $c, c_1, c_2, \ldots$ denote various inconsequential constants (which do not depend on the variables of interest). Everywhere $u_1, u_2, \ldots$ stand for independent standard normal variables.

$l_2(I)$ denotes the space of sequences $\{x_i : i \in I\}$ provided with the scalar product $(x,y)_{l_2} = \sum_{i\in I} x_i y_i$ and norm $\|x\|_2 = \left(\sum_{i\in I} x_i^2\right)^{1/2}$. The set of indices I depends on the context. Its continuous analog $L_2(0,1)$ consists of square-integrable functions on $(0,1)$ provided with the scalar product $(F,G)_{L_2} = \int_0^1 F(t)G(t)dt$ and norm $\|F\|_2 = \left(\int_0^1 F^2(t)dt\right)^{1/2}$. In case of the space $L_2((0,1)^2)$ of square-integrable functions on $(0,1)^2$ we use the same notation for the scalar product and norm.

The *discretization operator* $\delta_n : L_2(0,1) \to \mathbb{R}^n$ is defined by $(\delta_n F)_i = \sqrt{n}\int_{q_i} F(x)dx$, $i = 1, \ldots, n$, where $q_i = q_i^{(n)} = \left(\frac{i-1}{n}, \frac{i}{n}\right)$ are small intervals that partition $(0,1)$. Let $\{f_n\}$ be a sequence of vectors such that $f_n \in \mathbb{R}^n$ for each n. We say that $\{f_n\}$ is L_2-*approximable* if there exists a function $F \in L_2(0,1)$ such that $\|f_n - \delta_n F\|_2 \to 0,\ n \to \infty$. In this case we also say that $\{f_n\}$ is L_2-*close* to F.

These definitions easily generalize to a two-dimensional case. A 2-D analog of δ_n maps a function $K \in L_2((0,1)^2)$ to an $n \times n$ matrix with elements $(\delta_n K)_{ij} = n\int_{q_{ij}} K(x,y)dxdy, \quad i,j = 1,\ldots,n$, 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 $(0,1)^2$. Let $\{W_n\}$ be a sequence of matrices such that W_n is of size $n \times n$. Then we say that $\{W_n\}$ is L_2-*approximable* if there exists a function $K \in L_2((0,1)^2)$ such that $\|W_n - \delta_n K\|_2 \to 0,\ n \to \infty$. While L_2-approximability is sufficient for our purposes in the 1-D case, in the 2-D case we need to impose a stronger assumption, as the reader will see below.

2.2 Main Assumptions and Statements

Assumption 1 *(on the spatial matrices) There exists a function $K \in L_2((0,1)^2)$ such that (a) $\|W_n - \delta_n K\|_2 = o(1/\sqrt{n})$ and (b) K is symmetric and the eigenvalues* *of the integral operator* $(\mathscr{K}F)(x) = \int_0^1 K(x,y)F(y)\dot{y}$ *are summable:* $\sum_{i=1}^{\infty}|\lambda_i| < \infty$.

This condition implies [14] that the influence of a spatial unit on the others tends to zero and the total interaction among the units tends to infinity: $\max_{i,j}|w_{nij}| \to 0, \quad \sum_{i,j}|w_{nij}| \to \infty, \ n \to \infty$ (here and elsewhere a_{ij} denote the elements of a matrix A). A condition similar to Assumption 1 is widely used in Tanaka [18]. $\mathscr{K}$ is considered an operator in the space $L_2(0,1)$ of square-integrable functions on $(0,1)$. Its eigenvalues λ_i and eigenfunctions F_i are listed according to their multiplicity; the system of eigenfunctions is complete and orthonormal in $L_2(0,1)$. For a symmetric and square-integrable K, its eigenvalues are real and square-summable: $\sum_{i\geq 1}\lambda_i^2 < \infty$. The summability condition is stronger because $\left(\sum\lambda_i^2\right)^{1/2} \leq \sum|\lambda_i|$. Necessary and sufficient conditions (in terms of K) for summability of eigenvalues can be found in Gohberg & Kreĭn [19, Theorem 10.1]. The kernel can be decomposed into the series $K(x,y) = \sum_{i\geq 1}\lambda_i F_i(x)F_i(y)$ which converges in $L_2\left((0,1)^2\right)$. Therefore it can be approximated by its initial segments $K_L(x,y) = \sum_{i=1}^{L}\lambda_i F_i(x)F_i(y)$. This approximation plays an important role in the proof.

In the next assumption $\{e_t\}$ is a sequence of random variables adapted to an increasing sequence of σ-fields $\{\mathscr{F}_t\}$.

Assumption 2 (*on the error term*) (a) *The innovations* $\{e_t : t \in \mathbb{Z}\}$ *are martingale differences with respect to* σ*-fields* $\{\mathscr{F}_t : t \in \mathbb{Z}\}$ (*that is,* $\mathscr{F}_t \subset \mathscr{F}_{t+1}$, e_t *is* $\mathscr{F}_t$*-measurable,* $E(e_t|\mathscr{F}_{t-1}) = 0$ *for all* $t \in \mathbb{Z}$) *satisfying the following higher-order conditions*: *the second-order conditional moments are constant and equal,* $E(e_t^2|\mathscr{F}_{t-1}) = \sigma^2$ *for all* t*, the third conditional moments are constant* (*not necessarily equal*)*,* $E(e_t^3|\mathscr{F}_{t-1}) = c_t$*, and the fourth unconditional moments are uniformly bounded,* $\mu_4 = \sup_t Ee_t^4 < \infty$. (b) *The components of the error term* $V_n = (v_1, ..., v_n)'$ *are linear processes* $v_t = \sum_{j\in\mathbb{Z}}\psi_j e_{t-j}$, $t = 1, ..., n$, *with summable real coefficients* ψ_j.

The summability of ψ_j means that we deal with short-memory processes. For simplicity, the reader can think of $\{e_t\}$ as i.i.d. We denote $\alpha_\psi = \sum_{j\in\mathbb{Z}}|\psi_j| < \infty$, $\beta_\psi = \sum_{j\in\mathbb{Z}}\psi_j$ and $\nu(\lambda) = \frac{\lambda}{1-\rho\lambda}$, $S_n = I_n - \rho W_n$, $G_n = W_n S_n^{-1}$ when S_n^{-1} exists, where I_n is the identity matrix of size $n \times n$. G_n appears in the elements of the normal equation (9) and (10) below.

Mynbaev & Ullah [14] have shown that Assumption 1 and the condition

$$|\rho| < 1/\left(\sum_{i\geq 1}\lambda_i^2\right)^{1/2} \tag{4}$$

are sufficient for the existence of S_n^{-1}. For the asymptotic result they have imposed a stronger condition which we retain here:

Assumption 3 *The spatial parameter satisfies* $|\rho| < 1/\sum_{i\geq 1}|\lambda_i|$.

Theorem 1 *If Assumptions 1, 2 and 3 hold, then the OLS estimator for* (2) *satisfies* (3) *where* $\nu_i = \nu(\lambda_i),\ i \in \mathbb{N}$.

When the components $v_1, ..., v_n$ of V_n are i.i.d., this theorem becomes [14, Theorem 1] where the reader can find the results of Monte Carlo simulations.

Theorem 2 *If* W_n *is symmetric with eigenvalues* $\lambda_{ni},\ i = 1, ..., n$, *then Assumption 1 implies* $\sup_n \sum_{i=1}^n |\lambda_{ni}| < \infty$.

This theorem shows that [14, Assumption 4] is redundant and, therefore, their Theorem 3 (which treats the properties of a new two-step estimator) is applicable under the same conditions on the spatial matrices as in their Theorem 1.

We refer to (2) as *Submodel 1* of the mixed spatial model and to

$$Y_n = X_n\beta + V_n \tag{5}$$

as *Submodel 2*. The constructions below are based on the belief that what is known for Submodels 1 and 2 should be incorporated in the investigation of the main model (1). In the representation of the OLS estimator for (5) Anderson [20, Theorem 2.6.1] uses the matrix with normalized columns $H_n = X_n D_n^{-1}$ where

$$D_n = \mathrm{diag}[\|x_n^{(1)}\|_2, ..., \|x_n^{(k)}\|_2] \tag{6}$$

and $x_n^{(1)}, ..., x_n^{(k)}$ are the columns of X_n. See also Amemiya [21, Theorems 3.5.4, 3.5.5], who has relaxed the assumption on H_n and discussed the advantage of this normalization over the classical $\sqrt{n}$, and Mynbaev & Castelar [22], who have shown this normalization to be superior to any other (in the sense that if any other normalization works, then this one works too).

Assumption 4 (*on exogenous regressors*) (a) *The sequence of columns* $\{h_n^{(l)} : n \in \mathbb{N}\}$ *of the normalized regressor matrices* H_n *is* L_2*-close to* $M_l \in L_2(0,1)$, $l = 1,...,k$. (b) *The functions* $M_1,...,M_k$ *are linearly independent and, consequently, the determinant of the Gram matrix* $\Gamma_0 = \left((M_i,M_j)_{L_2}\right)_{i,j=1}^k$ *is positive, see Gantmacher* [23, Chapter IX, §5].

Assumptions 1, 2 and 4 allow one to prove a variant of Amemiya's [21, Theorem 3.5.4] for the OLS estimator for Submodel 2: $D_n(\widehat{\beta}-\beta) \xrightarrow{d} N(0,(\sigma\beta_\psi)^2\Gamma_0^{-1})$. While the condition on the regressors is stronger, the error term here is more general than in Amemiya's result.

DEFINITION OF THE NORMALIZER OF REGRESSORS. In the mixed spatial model (1) the regressor is $Z_n = (X_n, W_nY_n)$. For the *exogenous regressor* X_n we choose Anderson's normalizer D_n and for the *autoregressive part* W_nY_n we choose

$$d_n = d_n(\beta) = \max\left\{\|x_n^{(1)}\|_2|\beta_1|,...,\|x_n^{(k)}\|_2|\beta_k|,1\right\} \tag{7}$$

based on the analogy with the time series autoregression considered in Mynbaev [24]. Thus, Z_n is normalized by

$$\overline{D}_n = diag[D_n, d_n]. \tag{8}$$

DEFINITION OF THE ELEMENTS OF THE NORMAL EQUATION. The normalized regressor is $\overline{H}_n = Z_n\overline{D}_n^{-1}$ and the normal equation $Z_n'Z_n(\widehat{\theta}-\theta) = Z_n'V_n$ can be rearranged to $\overline{H}_n'\overline{H}_n\overline{D}_n(\widehat{\theta}-\theta) = \overline{H}_n'V_n$. Denoting $\kappa_n = \kappa_n(\beta) = \frac{1}{d_n}D_n\beta$, $\Phi_n = \overline{H}_n'\overline{H}_n$, $\zeta_n = \overline{H}_n'V_n$ we have $\overline{H}_n = \left(H_n, G_nH_n\kappa_n + \frac{1}{d_n}G_nV_n\right)$. The normal equation becomes $\Phi_n\overline{D}_n(\widehat{\theta}-\theta) = \zeta_n$ where

$$\zeta_n = \left((H_n'V_n)', \left(\kappa_n'H_n'G_n'V_n + \frac{1}{d_n}V_n'G_n'V_n\right)'\right)' \tag{9}$$

and Φ_n has the blocks $\Phi_{n11} = H_n'H_n$, $\Phi_{n12} = H_n'G_nH_n\kappa_n + \frac{1}{d_n}H_n'G_nV_n$, $\Phi_{n21} = \Phi_{n12}'$, and

$$\Phi_{n22} = \kappa_n'H_n'G_n'G_nH_n\kappa_n + \frac{2}{d_n}\kappa_n'H_n'G_n'G_nV_n + \frac{1}{d_n^2}V_n'G_n'G_nV_n. \tag{10}$$

According to Lemma 13 and (58) from Section 3, all parts of ζ_n and Φ_n not involving d_n and κ_n converge under Assumptions 1-4. We emphasize that on top of these assumptions we need just one more to prove the remaining results:

Assumption 5 *For all β, the limits $d=\lim_{n\to\infty} d_n \in [1,\infty]$ and $\kappa_i=\lim_{n\to\infty}\kappa_{ni}=\lim_{n\to\infty}\frac{\|x_n^{(i)}\|_2\beta_i}{d_n}\in[-1,1]$ exist .*

DEFINITIONS RELATED TO ANALYSIS AT INFINITY. **(a)** We employ infinite-dimensional matrices A of size $l\times m$ where one or both dimensions can be infinite. Matrices can extend downward or rightward but not upward or leftward. We consider only matrices with finite l_2-norms. Summation and multiplication are performed as usual and preserve this property because $||A+B||_2 \le ||A||_2+||B||_2$, $||AB||_2 \le ||A||_2||B||_2$. The last inequality ensures the validity of the associativity law for multiplication. **(b)** With a function $F\in L_2(0,1)$ we associate a vector $JF\in l_2$ of its Fourier coefficients $JF=((F,F_1)_{L_2},(F,F_2)_{L_2},...)'$. By Parseval's identity $(JF)'JG=(F,G)_{L_2}$ for any functions $F,G\in L_2(0,1)$. Hence, J is an isomorphism from $L_2(0,1)$ to l_2. **(c)** Denote $M=(M_1,...,M_k)'$. By Assumption 4 the matrix $X=(JM_1,...,JM_k)=JM'$ has square-summable and linearly independent columns. $||X||_2<\infty$ because the number of the columns is finite. It is easy to check that the operators

$$P=X(X'X)^{-1}X' \text{ and } Q=I-P \tag{11}$$

are symmetric and idempotent (for the identity operator I we don't use a matrix representation because $||I||_2=\infty$). **(d)** Denote $\nu_J=\mathrm{diag}[\nu(\lambda_1),\nu(\lambda_2),...]$ an infinite-dimensional diagonal matrix. It has been shown in [14] that Assumptions 1 and 3 imply

$$\sum_{i\ge 1}|\nu(\lambda_i)|<\infty, \tag{12}$$

so $||\nu_J||_2<\infty$. **(e)** Denote ν_J^i powers of ν_J and

$$\Gamma_i=X'\nu_J^iX,\ i=1,2,\ u=(u_1,u_2,...)', \tag{13}$$

$$\begin{aligned}\xi &= (\xi_1,\xi_2,\xi_3,\xi_4,\xi_5)'\\ &= |\sigma\beta_\psi|((X'u)',(X'\nu_Ju)',(X'\nu_J^2u)',|\sigma\beta_\psi|u'\nu_Ju,|\sigma\beta_\psi|u'\nu_J^2u)'.\end{aligned} \tag{14}$$

We cannot be sure that $X'u$ belongs to l_2 for every point in the sample space Ω. However, because of Assumption 4(a) the series $X'u$ converges in $L_2(\Omega)$. Similar remarks apply to other components of ξ. Note that $\xi_1,\ \xi_2,\ \xi_3$ are linear in standard normal variables and $\xi_{4,}\ \xi_5$ are quadratic.

