



HAL
open science

Efficiency of the OLS estimator in the vicinity of a spatial unit root

Federico Martellosio

► **To cite this version:**

Federico Martellosio. Efficiency of the OLS estimator in the vicinity of a spatial unit root. *Statistics and Probability Letters*, 2011, 10.1016/j.spl.2011.03.030 . hal-00753942

HAL Id: hal-00753942

<https://hal.science/hal-00753942>

Submitted on 20 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Efficiency of the OLS estimator in the vicinity of a spatial unit root

Federico Martellosio

PII: S0167-7152(11)00121-0
DOI: 10.1016/j.spl.2011.03.030
Reference: STAPRO 5965

To appear in: *Statistics and Probability Letters*

Received date: 20 November 2010

Revised date: 19 March 2011

Accepted date: 19 March 2011



Please cite this article as: Martellosio, F., Efficiency of the OLS estimator in the vicinity of a spatial unit root. *Statistics and Probability Letters* (2011), doi:10.1016/j.spl.2011.03.030

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Efficiency of the OLS Estimator in the Vicinity of a Spatial Unit Root

Federico Martellosio

University of Reading

March 19, 2011

Abstract

Previous results have indicated that the OLS estimator of the vector of regression coefficients can be nearly as efficient as the best linear unbiased estimator when the regression errors follow a spatial process with root in the vicinity of unity. Such results were derived under the assumption of a symmetric weights matrix, which simplifies the analysis considerably, but is very often not satisfied in applications. This paper provides nontrivial generalizations to the important case of nonsymmetric weights matrices.

Keywords: Efficiency of OLS; Linear regression; Spatial autocorrelation; Spatial unit root.

Mailing address: School of Economics, University of Reading, Whiteknights, Reading RG6 6AW, UK. Tel: +44 (0) 118 378 6033.

E-mail: f.martellosio@reading.ac.uk

1 Introduction

The efficiency of the OLS estimator of regression coefficients, relative to the best linear unbiased estimator, has been studied both for the case of a positive definite error variance matrix (e.g., Bloomfield and Watson, 1975, and references therein) and for the case of a singular error variance matrix (e.g., Liu, 2000, and references therein). Nonetheless, the specific case when the error variance matrix, or its inverse, approaches singularity has not been fully investigated. This limiting case is of interest in several statistical models, for instance time series ARMA processes with a near AR or MA unit root, fractionally integrated models with differencing parameter close to 0.5, and some spatial models. When the error variance matrix is singular, generalized matrix inversion is required to define the generalized least-squares (GLS) estimator. The main difficulty in assessing the relative efficiency of the OLS estimator as the error variance matrix approaches singularity is that the matrix generalized inverse is not continuous for a sequence of nonsingular matrices approaching a singular one (see e.g., Stewart and Sun, 1990, p. 136).

The present paper is concerned with a regression model whose errors follow a first-order Simultaneous Autoregressive (SAR(1)) process; see, e.g., Cliff and Ord (1973). More specifically, we focus on the case when the spatial autocorrelation parameter is close to the smallest positive value such that the inverse error variance matrix (or, equivalently, a scaled version of the error variance matrix) is singular. Due to some analogies with time series models, such a value has often been named a spatial unit root; see, e.g., Fingleton (1999), Lee and Yu (2008 and 2009), and Baltagi and Liu (2010).

In the context of a SAR(1) model, it is well known that the power of tests of residual autocorrelation may be very low in some circumstances (e.g., when the weights matrix is very dense), so that one may end up using the OLS estimator in the presence of strong error autocorrelation. As the autocorrelation parameter approaches the spatial unit root, the power of tests for spatial autocorrelation can even vanish (Krämer, 2005), so it is clearly important to understand how inefficient the OLS estimator can be. Several examples of empirical applications of SAR(1) models with autocorrelation parameter close to a spatial unit root are mentioned in Lee and Yu (2008) and Martellosio (2010). Spatial unit roots have also been analyzed in the case of SAR processes on regular

lattices; see, e.g., Bhattacharyya et al. (1997), Baran et al. (2004) and Paulauskas (2007).

The behavior of the relative efficiency of the OLS estimator in a regression model with SAR(1) errors has been previously studied by Krämer and Donninger (1987) and Tilke (1993). The results in those papers require the assumption that the spatial weights matrix is symmetric. This is unfortunate because such an assumption is very often not satisfied in applications—for example, it is not satisfied when \mathbf{W} is the adjacency matrix of an undirected graph, and is usually not satisfied when \mathbf{W} is a row-standardized matrix. The main objective of this paper is to extend those early contributions to the practically important case of nonsymmetric weights matrices.

In Section 2 we introduce the model and the measure of efficiency. In Section 3 we summarize the results available in the literature. Our extensions of those results are presented in Section 4. Section 5 concludes. Proofs are collected in the appendix.

2 The Model and the Measure of Efficiency

Consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1)$$

where \mathbf{X} is an $n \times k$ fixed regressor matrix of rank $k < n$, and \mathbf{u} is a $n \times 1$ error term with $E(\mathbf{u}) = \mathbf{0}$ and $\text{var}(\mathbf{u}) = \sigma^2 \mathbf{V}$, for a positive parameter σ^2 and some positive definite matrix \mathbf{V} . A SAR(1) process for \mathbf{u} is specified by

$$\mathbf{u} = \rho \mathbf{W} \mathbf{u} + \boldsymbol{\varepsilon}, \quad (2)$$

where ρ is a scalar unknown parameter, \mathbf{W} is a fixed $n \times n$ spatial weights matrix, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ random vector satisfying $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\text{var}(\boldsymbol{\varepsilon}) = \mathbf{I}_n$. In order for ρ to be identified we assume that \mathbf{W} is different from \mathbf{I}_n . Next, recall that a square matrix \mathbf{A} is said to be nilpotent if \mathbf{A}^q is equal to a zero matrix for some positive integer q ; e.g., Horn and Johnson (1985). We assume that \mathbf{W} satisfies the following condition.

