

Review

A Survey of Spatial Unit Roots

Badi H. Baltagi ^{1,*}  and Junjie Shu ²

¹ Center for Policy Research and Department of Economics, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244-1020, USA

² Department of Economics, Syracuse University, Syracuse, NY 13244-1020, USA; jshu03@syr.edu

* Correspondence: bbaltagi@syr.edu; Tel.: +1-315-443-3114

Abstract: This paper conducts a brief survey of spatial unit roots within the context of spatial econometrics. We summarize important concepts and assumptions in this area and study the parameter space of the spatial autoregressive coefficient, which leads to the idea of spatial unit roots. Like the case in time series, the spatial unit roots lead to spurious regression because the system cannot achieve equilibrium. This phenomenon undermines the power of the usual Ordinary Least Squares (OLS) method, so various estimation methods such as Quasi-maximum Likelihood Estimate (QMLE), Two Stage Least Squares (2SLS), and Generalized Spatial Two Stage Least Squares (GS2SLS) are explored. This paper considers the assumptions needed to guarantee the identification and asymptotic properties of these methods. Because of the potential damage of spatial unit roots, we study some test procedures to detect them. Lastly, we offer insights into how to relax the compactness assumption to avoid spatial unit roots, as well as the relationship between spatial unit roots and other models, such as the Spatial Dynamic Panel Data (SDPD) model and Lévy–Brownian motion.

Keywords: spatial correlation; spatial unit roots; nonstationarity; spurious spatial regression; panel data

MSC: 62-02



Citation: Baltagi, B.H.; Shu, J. A Survey of Spatial Unit Roots.

Mathematics **2024**, *12*, 1052. <https://doi.org/10.3390/math12071052>

Academic Editor: Manuel Alberto M. Ferreira

Received: 26 January 2024

Revised: 26 March 2024

Accepted: 29 March 2024

Published: 31 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is an extensive literature using spatial statistics that deals with cross-sectional correlation, and is popular in regional science, urban economics and geography, to mention a few; see Anselin [1] for a nice introduction to this literature. Unlike time-series, there is typically no unique natural ordering for cross-sectional data. Spatial dependence models may use a metric of economic distance that provides cross-sectional data with a structure similar to that provided by the time index in time series. Examples in economics usually involve spillover effects or externalities due to geographical proximity. For example, the productivity of public capital, like roads and highways, on the output of neighboring states. Also, the pricing of welfare in one state that pushes recipients to other states. In a linear regression model, this spatial correlation may be in the disturbances and is called the spatial error model (SEM), or modeled on the dependent variable itself and named the spatial autoregression (SAR) model, or the spatial lag model. Unlike the autoregressive model or lagged model in time series where there is a natural ordering across time and lagged values are well defined, in cross-sections, this is dealt with using neighbors whose shocks or disturbances are affected by their neighbor's shocks or disturbances in the SEM. For the SAR, the dependent variable is affected by neighbors, like house price being affected by neighboring house prices. This leads to the construction of a weight matrix that defines one's neighbors by distance or contagion; see Anselin [1]. So, rather than lagged house prices as in an autoregressive time series model, own house price is related to a weighted average of neighboring house prices. Spatial dependence has been extended from cross-section to panel data; see Chapter 13 of Baltagi [2] or Elhorst [3] for a textbook treatment of the subject.

OLS yields inconsistent estimates in the SAR model due to the endogeneity of the spatial lagged dependent variable and the disturbances. For the SEM model, OLS yields an unbiased but inefficient estimator. Because of the limitation of OLS, (Q)MLE is often used to estimate the spatial models as in Ord [4], Anselin [1] and Lee [5]. However, MLE is sometimes computationally intensive especially for large sample sizes because of the requirement to compute the Jacobian term in the likelihood function. Ord [4] proposes a simplified computational procedure only requiring the eigenvalues of the spatial weight matrix, but computing accurate eigenvalues is increasingly difficult for large n . Kelejian and Prucha [6] suggest a Generalized Method of Moments (GMM) estimator for the SEM with SAR structure. Alternatively, constructing instrumental variables (IVs) from the exogenous variables, Kelejian and Prucha [7] propose the GS2SLS estimator and Lee [8] discuss the best GS2SLS estimator by using more efficient IVs. Lee [9] considers a GMM estimator for the SAR models with exogenous variables and showed it is more efficient than the 2SLS estimator and is as efficient as the ML estimator asymptotically. Spatial panel data models (with dynamic terms) are also estimated using MLE, 2SLS or GMM as in Yu et al. [10], Baltagi and Liu [11], Kapoor et al. [12].

However, for all of these estimation methods considered above, they constrain the parameter space of the spatial coefficient to limit the degree of correlation between units. This is because when such spatial correlation is too strong, the spatial echoes passing through each unit do not die out and so the system cannot achieve an equilibrium. When the spatial weight matrix is row-normalized by convention, such a constraint requires the spatial coefficient to be smaller than 1 in absolute value. In this survey, we summarize the developments that relax this constraint and allow the spatial coefficient to equal or sufficiently approach unity, which is known as the spatial unit root. This is of practical relevance because there are many cases where the spatial coefficients are close to 1. For example, Keller and Shiue [13] detect the inter-regional trade of Chinese rice and find that rice prices for different provinces are highly related with spatial coefficients lying between 0.9 to 0.95.

When (near) spatial unit roots exist, the standard estimation procedures are not necessarily reliable and statistical inference is invalid. To remedy such a case, Fingleton [14] avoids the circularity of the spatial weight matrix and conducts a Monte Carlo simulation to explore the performance of OLS estimation. Alternatively, Lee and Yu [15] artificially let the spatial autoregression coefficient sufficiently approach the unit roots and derive the asymptotic behavior of QMLE and 2SLS estimators. The spatial unit roots have also been generalized to Spatial Dynamic Panel Data (SDPD) model by Yu and Lee [16].

In order to investigate the possible spatial unit roots, several test procedures have been proposed. Fingleton [14] suggests a “very high” value of the Moran’s I statistic could be useful for testing spatial unit roots. Lauridsen and Kosfeld [17,18] propose a two-stage Lagrange Multiplier (LM) tests that distinguish the spatial unit roots from the stationary positive spatial correlation. By the fact that spatial impulses do not die out under spatial unit roots and lead to explosive variance, Beenstock et al. [19] numerically calculate the critical value even under an irregular spatial weight matrix. These tests have been extensively used in the literature; see Yesilyurt and Elhorst [20], Olejnik [21], Machado et al. [22], Beenstock and Felsenstein [23].

This paper is organized as follows. Section 2 introduces some basic concepts in spatial econometrics. Next, the parameter space of the spatial autoregressive coefficient and the corresponding singular points are considered. Stationarity and spatial cointegration concepts derived from spatial unit roots are introduced. General assumptions and the corresponding implications in spatial econometrics are discussed. Section 3 introduces the potential problems with the existence of spatial unit roots: spurious and nonsense regression. Section 4 investigates estimation methods and inference under spatial unit roots. Section 5 discusses how to test for spatial unit roots. Section 6 discusses spatial unit roots in the SAR model while Section 7 concludes.

2. Basic Concepts in Spatial Econometrics

Spatial models study the spatial dependence between units. In practice, the spatial weight matrix is used to describe such dependence. Let W_n^* be the $n \times n$ contiguity-based spatial weight matrix, i.e., $w_{ij}^* = 1$ if unit i and j are contiguous and 0 otherwise. Also, the diagonal elements are set to 0 by convention. In practice, the spatial weight matrix is generally row-normalized by $w_{n,ij} = \frac{w_{n,ij}^*}{\sum_j w_{n,ij}^*}$, so the row sum of W_n will be one.

Different spatial model specifications have different implications. The SEM with spatial autoregressive (SAR) structure on the $n \times 1$ error vector u_n can be expressed as $u_n = \lambda_1 W_n u_n + \epsilon_n = (I_n - \lambda_1 W_n)^{-1} \epsilon_n$, where λ_1 is known as the spatial coefficient and satisfies some assumptions that will be introduced later, and ϵ_n is the $n \times 1$ independent and identically distributed (i.i.d.) innovations with variance σ_ϵ^2 . The error covariance matrix for the u_n with SAR structure is

$$\Omega_{SAR} = \mathbb{E}[u_n u_n'] = \sigma_\epsilon^2 (I_n - \lambda_1 W_n)^{-1} (I_n - \lambda_1 W_n')^{-1} = \sigma_\epsilon^2 (B_n' B_n)^{-1}, \tag{1}$$

where $u_n' = \epsilon_n' (I_n - \lambda_1 W_n)^{-1}$ and $B_n = I_n - \lambda_1 W_n$. Though W_n may be sparse, $(B_n' B_n)^{-1}$ is not necessarily so, thus the spatial covariance structure induced by such SEM model is classified as *global*. Conversely, a spatial moving average (SMA) specification for the error vector u_n can be expressed as $u_n = \lambda_2 W_n \epsilon_n + \epsilon_n = (I_n + \lambda_2 W_n) \epsilon_n$, and the corresponding covariance matrix will be

$$\Omega_{SMA} = \mathbb{E}[u_n u_n'] = \sigma_\epsilon^2 [I_n + \lambda_2 (W_n + W_n') + \lambda_2^2 W_n W_n'], \tag{2}$$

including only W_n and $W_n W_n'$ which are first and second order neighbors if W_n is defined as first-order contiguity. Hence, such a model is generally classified as *local*. See Baltagi et al. [24].

Kelejian and Prucha [7] consider a “cross-sectional (first-order) autoregressive spatial model with (first-order) autoregressive disturbances” (SARAR) and is labeled as spatial autoregressive combined (SAC) model. If the right-hand side includes both the independent variable and the spatially lagged dependent variable then it is termed the mixed regressive, spatial autoregression (MRSAR) or mixed SAR model. The spatial Durbin model (SDM) includes both the spatial lagged dependent variable and independent variables. A full model labeled as general nesting spatial (GNS) model given in Elhorst [3] is

$$\begin{aligned} Y_n &= \lambda_1 W_n Y_n + \alpha \iota_n + X_n \beta + W_n X_n \theta + u_n, \\ u_n &= \lambda_2 W_n u_n + \epsilon_n, \end{aligned} \tag{3}$$

where λ_1 is referred as spatial autoregressive coefficient (SAC) and λ_2 is called the spatial autocorrelation coefficient.

2.1. Parameter Space of the Spatial Autoregressive Coefficient

Consider the pure SAR model with data generating process (DGP):

$$Y_n = \lambda W_n Y_n + \epsilon_n, \tag{4}$$

where Y_n is an $n \times 1$ vector of observations on the dependent variable, W_n is an $n \times n$ spatial weight matrix and ϵ_n is an $n \times 1$ vector of disturbances which are assumed to be i.i.d. $(0, \sigma_\epsilon^2)$. The reduced form equation of Y_n can be written as

$$Y_n = S_n^{-1} \epsilon_n, \tag{5}$$

where $S_n = I_n - \lambda W_n$. So to guarantee the system achieves equilibrium, a crucial assumption is that the absolute value of the spatial coefficient λ is strictly less than 1 (see Kelejian and Prucha [6] (Assumption 2), Kelejian and Prucha [7] (Assumption 2)) to ensure the non-

singularity of S_n . This assumption follows from a sufficient condition for the invertibility matrix in Horn and Johnson [25] (Corollary 5.6.16, p. 351):

Theorem 1. *An $n \times n$ matrix A_n is nonsingular if there exists a matrix norm $\|\cdot\|$ such that $\|I_n - A_n\| < 1$. If this condition is satisfied, $A_n^{-1} = \sum_{k=0}^{\infty} (I_n - A_n)^k$.*

Thus, S_n is invertible if there exists a matrix norm such that $\|I_n - (I_n - \lambda W_n)\| = \|\lambda W_n\| < 1$. It is also well known that any norm of a matrix is larger than all of its eigenvalues. Let X_n be the eigenvector matrix and $\rho_i, i = 1, \dots, n$, be the eigenvalues of W_n , then

$$|\rho_i| \|X_n\| = \|\rho_i X_n\| = \|W_n X_n\| \leq \|W_n\| \|X_n\|. \tag{6}$$

So it is easy to see that $|\lambda| < \frac{1}{\|W_n\|}$ and therefore

$$\frac{1}{\rho_{min}} < \lambda < \frac{1}{\rho_{max}}, \tag{7}$$

because $\sum_i^n \rho_i = \text{tr}(W_n) = 0$ so that $\rho_{min} < 0 < \rho_{max}$. A useful result is given in Ord [4]:

Theorem 2. *If W_n has eigenvalues ρ_1, \dots, ρ_n , $|\rho I_n - W_n| = \prod_{i=1}^n (\rho - \rho_i)$. Then for $S_n = I_n - \lambda W_n$, $\det(S_n) = |I_n - \lambda W_n| = \prod_{i=1}^n (1 - \lambda \rho_i)$.*

Moreover, the log-likelihood function for Y_n , given $Y_n = y$ in (4) is

$$\ell(\lambda, \sigma^2) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} y' S_n' S_n y + \ln |S_n|, \tag{8}$$

and $|S_n| = \prod_{i=1}^n (1 - \lambda \rho_i) > 0$, of which a sufficient condition is $\lambda \rho_i < 1$, for all i . Again, since $\rho_{min} < 0 < \rho_{max}$, we obtain the range of λ that is given in (7).

However, either by Theorem 1 or 2, (7) is a sufficient condition for the invertibility of S_n . The singular points of S_n are $\frac{1}{\rho_1}, \dots, \frac{1}{\rho_n}$ by Theorem 2, and the number of these singular points are at most countably many as $n \rightarrow \infty$. This raises a problem as stated in Kelejian and Robinson [26]. These singular points can be determined generally by the n th polynomial numerically, and to avoid inconsistency, they should be removed from the possible values of λ . Griffith [27] (p. 19) states that this condition also ensures stationarity, but Kelejian and Robinson [26] give a counter example showing that it does not when W_n is a row-normalized double queen weight matrix.

When the matrix is row-normalized, Kelejian and Prucha [6] (Note 8, p. 120) show that $\rho_{max} = 1$ by Geršgorin’s theorem, and typically, $|\rho_{min}| < 1$ [26]. We assume $|\lambda| < 1$, and this is why λ is generally interpreted as the spatial autocorrelation coefficient similar to its counterpart in time series.

2.2. Stationarity, Order of Integration and Cointegration

Stationarity is a key assumption in time series. Similarly, in spatial econometrics, when the stationary assumption does not hold, spurious or nonsense regression appears, as will be shown in the next section. Stationarity is tightly connected with S_n . The formal definition of stationarity is given in Anselin [1] (p. 42):

Definition 1. *A process is strictly stationary if any finite subset $\{x_i, x_j, \dots, x_n\}$ from the stochastic process $\{x_i, i \in I\}$ has the same joint distribution as the subset $\{x_{i+s}, x_{j+s}, \dots, x_{n+s}\}$ for any s , where s represents an uniform shift in time, space or time-space.*

But we generally consider a weaker version, covariance stationarity. For the intuition of stationarity and the connection with the inverse of S_n , see Beenstock et al. [19]. Consider the pure SAR model in (4) and (5), the covariance matrix for Y_n is

$$\text{Var}(Y_n) = \mathbb{E}[Y_n Y_n'] = \sigma_\epsilon^2 S_n^{-1} S_n'^{-1} \equiv \sigma_\epsilon^2 \mathcal{B}_n, \tag{9}$$

where $\mathcal{B}_n = S_n^{-1} S_n'^{-1}$ and S_n^{-1} is defined in (5). Letting b_{kj} be the (k, j) element of the matrix \mathcal{B}_n , by normalizing $\sigma_\epsilon^2 = 1$, Y_k has variance b_{kk} and covariance b_{kj} with Y_j . By Definition 1, stationarity requires that b_{kk} and b_{kj} remain unchanged asymptotically (this implicitly assumes that both location j and k are far away from the edge). Note that by Theorem 1, we have

$$S_n^{-1} = \sum_{k=0}^{\infty} (I_n - S_n)^k = \sum_{k=0}^{\infty} (\lambda W_n)^k = I_n + \lambda W_n + \lambda^2 W_n^2 + \dots \tag{10}$$

So the stationarity assumption is equivalent to $\lambda^{m-k} W_n^{m-k} \rightarrow 0$, for all k as $m \rightarrow \infty$. Since $m < n$ and $m \rightarrow \infty$, m represents the “remote” area, and W_n^{m-k} is the $m - k$ step neighborhood of unit k . Thus, intuitively, stationarity requires that the shocks from far away locations will asymptotically not affect the epicenter area.