Theorem 3 *If Assumptions 1-5 hold, then the limit in distribution*

$$\mathrm{dlim}(\zeta_n,\Phi_n)=(\zeta,\Phi) \tag{15}$$

exists where, with Γ_0 defined in Assumption 4(b),

$$\zeta = \begin{pmatrix} \xi_1 \\ \kappa'\xi_2 + \frac{1}{d}\xi_4 \end{pmatrix}, \ \Phi = \begin{pmatrix} \Gamma_0 & \Gamma_1\kappa + \frac{1}{d}\xi_2 \\ \kappa'\Gamma_1 + \frac{1}{d}\xi_2' & \kappa'\Gamma_2\kappa + \frac{2}{d}\kappa'\xi_3 + \frac{1}{d^2}\xi_5 \end{pmatrix}. \tag{16}$$

For any sequence $\{u_i : i \in \mathbb{N}\}$ of independent standard normal variables denote $\overline{u}_L = (u_1, ..., u_L, 0, ...)'$ and define a *multicollinearity detector* by

$$\Xi = \text{plim}_{L\to\infty} \left\| Q\nu_J \left(X\kappa + \frac{|\sigma\beta_\psi|}{d}\overline{u}_L \right) \right\|_2^2. \tag{17}$$

DEFINITION OF EXTREME CASES. Theorem 1 means that the normalizer for Submodel 1 is unity. On the other hand, Anderson's normalizer for Submodel 2 uses l_2-norms of the columns of X_n. This is why we say that *the exogenous regressors dominate* if $d = \infty$. Since $d_n \geq 1$, it is natural to say that *the autoregressive term dominates* when $\kappa = 0$. Domination of the exogenous regressors and of the autoregressive term are mutually exclusive by Lemma 14(c) from Section 3.

By $\nu(\mathscr{K})$ we denote a function of the integral operator $\mathscr{K}$ defined using its spectral decomposition: if $F = \sum_{i=1}^{\infty}(F, F_i)_{L_2}F_i$, then $\nu(\mathscr{K})F = \sum_{i=1}^{\infty}\nu(\lambda_i)(F, F_i)_{L_2}F_i$. Also let $\mathfrak{M}$ be the linear span of $M_1, ..., M_k$.

Theorem 4 *Let Assumptions 1 through 5 hold. Then*

(a) *the limit* (17) *exists and the condition* $|\Phi| \neq 0$ *a.s. is equivalent to*

$$\Xi > 0 \text{ a.s.} \tag{18}$$

If the last condition is satisfied, then the OLS estimator converges in distribution

$$\overline{D}_n(\widehat{\theta} - \theta) \xrightarrow{d} \Phi^{-1}\zeta. \tag{19}$$

(b) *If the autoregressive term dominates, then*

$$\zeta = \begin{pmatrix} \xi_1 \\ \xi_4 \end{pmatrix}, \ \Phi = \begin{pmatrix} \Gamma_0 & \xi_2 \\ \xi_2' & \xi_5 \end{pmatrix}, \ \Xi = \text{plim}_{L\to\infty}(\sigma\beta_\psi)^2 ||Q\nu_J\overline{u}_L||_2^2. \tag{20}$$

and the condition $\beta_\psi \neq 0$ is necessary for (18).

(c) *If the exogenous regressors dominate, then*

$$\zeta = \begin{pmatrix} \xi_1 \\ \kappa'\xi_2 \end{pmatrix}, \ \Phi = \begin{pmatrix} \Gamma_0 & \Gamma_1\kappa \\ \kappa'\Gamma_1 & \kappa'\Gamma_2\kappa \end{pmatrix}, \ \kappa \neq 0, \tag{21}$$

$$\Xi = \mathrm{dist}^2(\nu(\mathscr{K})\kappa' M, \mathfrak{M}) \tag{22}$$

where $dist(x,\mathfrak{M})$ stands for the distance from point x to the subspace $\mathfrak{M}$. This means that linear independence of $\nu(\mathscr{K})\kappa' M$ and $M_1,...,M_k$ is necessary and sufficient for $|\Phi| \neq 0$. Further, if the constant Ξ is positive, then

$$\overline{D}_n(\widehat{\theta}-\theta) \xrightarrow{d} N(0,(\sigma\beta_\psi)^2\Phi^{-1}). \tag{23}$$

We call (18) an *invertibility criterion*. It is a precise condition under which it is possible to pass from convergence in distribution of the pair (15) to that of the ratio $\Phi_n^{-1}\zeta_n$. Thus, in case of the exogenous regressors domination the quadratic parts disappear from ζ and Φ; Φ becomes nonstochastic and the asymptotic distribution is normal. If the autoregressive term dominates, the linear parts vanish in ζ_2 and Φ_{22}. These are the traces of features of Submodels 1 and 2. None of these extreme cases involve ξ_3, which reflects interaction between the exogenous regressors and spatial lags.

REMARK 1. There are two important issues that will not be considered here in full because of their complexity.

(1) For the purposes of statistical inference, one needs to estimate the variance-covariance matrix of the vector $\Phi^{-1}\zeta$ from Theorem 4. The situation is relatively simple in case of exogenous regressors domination, when Φ is constant, $\Phi_n = \overline{H}_n'\overline{H}_n$ converges to Φ in probability and, hence, Φ_n^{-1} estimates Φ^{-1}. $(\sigma\beta_\psi)^2$ can be estimated by $\Phi_n^{-1}V(\zeta_n)$ (see the end of the proof of Theorem 4). Even in this case there is a problem because $\Phi_n = \overline{H}_n'\overline{H}_n$ depends, through d_n, on unknown β. This problem is partially alleviated by the fact that ζ_n depends on β in the same way. Therefore if some of $\|x_n^{(1)}\|_2,...,\|x_n^{(k)}\|_2$ tend to infinity and, for example, $\|x_n^{(1)}\|_2$ is the largest of these quantities and $\beta_1 \neq 0$, then $d_n = \|x_n^{(1)}\|_2|\beta_1|$ for all large n and the quantities that depend on β_1 in Φ_n and $V(\zeta_n)$ cancel out. If, on the other hand, all of $\|x_n^{(1)}\|_2,...,\|x_n^{(k)}\|_2$ are bounded, then $1 \leq d_n \leq const$, so that dependence on β is weak. In the general case, when Φ is stochastic, there is no simple link between estimates of Φ, $V(\zeta)$ and $V(\Phi^{-1}\zeta)$. At the moment I can suggest no constructive ideas on the matter and invite the profession to think about it.

(2) The second issue is consistency of the OLS estimator. Again, the problem deserves another paper, and only general considerations will be offered. Firstly, for a purely spatial model in [14] we have shown that, because of the presence of quadratic forms in standard normals in the asymptotic distribution, the consistency notion itself should be modified, from $\mathrm{plim}\hat{\rho} = \rho$ to $\mathrm{plim}\hat{\rho} = \rho + X$ where

$EX = 0$. Since the mixed spatial model inherits those quadratic forms, the situation for the problem at hand must be even more complex. Secondly, what is known for Submodel 2, about consistency [21, Theorem 3.5.1] and asymptotic normality [21, Theorem 3.5.4] of the OLS estimator, indicates that consistency and convergence in distribution are two essentially different problems that require different approaches and conditions. That is to say, trying to extract from Theorem 4 conditions sufficient for consistency may not be the best idea. Still, if one wishes to realize it, this is how. The components of $\widehat{\theta} - \theta$ converge with different rates. This can be written as $m_{ni}(\widehat{\theta}_i - \theta_i) \xrightarrow{d} \phi_i$, $i = 1, ..., k+1$, where m_{ni} are normalizing multipliers and ϕ_i are random variables. $m_{ni} \to 0$ means a swelling distribution, so in such cases $\widehat{\theta}_i - \theta_i$ does not converge in probability. If $m_{ni} \to \infty$, then $\widehat{\theta}_i - \theta_i$ behaves as $\frac{1}{m_{ni}}\phi_i$, which goes to 0 in probability. Finally, for i with $m_{ni} \equiv 1$ it suffices to impose conditions providing $\phi_i = 0$. The next example illustrates Theorem 4 and shows, in particular, that consistency can be obtained as its consequence.

Example Denote $l_m = (1, ..., 1)'$ (m unities) and $B_m = (l_m l_m' - I_m)/(m-1)$. The *Case matrices* [25] are defined by $W_n = I_r \otimes B_m$ where $n = rm$. Let us call $\widetilde{W}_n = I_r \otimes (l_m l_m')/(m-1)$ *pseudo-Case matrices*. In the model $Y_n = \beta l_n + \rho W_n Y_n + V_n$ with a constant term and Case matrix W_n the regressors are collinear because $W_n l_n = l_n$ and $Z_n = W_n(l_n, Y_n)$ is of rank at most 2. Therefore we consider $Y_n = \beta l_n + \rho \widetilde{W}_n Y_n + V_n$. The pseudo-Case matrix $\widetilde{W}_n$ satisfies Assumptions 1 and 3 [14, Lemma 7] with $\sum_{i\geq 1} |\lambda_i| \sim 2r - 1$, if r is fixed and $m \to \infty$. For simplicity, the components of the error vector V_n are assumed i.i.d. with mean 0 and variance σ^2. Application of Theorem 4 leads to the following conclusions. The conditions $d = \infty$ (exogenous regressors domination) and $\kappa = 0$ (autoregressive term domination) are mutually exclusive and together cover all possible β.

Table 1. Asymptotic distribution

$\beta = 0$	$\beta \neq 0$
$d = 1$, $\kappa = 0$ (autoregressive term domination)	$d = \infty$, $\lvert\kappa\rvert = 1$ (exogenous regressor domination)
If $r = 1$, there is asymptotic multicollinearity. If $r \geq 2$, $(\sqrt{n}(\hat{\beta} - \beta), \hat{\rho} - \rho) \xrightarrow{p} (0, 1-\rho)$.	For any natural r, there is asymptotic multicollinearity.

In this example ζ and Φ contain quadratic forms of standard normal variables but those forms cancel out in $\Phi^{-1}\zeta$. Still, the asymptotic distribution, when it exists, is not normal. In particular, $\hat{\beta}$ is consistent and $\hat{\rho}$ is not when $\beta = 0$, $r \geq 2$.

Computer simulations confirm the theoretical results. For pseudo-Case matrices the values of m,r have been fixed at $m = 200$, $r = 10$ giving $n = 1000$. Each of the values $\beta = -1,0,1$ was combined with 20 values of ρ from the segment $[-0.2,0.2]$, to see if there is deterioration of convergence at the boundary of the theoretical interval of convergence $\rho \in (-1/19,1/19)$. For each combination (β,ρ) 100 simulations were run. The ranges of sample means and sample standard deviations for the samples of size 100 are reported in the next table (for small values we indicate just the order of magnitude).

Table 2. Simulation results for pseudo-Case matrices

	$\beta=-1$	$\beta=0$	$\beta=1$
mean β	10^{-12}	10^{-12}	10^{-16}
st.d. β	10^{-11}	10^{-11}	10^{-15}
mean ρ	0.995	0.995	0.995
st.d. ρ	10^{-11}	10^{-11}	10^{-14}

As we see, the estimate of $\beta = 0$ is good, as predicted, and the estimates of $\beta = \pm 1$ are not. The estimate of ρ is always bad (closer to 1 than to the true ρ). To see the dynamics of $\hat{\rho}$ as m increases we combined $\beta = 0, \rho = -0.2$ with $m = 200,300,...,1000$. The corresponding values of $\hat{\rho}$ approach 1, starting from 0.9966 and monotonically increasing to 0.999. These simulations did not reveal any deterioration of convergence outside the theoretical interval, suggesting that the convergence may hold in a wider interval. For combinations $\beta = -1,0,1$ with $\rho = 0.2$ the null hypothesis of normal distribution for $\hat{\beta}$ and $\hat{\rho}$ is rejected (the p-value of Anderson-Darling statistic is less than 0.0001).

The results for Case matrices are reported in the following table:

Table 3. Simulation results for Case matrices

	$\beta=-1$	$\beta=0$	$\beta=1$
mean β	$[-1.7,-1.02]$	$[-0.92,-0.17]$	$[-0.008,0.014]$
st.d. β	$[0.54,1.15]$	$[0.57,1.63]$	$[0.02,0.05]$
mean $\hat{\rho}-\rho$	$[-0.82,-0.02]$	$[-0.95,-0.06]$	$[-0.79,-0.03]$
st.d. ρ	$[0.43,1.34]$	$[0.49,1.95]$	$[0.46,2.35]$

There is no definite pattern in these numbers, and for the combination $\beta = 0, \rho = -0.2$ an increase in m from 200 to 1000 did not improve the estimates. Finally, in cases $\beta = \pm 1$ both for the Case and pseudo-Case matrices the sample correlation between the estimates of β and ρ was at least 0.99 in absolute value.