Assumption 1 \mathbf{W} is a nonnegative and non-nilpotent matrix.

Assumption 1 rules out weights matrices that are, possibly after a re-ordering of the index set, strictly triangular. Indeed, if a square matrix is nonnegative and nilpotent,

then it must be permutationally similar to a strictly triangular matrix; e.g., Dietzenbacher (1996). Hence, Assumption 1 rules out the case of models admitting a unilateral representation, such as time-series autoregressive models.

Let λ_{\max} denote the spectral radius of \mathbf{W} . By the Perron-Frobenius theory (e.g., Horn and Johnson, 1985, Ch. 8) plus the fact that $\lambda_{\max} = 0$ if and only if \mathbf{W} is nilpotent, Assumption 1 implies that λ_{\max} is positive and is an eigenvalue of \mathbf{W} . Since λ_{\max} is real and nonzero, we can reparametrize model (2) by transforming from (ρ, \mathbf{W}) to $(\lambda_{\max}\rho, \lambda_{\max}^{-1}\mathbf{W})$. Hence, we take $\lambda_{\max} = 1$, without loss of generality and unless otherwise specified. The set of values of ρ that is empirically most relevant is $[0, 1)$. In this set

$$\mathbf{V} = [(\mathbf{I}_n - \rho\mathbf{W}')(\mathbf{I}_n - \rho\mathbf{W})]^{-1} =: \boldsymbol{\Sigma}(\rho). \quad (3)$$

Note that, since $\mathbf{I}_n - \mathbf{W}$ is singular, $\boldsymbol{\Sigma}(\rho)$ does not exist at the spatial unit root $\rho = 1$.

We denote by $\gamma_1(\rho), \dots, \gamma_s(\rho)$ the s distinct eigenvalues of $\boldsymbol{\Sigma}^{-1}(\rho)$, in increasing order of magnitude. Note that s does not depend on ρ except for a finite number of values of ρ (see, e.g., Kato, 1995, p. 64). Also, observe that, by continuity of the eigenvalues of $\boldsymbol{\Sigma}^{-1}(\rho)$ in ρ , $\lim_{\rho \rightarrow 1} \gamma_1(\rho) = \gamma_1(1) = 0$. Throughout our analysis and unless otherwise stated, we maintain the following technical assumption.

Assumption 2 $\lim_{\rho \rightarrow 1} \gamma_2(\rho) > 0$.

Assumption 2 is virtually always satisfied in applications of SAR(1) models. It requires that the eigenvalue $\gamma_1(1) = 0$ does not split into different eigenvalues of $\boldsymbol{\Sigma}^{-1}(\rho)$ in a neighborhood of $\rho = 1$. Without Assumption 2, the results to follow would become analytically more complicated.

A possible measure of the relative efficiency of the OLS estimator $\hat{\boldsymbol{\beta}}_{OLS} := (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, compared to the GLS estimator $\hat{\boldsymbol{\beta}}_{GLS} := (\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\rho)\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\rho)\mathbf{y}$, is the ratio of total variances

$$\eta := \frac{\text{tr}[\text{var}(\mathbf{X}\hat{\boldsymbol{\beta}}_{GLS})]}{\text{tr}[\text{var}(\mathbf{X}\hat{\boldsymbol{\beta}}_{OLS})]}.$$

This paper studies the limiting behavior of η as $\rho \rightarrow 1$ (from the left), in the context of the particular structure (3). Note that, as $\rho \rightarrow 1$, η has the same limit as $\eta^* := \text{tr}[\text{var}(\hat{\boldsymbol{\beta}}_{GLS})]/\text{tr}[\text{var}(\hat{\boldsymbol{\beta}}_{OLS})]$.¹ Examples of papers that have adopted η or η^* to analyze

¹Also, observe that, in order to analyze η , one may assume $\mathbf{X}'\mathbf{X} = \mathbf{I}_n$, because η depends on \mathbf{X} only through the column space of \mathbf{X} . Under that assumption, $\eta = \eta^*$ for any positive definite $\boldsymbol{\Sigma}(\rho)$.

the efficiency of $\hat{\beta}_{OLS}$ for various specifications of \mathbf{V} are Krämer (1980 and 1984), Stemmann and Trenkler (2000), Kleiber (2001), Song and Trenkler (2001), Jeske and Song (2003).

A word on notation: $\text{eig}_\lambda(\mathbf{Q})$ denotes the eigenspace of a square matrix \mathbf{Q} associated to the eigenvalue λ ; $\text{col}(\mathbf{Q})$ denotes the column space of a matrix \mathbf{Q} .

3 Previous Results

The behavior of the efficiency η as $\rho \rightarrow 1$ in a regression model with SAR(1) errors has been studied by Krämer and Donninger (1987), and subsequently by Tilke (1993), who pointed out that Krämer and Donninger's main results require symmetry of \mathbf{W} . The following theorem summarizes the available results.²

Theorem 3.1 (Krämer and Donninger, 1987; Tilke, 1993) *In a regression model with SAR(1) errors, $\lim_{\rho \rightarrow 1} \eta$ is:*

- (i) 1 if \mathbf{W} is symmetric and $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$;
- (ii) 0 if $\text{eig}_1(\mathbf{W}) \not\subseteq \text{col}(\mathbf{X})$, $\text{eig}_1(\mathbf{W}) \cap \text{col}(\mathbf{X}) = \{\mathbf{0}\}$, and a) \mathbf{W} is symmetric or b) $\dim(\text{eig}_1(\mathbf{W})) = 1$;
- (iii) in $(0, 1]$ if \mathbf{W} is symmetric and $\text{eig}_1(\mathbf{W}) \perp \text{col}(\mathbf{X})$.