Another two concepts tightly connected with unit roots and stationarity are the order of integration and cointegration. The order of integration is originally a concept in time series that describes the minimum number of differences that a non-stationary process needs to be (covariance) stationary. Cointegration, on the other hand, describes the minimum order of integration of a combination of two or more series with the same order of integration. The formal definitions of the order of integration and cointegration are given in Hamilton [28]:

Definition 2. A time series is integrated of order d , denoted $I(d)$, if $(1 - L)^d X_t$ is a stationary process, where L is the lag operator and $1 - L$ is the first difference.

Definition 3. Time series X and Y are cointegrated of order $I(d, b)$, if both of them are $I(d)$, and there exists a cointegrated vector (a, b) such that $aX + bY \sim I(d - b)$.

In time series, the lag operator L is defined by $LX_t = X_{t-1}$ because of the natural order of temporal data. In spatial econometrics, we regard W_n as the spatial lag operator and $I_n - W_n$ as the first order spatial difference, see Anselin [1] (pp. 22–26). We also use $I(d)$ and $I(a, b)$ to refer to spatial integration and spatial cointegration respectively.

More specifically, for the pure SAR model in (4) with a row-normalized weight matrix, if $\lambda = 1$, $Y_n \sim I(1)$ since ϵ_n is stationary. Also, suppose both Y_n and X_n are $I(1)$, but they have a long-term equilibrium relationship $Y_n = X_n \beta + \epsilon_n$, then obviously, (X_n, Y_n) are $I(1, 1)$ with cointegrated vector $(-\beta, 1)$.

2.3. Some Fundamental Assumptions

Different assumptions are made in spatial econometrics for different estimation methods. The most common ones are listed here, and the implications are explained.

Assumption 1. The disturbances $\epsilon_1, \dots, \epsilon_n$ are i.i.d. for all n (so uniformly) with zero mean and finite variance σ^2 . Additionally, fourth moments exist.

Assumption 2. The elements of the exogenous variables X_n are uniformly bounded for all n . The $\lim_{n \rightarrow \infty} \frac{X_n' X_n}{n}$ exists and is nonsingular.

Assumption 3. The matrix S_n is nonsingular.

The existence of up to the fourth moment of disturbances is needed to apply the central limit theorem for (a system) of the linear-quadratic form (see Kelejian and Prucha [29])

(p. 226) and Kelejian and Prucha [30] (p. 63)). The nonsingularity of S_n makes sure the system achieves an equilibrium as well as ensuring that the mean and variance of Y_n exist.

Assumption 4. *The matrices W_n and $S_n^{-1} = (I_n - \lambda W_n)^{-1}$ are uniformly bounded (UB) in both row and column sums for all n . (We say a matrix is UB in row (column) sums if its maximum row (column) sum is finite. This property preserves under finite matrix multiplication.)*

The UB condition for W_n implicitly assumes a limited number of neighbors for all units even as $n \rightarrow \infty$, so the weight matrix W_n is sparse for large n . This assumption is relaxed in Lee et al. [31] by introducing dominant (popular) units. In practice, the spatial units have a limited number of neighbors. Though sometimes $w_{n,ij}$ may be defined as the inverse of the distance between i, j physically or economically, $w_{n,ij}$ tends to be 0 between far away units as n increases. So in general this assumption is satisfied. The UB of $\{S_n^{-1}\}$ is to ensure the covariance matrix $\text{Var}(Y_n)$ in (9) is still UB, which limits the correlation between two different units since the UB property preserves under matrix multiplication.

Other assumptions to ensure identification conditions or the derivation of asymptotic distributions of estimators will be mentioned when needed.

3. Spurious Regression When (Near) Unit Roots Exist

The variance of Y_n explodes when unit roots exist, and OLS estimation may perform unsatisfactorily: the estimators are inconsistent, the test statistics do not have familiar distributions, and may even converge to a constant. These phenomena have been studied extensively in time series, and similar symptoms occur in spatial econometrics.

3.1. Spurious Regression of Driftless Series and Spatial Integration

Fingleton [14] studies unit roots and spatial cointegration in spatial econometrics. Using Monte Carlo simulations, he finds that spatial unit roots will lead to a spurious regression and proves that when two vectors are spatially cointegrated, even running a regression on the error-correction model yields inconsistent estimates. Beenstock et al. [19] distinguish between the terms spurious regression and nonsense regression and argue that, Fingleton [14] refers to nonsense regression instead of spurious regression. When Y_n and X_n are driftless random walks, the nonsense regression occurs because of the increased variances of Y_n and X_n over time. On the other hand, the spurious regression occurs when Y_n and X_n are independent random walks with drift, which causes their means to increase over time. See also Mur and Trivez [32].

To run the simulation, two independent pure SAR processes Y_n and X_n containing spatial unit roots are generated separately as in (5). But as discussed in Section 2.1, S_n^{-1} does not exist under a row-normalized weight matrix when $\lambda = 1$. To avoid this, Fingleton introduces the “unconnected central cell”, which manually sets one row of the spatial weight matrix equal to 0 to avoid circularity. This is a time-series analogy because there is always a starting point in temporal data ($t = 1$). By doing so, the singular point is slightly larger than 1, and the existence of S_n^{-1} is ensured [32]. Regressing X_n on Y_n , the t -statistic and coefficient of determination R^2 show the significance of the parameter between two unrelated variables when spatial unit roots exist. Letting e be the OLS residuals, Moran’s I , defined as

$$\begin{aligned}
 I_{Moran} &= \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{n,ij}} \frac{e' W_n e}{e' e} \\
 &= \frac{e' W_n e}{e' e} \quad (\text{when } W_n \text{ is row-normalized}),
 \end{aligned}
 \tag{11}$$

is the spatial version of the Durbin–Watson statistic, and thus is a measure of spatial autocorrelation. The simulation results of Moran’s I show a high level of positive spatial autocorrelation in the residuals and evidence for the presence of a spurious regression.

To remedy this situation, spatial differencing is introduced to the SAR process with unit roots:

$$\Delta Y_n = \gamma \Delta X_n + \epsilon_n, \tag{12}$$

where $\Delta Y_n = Y_n - W_n Y_n$ and $\Delta X_n = X_n - W_n X_n$. When both Y_n and X_n are spatial $I(1)$ processes, we have $\Delta Y_n = \epsilon_Y$ and $\Delta X_n = \epsilon_X$. The regression of the first-order spatial difference variable is equivalent to regressing two independent $I(0)$ processes, which should theoretically yield $\hat{\gamma} = 0$.

Next, spatially cointegrated series are considered. To generate $(X_n, Y_n) \sim I(1, 1)$, the “error-correction representation” is used. The idea is adopted from Robert [33] that the existence of error-correction representation is a necessary and sufficient condition for a cointegrated time series. The spatial analogy is

$$\begin{aligned} Y_n &= W_n Y_n + c(W_n X_n - W_n Y_n) + e_{1n}, \\ X_n &= W_n X_n + d(W_n X_n - W_n Y_n) + e_{2n}, \\ e_{1n} &\sim N(0, \sigma_1^2 I_n), \\ e_{2n} &\sim N(0, \sigma_2^2 I_n), \end{aligned} \tag{13}$$

where $W_n X_n - W_n Y_n$ is the equilibrium error assumed (for simplicity) stationary and hence the name “error-correction”. The spatial unit root series X_n and Y_n have a long-term equilibrium $X_n = Y_n$. Note that (13) has two equations and two unknowns and W_n is a noncircular matrix.

Moran’s I statistic may act as a useful indicator for cointegration because the cointegration regression (regress Y_n on X_n) involves endogenous variables. Also, the first-order regression is inappropriate because of omitted variable bias concerning the equilibrium error $W_n X_n - W_n Y_n$. Rearranging (13) yields the appropriate specifications:

$$\begin{aligned} Y_n - W_n Y_n &= c(W_n X_n - W_n Y_n) + e_{1n}, \\ X_n - W_n X_n &= d(W_n X_n - W_n Y_n) + e_{2n}. \end{aligned} \tag{14}$$

But OLS estimation for either c or d is inconsistent because of the presence of a spatially lagged dependent variable, which is different from the traditional time series counterpart.

3.2. Spurious Regression with Deterministic Trends

Fingleton [14] studies the effect of spatial unit roots by simulation while Mur and Trávez [32] show that the variance of the spatial unit roots series explodes. For the DGP in (4)

$$Y_n = \lambda W_n Y_n + \epsilon_n \implies Y_n = (I_n - \lambda W_n)^{-1} \epsilon_n = S_n^{-1} \epsilon_n, \tag{15}$$

since the contiguity-based spatial weight matrix is symmetric, W_n can be decomposed as $W_n = Q_n \Lambda_n Q_n^{-1}$ no matter whether it is row-normalized or not. Thus $S_n = I_n - \lambda W_n = Q_n (I_n - \lambda \Lambda_n) Q_n^{-1} = Q_n \Delta_n Q_n^{-1}$ where Λ_n is the eigenvalue matrix of W_n and thus Δ_n is also diagonal. So, Mur and Trávez [32] derive the variance of Y_n as

$$\begin{aligned} \text{Var}(Y_n) &= \sigma^2 (B_n' B_n)^{-1} \\ &= \sigma^2 Q_n \Delta_n^{-1} Q_n^{-1} (Q_n')^{-1} \Delta_n^{-1} Q_n' \\ &= \sigma^2 Q_n \Delta_n^{-1} (Q_n' Q_n)^{-1} \Delta_n^{-1} Q_n', \end{aligned} \tag{16}$$

with

$$Q_n \Delta_n^{-1} = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix} \text{diag} \left\{ \frac{1}{1-\lambda\rho_1}, \dots, \frac{1}{1-\lambda\rho_n} \right\} = \begin{bmatrix} \frac{q_{11}}{1-\lambda\rho_1} & \cdots & \frac{q_{1n}}{1-\lambda\rho_n} \\ \vdots & \ddots & \vdots \\ \frac{q_{n1}}{1-\lambda\rho_1} & \cdots & \frac{q_{nn}}{1-\lambda\rho_n} \end{bmatrix}. \tag{17}$$

Let m_{ij} be the element of row i and column j of the matrix $(Q_n' Q_n)^{-1}$, and the variance of the r -th element of Y_n is then

$$\text{Var}(Y_{n,r}) = \sigma^2 \sum_i^n \sum_j^n \frac{q_{ir} q_{jr}}{(1-\lambda\rho_i)(1-\lambda\rho_j)} m_{ij}. \tag{18}$$

If W_n is row-normalized, at least one $\rho_i = 1$, which means that if the spatial unit roots exist and $\lambda = 1$, $\text{Var}(Y_n)$ explodes. If W_n is not row-normalized so that Q_n is symmetric and orthogonal with $Q_n' Q_n = I_n$, $\text{Var}(Y_n) = \sigma^2 Q_n \Delta_n^{-2} Q_n$, then Mur and Trívéz [32] show the variance of the observation at r reduces to

$$\text{Var}(Y_{n,r}) = \sigma^2 \sum_{i=1}^n \frac{q_{ir}^2}{(1-\lambda\rho_i)^2} = \infty, \quad \text{when } \lambda = \frac{1}{\rho_i}. \tag{19}$$

But when λ is not one of the singular points of S_n , $\text{Var}(Y_n)$ is not necessarily a function of n , i.e., the variances of Y_n do not increase as the sample size grows [32]. This is in line with the discussion of stationarity in Section 2.2 and reveals the possible source for nonsense regression concerned with the spatial unit root SAR series [34] (p. 303).

Mur and Trívéz [32] focus on the spurious regression when a spatial deterministic trend exists and show that under such circumstances similar symptoms related to unit roots occur. Consider the DGP

$$Y_n = \delta t_n + \lambda W_n Y_n + \epsilon_n. \tag{20}$$

$$\implies Y_n = (I_n - \lambda W_n)^{-1} (\delta t_n + \epsilon_n) \tag{21}$$

$$= \delta t_n^\ddagger + S_n^{-1} \epsilon_n, \tag{22}$$

where t_n is an $n \times 1$ unit vector and $t_n^\ddagger = S_n^{-1} t_n$. Comparing the term δt_n^\ddagger in (22) with the time trends in a typical time series model, " t_n^\ddagger " is similar to the time trend " t ": t is different in terms of the relative position in time and the element of t_n^\ddagger is different in terms of its relative position in space. Also, the presence of such a trend term in the SAR process leads to spurious regression. Consider the simple regression

$$Y_{1n} = \alpha + Y_{2n} \beta + \mu_n, \tag{23}$$

where Y_{1n} and Y_{2n} are unrelated SAR processes generated by (20), respectively

$$Y_{1n} = \delta_1 t_n + \lambda_1 W_{n1} Y_{1n} + \epsilon_{1n} \implies Y_{1n} = \delta_1 S_{1n}^{-1} t_n + S_{1n}^{-1} \epsilon_{1n}, \tag{24}$$

$$Y_{2n} = \delta_2 t_n + \lambda_2 W_{n2} Y_{2n} + \epsilon_{2n} \implies Y_{2n} = \delta_2 S_{2n}^{-1} t_n + S_{2n}^{-1} \epsilon_{2n}.$$

Assuming ϵ_{1n} and ϵ_{2n} are independent white noises, we expect the estimate of β in (23) to be 0. However, this is generally not the case which means spurious regression occurs. This can be seen from the fact that the correlation coefficient between Y_{1n} and Y_{2n} given in Mur and Trívéz [32]

$$r_{Y_{1n}, Y_{2n}} = \frac{\sum_r (Y_{1n,r} - \bar{Y}_{1n})(Y_{2n,r} - \bar{Y}_{2n})}{\sqrt{\sum_r (Y_{1n,r} - \bar{Y}_{1n})^2 \sum_r (Y_{2n,r} - \bar{Y}_{2n})^2}} \rightarrow 1, \quad n \rightarrow \infty. \tag{25}$$

Though Fingleton argues that a high value of Moran’s I statistics may be a good indicator for the existence of spatial unit roots and spatial cointegration, he cannot distinguish between them, or even from the (genuine) positive spatial autocorrelation case. Some testing methods are developed and summarized in Section 5. The trend SAR series proposed by Mur and Trívez [32] seems to receive less attention, which may be due to the fact that when the mixed SAR process contains only a constant exogenous variable and W_n is row-normalized, multicollinearity occurs; see Kelejian and Prucha [7] (p. 105), Lee [35] (p. 258), Lee [5] (p. 1907).

3.3. Spurious Regression under the near Unit Roots with a Row-Normalized, Circular Weight Matrix

The spurious regression considered in the previous two sections is under a row-normalized, noncircular weight matrix, which implicitly assumes an unconnected central unit in the spatial system, and is too restrictive to be used in empirical applications. Thus, Lee and Yu [34] study the spurious regression under a circular, row-normalized spatial weight matrix. The DGP process for the (mixed) SAR series is

$$Y_{jn} = \lambda_{jn}W_{jn}Y_{jn} + Z_{jn}\gamma_j + \epsilon_{jn}, \quad j = 1, \dots, m. \tag{26}$$

In this case, $\lambda_{jn} \neq 1$, since the unit roots are singular points of W_{jn} . They study the consequence when λ_n approaches 1, namely

$$\lambda_{n0} = 1 - \frac{1}{\psi_n}, \tag{27}$$

where $\psi_n \rightarrow \infty$ as $n \rightarrow \infty$.

3.3.1. Decomposition of Y_n

Though the variance of Y_n explodes as $\lambda_n \rightarrow 1$, Y_n can be decomposed into a stable part and an unstable part by the decomposition of the weight matrix W_n . This decomposition given in Lee and Yu [15] is used in Yu et al. [10,36], Yu and Lee [16] to study unit roots in a spatial dynamic panel data (SDPD) model. Because of its importance, this procedure is summarized here. See Baltagi et al. [37] and Lee and Yu [15,34] for more information.