3 PROOFS OF THE MAIN RESULTS

Roadmap. Since the proofs contained herein deviate from the conventional approach in several respects, some ideas are highlighted below. **(a)** One of the main results of [15], stated here as Lemma 5(g), is a general central limit theorem (CLT) for weighted sums of short-memory processes. For the purposes of Lemma 10 it had to be generalized as Lemma 9, to allow $\beta_\psi = 0$ and $|\Gamma_0| = 0$. **(b)** The importance of the operator T_n defined in Lemma 5(d) is explained by the fact that linear processes are defined as convolutions and induce the operator T_n. **(c)** Some results in the asymptotic theory are obtained by what can be termed a *perturbation argument*. Two examples are: (1) Find the asymptotic distribution of the OLS estimator when the normalized regressors are exact images under the discretization mapping δ_n and then extend the result to those regressors that can be approximated by exact images (which takes one to the L_2-approximability notion). (2) Find the asymptotic distribution of the estimator for "good" (say, i.i.d.) errors and then generalize to more flexible specifications (say, linear processes). For the problem under consideration, the second type of perturbation is easier done at the CLT level than when analyzing the OLS estimator. Lemmas 7 and 8 are a part of the corresponding perturbation argument. **(d)** Assumption 1(b) means that the integral operator $\mathscr{K}$ is *nuclear*. The main properties of such operators are taken from [19]. Lemma 5, parts (a), (b), (c) and (h), and Lemmas 11 and 12 provide the necessary links between the operator theory and the theory of functions. **(e)** The study of nonlinear functions $s(W_n)$ and $\nu(\mathscr{K})$ was an important innovation of [14]. Most results of that study are stated as Lemma 6. It is because of them that the present theory is free from any assumptions involving G_n (which are high-level and abound in the previous papers). **(f)** The idea of a double-index approximation, as in (37), has been borrowed by Mynbaev [15] from [26]. It has been employed further in [14] and, in a more ornate manner, here in Lemma 10. Lemma 10 combines convergence to a normal vector, of type [15, Theorem 4.1], with convergence to a quadratic form, of type [15, Theorem 4.2]. **(g)** A limit in distribution to a variable other than normal can be obtained in two ways: (1) by applying a CLT on convergence to a nonnormal distribution, or (2) by combining a CLT on convergence to a normal distribution with the continuous mapping theorem. The second way in the context of spatial models has been followed in [14] and here in Theorems 1 and 3. Lemmas 13 through 18 demonstrate the technical obstacles awaiting an adventurous researcher on this path. See the long explanation before Lemma 15 for details. Notice the multiple applications of Lemma 8, too. After experimenting with variance we found that using Lemma 8 for bounding asymptotically negligible terms is better. **(h)** Since Φ in Theorem 3 may be stochastic, the use of the continuous mapping theorem (CMT) to prove convergence in distribution of $\Phi_n^{-1}\zeta_n$ in Theorem 4 is inevitable. The CMT is applicable when $|\Phi|>0$ a.e. This condition is expressed in an equivalent and somewhat simpler form using the multicollinearity detector. The analysis at infinity has been developed in order to apply matrix algebra to the derivation and study of the multicollinearity detector.

3.1 Convenience List

In this section for the reader's convenience we provide a list of necessary facts. Some of them are pretty simple. 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* $\Delta_n:\mathbb{R}^n\to L_2(0,1)$ takes a vector $x\in\mathbb{R}^n$ to a step function $\Delta_n x=\sqrt{n}\sum_{i=1}^n x_i 1_{q_i}$.

Lemma 5 (a) *Δ_n preserves scalar products*: $(\Delta_n x,\Delta_n y)_{L_2}=(x,y)_{l_2}$ *for all $x,y\in\mathbb{R}^n$ and n.*

(b) *The discretization operators are uniformly bounded: $||\delta_n F||_2\le||F||_2$.*

(c) *The product $\delta_n\Delta_n$ coincides with the identity operator in $\mathbb{R}^n$ and the product $\Delta_n\delta_n$ coincides with the Haar projector defined by $P_nF=n\sum_{i=1}^n\int_{q_i}F(x)dx1_{q_i}$.*

(d) *For the convolution operator $T_n:\mathbb{R}^n\to l_p(\mathbb{Z})$ defined by $(T_nx)_j=\sum_{t=1}^n x_t\psi_{t-j}$, $j\in\mathbb{Z}$, $x\in\mathbb{R}^n$ one has*

$$x'V_n=\sum_{j\in\mathbb{Z}}e_j(T_nx)_j. \tag{24}$$

(e) *If the sequence $\{\psi_j\}$ is summable, then*

$$||T_nx||\le\sqrt{3}\alpha_\psi||x||_2. \tag{25}$$

(f) *If $\{f_n\}$ is L_2-close to F and $\{g_n\}$ is L_2-close to G, then*

$$f_n'g_n\to(F,G)_{L_2},\ ||f_n||_2\to||F||_2. \tag{26}$$

(g) *Under Assumptions 2, 4 and $\beta_\psi\ne0$ one has $H_n'V_n\xrightarrow{d}N(0,(\sigma\beta_\psi)^2\Gamma_0)$.*

(h) *Denote by* δ_n^1 *and* δ_n^2 *the 1-D and 2-D discretization operators, respectively. Then for a product* $F(x,y) = G(x)H(y)$ *one has* $(\delta_n^2 F)_{st} = (\delta_n^1 G)_s (\delta_n^1 H)_t$, *for* $s,t = 1,...,n$.

(i) *Denote* $\mathscr{E}_n : \mathbb{R}^n \to l_2(\mathbb{Z})$ *the extension operator defined by* $\mathscr{E}_n(x_1,...,x_n)' = (...,0,x_1,...,x_n,0...)'$, $x \in \mathbb{R}^n$. *If* $\{f_n\}$ *is* L_2*-close to* $F \in L_2(0,1)$ *and* $\alpha_\psi < \infty$, *then* $||T_n f_n - \beta_\psi \mathscr{E}_n f_n||_2 \to 0$.

Proof. Statements (a)-(c) are straightforward. (d)-(h) have been proved in Mynbaev [15] (see equation (4.13), Theorems 2.2(a), 3.1(b), 3.4(c) and 4.1). (i) follows from Mynbaev [15, Theorem 3.1(c)] and the identity $||T_n f_n - \beta_\psi E_n f_n||_2^2 = \sum_{t=1}^n (T_n f_n - \beta_\psi f_n)_t^2 + \sum_{t<1} (T_n f_n)_t^2 + \sum_{t>n} (T_n f_n)_t^2$. ■

In the following lemma in the beginning of the statements in brackets we provide references to [14].

Lemma 6 (a) [Lemma 6a)] *One has* $\lim_{n\to\infty} \|W_n\|_2 = \lim_{n\to\infty} \|\delta_n K\|_2 = \|K\|_2$ *under Assumption 1.*

(b) [Equations (3.35), (3.36)] *Let Assumption 1 and* (4) *hold. For any square matrix such that* $|\rho|\|A\|_2 < 1$ *put* $s(A) = \sum_{l=0}^\infty \rho^l A^{l+1}$. *Then there exists* $n_0 > 0$ *such that*

$$\|s(W_n) - s(\delta_n K)\|_2 = \left\|s(W_n)' - s(\delta_n K)\right\|_2 \le c \|W_n - \delta_n K\|_2 \ for n \ge n_0, \tag{27}$$

$$\sup_{n\ge n_0} \|s(W_n)\|_2 < \infty, \ \sup_{n\ge n_0} \|s(\delta_n K)\|_2 < \infty. \tag{28}$$

(c) [Lemma 6b)] *Let* $\{F_i\} \subset L_2(0,1)$ *be any orthonormal system and let* $i = (i_1,...,i_{l+1})$ *be a collection of positive integers. Define a chain product* μ_{ni} *by* $\mu_{ni} = \prod_{j=1}^l (\delta_n F_{i_j}, \delta_n F_{i_{j+1}})_{l_2}$ *if* $l > 0$ *and* $\mu_{ni} = 1$ *if* $l = 0$ *and put* $\mu_{\infty i} = 1$ *if* ($i_1 = i_2 = ... = i_{l+1}$ *and* $l > 0$) *or* ($l = 0$) *and* $\mu_{\infty i} = 0$ *otherwise. Then for all* i

$$\lim_{n\to\infty} \mu_{ni} = \mu_{\infty i}. \tag{29}$$

(d) [Section: Estimating gammas] *The functions* μ_{ni} *allow us to write elements of the series* $s(\delta_n K_L)$ *and* $s^2(\delta_n K_L)$ *in a relatively compact form*

$$(s(\delta_n K_L))_{st} = \sum_{p\ge 0} \rho^p \sum_{i_1,...,i_{p+1}\le L} \prod_{j=1}^{p+1} \lambda_{i_j} \mu_{ni} (\delta_n F_{i_1})_s (\delta_n F_{i_{p+1}})_t, \tag{30}$$

$$(s^2(\delta_n K_L))_{st} = \sum_{p\geq 0}\rho^p(p+1)\sum_{i_1,...,i_{p+2}\leq L}\prod_{j=1}^{p+2}\lambda_{i_j}\mu_{ni}(\delta_n F_{i_1})_s(\delta_n F_{i_{p+2}})_t,$$
$$s,t=1,...,n. \tag{31}$$

(e) *For $|\rho\lambda_i|<1$ one has expansions*

$$\nu(\lambda_i)=\sum_{p\geq 0}\rho^p\lambda_i^{p+1},\ \nu^2(\lambda_i)=\sum_{p\geq 0}\rho^p(p+1)\lambda_i^{p+2}. \tag{32}$$

(f) [Equations (3.43), (3.44)] *Under Assumptions 1 and 3 the inequalities*

$$\sup_{n\in\mathbb{N},\ 1\leq L\leq\infty}\|s(\delta_n K_L)\|_2<\infty,\ \sup_n\|s(\delta_n K)-s(\delta_n K_L)\|_2\leq c\sum_{i>L}|\lambda_i|, \tag{33}$$

are true, where c does not depend on L.

(g) [Equation (3.54)] *Under Assumption 3 one has an equivalence*

$$\sum_{i\geq 1}|\lambda_i|<\infty \text{ if and only if } \sum_{i\geq 1}|\nu(\lambda_i)|<\infty. \tag{34}$$

3.2 Proof of Theorem 1

Lemma 7 *Denote $\mu_{pqrs}=Ee_pe_qe_re_s$ for $p,q,r,s\in\mathbb{Z}$. If the m.d. array $\{e_t,\mathscr{F}_t\}$ satisfies conditions stipulated in Assumption 2(a), then $\mu_{pqrs}=\sigma^4$ if $[(p=q)\neq(r=s)]$ or $[(p=r)\neq(q=s)]$ or $[(p=s)\neq(q=r)]$ and $\mu_{pqrs}=Ee_p^4$ if $p=q=r=s$. In all other cases $\mu_{pqrs}=0$.*

Proof. Without loss of generality we can order the indices: $p\leq q\leq r\leq s$. (i) If $s>r$, by the m.d. property $\mu_{pqrs}=E[e_pe_qe_rE(e_s|\mathscr{F}_{s-1})]=0$. (ii) If $s=r>q$, then by orthogonality of m.d.'s and the second moment condition $\mu_{pqrs}=E[e_pe_qE(e_r^2|\mathscr{F}_{r-1})]=\sigma^2Ee_pe_q=\sigma^4$ if $p=q$ and $\mu_{pqrs}=0$, if $p<q$. (iii) If $s=r=q>p$, then $\mu_{pqrs}=E[e_pE(e_q^3|\mathscr{F}_{q-1})]=c_qEe_p=0$. (iv) In case $s=r=q=p$ one has $\mu_{pqrs}=Ee_p^4$.