A substantial generalization of Theorem 3.1 was attempted by Krämer and Baltagi (1996), henceforth KB, by considering a general \mathbf{V} belonging to some subset \mathcal{S} of the cone of symmetric positive definite matrices. More precisely, KB consider, in the context of model (1), the following condition. (Recall that convergence of matrices does not depend on the choice of norm, and that convergence with respect to any norm is equivalent to componentwise convergence.)

Condition 3.2 *There are a scalar function $g(\mathbf{V})$ and a matrix $\tilde{\mathbf{V}}$ belonging to the closure of \mathcal{S} such that $g(\mathbf{V})\mathbf{V} \rightarrow \tilde{\mathbf{V}}$ and $\text{col}(\tilde{\mathbf{V}}) \subseteq \text{col}(\mathbf{X})$.*

In their Corollary 1, KB state that, under Condition 3.2, $\lim_{g(\mathbf{V})\mathbf{V} \rightarrow \tilde{\mathbf{V}}} \eta = 1$ (see also their Theorem 1, which corresponds to the case $g(\mathbf{V}) = 1$). Unfortunately, while correct

²Observe that Assumption 2 is satisfied when \mathbf{W} is symmetric or $\dim(\text{eig}_1(\mathbf{W})) = 1$, and hence is satisfied in all cases in Theorem 2.

in some interesting cases, KB's claim does not hold true in general. Counterexamples can be found precisely in the class of SAR(1) processes, as we show next.

Tilke (1993) observed that there are SAR(1) processes with nonsymmetric \mathbf{W} such that $\lim_{\rho \rightarrow 1} \eta \neq 1$. We now show that such processes constitute counterexamples to KB's claim. This can be done using the following result, where \mathbf{P}_U denotes the orthogonal projector onto a subspace U of \mathbb{R}^n .

Lemma 3.3 *As $\rho \rightarrow 1$, $\gamma_1(\rho)\boldsymbol{\Sigma}(\rho) \rightarrow \mathbf{P}_{\text{eig}_1(\mathbf{W})}$.*

Lemma 3.3 implies that any SAR(1) process satisfies Condition 3.2 as long as $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$ (with $\mathbf{V} = \boldsymbol{\Sigma}(\rho)$, $g(\mathbf{V}) = \gamma_1(\rho)$, $\tilde{\mathbf{V}} = \mathbf{P}_{\text{eig}_1(\mathbf{W})}$, $\text{col}(\tilde{\mathbf{V}}) = \text{eig}_1(\mathbf{W})$). This contradicts KB's claim because, as noted above, $\lim_{\rho \rightarrow 1} \eta$ is not necessarily 1 if \mathbf{W} is nonsymmetric. As a simple example, consider the case when \mathbf{W} is the row-standardized version of the adjacency matrix of a star graph, and the regression contains only an intercept. By straightforward computation, $\eta = (n-1)(\rho+1)^2 / [(n^2 - 3(n-1))\rho^2 + 2(n-1)\rho + n - 1]$, and $\lim_{\rho \rightarrow 1} \eta = 4(n-1)/n^2$, which approaches 0 as $n \rightarrow \infty$; see Figure 1.

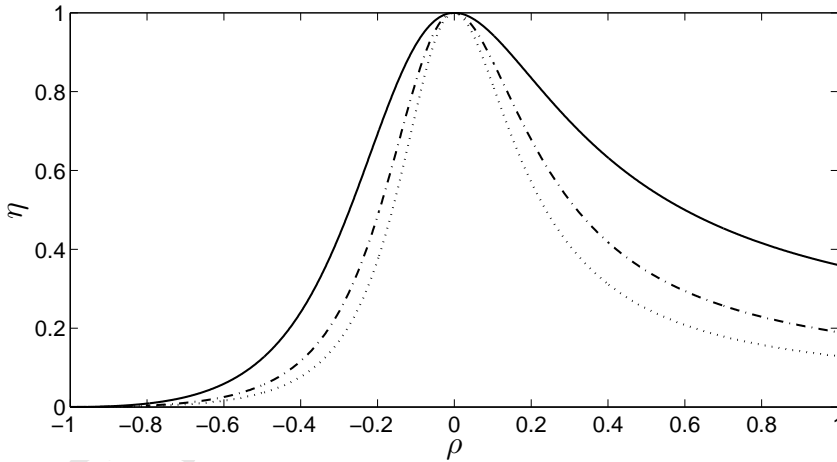


Figure 1: The relative efficiency of the OLS estimator of the constant mean for a SAR(1) process, when \mathbf{W} is the row-standardized version of the adjacency matrix of a star graph on $n = 10$ (solid line), 20 (dashed line), 30 (dotted line) vertices.

Counterexamples to KB's claim can also be found when the regressors follow a first-order spatial moving average models $\mathbf{u} = \boldsymbol{\varepsilon} + \rho\mathbf{W}\boldsymbol{\varepsilon}$ with nonsymmetric \mathbf{W} (for ρ approaching a value such that \mathbf{V} tends to a singular matrix).

4 Main Results

Our first step towards strengthening Theorem 3.1 is the following result, where the symmetry condition in Theorem 3.1(i) is replaced by the weaker condition that $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}')$.

Theorem 4.1 *In a regression model with SAR(1) errors, $\lim_{\rho \rightarrow 1} \eta = 1$ if $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}') \subseteq \text{col}(\mathbf{X})$.*

Two simple corollaries of Theorem 4.1 are given next. Recall that a matrix \mathbf{A} is *irreducible* if the graph with adjacency matrix \mathbf{A} (that is, the graph with n vertices and an edge from vertex i to vertex j if and only if $(\mathbf{A})_{ij} \neq 0$) has a path from any vertex i to any vertex j ; see, e.g., Horn and Johnson (1985).