Theorem 3. *Suppose that W_n is a row-normalized weight matrix from a symmetric matrix C_n , i.e., $W_n = \Lambda_n^{-1}C_n$, where Λ_n is a diagonal matrix with its diagonal elements formed by the row sums of C_n . Then (i) the eigenvalues of W_n are all real; and (ii) W_n is diagonalizable.*

(i) can be easily seen from the fact that all symmetric matrices have real eigenvalues. For (ii)

$$W_n = \Lambda_n^{-1}C_n \implies \Lambda_n^{\frac{1}{2}}W_n\Lambda_n^{-\frac{1}{2}} = \Lambda_n^{-\frac{1}{2}}C_n\Lambda_n^{-\frac{1}{2}}. \tag{28}$$

Let D_n^* be the eigenvalue matrix of $\Lambda_n^{-\frac{1}{2}}C_n\Lambda_n^{-\frac{1}{2}}$, and R_n^* be the corresponding orthogonal eigenvector matrix, i.e., $\Lambda_n^{-\frac{1}{2}}C_n\Lambda_n^{-\frac{1}{2}} = R_n^*D_n^*R_n^{*'}$. Lee and Yu [15] show W_n can be expressed as:

$$\begin{aligned} W_n &= \Lambda_n^{-\frac{1}{2}}\left(R_n^*D_n^*R_n^{*'}\right)\Lambda_n^{\frac{1}{2}} \\ &= \left(\Lambda_n^{-\frac{1}{2}}R_n^*\right)D_n^*\left(R_n^{*-1}\Lambda_n^{\frac{1}{2}}\right) \\ &= \left(\Lambda_n^{-\frac{1}{2}}R_n^*\right)D_n^*\left(\Lambda_n^{-\frac{1}{2}}R_n^*\right)^{-1}. \end{aligned} \tag{29}$$

Let $R_n = \Lambda_n^{-\frac{1}{2}} R_n^*$, $D_n = D_n^*$. By definition, R_n is the eigenvector of W_n and $D_n = D_n^*$ is the corresponding eigenvalue, so the eigendecomposition of W_n is

$$W_n = R_n D_n R_n^{-1}. \tag{30}$$

Moreover, the largest eigenvalues of a row-normalized matrix are 1 in absolute value. Without loss of generality, Lee and Yu [15] assume there are m_n eigenvalues equal to 1 and let

$$D_n = \text{diag}\{1_{m_n}, d_{n,m_n+1}, \dots, d_{nn}\}, \tag{31}$$

where 1_{m_n} is $1 \times m_n$ vector of ones and $|d_{ni}| \leq 1$ for all i . So the eigenvalue matrix D_n can be decomposed into two parts:

$$D_n = J_n + \tilde{D}_n, \tag{32}$$

where $J_n = \text{diag}\{1_{m_n}, 0, \dots, 0\}$ and $\tilde{D}_n = \text{diag}\{0, \dots, 0, d_{n,m_n+1}, \dots, d_{nn}\}$. Accordingly,

$$\begin{aligned} W_n &= R_n D_n R_n^{-1} = R_n (J_n + \tilde{D}_n) R_n^{-1} \\ &= R_n J_n R_n^{-1} + R_n \tilde{D}_n R_n^{-1} = W_n^u + \tilde{W}_n, \end{aligned} \tag{33}$$

where $W_n^u = R_n J_n R_n^{-1}$ and $\tilde{W}_n = R_n \tilde{D}_n R_n^{-1}$.

Lee and Yu [15] note that $J_n J_n = J_n$ and $J_n \tilde{D}_n = 0$, so

$$W_n^u W_n^u = W_n^u, \quad W_n^u \tilde{W}_n = 0, \quad W_n W_n^u = W_n^u. \tag{34}$$

Denoting $S_n(\lambda) = I_n - \lambda W_n = R_n (I_n - \lambda D_n) R_n^{-1}$, and $S_n = S_n(\lambda_{n0})$ where λ_{n0} is the true value of λ , they obtain $S_n^{-1}(\lambda) = R_n (I_n - \lambda D_n)^{-1} R_n^{-1}$ and thus

$$I_n - \lambda_{n0} D_n = \text{diag}\{(1 - \lambda_{n0}), \dots, (1 - \lambda_{n0}), (1 - \lambda_{n0} d_{n,m_n+1}), \dots, (1 - \lambda_{n0} d_{nn})\}, \tag{35}$$

$$\implies (I_n - \lambda_{n0} D_n)^{-1} = \text{diag}\left\{\psi_n, \dots, \psi_n, \frac{1}{1 - \lambda_{n0} d_{n,m_n+1}}, \dots, \frac{1}{1 - \lambda_{n0} d_{nn}}\right\}, \tag{36}$$

since $\lambda_{n0} = 1 - \frac{1}{\psi_n}$. Similarly,

$$(I_n - \lambda_{n0} \tilde{D}_n)^{-1} = \text{diag}\left\{1_{m_n}, \frac{1}{1 - \lambda_{n0} d_{n,m_n+1}}, \dots, \frac{1}{1 - \lambda_{n0} d_{nn}}\right\}. \tag{37}$$

Comparing the first m_n diagonals of $(I_n - \lambda_{n0} D_n)^{-1}$ and $(I_n - \lambda_{n0} \tilde{D}_n)^{-1}$, it is easy to obtain $(I_n - \lambda_{n0} D_n)^{-1} = \psi_n \lambda_{n0} J_n + (I_n - \lambda_{n0} \tilde{D}_n)^{-1}$, thus

$$S_n^{-1}(\lambda_{n0}) = \psi_n \lambda_{n0} W_n^u + (I_n - \lambda_{n0} \tilde{W}_n)^{-1}. \tag{38}$$

Denote $G_n = W_n S_n^{-1}$, by (38) and (34):

$$G_n = \psi_n \lambda_{n0} W_n^u + W_n (I_n - \lambda_{n0} \tilde{W}_n)^{-1}. \tag{39}$$

Thus, Lee and Yu [15] decompose $Y_{jn} = S_{jn}^{-1}(\lambda_{jn})(Z_{jn} \gamma_j + \epsilon_{jn})$ into two parts by (38):

$$Y_{jn} = \psi_{jn} Y_{jn}^u + \tilde{Y}_{jn}, \tag{40}$$

where $Y_{jn}^u = \lambda_{jn} W_{jn}^u (Z_{jn} \gamma_j + \epsilon_{jn})$ and $\tilde{Y}_{jn} = (I_n - \lambda_{jn} \tilde{W}_{jn})^{-1} (Z_{jn} \gamma_j + \epsilon_{jn})$.

It can be seen from (38) that when $\lambda_{n0} \rightarrow 1$, $S_n^{-1}(\lambda_{n0})$ is ill-conditioned because $\psi_n \rightarrow \infty$ and hence the variance of Y_n explodes. This is caused by the first unstable term in (40) (the second term is stable). Thus, Y_{jn} is of order ψ_{jn} , which may grow too fast to yield useful asymptotic analysis. Thus, a rate-adjusted factor $\frac{1}{\psi_{jn}}$ is needed to maintain

a controllable rate. A similar idea applies to QMLE and 2SLS methods for estimation as shown later.

3.3.2. Spurious Regression of OLS under Near Unit Root

To study spurious regression, following Fingleton [14], Lee and Yu [34] consider the DGP that is similar to (26) but without exogenous variables:

$$Y_{jn} = \lambda_{jn} W_{jn} Y_{jn} + \epsilon_{jn}, \quad j = 1, \dots, m. \tag{41}$$

Denote $Y_{-1,n} = [Y_{2n}, \dots, Y_{mn}]$, $X_n = [\iota_n, Y_{-1,n}]$ where ι_n is the $n \times 1$ vector of ones. Let λ be a scalar, $\beta = (\beta_2, \dots, \beta_m)'$ be an $(m - 1) \times 1$ vector and $\delta \equiv (\lambda, \beta)'$. OLS, which may yield spurious regression, is then

$$Y_{1n} = \alpha \iota_n + Y_{-1,n} \beta + V_n = X_n \delta + V_n, \tag{42}$$

where V_n is an $n \times 1$ vector of disturbances. To make sure the variable of interest is under controllable order, scale Y_{jn} and $S_{jn}^{-1} S_{jn}'^{-1}$ as

$$Y_{jn}^* = \frac{1}{\psi_{jn}} Y_{jn}, \quad S_{jn} = \frac{1}{\psi_{jn}^2} S_{jn}^{-1} S_{jn}'^{-1}, \quad j = 1, \dots, m. \tag{43}$$

Lee and Yu [34] introduce a sufficient condition for Assumption 4 that ensures the UB of S_n^{-1} , by (38):

Assumption 5. $(I_n - \lambda_{jn} \tilde{W}_{jn})^{-1}$ and W_{jn}^u are UB.

Under Assumption 5, S_{jn} is UB. To study the properties of OLS estimation, it is sufficient to show the asymptotic behaviors of $\frac{1}{\sqrt{n}} X_n^* Y_{1n}^*$ and $\frac{1}{n} Y_{in}^{*'} Y_{jn}^*$, where $X_n^* = [\iota_n, Y_{-1,n}^*]$. Proofs of these properties are about orders of matrices and random vectors as well as first and second moments of quadratic forms (some useful lemmas can be found at <https://www.asc.ohio-state.edu/lee.1777/wp/sar-qml-r-appen-04feb.pdf>, accessed on 28 March 2024). The OLS estimates for α and β : $(\hat{\alpha}_n, \hat{\beta}_n)'$ = $(X_n' X_n)^{-1} (X_n' Y_{1n})$ can be expressed in terms of $\frac{1}{\sqrt{n}} X_n^* Y_{1n}^*$ and $\frac{1}{\sqrt{n}} X_n^* X_n^*$

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix} = \psi_{1n} Y_m^{-1} \left(\frac{1}{n} X_n^* X_n^* \right)^{-1} \left(\frac{1}{\sqrt{n}} X_n^* Y_{1n}^* \right), \tag{44}$$

where

$$Y_m = \begin{pmatrix} 1 & 0 \\ 0 & Y_{2m} \end{pmatrix} \quad \text{and} \quad Y_{2m} = \begin{pmatrix} \psi_{2n} & 0 & \dots & 0 \\ 0 & \psi_{3n} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_{mn} \end{pmatrix}. \tag{45}$$

Lee and Yu [34] notice that the scaling factor Y_m is needed because the columns of X_n have different orders $(1, \psi_{2n}, \dots, \psi_{mn})$ and Y_{1n} is of order ψ_{1n} . Based on this fact, they give the following result:

$$\frac{1}{n} X_n^* X_n^* = D_{n,xx}^* + O_p \left(\frac{1}{\sqrt{n}} \right), \tag{46}$$

where $D_{n,xx}^* \equiv \text{diag} \left\{ 1, \sigma_2^2 \frac{1}{n} \text{tr}(S_{2n}), \dots, \sigma_m^2 \frac{1}{n} \text{tr}(S_{mn}) \right\}$, $O_p \left(\frac{1}{\sqrt{n}} \right)$ means the remaining terms of $\frac{1}{n} X_n^* X_n^*$ are at most of order $\frac{1}{\sqrt{n}}$ and S_{jn} symbols are defined in (43), for all $j = 2, \dots, m$.

Moreover, $\frac{1}{\sqrt{n}}X_n^{*'}Y_{1n}^*$ has limiting variance matrix:

$$\Sigma_m^* = \sigma_1^2 \text{diag} \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} l_n' \mathbb{S}_{1n} l_n, \sigma_2^2 \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(\mathbb{S}_{2n} \mathbb{S}_{1n}), \dots, \sigma_m^2 \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(\mathbb{S}_{mn} \mathbb{S}_{1n}) \right\}. \tag{47}$$

Lee and Yu [34] also adjust Assumption 2 to ensure $\frac{1}{n}X_n^{*'}X_n^*$ is of full rank when $n \rightarrow \infty$ and obtain the asymptotic distributions of the OLS estimators:

Assumption 6. $\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(\mathbb{S}_{jn}) \neq 0$ for $j = 1, 2, \dots, m$.

Denoting

$$D_{xx}^* \equiv \lim_{n \rightarrow \infty} D_{n,xx}^* = \text{diag} \left\{ 1, \sigma_2^2 \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(\mathbb{S}_{2n}), \dots, \sigma_m^2 \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(\mathbb{S}_{mn}) \right\}, \tag{48}$$

then

$$\frac{\sqrt{n}}{\psi_{1n}} Y_m (\hat{\lambda}_n, \hat{\beta}'_n)' \xrightarrow{d} N \left(0, [D_{xx}^*]^{-1} \Sigma_m^* [D_{xx}^*]^{-1} \right). \tag{49}$$

Since $Y_{-1,n}$ is independent of Y_{1n} , one may expect insignificant $\hat{\beta}$ in (42). However, whether $\hat{\beta}_j$ converges to 0 in probability or not depends on the factor $\frac{\sqrt{n}}{\psi_{1n}} \psi_{jn}$ for $2 \leq j \leq m$ (note that Y_m in (49) is diagonal): if $\frac{\sqrt{n}}{\psi_{1n}} \psi_{jn} \rightarrow \infty$, $\hat{\beta}_{jn} \xrightarrow{p} 0$ is $\frac{\sqrt{n}}{\psi_{1n}} \psi_{jn}$ -consistent; if $\frac{\sqrt{n}}{\psi_{1n}} \psi_{jn} \rightarrow c < \infty$, $\hat{\beta}_{jn}$ is asymptotically normal because its limiting variance does not converge to 0; if $\frac{\sqrt{n}}{\psi_{1n}} \psi_{jn} \rightarrow \infty$, $\hat{\beta}_{jn}$ is not bounded in probability and diverges. Intuitively, spurious regression will not occur if ψ_{1n} approaches ∞ faster than ψ_{jn} , or equivalently, λ_{1n} approaches 1 more quickly than λ_{jn} [34].

3.3.3. Other Test Statistics

It is also important to discuss the statistical properties of other test statistics based on (42), which could be potentially useful for distinguishing spurious regression. Lee and Yu [34] give the following theorem as a prerequisite:

Theorem 4. Under Assumptions 1 and 4, for any nonstochastic UB square matrix B_n ,

$$\frac{1}{n} Y_{1n}^{*'} B_n P_n Y_{1n}^* = O_p \left(\frac{1}{n} \right). \tag{50}$$

where $P_n = X_n (X_n' X_n)^{-1} X_n'$ is the projection matrix of X_n

Based on this theorem, Lee and Yu [34] show the order of the estimated variance of the disturbances $\hat{\sigma}_n^2$ is

$$\frac{1}{\psi_{1n}^2} \hat{\sigma}_n^2 = \frac{1}{\psi_{1n}^2} \left(\frac{e_n' e_n}{n - m} \right) = \sigma_1^2 \text{tr}(\mathbb{S}_{1n}) + O_p \left(\frac{1}{\sqrt{n}} \right), \tag{51}$$

and

$$t_{\beta_j} \xrightarrow{d} N \left(0, \lim_{n \rightarrow \infty} \frac{n \cdot \text{tr}(\mathbb{S}_{jn} \mathbb{S}_{1n})}{\text{tr}(\mathbb{S}_{jn}) \cdot \text{tr}(\mathbb{S}_{1n})} \right), \quad \text{for every } 2 \leq j \leq m, \tag{52}$$

$$F \xrightarrow{d} \frac{1}{(m - 1)} U_m' U_m, \tag{53}$$

where U_m is an $m - 1$ random vector with its elements $u_{jm} \sim N\left(0, \lim_{n \rightarrow \infty} \frac{n \text{tr}(S_{jn} S_{1n})}{\text{tr}(S_{jn}) \text{tr}(S_{1n})}\right)$. Thus t_{β_j} does not have a familiar asymptotic standard normal distribution, and the F -statistic has no familiar $\chi^2(m - 1)$ distribution either.

Even though the t - and F -statistic are not reliable, Lee and Yu [34] suggest the combination of R^2 and Moran's I could be a good indicator for spurious regression under near unit roots. Let $M_n^0 = I_n - \frac{1}{n} l_n l_n'$, where l_n is an $n \times 1$ vector of ones, the coefficient of determination R^2 is

$$R^2 := 1 - \frac{e_n' e_n}{Y_{1n}' M_n^0 Y_{1n}} = 1 - \frac{e_n' e_n / (n \psi_{1n}^2)}{Y_{1n}' M_n^0 Y_{1n} / (n \psi_{1n}^2)} = O_p\left(\frac{1}{n}\right) \xrightarrow{p} 0. \tag{54}$$

And the Moran's I statistic is

$$I_{\text{Moran}} := \frac{\frac{1}{n} e_n' W_n e_n}{\frac{1}{n} e_n' e_n} = \frac{\text{tr}\left(S_{1n}^{*-1} W_n S_{1n}^{*-1}\right)}{\text{tr}\left(S_{1n}^{*-1} S_{1n}^{*-1}\right)} + O_p\left(\frac{1}{\sqrt{n}}\right) \tag{55}$$

$$\xrightarrow{p} 1, \quad \text{when } W_n = W_{1n}.$$

3.3.4. Constant Terms in the DGP of Y_{jn} 's

Lee and Yu [34] also study the constant term and unit roots at the same time. It could be shown that the estimation of β is the same as in (49) after reparameterization. Consider the DGP of series Y_{jn}^c with near unit roots and a constant term c_{jn} as

$$Y_{jn}^c = \lambda_{jn} W_{jn} Y_{jn}^c + c_{jn} l_n + \epsilon_{jn}, \quad \text{for every } j = 1, \dots, m. \tag{56}$$

Regress Y_{1n}^c on l_n and $Y_{-1,n}^c$:

$$Y_{1n}^c = \alpha^c l_n + Y_{-1,n}^c \beta^c + V_n^c. \tag{57}$$

where α^c is a constant and $Y_{-1,n}^c$ is similarly defined as Y_{-1} above. Since W_{jn} is row-normalized, $W_{jn} l_n = l_n$, we have $Y_{jn}^c = S_{jn}^{-1}(c_{jn} l_n + \epsilon_{jn}) = \psi_{jn} c_{jn} l_n + S_{jn}^{-1} \epsilon_{jn} = \psi_{jn} c_{jn} l_n + Y_{jn}$. So (57) could be rewritten as:

$$Y_{1n} = \left(\alpha^c - \psi_{1n} c_{1n} + \sum_{j=2}^m \psi_{jn} c_{jn} \beta_{jn}^c \right) l_n + Y_{-1,n} \beta^c + V_n^c. \tag{58}$$

Compare (58) and (42), OLS estimation of β^c is the same as that of β and thus the corresponding statistics would be the same.