Because of the assumed ordering the cases $[(p=r)\neq(q=s)]$ and $[(p=s)\neq(q=r)]$ are impossible. The case $[(p=q)\neq(r=s)]$ is covered in (ii), while $p=q=r=s$ is contained in (iv). In "all other cases" $s>r\geq q\geq p$ or $s\geq r\geq q>p$ should be true. The equality of μ_{pqrs} to zero then follows from (i), (ii) and (iii). ■

Lemma 8 *For an $n \times n$ matrix A denote $N(A) = [E(V_n'AV_n)^2]^{1/2}$. Under Assumption 2 for any matrices A, B such that the product AB is of size $n \times n$ one has $N(AB) \le c||A||_2||B||_2$.*

Proof. Denoting $a^1, ..., a^k$ the columns of A and $b_1, ..., b_k$ the rows of B, by (24) we get $(a^l)'V_n = \sum_i e_i (T_n a^l)_i,\ b_l V_n = \sum_j e_j (T_n b_l)_j,\ l = 1, ..., k$. Hence, $V_n'ABV_n = (A'V_n)'BV_n = \sum_{i,j\in\mathbb{Z}} e_i e_j \sum_{l=1}^k (T_n a^l)_i (T_n b_l)_j$. By Lemma 7

$$
\begin{aligned}
E(V_n'ABV_n)^2 &= \sum_{l,m=1}^{k} \sum_{i_1,i_2,j_1,j_2\in\mathbb{Z}} E e_{i_1} e_{i_2} e_{j_1} e_{j_2} (T_n a^l)_{i_1} (T_n b_l)_{j_1} (T_n a^m)_{i_2} (T_n b_m)_{j_2} \\
&= \sigma^4 \sum_{l,m=1}^{k} \sum_{i,j\in\mathbb{Z}} [(T_n a^l)_i (T_n a^m)_i (T_n b_l)_j (T_n b_m)_j \\
&\quad + (T_n a^l)_i (T_n b_l)_i (T_n a^m)_j (T_n b_m)_j \\
&\quad + (T_n a^l)_i (T_n b_m)_i (T_n a^m)_j (T_n b_l)_j] + \sum \\
&= \sigma^4 \sum_{l,m=1}^{k} [(T_n a^l, T_n a^m)_{l_2} (T_n b_l, T_n b_m)_{l_2} \\
&\quad + (T_n a^l, T_n b_l)_{l_2} (T_n a^m, T_n b_m)_{l_2} \\
&\quad + (T_n a^l, T_n b_m)_{l_2} (T_n a^m, T_n b_l)_{l_2}] + \sum
\end{aligned}
$$

where $\sum = \sum_{l,m=1}^k \sum_{i\in\mathbb{Z}} E e_i^4 (T_n a^l)_i (T_n a^m)_i (T_n b_l)_i (T_n b_m)_i$. By the Cauchy-Schwartz inequality, (25) and Assumption 2 $|(T_n x, T_n y)_{l_2}| \le ||T_n x||_2 ||T_n y||_2 \le c_1 ||x||_2 ||y||_2$, $\sup_j |(T_n x)_j| \le ||T_n x||_2 \le c_2 ||x||_2,\ \sup_i E e_i^4 < \infty$. Therefore

$$
\begin{aligned}
E(V_n'ABV_n)^2 &\le c_3 \sum_{l,m=1}^{k} ||a^l||_2 ||a^m||_2 ||b_l||_2 ||b_m||_2 \\
&\quad + c_4 \sum_{l,m=1}^{k} \sup_i |(T_n a^l)_i| \sup_i |(T_n b_l)_i| \sum_i |(T_n a^m)_i (T_n b_m)_i| \\
&\le c_3 \left(\sum_{l=1}^{k} ||a^l||_2 ||b_l||_2 \right)^2 + c_5 \sum_{l,m=1}^{k} ||a^l||_2 ||b_l||_2 ||T_n a^m||_2 ||T_n b_m||_2 \\
&\le c_6 \left(\sum_{l=1}^{k} ||a^l||_2^2 \right)^{1/2} \left(\sum_{l=1}^{k} ||b_l||_2^2 \right)^{1/2} = c_6 ||A||_2^2 ||B||_2^2.
\end{aligned}
$$

■

Lemma 9 *If Assumptions 2 and 4(a) hold, then* $H_n'V_n \xrightarrow{d} N(0,(\sigma\beta_\psi)^2\Gamma_0)$.

Proof. (i) Lemma 5(g) has this statement under the additional conditions $|\Gamma_0| > 0,\ \beta_\psi \neq 0$. Let $|\Gamma_0| > 0,\ \beta_\psi = 0$. By Assumptions 2(a) and 4(a), Lemma 5(i) and (24)

$$E|h_n^{(l)\prime}V_n|^2 = E\left|\sum_{j\in\mathbb{Z}} e_j(T_nh_n^{(l)})_j\right|^2 = \sigma^2\sum_{j\in\mathbb{Z}}|(T_nh_n^{(l)})_j|^2 = \sigma^2||T_nh_n^{(l)}||_2^2 \to 0.$$

Hence, $H_n'V_n \xrightarrow{d} 0$ and $H_n'V_n \xrightarrow{d} N(0,(\sigma\beta_\psi)^2\Gamma_0)$.

(ii) We can let β_ψ be arbitrary and consider linearly dependent $M_1,...,M_k$. We can number these functions in such a way that the first l are independent and the last $k-l$ are their linear combinations:

$$M_j = \sum_{i=1}^{l} c_{ji}M_i,\ j = l+1,...,k. \tag{35}$$

(35) implies

$$S = CR,\ M = \begin{pmatrix} R \\ CR \end{pmatrix},\ R \equiv \begin{pmatrix} M_1 \\ \dots \\ M_l \end{pmatrix},\ S \equiv \begin{pmatrix} M_{l+1} \\ \dots \\ M_k \end{pmatrix},\ C \equiv \begin{pmatrix} c_{l+1,1} & \dots & c_{l+1,l} \\ \dots & \dots & \dots \\ c_{k,1} & \dots & c_{k,l} \end{pmatrix}$$

$$\Gamma_0 = (M,M')_{L_2} = \int_0^1 \begin{pmatrix} RR' & RR'C' \\ CRR' & CRR'C' \end{pmatrix} \underset{\cdot}{x} = \begin{pmatrix} \Gamma_R & \Gamma_RC' \\ C\Gamma_R & C\Gamma_RC' \end{pmatrix}$$

where $\Gamma_R = (R,R')_{L_2}$ is the Gram matrix of R. The proof will be complete if we show that $H_n'V_n$ converges in distribution to a normal vector with variance of this structure. With $r_n = (h_n^1,...,h_n^l),\ s_n = (h_n^{l+1},...,h_n^k)$ the matrix H_n is partitioned accordingly into $H_n = (r_n,s_n)$. Put $\widetilde{s}_n = r_nC',\ \widetilde{H}_n = (r_n,\widetilde{s}_n)$. Then

$$\widetilde{H}_n'V_n = \begin{pmatrix} r_n'V_n \\ \widetilde{s}_n'V_n \end{pmatrix} = \begin{pmatrix} r_n'V_n \\ Cr_n'V_n \end{pmatrix} = \begin{pmatrix} I \\ C \end{pmatrix} r_n'V_n.$$

By part (i) of this proof $r_n'V_n \xrightarrow{d} U$ with U distributed as $N(0,(\sigma\beta_\psi)^2\Gamma_R)$. Therefore $\widetilde{H}_n'V_n \xrightarrow{d} N(0,(\sigma\beta_\psi)^2\Gamma_0)$.

Now it suffices to show that $\text{plim}(H_n'V_n - \widetilde{H}_n'V_n) = 0$ or, since the first blocks in these vectors are the same, that $\text{plim}(s_n'V_n - Cr_n'V_n) = 0$. The jth component of

$s_n'V_n - Cr_n'V_n$ is $(h_n^{(j)\prime} - c_j r_n')V_n,\ j = l+1, ..., k$, where c_j denotes the jth row of C. By Hölder's inequality and Lemma 8

$$\begin{aligned} E|(h_n^{(j)\prime} - c_j r_n')V_n|^2 &= EV_n'(h_n^{(j)\prime} - c_j r_n')'(h_n^{(j)\prime} - c_j r_n')V_n \\ &\le \{E[V_n'(h_n^{(j)\prime} - c_j r_n')'(h_n^{(j)\prime} - c_j r_n')V_n]^2\}^{1/2} \\ &= N((h_n^{(j)\prime} - c_j r_n')'(h_n^{(j)\prime} - c_j r_n')) \le c||h_n^{(j)} - r_n c_j'||_2^2. \end{aligned}$$

Applying the discretization operator to both sides of (35) we get

$$\begin{aligned} ||h_n^{(j)} - r_n c_j'||_2 &\le ||h_n^{(j)} - \delta_n M_j||_2 + ||\delta_n M_j - \sum c_{ji} h_n^{(i)}||_2 \\ &\le ||h_n^{(j)} - \delta_n M_j||_2 + \sum |c_{ji}|||\delta_n M_i - h_n^{(i)}||_2 \to 0 \end{aligned}$$

by the L_2-approximability assumption. ■

For any natural n, L denote

$$U_{nL} = \begin{pmatrix} H_n' V_n \\ (\delta_n F_1)' V_n \\ \cdots \\ (\delta_n F_L)' V_n \end{pmatrix},\ X_{nL} = \begin{pmatrix} H_n' V_n \\ \sum_{i=1}^L \nu(\lambda_i)(M, F_i)_{L_2} U_{nL,k+i} \\ \sum_{i=1}^L \nu^2(\lambda_i)(M, F_i)_{L_2} U_{nL,k+i} \\ \sum_{i=1}^L \nu(\lambda_i) U_{nL,k+i}^2 \\ \sum_{i=1}^L \nu^2(\lambda_i) U_{nL,k+i}^2 \end{pmatrix}. \tag{36}$$

The limiting behavior of X_{nL}, as $n \to \infty$, is described in terms of

$$\xi_L = |\sigma \beta_\psi| \begin{pmatrix} \sum_{i=1}^\infty (M, F_i)_{L_2} u_i \\ \sum_{i=1}^L \nu(\lambda_i)(M, F_i)_{L_2} u_i \\ \sum_{i=1}^L \nu^2(\lambda_i)(M, F_i)_{L_2} u_i \\ |\sigma \beta_\psi| \sum_{i=1}^L \nu(\lambda_i) u_i^2 \\ |\sigma \beta_\psi| \sum_{i=1}^L \nu^2(\lambda_i) u_i^2 \end{pmatrix},\ 1 \le L < \infty.$$

Lemma 10 *Let Assumptions 1, 2 and 4 hold and let* (12) *be true. Then*

$$\text{dlim}_{n\to\infty} X_{nL} = \xi_L \ for\ all\ L < \infty,\ \text{plim}_{L\to\infty} \xi_L = \xi. \tag{37}$$

Proof. The matrix $(H_n,\ \delta_n F_1,\ ...,\ \delta_n F_L)$ satisfies all conditions of Lemma 9. Therefore U_{nL} converges in distribution to a normal vector with zero mean and variance $(\sigma\beta_\psi)^2\Gamma$ where Γ is the Gram matrix of the system $\{M_1, ..., M_k, F_1, ..., F_L\}$. Putting $F^{(L)} = (F_1, ..., F_L)'$ we see that

$$\Gamma = \left(\Gamma_{ij}\right)_{i,j=1}^2,\ \Gamma_{11} = (M, M')_{L_2} = \sum_{i=1}^\infty (M, F_i)_{L_2} (M', F_i)_{L_2},$$

$$\Gamma_{12} = (M, F^{(L)\prime})_{L_2} = ((M,F_1)_{L_2}, ..., (M,F_L)_{L_2}),\ \Gamma_{21} = \Gamma_{12}',$$
$$\Gamma_{22} = (F^{(L)}, F^{(L)\prime})_{L_2} = I$$

(the upper left block is obtained by Parseval's identity; the lower right block is a consequence of orthonormality of $\{F_i\}$). If we take a sequence of independent standard normals u_i and define U_L by $U_L = |\sigma\beta_\psi| (\sum_{i=1}^{\infty}(M', F_i)_{L_2} u_i, u_1, ..., u_L)'$, then it will be normal, have zero mean and variance Γ. Hence, $U_{nL} \xrightarrow{d} U_L$. X_{nL}, being a continuous function of U_{nL}, converges in distribution to the same function of U_L. Keeping in mind that the relationship $H_n' V_n \xrightarrow{d} N(0, (\sigma\beta_\psi)^2 (M, M')_{L_2})$ is equivalent to $H_n' V_n \xrightarrow{d} |\sigma\beta_\psi| \sum_{i=1}^{\infty}(M, F_i)_{L_2} u_i$ and that $U_{nL,k+i} \xrightarrow{d} u_i$ we get the first equation in (37). The second equation is obvious because all components of ξ_L converge to those of ξ in $L_1(\Omega)$ and in probability. ■

Proof of Theorem 1 Our Theorem 1 is proved by making in the proof of [14, Theorem 1] the following changes: (a) Instead of their δ_{nL} and Δ_L use the last two components of X_{nL} and ξ_L: $\delta_{nL} = \sum_{i=1}^{L} U_{nL,k+i}^2 a_i$, $\Delta_L = (\sigma\beta_\psi)^2 \sum_{i=1}^{L} u_i^2 a_i$ where $a_i = \mathrm{diag}[\nu(\lambda_i), \nu^2(\lambda_i)]$. (b) Instead of their Lemma 5d) apply our Lemma 8. (c) Their U_{nL} corresponds to the vector of the last L coordinates of our U_{nL}. (d) Replace their Lemma 7 by our Lemma 10. (e) Minor changes in estimating β_{nL}, γ_{nL} are necessary. Instead of describing those changes we refer to Lemmas 17 and 18 below where the proofs are given in full. ▮

3.3 Proof of Theorem 2

Lemma 11 *Let P_n, Δ_n^1 and δ_n^1 be the 1-D Haar projector, interpolation operator and discretization operator, respectively. Denote δ_n^2 the 2-D discretization operator and $\mathscr{K}_n = P_n \mathscr{K} P_n$. Then $\mathscr{K}_n = \Delta_n^1(\delta_n^2 K)\delta_n^1$.*

Proof. By Lemma 5(c) we have $\mathscr{K}_n = \Delta_n^1(\delta_n^1 \mathscr{K} \Delta_n^1)\delta_n^1$ and the statement will follow if we establish $\delta_n^1 \mathscr{K} \Delta_n^1 = \delta_n^2 K$. The definitions give:

$$[\delta_n^1(\mathscr{K}\Delta_n^1 f)]_s = \left[\delta_n^1 \int_0^1 K(\cdot,y)(\Delta_n^1 f)(y)\underset{\cdot}{\mathrm{y}}\right]_s$$
$$= \sqrt{n}\sum_{t=1}^{n}\left[\delta_n^1 \int_{q_t} K(\cdot,y)\underset{\cdot}{\mathrm{y}}\right]_s f_t = n\sum_{t=1}^{n}\int_{q_s}\left(\int_{q_t} K(x,y)\underset{\cdot}{\mathrm{y}}\right)\underset{\cdot}{\mathrm{x}} f_t$$
$$= n\sum_{t=1}^{n}\int_{q_{st}} K(x,y)\underset{\cdot}{\mathrm{x}}\underset{\cdot}{\mathrm{y}} f_t = [(\delta_n^2 K)f]_s,\ s = 1,...,n.$$