Corollary 4.2 *In a regression model with an intercept and SAR(1) errors, $\lim_{\rho \rightarrow 1} \eta = 1$ if \mathbf{W} is irreducible and doubly stochastic.*

Another class of matrices to which Theorem 4.1 applies is that of (nonnegative) *normal* matrices. Recall that a real matrix is said to be normal if it commutes with its transpose.

Corollary 4.3 *In a regression model with SAR(1) errors, $\lim_{\rho \rightarrow 1} \eta = 1$ if \mathbf{W} is normal and $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$.*

As an illustration of Corollaries 4.2 and 4.3, consider a (circulant) weights matrix with ones in the first lower diagonal and in position $(1, n)$, and zeroes everywhere else. Since such a weights matrix is irreducible, doubly stochastic, normal, and has $\text{eig}_1(\mathbf{W}) = \text{span}\{\boldsymbol{\iota}_n\}$, where $\boldsymbol{\iota}_n$ denotes the $n \times 1$ vector of all ones, either of the two corollaries implies that $\lim_{\rho \rightarrow 1} \eta = 1$ as long as an intercept is included in the regression.

Theorem 4.1 represents a significant generalization of Theorem 3.1(i), but is not entirely satisfactory because the condition $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}')$ is still far from being necessary. Considerable progress can be made by restricting attention to diagonalizable \mathbf{W} 's. Diagonalizability is certainly satisfied, for example, in the case of the most common way of specifying weights matrices, which consists of row-standardizing a symmetric matrix. Indeed, a weights matrix $\mathbf{D}^{-1}\mathbf{A}$, where \mathbf{A} is an $n \times n$ symmetric matrix with positive row-sums and \mathbf{D} is the diagonal matrix with main diagonal equal to $\mathbf{A}\boldsymbol{\iota}_n$, is similar to the symmetric matrix $\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, and is therefore diagonalizable.

Theorem 4.4 *In a regression model with SAR(1) errors, $\lim_{\rho \rightarrow 1} \eta = 1$ if \mathbf{W} is diagonalizable and, for each eigenvalue λ of \mathbf{W} , $\text{eig}_\lambda(\mathbf{W}') \perp \text{eig}_1(\mathbf{W}')$ or $\text{eig}_\lambda(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$.*

The condition in Theorem 4.4 requires $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$, because obviously $\text{eig}_1(\mathbf{W}')$ cannot be orthogonal to itself. Note that $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$ is sufficient for $\lim_{\rho \rightarrow 1} \eta = 1$ if \mathbf{W} is diagonalizable with $\text{eig}_\lambda(\mathbf{W}') \perp \text{eig}_1(\mathbf{W}')$ for all eigenvalues $\lambda \neq 1$ of \mathbf{W} . This is particularly relevant when $\text{eig}_1(\mathbf{W}) = \text{span}\{\boldsymbol{\iota}_n\}$ and the regression contains an intercept. On the other hand, if \mathbf{W} is diagonalizable but $\text{eig}_\lambda(\mathbf{W}')$ is not orthogonal to $\text{eig}_1(\mathbf{W}')$ for at least one eigenvalue $\lambda \neq 1$, then $\lim_{\rho \rightarrow 1} \eta = 1$ is unlikely to occur in practice.

Our next task is to generalize Theorem 3.1(ii).

Theorem 4.5 *In a regression model with SAR(1) errors, $\lim_{\rho \rightarrow 1} \eta = 0$ if and only if $\text{eig}_1(\mathbf{W}) \not\subseteq \text{col}(\mathbf{X})$ and $\text{eig}_1(\mathbf{W}) \cap \text{col}(\mathbf{X}) = \{\mathbf{0}\}$.*

Compared to Theorem 3.1(ii), Theorem 4.5 does not require the assumption that \mathbf{W} is symmetric or $\dim(\text{eig}_1(\mathbf{W})) = 1$. By the Perron-Frobenius Theorem (e.g., Horn and Johnson, 1985), $\dim(\text{eig}_1(\mathbf{W})) = 1$ if \mathbf{W} is irreducible, whereas $\dim(\text{eig}_1(\mathbf{W}))$ may be larger than 1 otherwise. For example, block diagonal weights matrices (often used in econometrics, see, e.g., Kelejian et al., 2006, and Liu and Lee, 2010) are reducible.

Theorem 4.5 delivers important information for the very common case when \mathbf{W} is a (symmetric or nonsymmetric) row-stochastic matrix and an intercept is included in the regression. In that case, $\boldsymbol{\iota}_n \in \text{eig}_1(\mathbf{W}) \cap \text{col}(\mathbf{X})$, and hence η cannot vanish as $\rho \rightarrow 1$.

Finally, it is worth pointing out that when the conditions in Theorems 4.1, 4.4, or 4.5 are not satisfied, the limiting efficiency can be anywhere in $(0, 1]$, depending on \mathbf{W} and \mathbf{X} .

Remark 4.6 Corollary 4.2 can be easily extended to weights matrices that are doubly stochastic but not necessarily irreducible. Since any doubly stochastic matrix is permutationally similar to a direct sum of doubly stochastic and irreducible matrices (e.g., Sinkhorn, 1968), it follows that there is an ordering of the observational units such that $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}') = \text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_p\}$, where \mathbf{e}_i is a vector with ones in correspondence of the i -th diagonal block of \mathbf{W} and zeros elsewhere. Corollary 4.2 continues to hold without the assumption of irreducibility, provided that the condition

that an intercept is included in the regression is replaced by the condition that $\mathbf{e}_1, \dots, \mathbf{e}_p$ are in $\text{col}(\mathbf{X})$ (as in a fixed effects model). Corollary 4.2 corresponds to the particular case $p = 1$.