3.4. "Spurious" Regression with Equal Weights

Baltagi and Liu [38] show that under the special case where the spatial weight matrix is row-normalized and with equal spatial weights, i.e.,

$$w_{n,ij} = \begin{cases} 0 & , \text{ if } i = j, \\ \frac{1}{n-1} & , \text{ if } i \neq j. \end{cases} \tag{59}$$

spurious regression will not occur. This spatial weight matrix "is naturally suggested if all units are neighbors to each other and there is no other natural or observable measure of distance [39]", such as interactions between students in a class or workers in a firm, etc. Without loss of generality, the DGP is assumed to be

$$\begin{aligned} Y_n &= \lambda_1 W_n Y_n + \epsilon_{1n}, \\ X_n &= \lambda_2 W_n X_n + \epsilon_{2n}. \end{aligned} \tag{60}$$

Consider the regression

$$Y_n = \alpha \iota_n + X_n \beta + V_n. \tag{61}$$

By the Frisch–Waugh–Lovell Theorem, the OLS estimation of β is $\hat{\beta} = (X_n' E_n X_n)^{-1} X_n' E_n Y_n$, where $E_n = I_n - \frac{J_n}{n}$ and J_n is an $n \times n$ matrix of ones. Kelejian and Prucha [39] show the inverse of the matrix $S_{nk} = I_n - \lambda_k W_n, k = 1, 2$, is

$$S_{nk}^{-1} = \delta_{k1} J_n - \delta_{k2} I_n, \tag{62}$$

where $\delta_{k1} = \frac{\lambda_k}{(n-1+\lambda_k)(1-\lambda_k)}$ and $\delta_{k2} = \frac{n-1}{n-1+\lambda_k}$. Using the fact that $E_n J_n = 0$, Baltagi and Liu [38] show $\frac{1}{n} X_n' E_n X_n \xrightarrow{p} \sigma_2^2$ and $\frac{1}{\sqrt{n}} X_n' E_n Y_n \xrightarrow{d} N(0, \sigma_1^2 \sigma_2^2)$, so the asymptotic distribution of $\hat{\beta}$ is given by

$$\sqrt{n} \hat{\beta} = \left(\frac{1}{n} X_n' E_n X_n \right)^{-1} \left(\frac{1}{\sqrt{n}} X_n' E_n Y_n \right) \xrightarrow{d} N(0, \sigma_1^2 / \sigma_2^2). \tag{63}$$

The asymptotic distribution of $\hat{\beta}$ does not depend on λ_1 and λ_2 and is \sqrt{n} consistent (compared with (49)), which means that the spurious regression does not occur.

4. Estimation and Inference

The spurious regression studied in the previous section suggests that OLS is not a good estimator in the SAR model with spatial unit roots. QMLE and 2SLS are alternative methods of estimation. This section briefly reviews these estimators and their performance under near unit roots.

When the error term of the DGP has a spatial autoregressive structure, OLS and (feasible) generalized least squares ((F)GLS) estimators are consistent. The efficiency of the OLS estimator was considered by Krämer and Donninger [40], Tilke [41] for the symmetric spatial weight matrices, and generalized by Krämer and Baltagi [42] with a broader covariance matrix. But the symmetry of the weight matrix is too restrictive to be used in practice, so Martellosio [43] generalizes this to nonsymmetric weights matrices. The efficiency of the OLS estimator is defined as $\eta := \frac{\text{tr}[\text{var}(X\hat{\beta}_{GLS})]}{\text{tr}[\text{var}(X\hat{\beta}_{OLS})]}$. But these papers generally focus on the relationship between W_n and X , for example, when the column space of W_n is contained by the column space of X . Following Lee and Yu [34], Baltagi et al. [37] derive the asymptotic properties of OLS, (F)GLS of β and point out important differences from conventional theory based on stationary spatial error.

A special SAR model has also been considered in regular lattices under spatial unit roots with the form

$$Y_{k,l} = \sum_{i=0}^{p_1} \sum_{j=0}^{p_2} \alpha_{i,j} Y_{k-i,l-j} + \epsilon_{k,l}, \quad \alpha_{0,0} = 0. \tag{64}$$

The simplest case of this special SAR model is the doubly geometric spatial autoregressive process:

$$Y_{k,l} = \alpha Y_{k-1,l} + \beta Y_{k,l-1} - \alpha\beta Y_{k-1,l-1} + \epsilon_{k,l}. \tag{65}$$

It is called “unilateral” because only the previous units have effects on the latter ones and have a lower triangular weight matrix. It can be considered as the combination of two autoregressive (AR) models [44]. This model has been widely used in the area of image processing, agriculture trials, digital filtering, and other different fields. Model (64) is unstable when either $|\alpha| \geq 1$ or $|\beta| \geq 1$ because of the existence of spatial unit roots [45,46].

A more complicated special case of the unilateral model is

$$Y_{k,l} = \alpha Y_{k-1,l} + \beta Y_{k,l-1} + \gamma Y_{k-1,l-1} + \epsilon_{k,l}. \tag{66}$$

Because of its simple form, the estimation and inference of (64) can be derived without too many assumptions, as will be seen shortly. Moreover, when spatial unit roots exist, the limit of the variance of $Y_{k,l}$ are analytically obtained in Baran [47] when the parameters (α, β, γ) are located on the interior, on the edges, and on the vertices of the domain of stability. Paulauskas [48] explicitly shows that the growth rate of the variance of Y is different in dimensions $d = 2, 3$ and $d = 4$. Though this approach studies spatial unit roots from a different angle than that of Fingleton [14], it points to some similarity as in a recent paper [49] that will be discussed in Section 6.

Another possible way to remedy the problem caused by spatial unit roots in the SAR, SEM model is to relax the compactness assumption. When some parameters approach the boundary of the parameter space, consistency of extremum estimators could be obtained with compact parameter spaces. Thus the compactness assumption is standard in spatial econometrics because of the existence of the singular point $\frac{1}{\rho_{max}}$, see proofs in Kelejian and Prucha [6] Lee [5], Gupta [50]. But such an assumption is also restrictive in the sense that if we choose an arbitrary compact set on the open parameter space, the true global optimizer may be exclusive, especially for near unit root cases. A recent paper by Liu et al. [51] generalizes Theorem 2.7 in Newey and McFadden [52] (p. 2133), which relaxes compactness when the objective function of an extremum estimator is concave and allows the non-stochastic objective functions to depend on the sample size n . This generalization is suitable for spatial econometrics models because the sample observations are usually in triangular arrays. (A triangular array is a doubly indexed sequence of numbers or polynomials. Each row of the array is only as long as the row's index. For example, the i th row contains only i elements.) The consistency of the QMLE of the SAR model and the MLE of the SAR Tobit model are investigated. But a closed-form solution is not obtained. On the other hand, Gupta [53] proposes a Newton-step computational algorithm of QMLE for a large-parameter-size SAR model, which is free from the compactness assumptions. Under the normality assumption, it has the same asymptotic efficiency as MLE, but has a closed-form solution and is computationally simple.

4.1. QMLE and 2SLS Methods for the (Mixed) SAR Model

4.1.1. Quasi-Maximum Likelihood Estimation Method

Lee [5] investigates the asymptotic distribution of the QMLE estimator of the mixed SAR model, which is the starting point for further analysis when spatial unit roots exist. This analysis is based on the discussion of the singularity of the information matrix of the log-likelihood function. Especially when the information matrix is singular, a scaling factor $\frac{1}{\sqrt{h_n}}$ is needed, where $\frac{1}{h_n}$ is the order of the elements of the spatial weight matrix W_n and thus $\frac{1}{\sqrt{h_n}}$ is the order elements of $G_n = W_n S_n^{-1}$. We have seen in (39) that the order of elements of G_n is ψ_n in the near unit roots case; thus, a similar scaling factor will be needed.

The (mixed) SAR model under consideration is

$$Y_n = X_n \beta + \lambda W_n Y_n + V_n, \tag{67}$$

with its reduced form

$$\begin{aligned} Y_n &= S_n^{-1}(X_n \beta_0 + V_n) \\ &= X_n \beta_0 + \lambda_0 G_n X_n \beta_0 + S_n^{-1} V_n, \end{aligned} \tag{68}$$

since $I_n + \lambda_0 G_n = S_n^{-1}$.

Also, Lee [5] imposes a weaker assumption about the spatial weight matrix and derives the information matrix.

Assumption 7. *The elements $w_{n,ij}$ of W_n are at most of order h_n^{-1} , denoted by $O(1/h_n)$, uniformly in all i, j where the rate sequence $\{h_n\}$ can be bounded or divergent. The ratio $h_n/n \rightarrow 0$ as n goes to infinity.*

Let $V_n(\delta) = Y_n - X_n\beta - \lambda W_n Y_n$, where $\delta = (\beta', \lambda)'$, then the log-likelihood function of (67) is

$$\ln L_n(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} V_n'(\delta) V_n(\delta), \tag{69}$$

where $\theta = (\beta', \lambda, \sigma^2)'$. The information matrix is

$$\mathbb{E} \left(\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta'} \right) = -\mathbb{E} \left(\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'} \right) + \Omega_{\theta, n}, \tag{70}$$

where

$$-\mathbb{E} \left(\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'} \right) = \begin{pmatrix} \frac{1}{\sigma_0^2 n} X_n' X_n & \frac{1}{\sigma_0^2 n} X_n' (G_n X_n \beta_0) & 0 \\ \frac{1}{\sigma_0^2 n} (G_n X_n \beta_0)' X_n & \frac{1}{\sigma_0^2 n} (G_n X_n \beta_0)' (G_n X_n \beta_0) + \frac{1}{n} \text{tr}(G_n^s G_n) & \frac{1}{\sigma_0^2 n} \text{tr}(G_n) \\ 0 & \frac{1}{\sigma_0^2 n} \text{tr}(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}. \tag{71}$$

with $G_n^s = G_n + G_n'$. The existence of the extra $\Omega_{\theta, n}$ is because V_n is not necessarily normally distributed.

To ensure the asymptotic distribution of QMLE $\hat{\theta}$ exists, $\Sigma_\theta = -\lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'} \right)$ must be well defined. Lee [5] proves the nonsingularity of Σ_θ can be guaranteed by the fact that there does not exist a nonzero vector $\lambda = (\lambda_1', \lambda_2, \lambda_3)'$ such that a linear combination of columns of Σ_θ is 0. This condition could be simplified as: there does not exist a $\lambda_2 \neq 0$, such that

$$\left\{ \lim_{n \rightarrow \infty} \frac{1}{n\sigma_0^2} (G_n X_n \beta_0)' M_n(G_n X_n \beta_0) + \lim_{n \rightarrow \infty} \frac{1}{n} \left[\text{tr}(G_n' G_n) + \text{tr}(G_n^2) - \frac{2}{n} \text{tr}^2(G_n) \right] \right\} \lambda_2 = 0. \tag{72}$$

Since each term in (72) is greater or equal to 0 (the first term is non-negative because it is symmetric; for the second term, $\text{tr}(G_n G_n') + \text{tr}(G_n^2) - \frac{2}{n} \text{tr}^2(G_n) = \frac{1}{2} \text{tr} \left[(C_n' + C_n)(C_n' + C_n) \right] = \frac{1}{2} \text{tr}(C_n^s' C_n^s) \geq 0$ where $C_n = G_n - \frac{\text{tr}(G_n)}{n} I_n$ and $C_n^s = C_n' + C_n$), Lee [5] studies the singularity of the information matrix in terms of these two terms, respectively. For the first term, by the partition matrix formula, $\lim_{n \rightarrow \infty} \frac{1}{n} (X_n, G_n X_n \beta_0)' (X_n, G_n X_n \beta_0)$ is nonsingular if and only if $\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n$ and $\lim_{n \rightarrow \infty} \frac{1}{n} (G_n X_n \beta_0)' M_n(G_n X_n \beta_0)$ are nonsingular. Moreover, under Assumption 7, if $G_n X_n \beta_0$ and X_n are independent, one sufficient condition for the nonsingularity of Σ_θ could be:

Assumption 8. $\lim_{n \rightarrow \infty} \frac{1}{n} (X_n, G_n X_n \beta_0)' (X_n, G_n X_n \beta_0)$ exists and is nonsingular.

However, Lee [5] states that if $G_n X_n \beta_0$ and X_n are linearly dependent, for example, when W_n is row-normalized and the relevant regressor is only the constant term, Assumption 8 should be adjusted to guarantee the second term in (72) is greater than 0:

Assumption 9. $\lim_{n \rightarrow \infty} \frac{1}{n} (G_n X_n \beta_0)' M_n(G_n X_n \beta_0) = 0$ and the $\{h_n\}$ is a bounded sequence and, for any $\lambda \neq \lambda_0$,

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \ln \left| \sigma_0^2 S_n^{-1} S_n'^{-1} \right| - \frac{1}{n} \ln \left| \sigma_n^2(\lambda) S_n^{-1}(\lambda) S_n'^{-1}(\lambda) \right| \right) \neq 0. \tag{73}$$

Then under Assumption 7 and either 8 or 9, the asymptotic distribution of QMLE $\hat{\theta}$ will be

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\left(0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1} \Omega_\theta \Sigma_\theta^{-1}\right), \tag{74}$$

where $\Omega_\theta = \lim_{n \rightarrow \infty} \Omega_{\theta,n}$, $\Sigma_\theta = -\lim_{n \rightarrow \infty} \mathbb{E}\left(\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'}\right)$.

The above results are based on Σ_θ being invertible of which a necessary condition is that $\{h_n\}$ is a bounded sequence. However, when $\lim_{n \rightarrow \infty} h_n = \infty$, Σ_θ will become singular because $\frac{1}{n} \text{tr}[(C'_n + C_n)(C'_n + C_n)'] = O\left(\frac{1}{h_n}\right) = o(1)$. Also, the singularity of the information matrix implies that the score function will be too flat to be useful and thus an adjustment of the rate should be imposed as in Lee [5]

Assumption 10. The $\{h_n\}$ is a divergent sequence, elements of $M_n(G_n X_n \beta_0)$ have the uniform order $O\left(\frac{1}{\sqrt{h_n}}\right)$, and $\lim_{n \rightarrow \infty} \left(\frac{h_n}{n}\right) (G_n X_n \beta_0)' M_n(G_n X_n \beta_0) = c$ with $0 \leq c < \infty$. Under this situation, either (a) $c > 0$, or (b) if $c = 0$

$$\lim_{n \rightarrow \infty} \left(\frac{h_n}{n} \ln \left| \sigma_0^2 S_n^{-1} S_n'^{-1} \right| - \frac{h_n}{n} \ln \left| \sigma_n^2(\lambda) S_n^{-1}(\lambda) S_n'^{-1}(\lambda) \right| \right) \neq 0, \tag{75}$$

whenever $\lambda \neq \lambda_0$.