Since f is arbitrary, this proves the desired equation. $\blacksquare$

Lemma 12 *If K is symmetric and integrable, then the point spectrum of $\mathscr{K}_n$ coincides with that of $\delta_n^2 K$, meaning that their nonzero eigenvalues, repeated according to their multiplicities, are the same.*

Proof. If $\mathscr{K}_n F = \lambda F$, then by Lemmas 5(c) and 11 one can see that $\lambda \delta_n^1 F = \delta_n^1 \mathscr{K}_n F = (\delta_n^1 \Delta_n^1)(\delta_n^2 K)\delta_n^1 F = (\delta_n^2 K)\delta_n^1 F$ and, hence, $f = \delta_n^1 F$ is an eigenvector of $\delta_n^2 K$ corresponding to λ. Conversely, if $(\delta_n^2 K)f = \lambda f$, then we multiply both sides by Δ_n^1 and use Lemma 5(c) to substitute $f = \delta_n^1 \Delta_n^1 f$: $\Delta_n^1(\delta_n^2 K)\delta_n^1 \Delta_n^1 f = \lambda \delta_n^1 f$. Application of Lemma 11 shows that $F = \Delta_n^1 f$ is an eigenvector of $\mathscr{K}_n$ corresponding to λ. In both cases, when we go from $\mathscr{K}_n$ to $\delta_n^2 K$ and back, if G (or $g = \delta_n^1 G$) is another eigenvector corresponding to the same eigenvalue λ and orthogonal to F (or f, respectively), orthogonality is preserved by Lemma 5(a). This implies preservation of multiplicities. Note that since the image of $\mathscr{K}_n$ is finite-dimensional, the subspace of eigenvectors of $\mathscr{K}_n$ corresponding to $\lambda = 0$ is infinite-dimensional and mapped to $\{0\}$ by δ_n^1. $\blacksquare$

Proof of Theorem 2 We need a series of definitions and facts from Gohberg & Kreĭn [19]. Let A be a compact linear operator in a Hilbert space H. If A is self-adjoint, then $\{\lambda_n(A) : n \in \mathbb{N}\}$ denotes its sequence of eigenvalues counted with their multiplicities and numbered in the order of decreasing absolute values. If A is not necessarily self-adjoint, then its s-numbers are defined by $s_n(A) = \lambda_n((A^*A)^{1/2})$. The expression $||A||_{\sigma_1} = \sum_{j=1}^{\infty} s_j(A)$ is a norm [19, p.92]. For self-adjoint operators $s_j(A) = |\lambda_j(A)|$ [19, p.27], so

$$||A||_{\sigma_1} = \sum_{j=1}^{\infty} |\lambda_j(A)|. \tag{38}$$

For any bounded operators B and C

$$s_j(BAC) \le ||B|| s_j(A) ||C|| \tag{39}$$

[19, §2.1]. If for some orthonormal basis $\{\phi_j\}$ one has $\sum_j ||A\phi_j|| < \infty$, then $||A||_{\sigma_1} \le \sum_j ||A\phi_j||$ [19, §7.8]. If A is a square matrix of order n, then plugging the jth unit vector $\phi_j = (0,...,0,1,0,...,0)' \in \mathbb{R}^n$ in the last inequality produces $A\phi_j = (a_{1j},...,a_{nj})'$ (jth column) and

$$||A||_{\sigma_1} \le \sum_j ||(a_{1j},...,a_{nj})||_2 \le \sqrt{n}||A||_2. \tag{40}$$

Now we can proceed with the proof. $\mathscr{K}_n$ is self-adjoint because $\mathscr{K}$ and P_n are. By Lemma 12 $||\mathscr{K}_n||_{\sigma_1} = ||\delta_n^2 K||_{\sigma_1}$. Hence, by (38), (40) and Assumption 1

$$\left| \sum_{i=1}^{n} |\lambda_{ni}| - ||\mathscr{K}_n||_{\sigma_1} \right| = \left| ||W_n||_{\sigma_1} - ||\delta_n^2 K||_{\sigma_1} \right|$$
$$\leq ||W_n - \delta_n^2 K||_{\sigma_1} \leq \sqrt{n}||W_n - \delta_n^2 K||_2 \to 0.$$

Since $||P_n|| \leq 1$ by Lemma 5, parts (a) and (b), bound (39) and Assumption 1 lead to

$$||\mathscr{K}_n||_{\sigma_1} = \sum_{j=1}^{\infty} s_j(P_n \mathscr{K} P_n) \leq \sum_{j=1}^{\infty} s_j(\mathscr{K}) = \sum_{j=1}^{\infty} |\lambda_j(\mathscr{K})| < \infty.$$

The last two displayed equations prove the theorem. ∎

3.4 Proof of Theorem 3

Lemma 13 *If Assumptions 1, 3 and 4(a) hold, then* (a) $\lim_{n\to\infty} H_n' H_n = \Gamma_0$, (b) $\lim_{n\to\infty} H_n' G_n H_n = \Gamma_1 = \lim_{n\to\infty} H_n' G_n' H_n$, (c) $\lim_{n\to\infty} H_n' G_n' G_n H_n = \Gamma_2$.

Proof. (a) directly follows from Assumption 4(a), (26) and the definition of X:

$$\lim_{n\to\infty} (H_n' H_n)_{lm} = \lim_{n\to\infty} h_n^{(l)\prime} h_n^{(m)} = \int_0^1 M_l(x) M_m(x) \underset{\cdot}{x}$$
$$= \sum_{j=1}^{\infty} (M_l, F_j)_{L_2} (M_m, F_j)_{L_2} = (X'X)_{lm}.$$

(b) The elements of the matrix $H_n' G_n H_n$ are $h_n^{(l)\prime} G_n h_n^{(m)}$, $1 \leq l, m \leq k$. For any l, m $h_n^{(l)\prime} G_n h_n^{(m)} = h_n^{(l)\prime} [s(W_n) - s(\delta_n K)] h_n^{(m)} + h_n^{(l)\prime} s(\delta_n K) h_n^{(m)}$. Here the first term tends to zero by Assumption 1(a), (27) and (26): $|h_n^{(l)\prime} [s(W_n) - s(\delta_n K)] h_n^{(m)}| \leq c \|h_n^{(l)}\|_2 \|W_n - \delta_n K\|_2 \|h_n^{(m)}\|_2 \to 0$. For the second term (30) with $L = \infty$ gives

$$h_n^{(l)\prime} s(\delta_n K) h_n^{(m)} = \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\ldots,i_{p+1}=1}^{\infty} \prod_{j=1}^{p+1} \lambda_{i_j} \mu_{ni} (\delta_n F_{i_1}, h_n^{(l)})_{l_2} (\delta_n F_{i_{p+1}}, h_n^{(m)})_{l_2}.$$

Here the series converge uniformly in l, m, n because the scalar and chain products μ_{ni} are uniformly bounded and (see Lemmas 5(b) and 5(f) and Assumption 3)

$$|h_n^{(l)\prime} s(\delta_n K) h_n^{(m)}| \leq c \sum_{p=0}^{\infty} |\rho|^p \sum_{i_1,\ldots,i_{p+1}=1}^{\infty} |\lambda_{i_1} \ldots \lambda_{i_{p+1}}|$$

$$= c\sum_{p=0}^{\infty}\left(|\rho|\sum_{i=1}^{\infty}|\lambda_i|\right)^p\sum_{i=1}^{\infty}|\lambda_i| < \infty.$$

Besides, by (29) and (26) we have element-wise convergence, so

$$h_n^{(l)\prime}s(\delta_n K)h_n^{(m)} \to \sum_{p=0}^{\infty}\rho^p\sum_{i_1,\dots,i_{p+1}=1}^{\infty}\prod_{j=1}^{p+1}\lambda_{i_j}\mu_{\infty i}(F_{i_1},M_l)_{L_2}(F_{i_{p+1}},M_m)_{L_2}$$

($\mu_{\infty i}$ vanishes outside the diagonal $i_1 = \dots = i_{p+1}$)

$$= \sum_{p=0}^{\infty}\rho^p\sum_{i=1}^{\infty}\lambda_i^{p+1}(F_i,M_l)_{L_2}(F_i,M_m)_{L_2}$$

(using (32) and the definition of X)

$$= \sum_{i=1}^{\infty}\nu(\lambda_i)(F_i,M_l)_{L_2}(F_i,M_m)_{L_2} = (X'\nu_J X)_{lm}.$$

We have proved the first equation in (b). The second equation in (b) follows from the first equation and the fact that $|h_n^{(l)\prime}(G_n - G_n')h_n^{(m)}| \le \|h_n^{(l)}\|_2(\|G_n - s(\delta_n K)\|_2 + \|G_n' - s(\delta_n K)\|_2)\|h_n^{(m)}\|_2 \to 0$ (recall that $G_n = s(W_n)$). ■

The next lemma partially answers the question of what Assumption 5 means in terms of the regressors and β.

Lemma 14 *Under Assumption 5 the following is true*: **(a)** *If* $\beta_i = 0$*, then* $x_n^{(i)}$ *is arbitrary and* $\kappa_i = 0$*.* **(b)** *Let* $\beta_i \ne 0$*. Then* (b$_1$) $\kappa_i = 0$ *is equivalent to* $\|x_n^{(i)}\|_2 = o(d_n)$ *and* (b$_2$) $\kappa_i \ne 0$ *is equivalent to* $\|x_n^{(i)}\|_2/d_n \to c_i > 0$ *where* c_i *is some constant.* **(c)** *Conditions*

$$\max_i |\kappa_i| < 1 \text{ and } d > 1 \tag{41}$$

are mutually exclusive. In particular, conditions $\kappa = 0$ *and* $d = \infty$ *are mutually exclusive.* **(d)** $\kappa = 0$ *if and only if either* (d$_1$) $\beta = 0$ *or* (d$_2$) $\beta \ne 0$ *and* $\lim_{n\to\infty}\|x_n^{(i)}\|_2 = 0$ *for any* i *such that* $\beta_i \ne 0$*. In either case* $d_n = 1$ *for all large* n *and* $d = 1$*.*

Proof. **(a)** is obvious. **(b)** If $\beta_i \neq 0$, then $\|x_n^{(i)}\|_2 = \kappa_{ni} d_n / \beta_i$. This equation implies (b_1) and (b_2). **(c)** Suppose that (41) is true and denote $\varepsilon = 1 - \max_i |\kappa_i|$. Then for all large n one has $d_n = \max\{\|x_n^{(1)}\|_2 |\beta_1|, ..., \|x_n^{(k)}\|_2 |\beta_k|\} > 1$ and $|\kappa_{ni}| = \|x_n^{(i)}\|_2 |\beta_i| / d_n \leq 1 - \varepsilon/2$. This leads to a contradiction: $d_n \leq (1-\varepsilon/2) d_n$. **(d)** Let $\kappa = 0$. 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 $\|x_n^{(i)}\|_2 = o(d_n)$. By (c) the assumption $\kappa = 0$ excludes the possibility $d = \infty$. Hence, $d < \infty$ and $\|x_n^{(i)}\|_2 = o(d_n)$ is equivalent to $\|x_n^{(i)}\|_2 = o(1)$. Since this is true for any i with $\beta_i \neq 0$, we have $d_n = 1$ for all large n and, consequently, $d = 1$. We have proved (d_2). Conversely, if (d_1) is true, then trivially $\kappa = 0$. If (d_2) is true, then $d_n = 1$ for all large n and $\kappa_{ni} = \|x_n^{(i)}\|_2 \beta_i \to 0$ for any i such that $\beta_i \neq 0$. Hence, $\kappa = 0$. ■

Now we define auxiliary random vectors used in the proof of Theorem 3. All random components contained in ζ_n (9) and Φ_n (10) are dumped into one vector

$$\begin{aligned}\mathscr{A}_n &= ((\mathscr{A}_{n1})', (\mathscr{A}_{n2})', (\mathscr{A}_{n3})', \mathscr{A}_{n4}, \mathscr{A}_{n5})' \\ &= ((H_n' V_n)', (H_n' G_n V_n)', (H_n' G_n' G_n V_n)', V_n' G_n' V_n, V_n' G_n' G_n V_n)'.\end{aligned}$$

$H_n' G_n' V_n$ (which is a part of ζ_n) is not included because $\mathrm{plim}(H_n' G_n V_n - H_n' G_n' V_n) = 0$, see Lemma 15(b) below. The first three components of $\mathscr{A}_n$ are $k \times 1$ and linear in V_n, whereas the last two are (scalar) quadratic forms of V_n.