5 Final Remarks

We have analyzed the efficiency of the OLS estimator of all regression coefficients, when the errors follow a spatial autoregressive process with autocorrelation parameter approaching unity. When the number of regressors is larger than one, there is necessarily some arbitrariness in the choice of the measure of efficiency, as no single measure exists. In this paper, we have considered a measure that has often been used in the literature—the ratio of total variances. It should be pointed out that the results obtained using that measure do not apply when one is interested in only a subset of the regression coefficients (see, e.g., Krämer, 1980, p. 1006, and Dielman and Pfaffenberger, 1989). Finally note that we have confined ourselves to the efficiency of the regression coefficients; for a discussion of standard error estimation in a spatial regression context see Cordy and Griffith (1993).

Appendix A Proofs

We first give two auxiliary lemmata, and then the proof of all results in the paper. We note that, as it is clear from its proof, the first lemma does not require Assumption 2.

Lemma A.1 $\text{eig}_0(\boldsymbol{\Sigma}^{-1}(1)) = \text{eig}_1(\mathbf{W})$.

Proof. First observe that $\text{eig}_1(\mathbf{W}) \subseteq \text{eig}_0(\boldsymbol{\Sigma}^{-1}(1))$, because, for any $\mathbf{f} \in \text{eig}_1(\mathbf{W})$, $\mathbf{W}\mathbf{f} = \mathbf{f}$ and hence $(\mathbf{I}_n - \mathbf{W}')(\mathbf{I}_n - \mathbf{W})\mathbf{f} = \mathbf{0}$. To prove the lemma, we only need to show that $\dim(\text{eig}_0(\boldsymbol{\Sigma}^{-1}(1))) = \dim(\text{eig}_1(\mathbf{W}))$. For any matrix \mathbf{B} , $\text{rank}(\mathbf{B}'\mathbf{B}) = \text{rank}(\mathbf{B})$ and $\text{rank}(\mathbf{B}) + \dim(\text{null}(\mathbf{B})) = n$, where $\text{null}(\mathbf{B})$ denotes the null space of \mathbf{B} (e.g., Horn and Johnson, 1985). It follows that $\dim(\text{eig}_0(\boldsymbol{\Sigma}^{-1}(1))) = n - \text{rank}((\mathbf{I}_n - \mathbf{W}')(\mathbf{I}_n - \mathbf{W})) = \dim(\text{null}(\mathbf{W} - \mathbf{I}_n)) = \dim(\text{eig}_1(\mathbf{W}))$. ■

Lemma A.2 Consider the matrix $\boldsymbol{\Sigma}(\rho)$ in (3). If $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}')$ then there exists a left neighborhood of $\rho = 1$ where $\text{eig}_{\gamma_1}(\boldsymbol{\Sigma}^{-1}(\rho)) = \text{eig}_1(\mathbf{W})$.

Proof. If $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}')$, then, for any $\mathbf{f} \in \text{eig}_1(\mathbf{W})$, $(\mathbf{I}_n - \rho\mathbf{W}')(\mathbf{I}_n - \rho\mathbf{W})\mathbf{f} = (1 - \rho)^2 \mathbf{f}$, which implies

$$\text{eig}_1(\mathbf{W}) \subseteq \text{eig}_{(1-\rho)^2}(\boldsymbol{\Sigma}^{-1}(\rho)), \quad (\text{A.1})$$

for any ρ . Observe that $\lim_{\rho \rightarrow 1} (1 - \rho)^2 = \gamma_1(1) = 0$ and recall that the eigenvalues of $\boldsymbol{\Sigma}(\rho)$ are continuous functions of ρ . By Assumption 2 it follows that there must be a left neighborhood of $\rho = 1$ where $(1 - \rho)^2$ is the smallest eigenvalue of $\boldsymbol{\Sigma}^{-1}(\rho)$ with a constant (i.e., not depending on ρ) multiplicity. For ρ in such a neighborhood of $\rho = 1$, $\text{eig}_{\gamma_1}(\boldsymbol{\Sigma}^{-1}(\rho)) = \text{eig}_1(\mathbf{W})$ by Lemma A.1 and expression (A.1). ■

Proof of Lemma 3.3. Consider the spectral decomposition

$$\boldsymbol{\Sigma}(\rho) = \sum_{i=1}^s \varphi_i(\rho) \mathbf{P}_{\text{eig}_{\varphi_i}(\boldsymbol{\Sigma}(\rho))}, \quad (\text{A.2})$$

where $\varphi_i(\rho)$, i, \dots, s , are the distinct eigenvalues of $\boldsymbol{\Sigma}(\rho)$, ordered in increasing order of magnitude. Since $\varphi_1(\rho) = 1/\gamma_s(\rho)$, we have

$$\gamma_1(\rho) \boldsymbol{\Sigma}(\rho) = \sum_{i=1}^{s-1} \gamma_1(\rho) \varphi_i(\rho) \mathbf{P}_{\text{eig}_{\varphi_i}(\boldsymbol{\Sigma}(\rho))} + \mathbf{P}_{\text{eig}_{\varphi_s}(\boldsymbol{\Sigma}(\rho))}.$$

Recall that as $\rho \rightarrow 1$ $\boldsymbol{\Sigma}^{-1}(\rho)$ tends to a singular matrix. Thus $\gamma_1(\rho) \rightarrow 0$, and, by Assumption 2, each $\varphi_i(\rho)$, $i = 1, \dots, s - 1$, has a finite limit. By Kato (1995), Chapter 2, Assumption 2 also implies that, as $\rho \rightarrow 1$, $\mathbf{P}_{\text{eig}_{\gamma_1}(\boldsymbol{\Sigma}^{-1}(\rho))} \rightarrow \mathbf{P}_{\text{eig}_0(\boldsymbol{\Sigma}^{-1}(1))}$. The proof is now completed, because $\mathbf{P}_{\text{eig}_{\gamma_1}(\boldsymbol{\Sigma}^{-1}(\rho))} = \mathbf{P}_{\text{eig}_{\varphi_s}(\boldsymbol{\Sigma}(\rho))}$, and $\mathbf{P}_{\text{eig}_0(\boldsymbol{\Sigma}^{-1}(1))} = \mathbf{P}_{\text{eig}_1(\mathbf{W})}$ by Lemma A.1. ■