Lee [5] gives the asymptotic distributions of the QMLE under this rate-adjusted assumption:

$$\begin{aligned} \sqrt{\frac{n}{h_n}}(\hat{\lambda}_n - \lambda_0) &\xrightarrow{D} N\left(0, \sigma_\lambda^2\right), \\ \sqrt{\frac{n}{h_n}}(\hat{\beta}_n - \beta_0) &\xrightarrow{D} N\left(0, \sigma_\lambda^2 \lim_{n \rightarrow \infty} (X'_n X_n)^{-1} X'_n (G_n X_n \beta_0) (G_n X_n \beta_0)' X_n (X'_n X_n)^{-1}\right), \\ \sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (v_i^2 - \sigma_0^2) + o_P(1) \xrightarrow{D} N\left(0, \mu_4 - \sigma_0^4\right). \end{aligned} \tag{76}$$

4.1.2. Generalized Spatial Two-Stage Least Squares (GS2SLS) Method

Kelejian and Prucha [7] proposed the GS2SLS method for the “cross-sectional (first-order) autoregressive spatial model with (first-order) autoregressive disturbances”:

$$\begin{aligned} Y_n &= X_n \beta + \lambda W_n Y_n + u_n & |\lambda| < 1 \\ &= Z_n \delta + u_n, \\ u_n &= \rho M_n u_n + \epsilon_n, & |\rho| < 1, \end{aligned} \tag{77}$$

where $Z_n = (X_n, W_n Y_n)$ and $\delta = (\beta', \lambda)'$. Since $W_n Y_n$ is endogenous, it should be instrumented. Assume ρ is known (a consistent estimator is given by Kelejian and Prucha [6]), a Cochrane–Orcutt (CO)-type transformation applied to (77) yield the transformed regression

$$\begin{aligned} Y_n^* &= Z_n^* \delta + \epsilon_n \\ &= (I_n - \rho M_n) Z_n \delta + \epsilon_n \\ &= (Z_n - \rho M_n Z_n) \delta + \epsilon_n. \end{aligned} \tag{78}$$

$M_n Z_n$ should be instrumented. The ideal instruments are of course

$$\mathbb{E}[Z_n] = (X_n, W_n \mathbb{E}[Y_n]) \quad \text{and} \quad \mathbb{E}[M_n Z_n] = (M_n X_n, M_n W_n \mathbb{E}[Y_n]). \tag{79}$$

Then, by (10), $\mathbb{E}[Y_n] = (I_n - \lambda W_n)^{-1} X_n \beta = [\sum_{i=0}^{\infty} \lambda^i W_n^i] X_n \beta$, the ideal instruments are:

$$H_n = \left[X_n, W_n^1 X_n, W_n^2 X_n, \dots, M_n X_n, M_n W_n^1 X_n, M_n W_n^2 X_n, \dots \right]. \tag{80}$$

In practice, $H_n = [X_n, W_n^1 X_n, W_n^2 X_n, M_n X_n, M_n W_n^1 X_n, M_n W_n^2 X_n]$ is used.

Assumptions about the instrument matrix are made to ensure that $\lim_{n \rightarrow 1} \frac{1}{n} H_n' Z_n$ exists and is of full rank. In the near unit roots case, similar to the QMLE method, scaling factors are needed to guarantee this property as we will see later. With these assumptions, the GS2SLS procedure has three steps, as in Kelejian and Prucha [7], as follows:

1. Run 2SLS on $Y_n = Z_n \delta + u_n$ with instruments H_n . This yields $\tilde{\delta}_n = (\hat{Z}_n' \hat{Z}_n)^{-1} \hat{Z}_n' Y_n$, where $\hat{Z}_n = P_{H_n} Z_n = (X_n, \widehat{W_n Y_n})$, P_{H_n} is the projection matrix of H_n , and $\tilde{\delta}_n = \delta + O_p(n^{-\frac{1}{2}})$.
2. Estimate ρ by Kelejian and Prucha [6] according to the GMM system: $\rho_n = G_n \lambda + V_n$ where $\lambda = (\rho, \rho^2, \sigma^2)'$, then solve $\tilde{\lambda} = G_n' g_n$, or by $\tilde{\lambda} = \operatorname{argmin}_{\rho, \sigma^2} V_n' V_n$. Both $\tilde{\lambda}$ and $\tilde{\tilde{\lambda}}$ are consistent, but $\tilde{\tilde{\lambda}}$ is more efficient.
3. Assuming ρ is known, run 2SLS on the CO transformed regression (78) with instruments H_n yields $\hat{\delta}_n = (\hat{Z}_n^* \hat{Z}_n^*)^{-1} \hat{Z}_n^* Y_n^*$, where

$$\begin{aligned} \hat{Z}_n^* &= P_{H_n} Z_n^* = P_{H_n} [X_n - \rho M_n X_n, W_n Y_n - \rho M_n W_n Y_n] \\ &= [X_n - \rho M_n X_n, W_n Y_n - \widehat{\rho M_n W_n Y_n}]. \end{aligned} \tag{81}$$

By replacing ρ by its consistent estimation $\hat{\rho}_n$ (in Step 2). The feasible 2SLS estimator is

$$\hat{\delta}_{F,n} = [\hat{Z}_n^* (\hat{\rho}_n)' \hat{Z}_n^* (\hat{\rho}_n)]^{-1} \hat{Z}_n^* (\hat{\rho}_n) Y_n (\hat{\rho}_n). \tag{82}$$

Obviously, this procedure is for a SARAR model. If it is a SAR model, only step 1 is needed and $H_n = [X_n, W_n^1 X_n, W_n^2 X_n, \dots]$.

4.1.3. Best Generalized Spatial Two Stage Least Squares (BGS2SLS) Estimators

Lee [8] does not drop the higher-order terms in (80) but use the fact that $W_n \mathbb{E}[Y_n] = W_n (I_n - \lambda W_n)^{-1} X_n \beta = G_n X_n \beta$ by the definition of G_n and proposes the best instrument:

$$\begin{aligned} H_n^* &= (I_n - \rho M_n) [X_n, W_n (I_n - \lambda W_n)^{-1} X_n \beta] \\ &= (I_n - \rho M_n) [X_n, G_n X_n \beta]. \end{aligned} \tag{83}$$

With the corresponding simple instrumental variable estimator, the BGS2SLS is

$$\hat{\delta}_{B,n} = [H_n^{*'} Z_n]^{-1} H_n^{*'} Y_n. \tag{84}$$

If there is no SAR structure in the disturbance term, the best instrument is

$$H_n^* = [X_n, G_n X_n \beta], \tag{85}$$

and β in (85) could be replaced by any consistent estimator such as the KP-GS2SLS estimator.

Compared with (80), H_n^* does not drop off the higher-order terms and is numerically equivalent to the ideal instrument, which in turn yields asymptotically optimal instrumental variable estimators.

4.2. Near Unit Roots in the SAR Model

Lee and Yu [15] study the asymptotic distribution of QMLE and 2SLS estimators of the SAR model by decomposing Y_n (see Section 3.3.1). The model is given as

$$Y_n = \lambda_{n0} W_n Y_n + X_n \beta_0 + \epsilon_n, \tag{86}$$

where $\lambda_{n0} = 1 - \frac{1}{\psi_n}$. And the reduced form, again, is

$$Y_n = S_n^{-1}(X_n\beta_0 + V_n) = X_n\beta_0 + \lambda_0 G_n X_n \beta_0 + S_n^{-1} V_n. \tag{87}$$

4.2.1. QMLE

Obviously, the generated regressor $G_n X_n \beta_0$ is explosive because of ψ_n in G_n , see (39). This is very similar to the case in Assumption 10 (a) with $\lim_{n \rightarrow \infty} h_n = \infty$ in Section 4.1.1. This implies that the convergence rate of estimators of ρ_{n0} is not the usual \sqrt{n} case as in (76).

Additional assumptions are made in Lee and Yu [15] to control the magnitude of the unstable part of W_n (Assumption 11), and specify the identification condition (Assumption 12), which are adjusted from Assumption 8:

Assumption 11. (1) $\lim_{n \rightarrow \infty} \frac{m_n}{n} > 0$; (2) $\lim_{n \rightarrow \infty} n^{-1} \text{tr}(W_n^u W_n^u)' > 0$; (3) for any finite constant c , $\lim_{n \rightarrow \infty} n^{-1} \text{tr}[(I_n - cW_n^u)'(I_n - cW_n^u)] > 0$.

Assumption 12. $\beta_0' \lim_{n \rightarrow \infty} (n^{-1} X_n' W_n^u M_n W_n^u X_n) \beta_0 > 0$ holds.

Assumption 11 (1) (2) guarantee that m_n is not too small compared to n and (3) implies that it is not too large. Assumption 12 is equivalent to (see Lee and Yu [15] (p. 338, Lemma 1 (7))) $\lim_{n \rightarrow \infty} \frac{1}{n\psi_n^2} (G_n X_n \beta_0)' M_n (G_n X_n \beta_0)$, which is a rate-adjusted version of Assumption 8, that ensures the identification uniqueness and implies nonsingularity of Σ_θ . For a detailed discussion of these two assumptions, see Baltagi et al. [37] (p. 6). Also, since the adjusted rate is $\frac{1}{\psi_n^2}$, QMLE $\hat{\lambda}_n$ would be $\sqrt{n}\psi_n$ -consistent, see (76) derived under Assumption 10.

The information matrix will be the same as in (70). In Section 3.3.2, when studying the spurious regression of OLS, we mentioned that the scaling factor Y_m is needed in terms of the order of X_n . Here, for the QMLE, similar scaling factor will also be introduced because elements of Σ_n and Ω_n have different orders as the existence of G_n . The second column and row of Σ_n and Ω_n , which are the derivatives with respect to the spatial coefficient λ_n contain G_n , thus, they have to be scaled by a factor $\frac{1}{\psi_n}$. Specifically, the (2, 2) element should be scaled by a factor $\frac{1}{\psi_n^2}$. This can be done by a left and right multiplying matrix $Y_{\theta,n}^{-1}$, where

$$Y_{\theta,n} = \begin{pmatrix} Y_{\delta,n} & 0_{(k+1) \times 1} \\ 0_{1 \times (k+1)} & 1 \end{pmatrix} \quad \text{and} \quad Y_{\delta,n} = \begin{pmatrix} I_k & 0_{k \times 1} \\ 0_{1 \times k} & \psi_n \end{pmatrix}. \tag{88}$$

Thus, Lee and Yu [15] give

$$\sqrt{n} Y_{\theta,n} (\hat{\theta}_n - \theta_{n0}) = - \left(Y_{\theta,n}^{-1} \frac{1}{n} \frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial \theta \partial \theta'} Y_{\theta,n}^{-1} \right)^{-1} Y_{\theta,n}^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{n0})}{\partial \theta}. \tag{89}$$

Let $\Sigma = \lim_{n \rightarrow \infty} Y_{\theta,n}^{-1} \Sigma_n Y_{\theta,n}^{-1}$ and $\Omega = \lim_{n \rightarrow \infty} Y_{\theta,n}^{-1} \Omega_n Y_{\theta,n}^{-1}$ and assume they exist, the asymptotic distribution of $\hat{\theta}_n$ is

$$\sqrt{n} Y_{\theta,n} (\hat{\theta}_n - \theta_{n0}) \xrightarrow{D} N(0, \Sigma^{-1} + \Sigma^{-1} \Omega \Sigma^{-1}). \tag{90}$$

Recently, Rossi and Lieberman [54] combine the near unit roots with a similarity-based weighted matrix and study the consistency of the QMLE estimator when the spatial coefficient $0 \leq \lambda < 1$ and $\lambda = 1$, by allowing uncentered units. The element of similarity-based weighted matrix is $w_{i,j} = \frac{s(x_i, x_j; w_0)}{\sum_{j \neq i} s(x_i, x_j; w_0)}$, where $s(x_i, x_j; w_0)$ is some function that measures the similarity between unit i and j according to some parameter w_0 . The parameters they are most interested in are $\theta_2 = (\lambda, w')'$. They establish the connection between the λ and the

order of uniform absolute row-sum norm of S_0^{-1} , $\|S_0^{-1}\|_\infty = \|S^{-1}(\theta_{20})\|_\infty = O(n^\gamma)$. This means that $\lambda_0 = 1 \Leftrightarrow \gamma = 1$ [54] (p. 11, Proposition 1). Recall that $\text{Var}(Y_n) = \sigma_\epsilon^2 S_n^{-1} S_n'^{-1}$, so when $\gamma = 0$, the variance is independent of n , corresponding to the standard SAR setup ($\lambda < 1$ and fixed). In the case $0 < \gamma < 1$, the variance increases with the sample size but with a lower speed, which is the case that we have seen when studying near unit root; but when $\gamma = 1$, $\lambda = 1$, the variance increases so fast that the non-standard limit distribution of $\theta = (\lambda, w)$ has to be established on a case-by-case basis, according to the resulting S_n^{-1} . Their result is much more complicated than that of Lee and Yu [15] because of the introduction of similarity structure in the weight matrix, but are much more flexible since now W_n is no longer fixed but data-driven, and is potentially more useful in empirical work.

4.2.2. GS2SLS and BGS2SLS

Lee and Yu [15] derive the 2SLS estimators and their asymptotic distributions using the procedures mentioned above. Using instruments defined in (80), the GS2SLS estimator of δ_{n0} is

$$\hat{\delta}_{n,2sls} = (Z_n' P_{H_n} Z_n)^{-1} Z_n' P_{H_n} Y_n. \tag{91}$$

Since $Z_n = [X_n, W_n Y_n]$ contains W_n , the adjustment by $Y_{\delta,n}$ is needed. The asymptotic properties of the GS2SLS estimators of λ_{n0} and β_0 are obtained as follows:

$$Y_{\delta,n} \sqrt{n} (\hat{\delta}_{n,2sls} - \delta_{n0}) = \begin{pmatrix} \sqrt{n} (\hat{\beta}_{n,2sls} - \beta_0) \\ \sqrt{n} \psi_n (\hat{\lambda}_{n,2sls} - \lambda_{n0}) \end{pmatrix} \xrightarrow{d} N(0, \Phi_{2sls}) \tag{92}$$

where $\Phi_{2sls} = \sigma_0^2 \left(\lim_{n \rightarrow \infty} \frac{1}{n} Y_{\delta,n}^{-1} Z_n' P_{H_n} Z_n Y_{\delta,n}^{-1} \right)^{-1}$. So the GS2SLS estimator $\hat{\lambda}_{n,2sls}$ of λ_{n0} is $\sqrt{n} \psi_n$ -consistent, which is higher than the usual \sqrt{n} rate in Kelejian and Prucha [7] and Lee [8], but $\hat{\beta}_{n,2sls}$ has the usual \sqrt{n} rate of convergence.

Choosing the instrument $H_n^* = [X_n, G_n X_n \beta_0]$ as in (85), the BGS2SLS estimator is $\hat{\delta}_{n,b2sls} = (H_n^{*'} Z_n)^{-1} H_n^{*'} Y_n$ with asymptotic distribution:

$$Y_{\delta,n} \sqrt{n} (\hat{\delta}_{n,b2sls} - \delta_{n0}) \xrightarrow{d} N(0, \Phi_{b2sls}) \tag{93}$$

where $\Phi_{b2sls} = \sigma_0^2 \left(\lim_{n \rightarrow \infty} \frac{1}{n} Y_{\delta,n}^{-1} H_n^{*'} H_n^* Y_{\delta,n}^{-1} \right)^{-1}$. Since $\Phi_{b2sls} - \Phi_{2sls}$ is negative semidefinite, the BGS2SLS estimator is more efficient.

The above result is based on the fact that $G_n X_n \beta$ and X_n are independent, which makes sure that the instrument matrix is of full rank, otherwise the 2SLS estimator will be inconsistent [9]. Liu [55] shows that even though $G_n X_n \beta$ and X_n are linearly dependent, i.e., $G_n X_n \beta = X_n c_n$, where c_n is a nonzero vector, we still have $\hat{\rho}_{2SLS} - \rho = O_p\left(\frac{1}{\psi_n}\right)$, $\hat{\beta}_{k,2SLS} - \beta_k = O_p\left(\frac{c_{kn}}{\psi_n}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)$ under near unit roots case; and $\hat{\rho}_{2SLS} - \rho = O_p(1)$, $\hat{\beta}_{k,2SLS} - \beta_k = O_p(c_{kn}) + O_p\left(\frac{1}{\sqrt{n}}\right)$ under the regular case, as long as $c_n = o(1)$. This is equivalent to $G_n X_n \beta$ and X_n , which are asymptotically independent.

To provide guidelines for empirical studies, Lee and Yu [15] conduct simulations to compare the performance of QMLE and 2SLS methods. QMLEs are relatively robust whether the error term is normally distributed or not. Moreover, as n increases (and the spatial coefficient is closer to the spatial unit roots), the QMLEs perform better than the 2SLS estimators because of smaller variances. One interesting phenomenon is that the best 2SLS estimators are even worse than the regular 2SLS estimators in some cases, which violates the theoretical result as shown in (93). One possible reason for this is that the best 2SLS estimator requires an initial consistent estimator by construction (see (85)) and under spatial unit roots, such an initial estimator may not be accurately calculated.