$\mathscr{A}_n$ is represented as

$$\mathscr{A}_n = \alpha_n + \beta_{nL} + \gamma_{nL} + X_{nL} \tag{42}$$

where the vectors at the right-hand side have blocks conformable with those of $\mathscr{A}_n$. X_{nL} has been defined in (36) and represents the main part of $\mathscr{A}_n$. The other three vectors will be shown to be negligible in some sense and are defined by

$$\alpha_n = \begin{pmatrix} 0 \\ H_n'[G_n - s(\delta_n K)]V_n \\ H_n'[G_n' G_n - s^2(\delta_n K)]V_n \\ V_n'[G_n' - s(\delta_n K)]V_n \\ V_n'[G_n' G_n - s^2(\delta_n K)]V_n \end{pmatrix}, \ \beta_{nL} = \begin{pmatrix} 0 \\ H_n'[s(\delta_n K) - s(\delta_n K_L)]V_n \\ H_n'[s^2(\delta_n K) - s^2(\delta_n K_L)]V_n \\ V_n'[s(\delta_n K) - s(\delta_n K_L)]V_n \\ V_n'[s^2(\delta_n K) - s^2(\delta_n K_L)]V_n \end{pmatrix}, \ \gamma_{nL} = \begin{pmatrix} 0 \\ H_n' s(\delta_n K_L) V_n \\ H_n' s^2(\delta_n K_L) V_n \\ V_n' s(\delta_n K_L) V_n \\ V_n' s^2(\delta_n K_L) V_n \end{pmatrix} - \begin{pmatrix} 0 \\ X_{nL2} \\ X_{nL3} \\ X_{nL4} \\ X_{nL5} \end{pmatrix}. \tag{43}$$

Intuitive explanations: for α_n, if W_n is close to $\delta_n K$, then $G_n' = s(W_n')$ and $G_n' G_n = s(W_n') s(W_n)$ should be close to $s(\delta_n K)$ and $s^2(\delta_n K)$, resp.; the definition of β_n reflects approximation of K by its segments, and γ_n is a small correction needed to obtain a continuous function of an asymptotically normal vector. In α_n, β_{nL}, γ_{nL} the first blocks are null because Lemma 9 is directly applicable to the first block of $\mathscr{A}_n$.

Lemma 15 (a) *For any $n \times n$ matrix A_n one has* $(E\|H_n' A_n V_n\|_2^2)^{1/2} \le c\|H_n\|_2 \|A_n\|_2$ *provided that Assumption 2 is met.* (b) *If, additionally, Assumptions 1 and 4*(*a*) *hold and ρ satisfies* (4), *then* $\operatorname{plim}(H_n' G_n V_n - H_n' G_n' V_n) = 0$.

Proof. (a) Use the partition of H_n into its columns:

$$E\|H_n' A_n V_n\|_2^2 = E\sum_{l=1}^{k}\left(h_n^{(l)\prime} A_n V_n\right)^2 = \sum_{l=1}^{k} E V_n' A_n' h_n^{(l)} h_n^{(l)\prime} A_n V_n$$

(apply Hölder's inequality and Lemma 8)

$$\begin{aligned} &\le \sum_{l=1}^{k}[E(V_n' A_n' h_n^{(l)} h_n^{(l)\prime} A_n V_n)^2]^{1/2} = \sum_{l=1}^{k} N(A_n' h_n^{(l)} h_n^{(l)\prime} A_n) \\ &\le c\sum_{l=1}^{k} ||A_n' h_n^{(l)}||_2 ||h_n^{(l)\prime} A_n||_2 \le c\sum_{l=1}^{k} ||A_n||_2^2 ||h_n^{(l)}||_2^2 = c\|H_n\|_2^2 \|A_n\|_2^2. \end{aligned}$$

(b) (26) and Assumption 4(a) imply

$$\lim_{n\to\infty} \|H_n\|_2^2 = \sum_{l=1}^{k} ||M_l||_2^2. \tag{44}$$

Hence, by part (a) of this lemma, (27) and Assumption 1

$$\begin{aligned} (E\|H_n'(G_n' - G_n)V_n\|_2^2)^{1/2} &\le c\|H_n\|_2 \|G_n' - G_n\|_2 \\ &\le c\|H_n\|_2(\left\|G_n' - s(\delta_n K)\right\|_2 + \|s(\delta_n K) - G_n\|_2) \\ &\le c\,\|W_n - \delta_n K\|_2 \to 0. \end{aligned}$$

■

Lemma 16 *Suppose Assumptions 1, 2 and 4*(*a*) *are met and let ρ satisfy* (4). *Then* $E\|\alpha_n\|_2^2 = o(1)$.

Proof. Since the number of components of α_n is finite and $\alpha_{n1} = 0$, it suffices to prove $E\|\alpha_{nj}\|_2^2 = o(1),\ j = 2, ..., 5$. Assumptions of this lemma allow us to use (27), Lemma 15(a) and (44):

$$\begin{aligned} \left(E\|\alpha_{n2}\|_2^2\right)^{1/2} &= (E\|H_n'[G_n - s(\delta_n K)]V_n\|_2^2)^{1/2} \\ &\le c\,\|H_n\|_2 \,\|G_n - s(\delta_n K)\|_2 \le c_1\,\|W_n - \delta_n K\|_2 \to 0. \end{aligned}$$

Similarly, applying also (28),

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

In the next two cases in place of Lemma 15(a) we use Lemma 8:

$$\begin{aligned}\left(E\|\alpha_{n4}\|_2^2\right)^{1/2}&=N(I[G_n'-s(\delta_nK)])\\&\le c\sqrt{n}\|G_n'-s(\delta_nK)\|_2\le c_1\sqrt{n}\,\|W_n-\delta_nK\|_2\to 0,\\ \left(E\|\alpha_{n5}\|_2^2\right)^{1/2}&=N(G_n'G_n-s^2(\delta_nK))\\&\le N([G_n'-s(\delta_nK)]G_n)+N(s(\delta_nK)[G_n-s(\delta_nK)])\\&\le c[\|G_n'-s(\delta_nK)\|_2\|G_n\|_2+\|s(\delta_nK)\|_2\,\|G_n-s(\delta_nK)\|_2]\\&\le c_1\,\|W_n-\delta_nK\|_2\to 0.\end{aligned}$$

■

Lemma 17 *If Assumptions 1-3 and 4(a) hold, then* $\left(E\|\beta_{nL}\|_2^2\right)^{1/2}\le c\sum_{i>L}|\lambda_i|$, *where c does not depend on* n,L.

Proof. Like in the previous lemma, we need only consider the last four components of β_{nL}. By Lemma 15(a), (44) and the second bound in (33)

$$\begin{aligned}\left(E\|\beta_{nL2}\|_2^2\right)^{1/2}&=\{E\|H_n'[s(\delta_nK)-s(\delta_nK_L)]V_n\|_2^2\}^{1/2}\\&\le c\,\|H_n\|_2\,\|s(\delta_nK)-s(\delta_nK_L)\|_2\le c_1\sum_{i>L}|\lambda_i|.\end{aligned}$$

For β_{nL3} we also use the first estimate in (33):

$$\begin{aligned}\left(E\|\beta_{nL3}\|_2^2\right)^{1/2}&=(E\|H_n'[s^2(\delta_nK)-s^2(\delta_nK_L)]V_n\|_2^2)^{1/2}\\&\le c\|s^2(\delta_nK)-s^2(\delta_nK_L)\|_2\\&\le c(\|s(\delta_nK)\|_2+\|s(\delta_nK_L)\|_2)\,\|s(\delta_nK)-s(\delta_nK_L)\|_2\\&\le c_1\sum_{i>L}|\lambda_i|.\end{aligned}$$

The proofs for β_{nL4} and β_{nL5} given in [14] don't work for the current error structure. For any $1 \le L \le M \le \infty$ consider a segment $K_{L,M} = \sum_{j=L}^{M} \lambda_j F_j(x) F_j(y)$ of K. By Lemma 5(h)

$$(\delta_n^2 K_{L,M})_{st} = \sum_{j=L}^{M} \lambda_j (\delta_n^1 F_j)_s (\delta_n^1 F_j)_t, \tag{45}$$

so Lemma 5(b) gives

$$\begin{aligned} ||\delta_n^2 K_{L,M}||_2^2 &= \sum_{s,t=1}^{n} \sum_{i,j=L}^{M} \lambda_i \lambda_j (\delta_n^1 F_i)_s (\delta_n^1 F_i)_t (\delta_n^1 F_j)_s (\delta_n^1 F_j)_t \\ &= \sum_{i,j=L}^{M} \lambda_i \lambda_j (\delta_n^1 F_i, \delta_n^1 F_j)_{l_2}^2 \le \left(\sum_{j=L}^{M} |\lambda_j| \right)^{1/2}. \end{aligned} \tag{46}$$

Using Lemma 8 in the proof of [14, Equation (3.5)] we have

$$N(A^{k+1} - B^{k+1}) \le c||A - B||_2 (k+1)(\max\{||A||_2, ||B||_2\})^k \tag{47}$$

for any natural k and square matrices A, B of order n. Now we can proceed with bounding β_{nL4}:

$$\begin{aligned} \left(E|\beta_{nL4}|_2^2\right)^{1/2} &= \{E|V_n'[s(\delta_n K) - s(\delta_n K_L)]V_n|_2^2\}^{1/2} \\ &= N(s(\delta_n K) - s(\delta_n K_L)) \le N(\delta_n K - \delta_n K_L) \\ &+ \sum_{k>0} |\rho|^k N((\delta_n K)^{k+1} - (\delta_n K_L)^{k+1}). \end{aligned} \tag{48}$$

By Lemma 8 and (45)

$$\begin{aligned} N(\delta_n K - \delta_n K_L) &= N\left(\left(\sum_{j>L} \lambda_j (\delta_n F_j)_s (\delta_n F_j)_t \right)_{s,t=1}^{n} \right) \\ &= N\left(\sum_{j>L} \lambda_j \delta_n F_j (\delta_n F_j)' \right) \le \sum_{j>L} |\lambda_j| N\left(\delta_n F_j (\delta_n F_j)'\right) \\ &\le c \sum_{j>L} |\lambda_j| ||\delta_n F_j||_2^2 \le c \sum_{j>L} |\lambda_j|. \end{aligned} \tag{49}$$

For the remaining terms at the right of (48) by (47) we have

$$N((\delta_n K)^{k+1} - (\delta_n K_L)^{k+1})$$

$$\leq c(k+1)(\max\{||\delta_n K||_2, ||\delta_n K_L||_2\})^k ||\delta_n K - \delta_n K_L||_2$$
$$\leq c(k+1)\left(\max\left\{\sum_{j=1}^{\infty}|\lambda_j|, \sum_{j=1}^{L}|\lambda_j|\right\}\right)^k \sum_{j>L}|\lambda_j|$$
$$\leq c(k+1)\left(\sum_{j=1}^{\infty}|\lambda_j|\right)^k \sum_{j>L}|\lambda_j|. \tag{50}$$

Here we have applied three particular cases of (46). Putting together (48), (49) and (50) yields

$$\left(E|\beta_{nL4}|_2^2\right)^{1/2} \leq c\sum_{j>L}|\lambda_j| + c\sum_{k>0}(k+1)\left(|\rho|\sum_{j=1}^{\infty}|\lambda_j|\right)^k \sum_{j>L}|\lambda_j| = c_1\sum_{j>L}|\lambda_j|.$$

In the proof for β_{nL5} we apply Lemma 8 and (33):

$$\begin{aligned}\left(E|\beta_{nL5}|_2^2\right)^{1/2} &= \{E|V_n'[s^2(\delta_n K) - s^2(\delta_n K_L)]V_n|_2^2\}^{1/2} \\ &= N(s^2(\delta_n K) - s^2(\delta_n K_L)) \leq N(s(\delta_n K)[s(\delta_n K) - s(\delta_n K_L)]) \\ &+ N([s(\delta_n K) - s(\delta_n K_L)]s(\delta_n K_L)) \\ &\leq c(\|s(\delta_n K)\|_2 + \|s(\delta_n K_L)\|_2)\,\|s(\delta_n K) - s(\delta_n K_L)\|_2 \\ &\leq c_1\sum_{i>L}|\lambda_i|.\end{aligned}$$

■

Lemma 18 *If Assumptions 1 through 4 hold, then for any positive* (*small*) ε *and* (*large*) L *there exists* $n_0 = n_0(\varepsilon, L)$ *such that* $\sum_{j=1}^{5} E||\gamma_{nLj}||_2 \leq c\varepsilon$ *for all* $n \geq n_0$ *where* c *does not depend on* n *and* L*.*

Proof. Recall definitions (36) and (43): $\gamma_{nL1} = 0$,

$$\gamma_{nL2} = H_n' s(\delta_n K_L)V_n - \sum_{i=1}^{L}\nu(\lambda_i)(M, F_i)_{L_2}U_{nL,k+i},$$
$$\gamma_{nL3} = H_n' s^2(\delta_n K_L)V_n - \sum_{i=1}^{L}\nu^2(\lambda_i)(M, F_i)_{L_2}U_{nL,k+i},$$
$$\gamma_{nL4} = V_n' s(\delta_n K_L)V_n - \sum_{i=1}^{L}\nu(\lambda_i)U_{nL,k+i}^2,$$

$$\gamma_{nL5} = V_n' s^2(\delta_n K_L) V_n - \sum_{i=1}^{L} \nu^2(\lambda_i) U_{nL,k+i}^2.$$

Using (36) and (30) we write the lth component of $H_n' s(\delta_n K_L) V_n$ as

$$\begin{aligned} h_n^{(l)\prime} s(\delta_n K_L) V_n &= \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\dots,i_{p+1} \le L} \prod_{j=1}^{p+1} \lambda_{i_j} \mu_{ni} \sum_{s,t=1}^{n} (\delta_n F_{i_1})_s (h_n^{(l)})_s (\delta_n F_{i_{p+1}})_t v_t \\ &= \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\dots,i_{p+1} \le L} \prod_{j=1}^{p+1} \lambda_{i_j} \mu_{ni} (h_n^{(l)}, \delta_n F_{i_1})_{l_2} U_{nL,k+i_{p+1}}. \qquad (51) \end{aligned}$$

To rearrange the lth component of X_{nL2}, we use the first equation from (32) and the definition of $\mu_{\infty i}$ from Lemma 6(c):

$$\begin{aligned} (X_{nL2})_l &= \sum_{i=1}^{L} \nu(\lambda_i)(M_l, F_i)_{L_2} U_{nL,k+i} \\ &= \sum_{p=0}^{\infty} \rho^p \sum_{i=1}^{L} \lambda_i^{p+1} (M_l, F_i)_{L_2} U_{nL,k+i} \\ &= \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\dots,i_{p+1} \le L} \prod_{j=1}^{p+1} \lambda_{i_j} \mu_{\infty i} (M_l, F_{i_1})_{L_2} U_{nL,k+i_{p+1}}. \qquad (52) \end{aligned}$$