Proof of Theorem 4.1. On assuming without loss of generality that $\mathbf{X}'\mathbf{X} = \mathbf{I}_n$ (see footnote 1), one has $\eta = \text{tr}((\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\rho)\mathbf{X})^{-1})/\text{tr}(\mathbf{X}'\boldsymbol{\Sigma}(\rho)\mathbf{X})$. For any $n \times n$ positive definite matrix $\boldsymbol{\Omega}$, and for \mathbf{Z} a full rank $n \times (n - k)$ matrix such that $\mathbf{Z}'\mathbf{X} = \mathbf{O}$, $(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1} = \mathbf{X}'\boldsymbol{\Omega}\mathbf{X} - \mathbf{X}'\boldsymbol{\Omega}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Omega}\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Omega}\mathbf{X}$ (see Rao, 1967). Hence,

$$\eta = 1 - \frac{\text{tr}(\mathbf{X}'\boldsymbol{\Sigma}(\rho)\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}(\rho)\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}(\rho)\mathbf{X})}{\text{tr}(\mathbf{X}'\boldsymbol{\Sigma}(\rho)\mathbf{X})}, \quad (\text{A.3})$$

which we rewrite as

$$\eta = 1 - \frac{\text{tr}(\mathbf{A}\mathbf{B}^{-1}\mathbf{A}')}{\text{tr}(\mathbf{C})}, \quad (\text{A.4})$$

where $\mathbf{A} := \gamma_1^{1/2}(\rho)\mathbf{X}'\boldsymbol{\Sigma}(\rho)\mathbf{Z}$, $\mathbf{B} := \mathbf{Z}'\boldsymbol{\Sigma}(\rho)\mathbf{Z}$, and $\mathbf{C} := \gamma_1(\rho)\mathbf{X}'\boldsymbol{\Sigma}(\rho)\mathbf{X}$. By Lemma 3.3, $\mathbf{C} \rightarrow \mathbf{X}'\mathbf{P}_{\text{eig}_1(\mathbf{W})}\mathbf{X} = (\mathbf{P}_{\text{eig}_1(\mathbf{W})}\mathbf{X})'\mathbf{P}_{\text{eig}_1(\mathbf{W})}\mathbf{X}$. Thus, $\lim_{\rho \rightarrow 1} \text{tr}[\mathbf{C}] \neq 0$ as long as $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$. Using decomposition (A.2),

$$\mathbf{A} = \sum_{i=1}^{s-1} \gamma_1^{1/2}(\rho)\varphi_i(\rho)\mathbf{X}'\mathbf{P}_{\text{eig}_{\varphi_i}(\boldsymbol{\Sigma}(\rho))}\mathbf{Z} + \gamma_1^{-1/2}(\rho)\mathbf{X}'\mathbf{P}_{\text{eig}_{\varphi_s}(\boldsymbol{\Sigma}(\rho))}\mathbf{Z}.$$

If $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$ then $\mathbf{P}_{\text{eig}_1(\mathbf{W})}\mathbf{Z} = \mathbf{O}$. If, in addition, $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}')$ then, by Lemma A.2, there is a left neighborhood of $\rho = 1$ where $\mathbf{P}_{\text{eig}_{\varphi_s}(\boldsymbol{\Sigma}(\rho))}\mathbf{Z} = \mathbf{O}$. As $\rho \rightarrow 1$, $\gamma_1(\rho) \rightarrow 0$, and, by Assumption 2, each $\varphi_i(\rho)$, $i = 1, \dots, s-1$, has a finite limit. It follows that $\mathbf{A} \rightarrow \mathbf{O}$ if $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}') \subseteq \text{col}(\mathbf{X})$. Similarly, write

$$\mathbf{B} = \sum_{i=1}^{s-1} \varphi_i(\rho)\mathbf{Z}'\mathbf{P}_{\text{eig}_{\varphi_i}(\boldsymbol{\Sigma}(\rho))}\mathbf{Z} + \varphi_s(\rho)\mathbf{Z}'\mathbf{P}_{\text{eig}_{\varphi_s}(\boldsymbol{\Sigma}(\rho))}\mathbf{Z}.$$

Using again the fact that, when $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}') \subseteq \text{col}(\mathbf{X})$, there is a left neighborhood of $\rho = 1$ where $\mathbf{P}_{\text{eig}_{\varphi_s}(\boldsymbol{\Sigma}(\rho))}\mathbf{Z} = \mathbf{O}$, we can see that \mathbf{B} tends to a finite matrix as $\rho \rightarrow 1$. The theorem now follows from expression (A.4). ■

Proof of Corollary 4.2. If \mathbf{W} is irreducible and doubly stochastic, then $\text{eig}_1(\mathbf{W}') = \text{eig}_1(\mathbf{W}) = \text{span}\{\boldsymbol{\iota}_n\}$, because the vector $\boldsymbol{\iota}_n$ of all ones is an eigenvector of both \mathbf{W} and \mathbf{W}' , and because both $\text{eig}_1(\mathbf{W})$ and $\text{eig}_1(\mathbf{W}')$ have dimension 1 by the Perron-Frobenius Theorem. The result follows from Theorem 4.1. ■