4.3. Near Unit Roots in the SEM Model

Baltagi et al. [37] extend the study of near unit roots from SAR model to SEM model by considering the OLS, GLS and FGLS estimation and properties of the corresponding statistics. The model is given as

$$\begin{aligned} Y_n &= X_n\beta_0 + u_n, \\ u_n &= \lambda_{n0}W_nu_n + \epsilon_n, \end{aligned} \tag{94}$$

with $\lambda_{n0} = 1 - \frac{1}{\psi_n}$. Similar to Lee and Yu [15], u_n could be decomposed to

$$u_n = S_n^{-1}\epsilon_n = \psi_nW_n^u\epsilon_n + (I_n - \lambda_{n0}\tilde{W}_n)^{-1}\epsilon_n. \tag{95}$$

One more assumption that they impose is

Assumption 13. *The elements of X_n are nonstochastic and bounded, uniformly in n , $\lim_{n \rightarrow \infty} n^{-1}X_n'X_n$ exists and is nonsingular. $\lim_{n \rightarrow \infty} n^{-1}X_n'W_n^uW_n^{u'}X_n$ exists. Furthermore, $\lim_{n \rightarrow \infty} n^{-1}X_n'S_n'S_nX_n$ exists and is nonsingular.*

The OLS estimator $\hat{\beta}_{OLS} = (X_n'X_n)^{-1}X_n'Y_n$ has the asymptotic distribution when $\frac{\psi_n}{\sqrt{n}} \rightarrow 0$

$$\begin{aligned} \frac{\sqrt{n}}{\psi_n}(\hat{\beta}_{OLS} - \beta_0) &\xrightarrow{d} N\left(0, \sigma_0^2 \left(\lim_{n \rightarrow \infty} \frac{1}{n}X_n'X_n\right)^{-1} \right. \\ &\quad \left. \left(\lim_{n \rightarrow \infty} \frac{1}{n}X_n'W_n^uW_n^{u'}X_n\right) \left(\lim_{n \rightarrow \infty} \frac{1}{n}X_n'X_n\right)^{-1}\right), \end{aligned} \tag{96}$$

and when $\frac{\psi_n}{\sqrt{n}} \rightarrow c < \infty$, $\hat{\beta}_{OLS} - \beta_0 = O_p(1)$. Thus $\hat{\beta}_{OLS} = \beta_0 + O_p\left(\frac{\psi_n}{\sqrt{n}}\right)$, which is $\frac{\sqrt{n}}{\psi_n}$ -consistent and is slower than the stationary error term case.

Baltagi et al. [37] also study the asymptotic properties of the GLS and FGLS estimators. If λ_{n0} is known, $\hat{\beta}_{GLS} = (X_n'S_n'S_nX_n)^{-1}X_n'S_n'S_nY_n$, and

$$\sqrt{n}(\hat{\beta}_{GLS} - \beta_0) \xrightarrow{d} N\left(0, \sigma_0^2 \left(\lim_{n \rightarrow \infty} \frac{1}{n}X_n'S_n'S_nX_n\right)^{-1}\right), \tag{97}$$

which implies that $\hat{\beta}_{GLS}$ is robust for the near unit roots in the error term because it has \sqrt{n} rate of convergence. The feasible GLS (FGLS) could be achieved by replacing λ by a consistent estimator $\hat{\lambda}_n$, which yields

$$\hat{\beta}_{FGLS} = (X_n'\hat{S}_n'\hat{S}_nX_n)^{-1}X_n'\hat{S}_n'\hat{S}_nY_n, \tag{98}$$

where $\hat{S}_n = I_n - \hat{\lambda}_nW_n$. It can be seen that FGLS is identical to the QMLE: concentrated log likelihood function of (94) with respect to λ is

$$\ln L_n(\lambda) = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2} \ln \hat{\sigma}_n^2(\lambda) + \ln |S_n(\lambda)|, \tag{99}$$

where

$$\hat{\sigma}_n^2(\lambda) = n^{-1}u_n'S_n(\lambda)'\bar{P}_n(\lambda)S_n(\lambda)u_n, \tag{100}$$

$$\bar{P}_n(\lambda) = I_n - S_n(\lambda)X_n[X_n'S_n(\lambda)'S_n(\lambda)X_n]^{-1}X_n'S_n(\lambda)'. \tag{101}$$

The QMLE is of order $\psi_n(\hat{\lambda}_n - \lambda_{n0}) = o_p(1)$, which is ψ_n -consistent and

$$\begin{aligned} \hat{\beta}_n(\hat{\lambda}_n) &= \left[X_n' S_n(\hat{\lambda}_n)' S_n(\hat{\lambda}_n) X_n \right]^{-1} X_n' S_n(\hat{\lambda}_n)' S_n(\hat{\lambda}_n) Y_n, \\ \hat{\sigma}_n^2(\hat{\lambda}_n) &= \frac{1}{n} u_n' S_n(\hat{\lambda}_n)' \bar{P}_n(\hat{\lambda}_n) S_n(\hat{\lambda}_n) u_n. \end{aligned} \tag{102}$$

Comparing with (98), $\hat{\beta}_n(\hat{\lambda}_n)$ is a FGLS of β using $\hat{\lambda}_n$. Thus, the QMLE $\hat{\beta}_{ML}$ and the infeasible GLS estimator $\hat{\beta}_{GLS}$ have the same asymptotic distribution as shown before. Next, Baltagi et al. [37] consider the Wald test statistic for the null hypothesis $H_0 : R\beta = r$ for OLS, GLS and FGLS, where R is a $q \times k$ matrix of rank $q < k$ and r is $q \times 1$. For OLS,

$$\begin{aligned} W_{OLS} &:= (R\hat{\beta}_{OLS} - r)' \left[\hat{\sigma}_{OLS}^2 R(X_n' X_n)^{-1} R' \right]^{-1} (R\hat{\beta}_{OLS} - r) \\ &\xrightarrow{d} \zeta' \left[\sigma_0^2 \left(\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(W_n^{uu} W_n^{uu}) \right) R \left(\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n \right)^{-1} R' \right]^{-1} \zeta, \end{aligned} \tag{103}$$

where

$$\zeta \sim N\left(0, \sigma_0^2 R \left(\lim_{n \rightarrow \infty} n^{-1} X_n' X_n \right)^{-1} \left(\lim_{n \rightarrow \infty} n^{-1} X_n' W_n^{uu} W_n^{uu} X_n \right) \left(\lim_{n \rightarrow \infty} n^{-1} X_n' X_n \right)^{-1} R' \right), \tag{104}$$

does not have a standard χ^2 distribution, which is similar to the F -statistic shown above. However, the GLS Wald statistic

$$W_{GLS} = (R\hat{\beta}_{GLS} - r)' \left[\sigma_0^2 R(X_n' S_n' S_n X_n)^{-1} R' \right]^{-1} (R\hat{\beta}_{GLS} - r) \xrightarrow{d} \chi_k^2, \tag{105}$$

has a chi-squared limiting distribution.

Baltagi et al. [37] conduct extensive simulations. Using the root mean squared error (RMSE) as the evaluation criteria, the QML (FGLS) estimators perform uniformly better than the OLS estimator. In particular, when the spatial coefficient is sufficiently close to 1 and the sample size n increases, the RMSE of the OLS estimator grows dramatically. Together with the fact that the Wald test statistic based on the QML method has a standard Chi-squared distribution, QMLE is recommended when near spatial unit roots exist in the spatial error model.

4.4. Doubly Geometric Spatial Autoregressive Process

The main difference between the SAR and the doubly geometric spatial autoregressive models is that the spatial dependence form of the latter is clearly specified. However, for the SAR model, such dependence relies on the specification of W_n whose explicit form varies in different situations.

For model (65), based on the observation $\{X_{k,\ell} : 1 \leq k \leq m \text{ and } 1 \leq \ell \leq n\}$, Baran [47] shows that the asymptotic normality of the estimators $(\hat{\alpha}_{m,n}, \hat{\beta}_{m,n})$ is $\sqrt{mn} \begin{pmatrix} \hat{\alpha}_{m,n} - \alpha \\ \hat{\beta}_{m,n} - \beta \end{pmatrix} \xrightarrow{D} N(0, \Sigma_{\alpha,\beta})$ in the stable case ($|\alpha| < 1$ and $|\beta| < 1$), with some covariance matrix $\Sigma_{\alpha,\beta}$. For the unstable case ($\alpha_n \rightarrow 1, \beta_n \rightarrow 1$), using the martingale central limit theorem, Bhattacharyya et al. [56,57] show “one step Gauss-Newton” estimators are asymptotically normal with convergence rate $n^{3/2}$. This is different from the classical time series $AR(1)$, where the OLS estimator converges to a fraction of functionals of the standard Brownian motion: $T(a'_{OLS} - 1) \Rightarrow \frac{W(1)^2 - 1}{2 \int_0^1 W(r)^2 dr}$ [58] (p. 281).

Baran and Pap [59] consider the more complicated model as in (66). The model is stable if and only if $(\alpha, \beta, \gamma) \in S$, where S is the open tetrahedron with vertices $V := \{(1, 1, -1), (1, -1, 1), (-1, 1, 1), (-1, -1, -1)\}$. They also prove that the OLS estimator is asymptotically normally distributed with the convergence rate n when the model is stable, and $n^{3/2}$ otherwise. (The simpler model $Y_{k,l} = \alpha Y_{k-1,l} + \beta Y_{k,l-1} + \epsilon_{k,l}$, with possibly $\alpha = \beta$

was investigated in Baran et al. [60], Baran et al. [44], Baran and Pap [61] under stable and unstable cases. Under different settings, the limiting distribution of the OLS estimator is normal but has different rates of convergence.)

Roknossadati and Zarepour [62,63] study the limiting behavior of M-estimation for the near unit roots of model (65). The M-estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ of (α_n, β_n) is defined to minimize of the objective function

$$g(\alpha_n, \beta_n) = \sum_{i=2}^n \sum_{j=2}^n \lambda(Y_{ij} - \alpha_n Y_{i-1,j} - \beta_n Y_{i,j-1} + \alpha_n \beta_n Y_{i-1,j-1}), \quad (106)$$

for some convex function $\lambda(\cdot)$. Roknossadati and Zarepour [62] show that the self-normalized M-estimators are asymptotically normal, and when the series is stable, the convergence rate of M-estimators is still $n^{3/2}$, same as in Bhattacharyya et al. [56,57]. But if it is unstable, i.e., when the model has infinite variance innovations, the M-estimates have a higher consistency rate.

5. Tests for Spatial Unit Roots and Nonstationarity

Recognizing the possible consequences of spatial unit roots, it is necessary to test for it. In fact, in nonstationary cases, the estimator is inconsistent and diverges [64]. If the series contains spatial unit roots, one may employ the spatial first difference as recommended by Fingleton [14]: after the first-order difference, such series will be converted to a stationary one, otherwise it is over-differenced and spatial correlation still exists. Based on this idea, Lauridsen and Kosfeld [17,18] propose two-stage LM tests to check for spatial unit roots. However, such LM tests have a high power function because of $LM > LR > Wald$ in finite samples and are not useful for spatial cointegration since they mis-specify the regression in the second stage. A Wald test is proposed by Lauridsen and Kosfeld [65] but it does not have a usual χ^2 distribution so simulation has to be conducted before each test to obtain the critical values. A different approach introduced by Beenstock et al. [19] uses the fact that when spatial unit roots exist, the variance explodes and the spatial impulse does not die out as distance increases, so they iterate on the parameter space to find out the value of the unit roots (for irregular lattice) and then generate nonstationary series to conduct interval estimation.

Martellosio [66] derives the power properties of invariant tests, for example, $\frac{\hat{u}'Q\hat{u}}{\hat{u}'\hat{u}}$, where \hat{u} is the OLS residuals and Q is a fixed matrix. When $Q = W_n$, we obtain the Cliff–Ord test. When the regression contains only a constant, the Cliff–Ord test reduces to Moran’s test as introduced before, which is best locally invariant as shown by King [67]. It has been shown that for the SEM model, as $\lambda \uparrow \frac{1}{\rho_{max}}$, the test power vanishes. For the SAR model, as $\lambda \uparrow \frac{1}{\rho_{max}}$, the limiting power is either 0 or 1. Krämer [68] shows similar conclusions but focuses on the symmetric weight matrix. Martellosio [69] further shows the power of any test vanishes as spatial correlation increases for a set of regression spaces. Heteroskedasticity robust tests have been studied. For example, Born and Breitung [70], Baltagi and Yang [71] design diagnostic tests for SEM and SAR employing the outer product of gradients (OPG) variant of the LM test which are robust against heteroskedastic (and non-normal) errors. But these tests suffer from the same deficiency as in Martellosio [66] because such test is asymptotically equivalent to Moran’s I . Baltagi and Yang [72] have also shown that the standard LM test undergoes finite sample distortion when spatial dependence is heavy in both spatial and panel data settings. Recently Preinerstorfer [73] suggests some modified tests to avoid this “zero-power trap” phenomenon, which works well for small spatial autocorrelation, but still has limiting power smaller than 1 (only 0.619 by simulation). Thus, the invariant test of $I(0)$ null hypothesis is not satisfactory when spatial unit roots exist, and methodologies to determine it (Tests of the $I(1)$ null hypothesis) deserve more attention.

5.1. Two-Stage LM Test for the Sources of Spurious Spatial Regression

Lauridsen and Kosfeld [17] develop a two-stage LM test to distinguish between two possible sources for spurious regression. The first one is the existence of spatial (near) unit roots in the regressand and/or regressors as in Fingleton [14], Lee and Yu [34]; the second is that the spatial error term itself is nonstationary. So the LM tests are essentially testing if the spatial process is stable or not. The idea originates from the fact that Fingleton [14] suggests a high value of Moran’s I statistic as an indicator for both spatial nonstationarity and spurious regression, but we cannot distinguish between them or even distinguish between the nonstationarity and the positive spatial correlation among the error terms, which by definition imply a high value of Moran’s I . Specifically, we are trying to distinguish if (i) the X_n is a SAR process and we regressed Y_n on X_n or (ii) the model itself is SEM as $Y_n = X_n\beta + \epsilon_n$, where $\epsilon_n = \lambda_\epsilon W_n \epsilon_n + \mu_n$, because both (i) and (ii) can cause spatial autocorrelation.

There are at least three advantages of the LM tests [17]. First is that compared with Wald or LR, LM is usually simpler to compute because it is constructed under H_0 . Second is that with the LM test, it is possible to control for some omitted model features such as heterogeneity and autoregression, as in Anselin [74], which will be discussed later. The last one is that, other statistics may not have a standard asymptotic distribution, such as the OLS Wald type statistic as in Baltagi et al. [37]. The proposed two-stage LM test is based on the SEM model and all four possible results are summarized in Lauridsen and Kosfeld [17] (Table 1):

1. Under $H_0 : \lambda_\epsilon = 0$, the LM error statistic (LME) developed by Anselin [74] (p. 11, Equation (35)) is

$$LME = \frac{(e'W_n e / \sigma^2)^2}{\text{tr}(W_n^2 + W_n'W_n)} \sim \chi^2(1). \tag{107}$$

Thus large values of LME reject the null hypothesis, which implies either $0 < \lambda_\epsilon < 1$ or $\lambda_\epsilon = 1$.

2. The next step is to test if $H_0 : \lambda_\epsilon = 1$. This could be carried out by using the spatial differencing we introduced before. Under H_0 , $\Delta\epsilon = \mu$, thus the first order difference on the regression, $\Delta Y_n = \Delta X_n\beta + \mu_n$, yields i.i.d. error μ_n , which means the value of differenced LME (DLME) should be close 0 under H_0 . But if $\lambda_\epsilon < 1$, Δ represents overdifferencing, i.e., $\Delta\epsilon_n = (I_n - W_n)(I_n - \lambda_\epsilon W_n)^{-1}\mu_n$, so spatial correlation in the error term still exists, and we cannot reject H_0 .

Similar procedures could be used to investigate whether Y_n or any X_n are spatially nonstationary, as the case in Lee and Yu [34]. Letting Z_n be one of $Y_n, X_{n1}, X_{n2}, \dots$, Lauridsen and Kosfeld [17] suggest using the regression

$$\begin{aligned} Z_n &= \alpha t_n + \epsilon_n, \\ \Delta Z_n &= \alpha \Delta t_n + \Delta\epsilon_n = \Delta\epsilon_n, \end{aligned} \tag{108}$$

to obtain LME and DLME, respectively. Z_n is regressed on a constant term because there is no meaningful regressor but we still need the residuals.

Spatial cointegration could also be tested using this LM test. Thus, after determining Y_n and X_n are nonstationary, regress Y_n on X_n and ΔY_n on ΔX_n to obtain LME and DLME. The cointegration relation exists if LME is 0; and non-cointegration if LME is positive and DLME is 0; the limiting case of “near cointegration” occurs if LME and DLME are positive.