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

$$(\gamma_{nL2})_l = \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\dots,i_{p+1} \le L} \prod_{j=1}^{p+1} \lambda_{i_j} [\mu_{ni} (h_n^{(l)}, \delta_n F_{i_1})_{l_2} - \mu_{\infty i} (M_l, F_{i_1})_{L_2}] U_{nL,k+i_{p+1}}. \qquad (53)$$

Due to Lemmas 5(f) and 6(c) for any $\varepsilon, L > 0$ there exists $n_0 = n_0(\varepsilon, L)$ such that

$$|\mu_{ni} (h_n^{(l)}, \delta_n F_{i_1})_{l_2} - \mu_{\infty i} (M_l, F_{i_1})_{L_2}| < \varepsilon, \; n \ge n_0, \qquad (54)$$

for all i which appear in $(\gamma_{nL2})_l$. Besides, by Lemmas 5(b) and 8

$$\begin{aligned} E|U_{nL,k+i_{p+1}}| &\le (E|U_{nL,k+i_{p+1}}|^2)^{1/2} = [E(V_n' \delta_n F_{i_{p+1}} (\delta_n F_{i_{p+1}})' V_n)]^{1/2} \\ &\le \{E[V_n' \delta_n F_{i_{p+1}} (\delta_n F_{i_{p+1}})' V_n]^2\}^{1/4} \\ &= N(\delta_n F_{i_{p+1}} (\delta_n F_{i_{p+1}})')^{1/2} \le c \|\delta_n F_{i_{p+1}}\|_2 \le c. \qquad (55) \end{aligned}$$

The result of (53), (54) and (55) is the desired estimate of $(\gamma_{nL2})_l$:

$$E|(\gamma_{nL2})_l| \le c\varepsilon \sum_{p=0}^{\infty} |\rho|^p \sum_{i_1,\dots,i_{p+1}\le L} \prod_{j=1}^{p+1} |\lambda_{i_j}|$$
$$\le c\varepsilon \sum_{p=0}^{\infty} \left(|\rho| \sum_{j\ge 1} |\lambda_j|\right)^p \sum_{j\ge 1} |\lambda_j| = c_1\varepsilon,\ l = 1,\dots,k.$$

Similarly, using (31) instead of (30) and the second equation in (32) instead of the first one in the derivation of (52), we obtain a representation for the lth component of γ_{nL3}:

$$(\gamma_{nL3})_l = \sum_{p=0}^{\infty} \rho^p (p+1) \sum_{i_1,\dots,i_{p+2}\le L} \prod_{j=1}^{p+2} \lambda_{i_j} [\mu_{ni}(h_n^{(l)}, \delta_n F_{i_1})_{l_2}$$
$$- \mu_{\infty i}(M_l, F_{i_1})_{L_2}] U_{nL,k+i_{p+2}}.$$

Application of (54) and (55) finishes the proof for $(\gamma_{nL3})_l$.

Replacing in (51) $h_n^{(l)}$ by V_n gives

$$V_n' s(\delta_n K_L) V_n = \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\dots,i_{p+1}\le L} \prod_{j=1}^{p+1} \lambda_{i_j} \mu_{ni} U_{nL,k+i_1} U_{nL,k+i_{p+1}}.$$

Using (32) and properties of $\mu_{\infty i}$ yields

$$\sum_{i=1}^{L} \nu(\lambda_i) U_{nL,k+i}^2 = \sum_{p=0}^{\infty} \rho^p \sum_{i=1}^{L} \lambda_i^{p+1} U_{nL,k+i}^2$$
$$= \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\dots,i_{p+1}\le L} \prod_{j=1}^{p+1} \lambda_{i_j} \mu_{\infty i} U_{nL,k+i_1} U_{nL,k+i_{p+1}}.$$

The last two equations imply

$$\gamma_{nL4} = \sum_{p=0}^{\infty} \rho^p \sum_{i_1,\dots,i_{p+1}\le L} \prod_{j=1}^{p+1} \lambda_{i_j} (\mu_{ni} - \mu_{\infty i}) U_{nL,k+i_1} U_{nL,k+i_{p+1}}. \tag{56}$$

By Lemmas 5(b) and 8

$$(E|U_{nL,k+i_1} U_{nL,k+i_{p+1}}|^2)^{1/2}$$

$$
\begin{aligned}
&= [E(V_n'\delta_n F_{i_1}(\delta_n F_{i_{p+1}})'V_n)^2]^{1/2} \\
&= N(\delta_n F_{i_1}(\delta_n F_{i_{p+1}})') \le c||\delta_n F_{i_1}||_2||\delta_n F_{i_{p+1}}||_2 \le c_1. \qquad (57)
\end{aligned}
$$

Equations (56), (57) and Lemma 6(c) allow us to conclude that

$$
E|\gamma_{nL4}| \le c_1\varepsilon \sum_{p=0}^{\infty} |\rho|^p \sum_{i_1,\ldots,i_{p+1}\le L} \prod_{j=1}^{p+1} |\lambda_{i_j}| \le c_2\varepsilon.
$$

If in the derivation of (56) one replaces (30) by (31) and the first equation from (32) by the second one, then one gets

$$
\gamma_{nL5} = \sum_{p=0}^{\infty} \rho^p(p+1) \sum_{i_1,\ldots,i_{p+2}\le L} \prod_{j=1}^{p+2} \lambda_{i_j}(\mu_{ni}-\mu_{\infty i})U_{nL,k+i_1}U_{nL,k+i_{p+2}}.
$$

The rest of the proof is the same as for γ_{nL4}. ■

REMARK 2. Since H_n does not appear in the last two components of $\alpha_n,\ \beta_{nL}$ and γ_{nL}, the results of Lemmas 16-18 regarding those components do not depend on Assumption 4 and can be used in the proof of Theorem 1.

Proof of Theorem 3 Assumption 3 implies (4). By the Chebyshev inequality and Lemma 16 $\text{plim}_{n\to\infty}\alpha_n = 0$. By Lemma 17 $P(\|\beta_{nL}\|_2 > \varepsilon) \le \frac{1}{\varepsilon}(E\|\beta_{nL}\|_2^2)^{1/2} \le \frac{c}{\varepsilon}\sum_{i>L}|\lambda_i|$ where c does not depend on ε, n, L. Lemma 18 implies $\text{plim}_{n\to\infty}\gamma_{nL} = 0$ for any fixed L. The facts we have just listed and (42) show that for any fixed L $\limsup_{n\to\infty} P(\|\mathscr{A}_n - X_{nL}\|_2 > \varepsilon) \le \frac{c}{\varepsilon}\sum_{i>L}|\lambda_i|$. Equivalence (34) allows us to use Lemma 10. By Billingsley's [27, Theorem 4.2] we have

$$
\text{dlim}\mathscr{A}_n = \xi. \qquad (58)
$$

This relation and Lemma 15(b) ensure convergence in distribution of all parts of the pair (ζ_n, Φ_n) involving the error. Convergence in probability of all other (deterministic) parts of (ζ_n, Φ_n) is provided by Lemma 13 and Assumption 5. Thus, (ζ_n, Φ_n) converges in distribution. The expressions for the limit (16) and (16) are established by comparing the formulas for $\zeta_n,\ \Phi_n,\ \mathscr{A}_n,\ \xi,\ \kappa$ and d contained in (9), (10), (42), (14) and Assumption 5. ∎

3.5 Proof of Theorem 4

(a) Assumption 4(b) enables us to apply the standard fact for partitioned matrices (see Lütkepohl [28, Section A.10]) $|\Phi| = |\Gamma_0|\Delta$, where $\Delta = |\Phi_{22} - \Phi_{21}\Gamma_0^{-1}\Phi_{12}|$.

Here Φ_{ij} are the blocks of (16) and $\Phi_{11} = \Gamma_0$. From (16) we see that the random variable Δ is

$$\begin{aligned}\Delta &= \kappa'\Gamma_2\kappa + \frac{2}{d}\kappa'\xi_3 + \frac{1}{d^2}\xi_5 - \left(\kappa'\Gamma_1 + \frac{1}{d}\xi_2'\right)\Gamma_0^{-1}\left(\Gamma_1\kappa + \frac{1}{d}\xi_2\right) \\ &= \kappa'(\Gamma_2 - \Gamma_1\Gamma_0^{-1}\Gamma_1)\kappa + \frac{1}{d^2}(\xi_5 - \xi_2'\Gamma_0^{-1}\xi_2) + \frac{2}{d}\kappa'(\xi_3 - \Gamma_1\Gamma_0^{-1}\xi_2).\end{aligned}$$

Let $\overline{u}_L = (u_1, ..., u_L, 0, ...)' \in l_2$,

$$\overline{\xi}_L = |\sigma\beta_\psi|\left((X'\overline{u}_L)''\nu_J\overline{u}_L)''\nu_J^2\overline{u}_L)', |\sigma\beta_\psi|\overline{u}_L'\nu_J\overline{u}_L, |\sigma\beta_\psi|\overline{u}_L'\nu_J^2\overline{u}_L\right)'.$$

Obviously, components of $\overline{\xi}_L$ converge to those of ξ in $L_2(\Omega)$, so $\text{plim}_{L\to\infty}\overline{\xi}_L = \xi$ and $\Delta = \text{plim}\Delta_L$ where Δ_L is obtained from Δ by replacing the components of ξ with those of $\overline{\xi}_L$. Using also the definitions of matrices Γ_j (see Assumption 4(b) and (13)) we get

$$\begin{aligned}\Delta_L = {}& \kappa'[X'\nu_J^2X - X'\nu_JX(X'X)^{-1}X'\nu_JX]\kappa \\ &+ \frac{(\sigma\beta_\psi)^2}{d^2}[\overline{u}_L'\nu_J^2\overline{u}_L - (X'\nu_J\overline{u}_L)'(X'X)^{-1}X'\nu_J\overline{u}_L] \\ &+ \frac{2}{d}|\sigma\beta_\psi|\kappa'[X'\nu_J^2\overline{u}_L - X'\nu_JX(X'X)^{-1}X'\nu_J\overline{u}_L].\end{aligned}$$

Now rearrange this using (11):

$$\begin{aligned}\Delta_L &= (\nu_JX\kappa)'(I-P)\nu_JX\kappa + \frac{(\sigma\beta_\psi)^2}{d^2}(\nu_J\overline{u}_L)'(I-P)\nu_J\overline{u}_L \\ &\quad + \frac{2}{d}|\sigma\beta_\psi|(\nu_JX\kappa)'(I-P)\nu_J\overline{u}_L \\ &= \nu_JX\kappa)'Q^2\nu_JX\kappa + \frac{(\sigma\beta_\psi)^2}{d^2}(\nu_J\overline{u}_L)'Q^2\nu_J\overline{u}_L + \\ &\quad + \frac{|\sigma\beta_\psi|}{d}(\nu_JX\kappa)'Q^2\nu_J\overline{u}_L + \frac{|\sigma\beta_\psi|}{d}(\nu_J\overline{u}_L)'Q^2\nu_JX\kappa \\ &= \left\|Q\nu_J\left(X\kappa + \frac{|\sigma\beta_\psi|}{d}\overline{u}_L\right)\right\|_2^2.\end{aligned}$$

We arrive to the conclusion that the limit (17) exists and $|\Phi| = |\Gamma_0|\Xi$. Consequently, conditions $P(|\Phi| > 0) = 1$ and (18) are equivalent.

Convergence (19) follows from (15) and the invertibility condition (18) by the continuous mapping theorem.

(b) If the autoregressive term dominates, by Lemma 14(d) $d = 1$. Thus, (20) follows from (16) and (17) on putting $\kappa = 0,\ d = 1$.

(c) Equations (21) follow from (16) with $d = \infty$ (in this case $\kappa \neq 0$ by Lemma 14(c)). Let us prove (22). Since for any $x \in l_2$ the vector Px is a linear combination of $JM_1, ..., JM_k$, P projects l_2 onto the image $J\mathfrak{M}$ of $\mathfrak{M}$ under the mapping J. Hence, Q projects onto the subspace of l_2 orthogonal to $J\mathfrak{M}$ and $||Qx||_2^2$ is the squared distance from x to $J\mathfrak{M}$. Thus,

$$\Xi = \text{plim}_{L\to\infty} \|Q\nu_J X\kappa\|_2^2 = \|Q\nu_J X\kappa\|_2^2 = \text{dist}^2(\nu_J X\kappa, J\mathfrak{M}). \tag{59}$$

Since $JF_j = (0, ..., 0, 1, 0, ...)'$ (unity in the nth place), we have for any $F \in L_2(0,1)$

$$J\nu(\mathscr{K})F = \sum_{i\geq 1} \nu(\lambda_i)(F, F_i)_{L_2}(JF_i) = (\nu(\lambda_1)(F, F_1)_{L_2}, \nu(\lambda_2)(F, F_2)_{L_2}, ...) = \nu_J JF.$$

Hence, $J\nu(\mathscr{K}) = \nu_J J$. By linearity of J

$$\nu_J X\kappa = \nu_J \sum_{l=1}^{k} \kappa_l JM_l = \nu_J J\kappa' M = J\nu(\mathscr{K})\kappa' M. \tag{60}$$

We get (22): $\Xi = \text{dist}^2(J\nu(\mathscr{K})\kappa' M, J\dot{\mathfrak{M}}) = \text{dist}^2(\nu(\mathscr{K})\kappa' M, \dot{\mathfrak{M}})$ (recall that J is an isomorphism and apply (59) and (60)).