Proof of Corollary 4.3. Any normal \mathbf{W} can be diagonalized by a unitary matrix \mathbf{U} , say $\mathbf{W} = \mathbf{U}\boldsymbol{\Lambda}\overline{\mathbf{U}}'$, where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of \mathbf{W} , and $\overline{\mathbf{U}}$ denotes the conjugate matrix of \mathbf{U} . Since $\mathbf{U}\overline{\mathbf{U}}' = \overline{\mathbf{U}}'\mathbf{U} = \mathbf{I}_n$, it follows that the i -th column of \mathbf{U} and the i -th column of $\overline{\mathbf{U}}$ are, respectively, an eigenvector of \mathbf{W} and an eigenvector of \mathbf{W}' associated to the eigenvalue $\boldsymbol{\Lambda}_{i,i}$. This implies $\text{eig}_1(\mathbf{W}) = \text{eig}_1(\mathbf{W}')$, because all eigenvectors associated to the eigenvalue 1 can be taken to be real. The result follows from Theorem 4.1. ■

Proof of Theorem 4.4. Assume without loss of generality that $\mathbf{X}'\mathbf{X} = \mathbf{I}_n$, and consider expression (A.4). As in the proof of Theorem 4.1, $\lim_{\rho \rightarrow 1} \text{tr}[\mathbf{C}] \neq 0$. Let $\lambda_1, \dots, \lambda_q$ denote the q distinct eigenvalues of \mathbf{W} in non-decreasing order of modulus (so that $\lambda_q = 1$). If \mathbf{W} is diagonalizable,

$$(\mathbf{I}_n - \rho\mathbf{W})^{-1} = \sum_{i=1}^q \frac{1}{1 - \rho\lambda_i} \mathbf{G}_i, \quad (\text{A.5})$$

where \mathbf{G}_i is the (oblique) projector onto $\text{eig}_{\lambda_i}(\mathbf{W})$ along $\text{col}(\mathbf{W} - \lambda_i \mathbf{I}_n)$, $i = 1, \dots, q$. Under the condition $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$, $\mathbf{Z}'\mathbf{G}_q = \mathbf{O}$, and hence

$$\mathbf{D} := \mathbf{Z}'(\mathbf{I}_n - \rho\mathbf{W})^{-1} = \sum_{i=1}^{q-1} \frac{1}{1 - \rho\lambda_i} \mathbf{Z}'\mathbf{G}_i \quad (\text{A.6})$$

tends to a finite matrix as $\rho \rightarrow 1$. Also, note that \mathbf{D} has full rank as $\rho \rightarrow 1$, because \mathbf{Z}' has full rank and the rank of a matrix is unchanged upon multiplication by a nonsingular matrix. It follows that $\mathbf{B} = \mathbf{D}\mathbf{D}'$ tends to a finite and nonsingular matrix. Looking at expression (A.4), in order to complete the proof of part (i) it remains to be shown that $\mathbf{A} \rightarrow \mathbf{O}$ under the conditions stated in the theorem. Write $\mathbf{A} = \gamma_1^{1/2}(\rho)\mathbf{X}'(\mathbf{I}_n - \rho\mathbf{W})^{-1}(\mathbf{I}_n - \rho\mathbf{W}')^{-1}\mathbf{Z}$. If \mathbf{W} is diagonalizable, using (A.5) we have

$$\mathbf{A} = \left(\frac{\gamma_1^{1/2}(\rho)}{1 - \rho} \mathbf{X}'\mathbf{G}_q + \sum_{i=1}^{q-1} \frac{\gamma_1^{1/2}(\rho)}{1 - \rho\lambda_i} \mathbf{X}'\mathbf{G}_i \right) \left(\frac{1}{1 - \rho} \mathbf{G}'_q \mathbf{Z} + \sum_{i=1}^{q-1} \frac{1}{1 - \rho\lambda_i} \mathbf{G}'_i \mathbf{Z} \right).$$

This expression shows that $\mathbf{A} \rightarrow \mathbf{O}$ if $\mathbf{G}'_q \mathbf{Z} = \mathbf{O}$ (i.e., $\text{eig}_1(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$), and

$$\mathbf{X}'\mathbf{G}_q \mathbf{G}'_i \mathbf{Z} = \mathbf{O}, \quad (\text{A.7})$$

for all $i = 1, \dots, q-1$. Note that the product $\mathbf{X}'\mathbf{G}_q$ in (A.7) is nonzero if $\mathbf{Z}'\mathbf{G}_q = \mathbf{O}$. Thus, condition (A.7) is satisfied if $\mathbf{Z}'\mathbf{G}_i = \mathbf{O}$ or $\mathbf{G}_i \mathbf{G}'_q = \mathbf{O}$, that is, if $\text{eig}_{\lambda_i}(\mathbf{W}) \subseteq \text{col}(\mathbf{X})$ or $\text{eig}_{\lambda_i}(\mathbf{W}') \perp \text{eig}_1(\mathbf{W}')$. This completes the proof. ■

Proof of Theorem 4.5. Write

$$\eta = \frac{\gamma_1(\rho)\text{tr}((\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\rho)\mathbf{X})^{-1})}{\gamma_1(\rho)\text{tr}(\mathbf{X}'\boldsymbol{\Sigma}(\rho)\mathbf{X})}. \quad (\text{A.8})$$

By Lemma 3.3, as $\rho \rightarrow 1$ the denominator of (A.8) approaches $\text{tr}(\mathbf{X}'\mathbf{P}_{\text{eig}_1(\mathbf{W})}\mathbf{X}) = \text{tr}((\mathbf{X}\mathbf{P}_{\text{eig}_1(\mathbf{W})})'\mathbf{P}_{\text{eig}_1(\mathbf{W})}\mathbf{X})$, which is zero if and only if $\text{eig}_1(\mathbf{W}) \perp \text{col}(\mathbf{X})$. Let us now look at the numerator of (A.8). The matrix $\lim_{\rho \rightarrow 1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\rho)\mathbf{X} = \mathbf{X}'(\mathbf{I}_n - \mathbf{W}')(\mathbf{I}_n - \mathbf{W})\mathbf{X}$ is singular if and only if $\mathbf{X}'(\mathbf{I}_n - \mathbf{W}')(\mathbf{I}_n - \mathbf{W})\mathbf{X}\mathbf{v} = 0$ for some vector $\mathbf{v} \neq \mathbf{0}$, that is, if and only if there exists a nonzero linear combination of the columns of \mathbf{X} that belongs to $\text{null}(\mathbf{W} - \mathbf{I}_n) = \text{eig}_1(\mathbf{W})$. Thus, if $\text{eig}_1(\mathbf{W}) \cap \text{col}(\mathbf{X}) = \{\mathbf{0}\}$, then $\lim_{\rho \rightarrow 1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\rho)\mathbf{X}$ is nonsingular, which in turn implies that the numerator of (A.8) vanishes as $\rho \rightarrow 1$. ■