Moreover, Lauridsen and Kosfeld [18] generalize their two-stage LM test to account for unobserved heteroskedasticity. They specify the covariance matrix of ϵ_n , $\Omega = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$, have the diagonal element $\sigma_i^2 = f(Z_i, \lambda_Z)$, where Z_i is $P \times 1$ vector of observations of exogenous variables for region i , related to σ_i^2 via the $P \times 1$ vector

of parameters λ_Z . So the statistic in (107) should be adjusted as in Anselin [74] (p. 9, Equation (29)):

$$LMEH = \frac{(e'W_n e / \sigma^2)^2}{\text{tr}(W_n^2 + W_n'W_n)} + \frac{f'Z(Z'Z)^{-1}Z'f}{2} \sim \chi^2(P + 1), \tag{109}$$

with $f_i = \frac{e_i^2}{\sigma^2} - 1$ and Z as the $n \times P$ matrix containing the Z vectors that cause heteroskedasticity.

However, the Lauridsen and Kosfeld test procedure is not without problems. Beenstock et al. [19] point out that this procedure is not suitable for testing spatial cointegration since the second stage is misspecified. To see it more clearly, regress $\Delta Y_n = \beta \Delta X_n + v_n$. The LM procedure asserts that if v_n is not spatially correlated, then X_n and Y_n are spatially integrated; and if v_n is spatially correlated, X_n and Y_n are not spatially integrated because of overdifferencing. Nevertheless, regressing ΔY_n on ΔX_n is equivalent to regressing two white noise series, ϵ_Y and ϵ_X ; because Y and X are both $I(1)$. Hence, the corresponding residuals must be not spatially correlated as long as ϵ_Y and ϵ_X are independent, regardless of whether v_n is spatially correlated, nonstationary, or not.

5.2. A Wald Test for Spatial Nonstationarity

Lauridsen and Kosfeld [65] suggest a Wald post-test statistic. Based on MLE, under $H_0 : R\theta = q$, the general form of the Wald test is $W = (R\theta - q)'(RVR')^{-1}(R\theta - q)$, where $V = I^{-1}$ is the inverse of information matrix. If we specify the null hypothesis as $\lambda = 1$, then with $R = (0', 1, 0)$ and $q = 1$, we have $W = \frac{(\lambda - 1)^2}{V_\lambda}$, where V_λ is the diagonal element of V corresponding to λ . However, as mentioned before, Wald statistics may not have a standard distribution so simulations are conducted. Unlike Fingleton [14], to generate SAR series with spatial unit roots, Lauridsen and Kosfeld [65] do not introduce the noncircular matrix. Thus $\lambda = 1$ is a singular point of $(I_n - \lambda W_n)$ so the inverse does not exist. To solve this issue, they use the Moore–Penrose pseudoinverse.

According to Monte Carlo simulation, they find the critical limit of the Wald test under spatial nonstationarity is higher than the $\chi^2(1)$ distribution, especially for the 5th and 10th percentile.

5.3. Test Unit Roots and Cointegration in the Sense of Spatial Impulses

Beenstock et al. [19] come up with an innovative method to test spatial unit roots and spatial cointegration by considering the behavior of the variance and the spatial impulse. Also, they do not assume unconnected spatial units or row normalize the spatial weight matrix either. Thus, based on the topology of the unit neighborhood, the spatial unit roots are $\lambda^* = \frac{1}{n}$ in the regular lattices where n is the maximum and general number of neighbors of each unit. For example, $n = 2$, for bilateral space, $n = 4$ for rook lattice and $n = 8$ for queen lattice, with $\lambda^* = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ respectively. (“The weight matrix with first-order contiguity according to the rook criterion has the cells immediately above, below, to the right, and to the left, for a total of four neighboring cells. The weight matrix with first-order contiguity according to the queen criterion is eight cells immediately surrounding the central cell” [75] (p. 131). For the introduction of other types of the spatial weight matrices, see Kelejian and Robinson [26] (pp. 94–95).) And with spatial unit roots, $(I_n - \lambda^* W_n)^{-1}$ is still well-defined because of the existence of the edge effect, that is, there exist some units having fewer neighbors than n . However, even though $(I_n - \lambda^* W_n)^{-1}$ exists, the variance tends to explode even in finite sample space, which provides us with a way to determine the spatial unit roots for any arbitrary irregular lattices.

5.3.1. Spatial Impulse

The spatial impulse response is essentially the consequence of the shocks from one location to another. Intuitively, shocks should have no effect on the remote units if the

spatial data are stationary. Beenstock et al. [19] first consider the simplest SAR model in lateral space:

$$\begin{aligned}
 Y_{n,j} &= \lambda(Y_{n,j+1} + Y_{n,j-1}) + u_{n,j}, \quad j = -\infty, \dots, \infty, \\
 \implies (1 - \lambda^{-1}L + L^2)Y_{n,j} &= -\lambda^{-1}u_{n,j-1},
 \end{aligned}
 \tag{110}$$

where L denote a spatial lag operator such that $L^i Y_{n,j} = Y_{n,j-i}$. The auxiliary equation is

$$x^2 - \lambda^{-1}x + 1 = 0. \tag{111}$$

When the discriminant of the above equation is greater than 0, $0 < \lambda < \frac{1}{2}$, and there are two different solutions, $x_1 < 1 < x_2$, by Vieta's formula. Hence, Beenstock et al. [19] express $Y_{n,j}$ as

$$\begin{aligned}
 Y_{n,j} &= \frac{\lambda^{-1}}{x_1^{-1} - x_1} \left[\sum_{i=1}^{\infty} x_1^i u_{n,j-i} + \sum_{i=0}^{\infty} x_1^i u_{n,j+i} \right], \\
 \implies \frac{\partial Y_{n,j}}{\partial u_{n,j-i}} &= \frac{\partial Y_{n,j}}{\partial u_{n,j+i}} = \frac{\lambda^{-1}x_1^i}{x_1^{-1} - x_1},
 \end{aligned}
 \tag{112}$$

where (112) is known as the Wold representation that expresses $Y_{n,j}$ in terms of the shocks. The impulse from location $j - i$ to i tends to 0 because $x_1 < 1$. Also, x_1 varies with λ . When $\lambda = \lambda^* = \frac{1}{2}$, $x_1 = x_2 = 1$, so the impulse does not die out with distance and explodes. This fact can also be seen from

$$\text{Var}(Y_{n,j}) = \frac{\lambda^{-2}(1 + x_1^2)}{(1 - x_1^2)(x_1^{-1} - x_1)^2} \sigma_u^2. \tag{113}$$

If $0 < \lambda < \frac{1}{2}$, $\text{Var}(Y_{n,j})$ is finite and independent of j ; if $\lambda = \frac{1}{2}$, $x_1 = 1$, and $\text{Var}(Y_{n,j})$ is infinite.

For the bilateral space case, $\lambda^* = \frac{1}{n}$ are the spatial unit roots as shown before. Because of the edge effect, the singular point is strictly greater but approaches λ^* . This fact shows a downside of the row-normalized spatial weight matrix: it overstates the true weight of the unit at the edge of the lattice. For example, in a rook lattice, the units have three neighbors with weight $\frac{1}{3}$ at the edge, and four neighbors with weight $\frac{1}{4}$ in the center. The row-normalized procedure assigns a higher weight to the neighbors of edge units. This weight assignment is not necessarily reasonable and makes the spatial unit roots the same as the singular points. Moreover, without row-normalization, the edging units play the role of the unconnected unit as in Fingleton [14]. The general SAR model in bilateral space is $Y_n = \lambda W_n Y_n + u_n$ and the Wold representation is $Y_n = A_n u_n$, where $A_n = (I_n - \lambda W_n)^{-1} = I_n + \sum_{i=1}^{\infty} \lambda^i W_n^i$. Let the spatial impulse response be defined as $\frac{dY_{n,j}}{du_{n,j}} = a_{jj}$ and $\frac{dY_{n,j}}{du_{n,i}} = a_{ji}$. Analytical solutions of spatial impulse response in bilateral and higher dimension lattices are not obtained, but Beenstock et al. [19] expect a_{jj} to be positively related to the number of spatial units because of the larger spillover effect and a_{ji} varies inversely with the distance between i and j in the stationary case. If spatial unit roots exist, as in the lateral case, the impulse a_{ji} would not die out as the distance increases. This is supported by the simulation, though only the finite sample case could be simulated, see Beenstock and Felsenstein [76] (Figures 5.2 and 5.4). Compared with $\lambda < \lambda^*$, when $\lambda = \lambda^*$, it obviously shows a qualitative difference in the persistence of spatial impulses, as well as in the tendency for the explosion of the variance.

In the irregular lattices, the number of neighbors for the unit is undetermined generally, so the spatial unit roots, λ^* , cannot be calculated as the reciprocal of n . However, since the nonstationarity implies that spatial impulses do not disappear, one can find the empirical spatial unit roots by simulation.

The simulation method in Beenstock et al. [19] to calculate the critical value is pretty flexible and can be adapted to different models. For example, when both dynamic and spatial terms are included, Beenstock and Felsenstein [23] develop a similar procedure for testing cointegration in nonstationary panel data when estimating the spatial spillover effect in housing construction for Israel.

5.3.2. Spatial Unit Roots and Cointegration Tests

Knowing the spatial unit roots λ^* , Beenstock et al. [19] conduct Monte Carlo simulations to generate the artificial SAR series and use the MLE method to estimate SAC to obtain the corresponding distributions under different topologies (different sample size, criteria, etc.). Results show that the empirical distribution of SAC could be used to construct interval estimation and critical values for statistics that test spatial unit roots. For the spatial cointegration test, a similar procedure applies, but OLS estimation is used.

5.4. Some Applications

Kosfeld and Lauridsen [77] offer an application of the two-stage LM test in Section 5.1 to the income and productivity convergence in the German regional labor market. They find highly significant LME and DLME statistics (refer to formulas above like (107)) for all variables, which means the spatial unit roots are rejected. Yesilyurt and Elhorst [20] estimate the spatial interaction effects of inflation in Turkey. Because the regional inflation rates have a high tendency to co-move over time, they question whether the inflation rates of different regions are stationary in space. Using the two-stage LM statistics from (108), they find that the inflation curve is stationary in space. Olejnik [21] studies the income process of the extended European 25 based on the augmented Solow model taking into account the spatial autocorrelation effect. The stationarity of the error term as well as all variables in the model are investigated. No problem of spurious regression is found. Machado et al. [22] examine the spatial correlation of traffic accidents of vulnerable road users (such as pedestrians and cyclists) in big cities and detect the factors that contribute to these accidents. Because their study covers several cities, the model specifications may vary across different locations. Thus they use the two-stage LM statistic to choose the best model, see Machado et al. [22] (Table 4). Though the Wald post-test in Lauridsen and Kosfeld [65] is asymptotically equivalent to the LM test, “It is generally recommended to choose among these alternatives on the basis of computational ease [78] (p. 94)”.

6. Related Topics

Spatial panel data have been studied extensively. The spatial dependence is incorporated in the error component [12,75,79] or by spatial lag dependence [11,80]. See Baltagi [2] for a textbook discussion. Also, the panel data model can have time lagged dependent variables. If the panel data model includes both spatial and dynamic features, it is named as spatial dynamic panel data (SDPD) model by Yu et al. [10]. Yu et al. [10], Yu and Lee [16] and Yu et al. [36] study the QMLE estimator of the SDPD model under stable, unit roots, and spatial cointegration respectively. The concept of unit roots under the SDPD model is a combination of the spatial and dynamic one. To see this more clearly, Yu et al. [10] specify the model as

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt} \beta_0 + c_{n0} + V_{nt}, \tag{114}$$

where $Y_{nt} = (y_{1t}, y_{2t}, \dots, y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, \dots, v_{nt})'$ are $n \times 1$ column vectors. Since $S_n(\lambda) = I_n - \lambda W_n$, assuming S_n is invertable, the reduced form is

$$Y_{nt} = A_n Y_{n,t-1} + S_n^{-1} X_{nt} \beta_0 + S_n^{-1} c_{n0} + \alpha_{t0} S_n^{-1} l_n + S_n^{-1} V_{nt}, \tag{115}$$

where $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$. If the infinite sums are well-defined, then by continuous substitution

$$Y_{nt} = \sum_{h=0}^{\infty} A_n^h S_n^{-1} (c_{n0} + X_{n,t-h} \beta_0 + V_{n,t-h}). \tag{116}$$

So instead of focusing on the singular points of S_n , A_n should be considered, which contains λ , γ and ρ : the parameter of contemporaneous spatial effect, time lagged variable and time–spatial effect. A similar process as in Section 3.3.1, letting $\bar{w}_n = \text{diag}\{\bar{w}_{n1}, \bar{w}_{n2}, \dots, \bar{w}_{nn}\}$ be the eigenvalue matrix of W_n , Yu et al. [10] show that the eigenvalue matrix of A_n is $D_n = (I_n - \lambda_0 \bar{w}_n)^{-1}(\gamma_0 I_n + \rho_0 \bar{w}_n)$, which can be decomposed as $D_n = \frac{\gamma_0 + \rho_0}{1 - \lambda_0} J_n + \tilde{D}_n$. The power matrix of A_n follows as $A_n^h = \left(\frac{\gamma_0 + \rho_0}{1 - \lambda_0}\right)^h R_n J_n R_n^{-1} + B_n^h$ with $B_n^h = R_n \tilde{D}_n^h R_n^{-1}$ since the eigenvector matrix R_n is orthogonal and $J_n \tilde{D}_n = 0$. Thus, whether Y_{nt} is stable or not depends on the value of $\frac{\gamma_0 + \rho_0}{1 - \lambda_0}$ compared with 1. Consequently, the decomposition of Y_{nt} , which is a generalization of (40), can be expressed as

$$Y_{nt} = Y_{nt}^u + Y_{nt}^s + Y_{nt}^\alpha, \tag{117}$$

where Y_{nt}^s is a possible stable part, Y_{nt}^u is a possible unstable part, and Y_{nt}^α is the time effect part, see Lee and Yu [81] for details. A data transformation procedure is imposed by them to eliminate both the time effects and the possible unstable term. Based on their analysis, the eigenvalues of A_n , the asymptotic properties of QMLE and bias are derived. When eigenvalues of A_n are all less than 1 ($\gamma_0 + \lambda_0 + \rho_0 < 1$), or some equal to 1 ($\gamma_0 + \lambda_0 + \rho_0 = 1$ and $\gamma_0 \neq 1$), or all equal to 1 ($\gamma_0 + \lambda_0 + \rho_0 = 1$ and $\gamma_0 = 1$), the information matrix has different properties, see Yu and Lee [16] (Table 5).

Thus, the test for the unit eigenvalues of A_n is of great importance. Most attention is paid to the unit roots in the time dimension, i.e., $\gamma_0 = 1$ and equivalently, if $\lambda_0 + \rho_0 = 0$. Unit root tests in panel data under spatial dependence have been extensively studied, see Baltagi [2] (Section 12.3) for a summary. Also, the performance of different tests has been considered in Baltagi et al. [24]. The test for $H_0 : \gamma_0 + \rho_0 + \lambda_0 = 1$ has been investigated in Lee and Yu [81] (Section 14.3.4). Such a test works well when $\lambda_0 + \gamma_0 + \rho_0 < 1$. However, when $\lambda_0 + \gamma_0 + \rho_0 > 1$ and T is small, it is not reliable. Thus, further study of the unit root test for the SDPD model should be investigated.

Recently, another approach to describe strong spatial dependence has been proposed by Müller and Watson [49]. Since the spatial units are not neatly arranged, i.e., irregular lattice, they do not model the spatial dependence by SAR model but “posit a continuous parameter model of spatial variation [49]”. They use the Lévy–Brownian motion to define the spatial $I(1)$ process, $L(s), s \in R^d$, which is a generalization of the Wiener process that is widely used in time series in d dimensions (when $d = 2$, it could be regarded as a random walk on the plane). The advantage of the Lévy–Brownian motion is that such a process is isotropic, which means the relative variance between two locations is determined by the distance but not the orientation (see Anselin [1] (p. 42)). The functional central limit theorem (FCLT) is established to measure the asymptotic behavior of such a process. When regressing two independent $I(1)$ processes, spurious regression also occurs since classical, HAC-corrected and clustered standard errors F statistics diverge to infinity, which is similar to that in Fingleton [14]. To remedy this situation, “difference” regression is again considered. Unlike time series, Müller and Watson [49] introduce the isotropic differences that treat all directions symmetrically. That is, they regard the weighted values of the neighborhood as the “average” value of the current location, just as $W_n Y_n$ in the SAR model. And the difference transformation is defined as $y_i^* = \frac{1}{n} \sum_{\ell \neq i} \kappa_b(\lambda_n^{-1} |s_\ell - s_i|)(y_\ell - y_i)$, where κ_b is some weighting function. Their simulations show regressions using isotropic differences do not suffer from spurious regression problems and valid inference can be conducted. Some test procedures for $H_0 : I(1)$ and $H_0 : I(0)$ are also suggested.