Now we calculate

$$\begin{aligned} V(\zeta) &= (\sigma\beta_\psi)^2 E \begin{pmatrix} \xi_1 \\ \kappa'\xi_2 \end{pmatrix} \begin{pmatrix} \xi_1' & \xi_2'\kappa \end{pmatrix} \\ &= (\sigma\beta_\psi)^2 \begin{pmatrix} EX'uu'X & EX'uu'\nu_J X\kappa \\ E\kappa' X'\nu_J uu'X & E\kappa' X'\nu_J uu'\nu_J X\kappa \end{pmatrix} \\ &= (\sigma\beta_\psi)^2 \begin{pmatrix} X'X & X'\nu_J X\kappa \\ \kappa' X'\nu_J X & \kappa' X'\nu_J^2 X\kappa \end{pmatrix} = (\sigma\beta_\psi)^2 \Phi. \end{aligned}$$

This equation and (21) lead to (23). ∎

4 CONCLUSIONS

We have characterized the asymptotic distribution of the OLS estimator for a mixed spatial model under the scenario that each unit can be influenced by many neighbors. The main lesson learned is that the asymptotic distribution may be nonnormal. It can include quadratic forms in standard normal variables, and the limit in distribution of the matrix in the denominator of the estimator may not be a constant matrix. The suggested multicollinearity detection device in general is a random variable. In a situation like this several methodological recommendations can be made. Since various components of the estimator may have different rates of convergence (usually unknown *a priori* in practice), a self-adjusting normalizer of the regressors of Anderson's type should be used. A balanced choice of theoretical assumptions, sufficiently general and unequivocally feasible, is desirable. The expression of the asymptotic distribution should include automatic built-in switches which would select the appropriate distribution without the user having to fit his/her setup in particular theoretical assumptions. Our result calls for reconsideration of hypothesis testing procedures for the least squares estimation of the parameters in the mixed spatial model.

The present theory is far from being complete. Estimation of the variance-covariance matrix of the limit distribution, conditions for consistency and some other problems have to be addressed in the future research. On a more general note, if one is willing to adopt higher standards of verifiability and transparency of assumptions, asymptotic results for QML and MM also need to be reconsidered.

Appendix. GAUSS code

```
/* All parameters have to be chosen in Part 1.
Either Part 2 or 3 should be commented out.
For Part 3 (changing rho) numsim can be large (several hundred).
For Part 2 (changing m) matrix inversion takes a lot of time, and numsim is better
left at 100. */
    /* Part 1. Assign all parameters in this section */
/* clear input-output window and start the timer */
cls; t= time; print "start time =" t;
/* all parameters for spatial model are set here */
m_first=200;r=5;            @ parameters for (pseudo)Case matrix @
rho_first=-0.2;beta=0;      @ parameters for spatial model @
grid=8;step=100;            @ number and size of steps for changing m or rho @
numsim=100;                 @ number of simulations to run for one combination
of m, r, rho and beta @
sheet=8;                    @ sheet number of Excel file "Mixed_spatial_Monte_Carlo.xls"
```

to write the results @
switch =0;   @ The value of switch is either 1 for "pseudo-Case" or 0 for "Case" @

/* First part of info to write in Excel file. I know little about what this means */

```
field = 1;prec = 2;fmat = "%*.*lf;";
info = "m_1=" $+ ftos(m_first,fmat,field,prec)$+ "r="
$+ ftos(r,fmat,field,prec) $+ "rho_1=" $+ ftos(rho_first,fmat,field,prec)$+ "beta="
$+ ftos(beta,fmat,field,prec) $+ "grid=" $+ ftos(grid,fmat,field,prec)$+ "step="
$+ ftos(step,fmat,field,prec) $+ "numsim=" $+ ftos(numsim,fmat,field,prec);
/* allocating space for variables */
bias=matalloc(numsim,2);
bias_out=matalloc(grid,1);
mean_b=matalloc(grid,1);
std_b=matalloc(grid,1);
mean_r=matalloc(grid,1);
std_r=matalloc(grid,1);
declare cor_b_r;
    /* Part 2. changing m */
rho=rho_first;                              @ rho is fixed @
for j (1,grid,1);                           @ starting a loop for m @
m=m_first+step*j; n=m*r;                    @ model dimension @
l_n=ones(n,1);                              @ constant term @
if switch == 1;
W_n=eye(r) .*. (ones(m,m)/(m-1));           @ pseudo-Case matrix definition @
elseif switch ==0;
W_n=eye(r) .*. ((ones(m,m)-eye(m))/(m-1));@ Case matrix definition @
endif;
G_n=W_n*inv(eye(n)-rho*W_n);
for i (1, numsim, 1);
Vn= rndn(n,1);                              @ generating error vector @
Zn=l_n (G_n*(beta*l_n+Vn));                 @ together with G_n and Vn this defines
regressor matrix Z_n @
bias[i,.]=(inv(Zn'*Zn)*(Zn'*Vn))';          @ this gives bias_theta=theta_hat-theta @
endfor;
bias_b=submat(bias,0,1);                    @ extracting bias_beta from bias_theta @
mean_b[j]=(sumc(bias_b))/numsim;            @ mean of bias_beta @
std_b[j]=sqrt((bias_b-mean_b[j]*ones(numsim,1))'*(bias_b-mean_b[j]*ones(numsim,1))/(numsim-
```

```
1));@sample standard deviation of bias_beta@
bias_r=submat(bias,0,2);                    @ extracting bias_rho from bias_theta @
mean_r[j]=(sumc(bias_r))/numsim;       @ mean of bias_rho @
std_r[j]=sqrt((bias_r-mean_r[j]*ones(numsim,1))'*(bias_r-mean_r[j]*ones(numsim,1))/(numsim-
1)); @ sample standard deviation of bias_rho @
if j>grid-1;                                  @ "if" statement to retain bias and
sample correlation from the last run @
cor_b_r=(bias_b-mean_b[j]*ones(numsim,1))'*(bias_r-mean_r[j]*ones(numsim,1))/((numsim-
1)*std_b[j]*std_r[j]);@ sample correlation between bias_beta and bias_rho @
bias_out=bias;
endif;
endfor;
if switch == 1;
info="pseudo-Case matrix: changing m: " $+ info;@ info for output if pseudo-
Case matrix used @
elseif switch ==0;
info="Case matrix: changing m: " $+ info;@ info for output if Case matrix used
@
endif;
    /*
    Part 3. changing_rho
m=m_first;n=m*r;                          @ m and n are fixed @
l_n=ones(n,1);                            @ constant term @
I_n=eye(n);                               @ identity matrix of order n @
if switch == 1;
W_n=eye(r) .*. (ones(m,m)/(m-1));         @ pseudo-Case matrix definition @
elseif switch ==0;
W_n=eye(r) .*. ((ones(m,m)-eye(m))/(m-1));@ Case matrix definition @
endif;
for j (1,grid,1);                         @ starting a loop for rho @
rho=rho_first+step*j;
G_n=W_n*inv(I_n-rho*W_n);
for i (1, numsim, 1);
Vn= rndn(n,1);                            @ generating error vector @
Zn=l_n (G_n*(beta*l_n+Vn));               @ together with G_n and Vn this de-
fines regressor matrix Z_n @
bias[i,.]=(inv(Zn'*Zn)*(Zn'*Vn))';
endfor;
```

```
bias_b=submat(bias,0,1);                    @ extracting bias_beta from bias_theta
@
mean_b[j]=(sumc(bias_b))/numsim;            @ mean of bias_beta @
std_b[j]=sqrt((bias_b-mean_b[j]*ones(numsim,1))'*(bias_b-mean_b[j]*ones(numsim,1))/(numsim-
1));
@ sample standard deviation of bias_beta @
bias_r=submat(bias,0,2);                    @ extracting bias_rho from bias_theta
@
mean_r[j]=(sumc(bias_r))/numsim;            @ mean of bias_rho @
std_r[j]=sqrt((bias_r-mean_r[j]*ones(numsim,1))'*(bias_r-mean_r[j]*ones(numsim,1))/(numsim-
1));@ sample standard deviation of bias_rho @
if j>grid-1;                                @ ”if” statement to retain bias
and sample correlation from the last run @
cor_b_r=(bias_b-mean_b[j]*ones(numsim,1))'*(bias_r-mean_r[j]*ones(numsim,1))/((numsim-
1)*std_b[j]*std_r[j]);@ sample correlation between bias_beta and bias_rho @
bias_out=bias;
endif;
endfor;
if switch == 1;
info=”pseudo-Case matrix: changing rho: ” $+ info;@ info for output if pseudo-
Case matrix used @
elseif switch ==0;
info=”Case matrix: changing rho: ” $+ info;@ info for output if Case matrix used
@
endif;
*/
    /* Part 4. Write the results in ”Mixed_spatial_Monte_Carlo.xls” */
print info;
ret = xlsWrite(info, ”Mixed_spatial_Monte_Carlo.xls”, ”a1”, sheet, ””);
ret = xlsWrite(”mean_b”, ”Mixed_spatial_Monte_Carlo.xls”, ”a2”, sheet, ””);
ret = xlsWrite(”std_b”, ”Mixed_spatial_Monte_Carlo.xls”, ”b2”, sheet, ””);
ret = xlsWrite(”mean_r”, ”Mixed_spatial_Monte_Carlo.xls”, ”c2”, sheet, ””);
ret = xlsWrite(”std_r”, ”Mixed_spatial_Monte_Carlo.xls”, ”d2”, sheet, ””);
ret = xlsWrite(mean_b, ”Mixed_spatial_Monte_Carlo.xls”, ”a3”, sheet, ””);
ret = xlsWrite(std_b, ”Mixed_spatial_Monte_Carlo.xls”, ”b3”, sheet, ””);
ret = xlsWrite(mean_r, ”Mixed_spatial_Monte_Carlo.xls”, ”c3”, sheet, ””);
ret = xlsWrite(std_r, ”Mixed_spatial_Monte_Carlo.xls”, ”d3”, sheet, ””);
ret = xlsWrite(cor_b_r, ”Mixed_spatial_Monte_Carlo.xls”, ”e3”, sheet, ””);
```

```
ret = xlsWrite(bias_out, "Mixed_spatial_Monte_Carlo.xls", "f3", sheet, "");
print "execution time=" time-t;
```

References

[1] J. Paelinck, L. Klaassen, Spatial Econometrics, Saxon House, Farnborough, 1979.

[2] A.D. Cliff, J.K. Ord, Spatial Autocorrelation, Pion Ltd., London, 1973.

[3] L. Anselin, Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, The Netherlands, 1988.

[4] N. Cressie, Statistics for Spatial Data, Wiley, New York, 1993.

[5] L. Anselin, R. Florax, New Directions in Spatial Econometrics, Springer-Verlag, Berlin, 1995.

[6] J.K. Ord, Estimation Methods for Models of Spatial Interaction, J. Amer. Statist. Assoc., 70 (1975) 120–126.

[7] H.H. Kelejian, I.R. Prucha, A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, J. of Real Estate Finance and Econom., 17 (1998) 99–121.

[8] H.H. Kelejian, I.R. Prucha, A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, Internat. Econom. Rev., 40 (1999) 509–533.

[9] O.L. Smirnov, L. Anselin, Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach, Comput. Statist. Data Anal., 35 (2001) 301–319.

[10] L.F. Lee, Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models I: Spatial Autoregressive Processes, Unpublished Manuscript, Ohio State University, 2001.

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

[12] L.F. Lee, Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances, Econometric Rev., 22 (2003) 307–335.

[13] L.F. Lee, Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models, Econometrica, 72 (2004) 1899–1925.

[14] K.T. Mynbaev, A. Ullah, Asymptotic distribution of the OLS estimator for a purely autoregressive spatial model, J. Multivariate Analysis, 99 (2008) 245–277.

[15] K.T. Mynbaev, L_p-approximable Sequences of Vectors and Limit Distribution of Quadratic Forms of Random Variables, Adv. in Appl. Math., 26 (2001) 302–329.

[16] K.T. Mynbaev, Asymptotic Distribution of the OLS Estimator for a Mixed Regressive, Spatial Autoregressive Model. (2006) Unpublished. http://mpra.ub.uni-muenchen.de/4411/

[17] B.M. Pötscher, I.R. Prucha, Basic Structure of the Asymptotic Theory in Dynamic Nonlinear Econometric Models, Part I: Consistency and approximation concepts, Econometric Rev., 10 (1991) 125-216.

[18] K. Tanaka, Time Series Analysis: Nonstationary and Noninvertible Distribution Theory, Wiley, New York, 1996.

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

[20] T.W. Anderson, The Statistical Analysis of Time Series, Wiley, New York, 1971.

[21] T. Amemiya, Advanced Econometrics, Blackwell, Oxford, 1985.

[22] K.T. Mynbaev, I. Castelar, The Strengths and Weaknesses of L_2-approximable Regressors: Two Essays on Econometrics, V.1, Fortaleza, Expressão Gráfica, 2001.

[23] F.R. Gantmacher, The Theory of Matrices, V. 1, Providence, American Mathematical Society, Rhode Island, 1998.

[24] K.T. Mynbaev, Asymptotic Properties of OLS Estimates in Autoregressions with Bounded or Slowly Growing Deterministic Trends, Comm. Statist. Theory Methods, 35 (2006) 499–520.

[25] A.C. Case, Spatial Patterns in Household Demand, Econometrica, 59 (1991) 953–965.

[26] S. Nabeya, K. Tanaka, A General Approach to the Limiting Distribution for Estimators in Time Series Regression with Nonstable Autoregressive Errors, Econometrica, 58 (1990) 145–163.

[27] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.

[28] H. Lütkepohl, Introduction to Multiple Time Series Analysis, Springer, New York, 1991.