Acknowledgements

I would like to thank a referee for helpful comments, and Rocco Mosconi for discussion.

References

- Baltagi, B.H., Liu, L., 2010. Spurious spatial regression with equal weights. *Stat. Probabil. Lett.* 80, 1640-1642.
- Baran, S., Pap, G., van Zuijlen, M.C.A., 2004. Asymptotic inference for a nearly unstable sequence of stationary spatial AR models. *Stat. Probabil. Lett.* 69, 53-61.
- Bhattacharyya, B.B., Richardson, G.D., Franklin, L.A., 1997. Asymptotic inference for near unit roots in spatial autoregression. *Ann. Statist.* 25, 1709-1724.
- Bloomfield, P., Watson, G.S., 1975. The inefficiency of least squares. *Biometrika* 62, 121-128.
- Cliff, A.D., Ord, J.K., 1973. *Spatial Autocorrelation*. Pion, London.
- Cordy, C.B., Griffith, D.A., 1993. Efficiency of least squares estimators in the presence of spatial autocorrelation. *Commun. Stat. B-Simul.* 22, 1161-1179.
- Dielman, D.E., Pfaffenberger, R.C., 1989. Efficiency of ordinary least squares for linear models with autocorrelation. *J. Amer. Statist. Assoc.* 84, 248.
- Dietzenbacher, E., 1996. An algorithm for finding block-triangular forms. *Appl. Math. Comput.* 76, 161-171.
- Fingleton, B., 1999. Spurious spatial regression: Some Monte Carlo results with a spatial unit root and spatial cointegration. *J. Reg. Sci.* 39, 1-19.
- Horn, R., Johnson, C.R., 1985. *Matrix Analysis*. Cambridge University Press.
- Jeske, R., Song, C.R., 2003. Relative efficiency of OLSE and COTE for seasonal autoregressive disturbances. *Stat. Pap.* 44, 421-432.
- Kato, T., 1995. *Perturbation Theory for Linear Operators*. Springer.
- Kelejian, H.H., Prucha, I.R., Yuzevovich, Y., 2006. Estimation problems in models with spatial weighting matrices which have blocks of equal elements, *Journal of Regional Science* 46, 507-515.
- Kleiber, C., 2001. Finite sample efficiency of OLS in linear regression models with long-memory disturbances. *Econ. Lett.* 72, 131-136.
- Krämer, W., 1980. Finite sample efficiency of Ordinary Least Squares in the linear regression model with autocorrelated errors. *J. Amer. Statist. Assoc.* 75, 1005-1009.
- Krämer, W., 1984. High correlation among errors and the efficiency of ordinary least squares in linear models. *Stat. Pap.* 25, 135-142.
- Krämer, W., Donniger, C., 1987. Spatial autocorrelation among errors and relative efficiency of OLS in the linear regression model. *J. Amer. Statist. Assoc.* 82, 577-579.
- Krämer, W., 2005. Finite sample power of Cliff-Ord-type tests for spatial disturbance correlation in linear regression. *J. Statist. Plann. Infer.* 128, 489-496.
- Krämer, W., Baltagi, B., 1996. A general condition for an optimal limiting efficiency of OLS in the general linear regression model. *Econ. Lett.* 50, 13-17.
- Lee, L.F., Yu, J., 2008. Near unit root in the spatial autoregressive model. Manuscript, The Ohio State University.
- Lee, L.F., Yu, J., 2009. Spatial nonstationarity and spurious regression: the case with row-normalized spatial weights matrix. *Spat. Econ. Anal.* 4, 301-327.
- Liu, S., 2000. Efficiency comparisons between the OLSE and the BLUE in a singular linear model. *J. Statist. Plann. Infer.* 84, 191-200.
- Liu, X., Lee, L.F. 2010. GMM estimation of social interaction models with centrality, *Journal of Econometrics* 159, 99-115.

- Martellosio, F., 2010. Power properties of invariant tests for spatial autocorrelation in linear regression. *Econometric Th.* 26, 152-186.
- Paulauskas, V., 2007. On unit roots for spatial autoregressive models. *J. Multivariate Anal.* 98, 209-226.
- Rao, C.R., 1967. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, eds. L.M. Le Cam and J. Neyman, Berkeley: University of California Press, 355-372.
- Sinkhorn, R., 1968. Two results concerning doubly stochastic matrices. *Amer. Math. Monthly* 75, 632-634.
- Song, S.H., Trenkler, G., 2001. On the efficiency of the Cochrane-Orcutt estimator in the serially correlated error components regression model for panel data. *Commun. Stati. A-Theor.* 30, 195-207.
- Stemann, D., Trenkler, G., 2000. Some further results on the efficiency of the Cochrane-Orcutt-estimator. *J. Statist. Plann. Infer.* 88, 205-214.
- Stewart, G.W., Sun, J.-G., 1990. *Matrix Perturbation Theory*. Academic Press.
- Tilke, C., 1993. The relative efficiency of OLS in the linear regression model with spatially autocorrelated errors. *Stat. Pap.* 34, 263-270.