7. Conclusions

This paper briefly surveys spatial unit roots in spatial models. First, some fundamental concepts in spatial econometrics are introduced. Spatial unit roots in SAR and SEM models may lead to spurious regression. For the estimation and inference in the presence of spatial unit roots, QMLE and 2SLS methods are generally used and have satisfactory properties after scaling. The compactness assumption has been recently relaxed in spatial econometrics which potentially makes the spatial unit roots no longer a concern but its implication to concepts like stationarity, and spatial cointegration should be further investigated. The doubly geometric spatial autoregressive process, has been widely used in some scientific fields of which the most concern is about regular lattice. Similar to time series, exact orders of convergence for different estimators are obtained because of its simple specification. But this limits its application in economics where irregular lattice and different types of weight matrices are applied.

To detect possible spatial unit roots, as well as spatial cointegration, several test procedures have been proposed. Their applications are rather limited and depend heavily on simulations to obtain critical values. This could be explained by the fact that statistics under spatial scenarios generally do not have standard asymptotic distributions, not to mention the irregular lattice.

Lastly, some related topics were introduced. The idea of singular points is generalized in SDPD model because such a model includes the time lagged variable that is based on the traditional SAR model. However, the existing literature focuses on the temporal unit roots in the SDPD model. Recently, an innovative way to study spatial unit roots describes the underlying spatial process using Lévy–Brownian motion, which is a generalization and spatial analogy to the time series counterpart. The limitations of different approaches and further research were also discussed.

Author Contributions: Conceptualization, B.H.B. and J.S.; methodology, B.H.B. and J.S.; formal analysis, B.H.B. and J.S.; writing—original draft preparation, B.H.B. and J.S.; writing—review and editing, B.H.B. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We would like to thank the editors and four anonymous referees for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1988.
2. Baltagi, B.H. *Econometric Analysis of Panel Data*, 6th ed.; Springer: Cham, Switzerland, 2021.
3. Elhorst, J.P. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*; Springer: Heidelberg, Germany, 2014; Volume 479.
4. Ord, K. Estimation methods for models of spatial interaction. *J. Am. Stat. Assoc.* **1975**, *70*, 120–126. [[CrossRef](#)]
5. Lee, L.F. Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **2004**, *72*, 1899–1925. [[CrossRef](#)]
6. Kelejian, H.H.; Prucha, I.R. A generalized moments estimator for the autoregressive parameter in a spatial model. *Int. Econ. Rev.* **1999**, *40*, 509–533. [[CrossRef](#)]
7. Kelejian, H.H.; Prucha, I.R. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real Estate Financ. Econ.* **1998**, *17*, 99–121. [[CrossRef](#)]
8. Lee, L.F. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econom. Rev.* **2003**, *22*, 307–335. [[CrossRef](#)]
9. Lee, L.F. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J. Econom.* **2007**, *137*, 489–514. [[CrossRef](#)]
10. Yu, J.; De Jong, R.; Lee, L.F. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *J. Econom.* **2008**, *146*, 118–134. [[CrossRef](#)]
11. Baltagi, B.H.; Liu, L. Instrumental variable estimation of a spatial autoregressive panel model with random effects. *Econ. Lett.* **2011**, *111*, 135–137. [[CrossRef](#)]

12. Kapoor, M.; Kelejian, H.H.; Prucha, I.R. Panel data models with spatially correlated error components. *J. Econom.* **2007**, *140*, 97–130. [[CrossRef](#)]
13. Keller, W.; Shiue, C.H. The origin of spatial interaction. *J. Econom.* **2007**, *140*, 304–332. [[CrossRef](#)]
14. Fingleton, B. Spurious spatial regression: Some Monte Carlo results with a spatial unit root and spatial cointegration. *J. Reg. Sci.* **1999**, *39*, 1–19. [[CrossRef](#)]
15. Lee, L.F.; Yu, J. Near unit root in the spatial autoregressive model. *Spat. Econ. Anal.* **2013**, *8*, 314–351. [[CrossRef](#)]
16. Yu, J.; Lee, L.F. Estimation of unit root spatial dynamic panel data models. *Econom. Theory* **2010**, *26*, 1332–1362. [[CrossRef](#)]
17. Lauridsen, J.; Kosfeld, R. A test strategy for spurious spatial regression, spatial nonstationarity, and spatial cointegration. *Pap. Reg. Sci.* **2006**, *85*, 363–377. [[CrossRef](#)]
18. Lauridsen, J.; Kosfeld, R. Spatial cointegration and heteroscedasticity. *J. Geogr. Syst.* **2007**, *9*, 253–265. [[CrossRef](#)]
19. Beenstock, M.; Feldman, D.; Felsenstein, D. Testing for unit roots and cointegration in spatial cross-section data. *Spat. Econ. Anal.* **2012**, *7*, 203–222. [[CrossRef](#)]
20. Yesilyurt, F.; Elhorst, J.P. A regional analysis of inflation dynamics in Turkey. *Ann. Reg. Sci.* **2014**, *52*, 1–17. [[CrossRef](#)]
21. Olejnik, A. Using the spatial autoregressively distributed lag model in assessing the regional convergence of per-capita income in the EU25. *Pap. Reg. Sci.* **2008**, *87*, 371–385. [[CrossRef](#)]
22. Machado, C.A.S.; Giannotti, M.A.; Chiaravalloti Neto, F.; Tripodi, A.; Persia, L.; Quintanilha, J.A. Characterization of black spot zones for vulnerable road users in São Paulo (Brazil) and Rome (Italy). *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 858–882. [[CrossRef](#)]
23. Beenstock, M.; Felsenstein, D. Estimating spatial spillover in housing construction with nonstationary panel data. *J. Hous. Econ.* **2015**, *28*, 42–58. [[CrossRef](#)]
24. Baltagi, B.H.; Bresson, G.; Pirotte, A. Panel unit root tests and spatial dependence. *J. Appl. Econom.* **2007**, *22*, 339–360. [[CrossRef](#)]
25. Horn, R.A.; Johnson, C.R. *Matrix Analysis*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2012.
26. Kelejian, H.H.; Robinson, D.P. Spatial correlation: A suggested alternative to the autoregressive model. In *New Directions in Spatial Econometrics*; Anselin, L., Florax, R.J.G.M., Eds.; Springer: Berlin/Heidelberg, Germany, 1995; pp. 75–95.
27. Griffith, D.A. *Advanced Spatial Statistics: Special Topics in the Exploration of Quantitative Spatial Data Series*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 12.
28. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994.
29. Kelejian, H.H.; Prucha, I.R. On the asymptotic distribution of the Moran I test statistic with applications. *J. Econom.* **2001**, *104*, 219–257. [[CrossRef](#)]
30. Kelejian, H.H.; Prucha, I.R. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econom.* **2010**, *157*, 53–67. [[CrossRef](#)] [[PubMed](#)]
31. Lee, L.F.; Yang, C.; Yu, J. QML and efficient GMM estimation of spatial autoregressive models with dominant (popular) units. *J. Bus. Econ. Stat.* **2023**, *41*, 550–562. [[CrossRef](#)]
32. Mur, J.; Trivez, F.J. Unit roots and deterministic trends in spatial econometric models. *Int. Reg. Sci. Rev.* **2003**, *26*, 289–312. [[CrossRef](#)]
33. Robert, E.F. Co-integration and error correction: Representation, estimation, and testing. *Econometrica* **1987**, *55*, 251–276.
34. Lee, L.F.; Yu, J. Spatial nonstationarity and spurious regression: The case with a row-normalized spatial weights matrix. *Spat. Econ. Anal.* **2009**, *4*, 301–327. [[CrossRef](#)]
35. Lee, L.F. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econom. Theory* **2002**, *18*, 252–277. [[CrossRef](#)]
36. Yu, J.; de Jong, R.; Lee, L.F. Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration. *J. Econom.* **2012**, *167*, 16–37. [[CrossRef](#)]
37. Baltagi, B.H.; Kao, C.; Liu, L. The estimation and testing of a linear regression with near unit root in the spatial autoregressive error term. *Spat. Econ. Anal.* **2013**, *8*, 241–270. [[CrossRef](#)]
38. Baltagi, B.H.; Liu, L. Spurious spatial regression with equal weights. *Stat. Probab. Lett.* **2010**, *80*, 1640–1642. [[CrossRef](#)]
39. Kelejian, H.H.; Prucha, I.R. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Reg. Sci. Urban Econ.* **2002**, *32*, 691–707. [[CrossRef](#)]
40. Krämer, W.; Donniger, C. Spatial autocorrelation among errors and the relative efficiency of OLS in the linear regression model. *J. Am. Stat. Assoc.* **1987**, *82*, 577–579.
41. Tilke, C. The relative efficiency of OLS in the linear regression model with spatially autocorrelated errors. *Stat. Pap.* **1993**, *34*, 263–270. [[CrossRef](#)]
42. Krämer, W.; Baltagi, B. A general condition for an optimal limiting efficiency of OLS in the general linear regression model. *Econ. Lett.* **1996**, *50*, 13–17. [[CrossRef](#)]
43. Martellosio, F. Efficiency of the OLS estimator in the vicinity of a spatial unit root. *Stat. Probab. Lett.* **2011**, *81*, 1285–1291. [[CrossRef](#)]
44. Baran, S.; Pap, G.; van Zuijlen, M.C. Asymptotic inference for unit roots in spatial triangular autoregression. *Acta Appl. Math.* **2007**, *96*, 17–42. [[CrossRef](#)]

45. Basu, S.; Reinsel, G.C. A note on properties of spatial Yule-Walker estimators. *J. Stat. Comput. Simul.* **1992**, *41*, 243–255. [[CrossRef](#)]
46. Basu, S.; Reinsel, G.C. Properties of the spatial unilateral first-order ARMA model. *Adv. Appl. Probab.* **1993**, *25*, 631–648. [[CrossRef](#)]
47. Baran, S. On the variances of a spatial unit root model. *Lith. Math. J.* **2011**, *51*, 122–140. [[CrossRef](#)]
48. Paulauskas, V. On unit roots for spatial autoregressive models. *J. Multivar. Anal.* **2007**, *98*, 209–226. [[CrossRef](#)]
49. Müller, U.K.; Watson, M.W. *Spatial Unit Roots*; Princeton University: Princeton, NJ, USA, 2022.
50. Gupta, A. Estimation of spatial autoregressions with stochastic weight matrices. *Econom. Theory* **2019**, *35*, 417–463. [[CrossRef](#)]
51. Liu, T.; Xu, X.; Lee, L.F. Consistency without compactness of the parameter space in spatial econometrics. *Econ. Lett.* **2022**, *210*, 110224. [[CrossRef](#)]
52. Newey, W.K.; McFadden, D. Large sample estimation and hypothesis testing. In *Handbook of Econometrics*; Elsevier: Amsterdam, The Netherlands, 1994; Volume 4, Chapter 35, pp. 2111–2245.
53. Gupta, A. Efficient closed-form estimation of large spatial autoregressions. *J. Econom.* **2023**, *232*, 148–167. [[CrossRef](#)]
54. Rossi, F.; Lieberman, O. Spatial autoregressions with an extended parameter space and similarity-based weights. *J. Econom.* **2023**, *235*, 1770–1798. [[CrossRef](#)]
55. Liu, L. A note on 2SLS estimation of the mixed regressive spatial autoregressive model. *Econ. Lett.* **2015**, *134*, 49–52. [[CrossRef](#)]
56. Bhattacharyya, B.; Richardson, G.; Franklin, L. Asymptotic inference for near unit roots in spatial autoregression. *Ann. Stat.* **1997**, *25*, 1709–1724. [[CrossRef](#)]
57. Bhattacharyya, B.; Khalil, T.; Richardson, G. Gauss-Newton estimation of parameters for a spatial autoregression model. *Stat. Probab. Lett.* **1996**, *28*, 173–179. [[CrossRef](#)]
58. Phillips, P.C. Time series regression with a unit root. *Econometrica* **1987**, *55*, 277–301. [[CrossRef](#)]
59. Baran, S.; Pap, G. Parameter estimation in a spatial unilateral unit root autoregressive model. *J. Multivar. Anal.* **2012**, *107*, 282–305. [[CrossRef](#)]
60. Baran, S.; Pap, G.; Van Zuijlen, M.C. Asymptotic inference for a nearly unstable sequence of stationary spatial AR models. *Stat. Probab. Lett.* **2004**, *69*, 53–61. [[CrossRef](#)]
61. Baran, S.; Pap, G. On the least squares estimator in a nearly unstable sequence of stationary spatial AR models. *J. Multivar. Anal.* **2009**, *100*, 686–698. [[CrossRef](#)]
62. Roknossadati, S.; Zarepour, M. M-estimation for a spatial unilateral autoregressive model with infinite variance innovations. *Econom. Theory* **2010**, *26*, 1663–1682. [[CrossRef](#)]
63. Roknossadati, S.; Zarepour, M. M-estimation for near unit roots in spatial autoregression with infinite variance. *Statistics* **2011**, *45*, 337–348. [[CrossRef](#)]
64. Ahlgren, N.; Gerkman, L. Inference in unilateral spatial econometric models. *Bull. Int. Stat. Inst.* **2007**, *56*, 1–44.
65. Lauridsen, J.; Kosfeld, R. A Wald test for spatial nonstationarity. *Estud. Econ. Apl.* **2004**, *22*, 475–486.
66. Martellosio, F. Power properties of invariant tests for spatial autocorrelation in linear regression. *Econom. Theory* **2010**, *26*, 152–186. [[CrossRef](#)]
67. King, M.L. A small sample property of the Cliff-Ord test for spatial correlation. *J. R. Stat. Soc. Ser. B (Methodological)* **1981**, *43*, 263–264. [[CrossRef](#)]
68. Krämer, W. Finite sample power of Cliff-Ord-type tests for spatial disturbance correlation in linear regression. *J. Stat. Plan. Inference* **2005**, *128*, 489–496. [[CrossRef](#)]
69. Martellosio, F. Testing for spatial autocorrelation: The regressors that make the power disappear. *Econom. Rev.* **2012**, *31*, 215–240. [[CrossRef](#)]
70. Born, B.; Breitung, J. Simple regression-based tests for spatial dependence. *Econom. J.* **2011**, *14*, 330–342. [[CrossRef](#)]
71. Baltagi, B.H.; Yang, Z. Heteroskedasticity and non-normality robust LM tests for spatial dependence. *Reg. Sci. Urban Econ.* **2013**, *43*, 725–739. [[CrossRef](#)]
72. Baltagi, B.H.; Yang, Z. Standardized LM tests for spatial error dependence in linear or panel regressions. *Econom. J.* **2013**, *16*, 103–134. [[CrossRef](#)]
73. Preinerstorfer, D. How to avoid the zero-power trap in testing for correlation. *Econom. Theory* **2021**, *39*, 1292–1324. [[CrossRef](#)]
74. Anselin, L. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geogr. Anal.* **1988**, *20*, 1–17. [[CrossRef](#)]
75. Baltagi, B.H.; Song, S.H.; Koh, W. Testing panel data regression models with spatial error correlation. *J. Econom.* **2003**, *117*, 123–150. [[CrossRef](#)]
76. Beenstock, M.; Felsenstein, D. Unit root and cointegration tests in spatial cross-section data. In *The Econometric Analysis of Non-Stationary Spatial Panel Data*; Advances in Spatial Science; Springer: Berlin/Heidelberg, Germany, 2019; Chapter 5, pp. 97–127.
77. Kosfeld, R.; Lauridsen, J. Dynamic spatial modelling of regional convergence processes. *Empir. Econ.* **2004**, *29*, 705–722. [[CrossRef](#)]
78. Vaona, A. Spatial autocorrelation and the sensitivity of RESET: A simulation study. *J. Geogr. Syst.* **2010**, *12*, 89–103. [[CrossRef](#)]
79. Fingleton, B. A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. *Empir. Econ.* **2008**, *34*, 35–57. [[CrossRef](#)]

-
80. Baltagi, B.H.; Liu, L. Testing for random effects and spatial lag dependence in panel data models. *Stat. Probab. Lett.* **2008**, *78*, 3304–3306. [[CrossRef](#)]
 81. Lee, L.F.; Yu, J. A unified transformation approach for the estimation of spatial dynamic panel data models: Stability, spatial cointegration and explosive roots. In *Handbook on Empirical Economics and Finance*; Ullah, A., Giles, D., Eds.; Taylor & Francis Group: New York, NY, USA, 2011; Chapter 13, pp. 397–434.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.