Spatial Nonstationarity and Spurious Regression: The Case with Row-Normalized Spatial Weights Matrix

March 26, 2009

Abstract

This paper investigates the spurious regression in the spatial setting where the regressant and regressors may be generated from possible nonstationary spatial autoregressive processes. Under the near unit root specification with a row-normalized spatial weights matrix, it is shown that the possible spurious regression phenomena in the spatial setting are relatively weaker than those in the nonstationary time series scenario. The regression estimates might or might not converge to 0. The divergence might occur only when the regressant has a near unit root much closer to unity than that of the regressor. For the t and F statistics, there could be over-rejection of the null of uncorrelatedness under certain situations, but they do not diverge. However, the coefficient of determination R^2 converges to 0, which provides strong evidence of the spurious regression even when t and F statistics are large. The Moran I test may reject the null of no spatial dependence in the least squares residual. Simulation results about different statistics are in line with the theoretical results we derive in this paper.

JEL classification: C13; C23; R15
Keywords: Near Unit Root, Spatial Nonstationarity, Spurious Regression

1 Introduction

Spatial econometrics deals with spatial correlations among economic units. The spatial autoregressive (SAR) model by Cliff and Ord (1973) has received the most attention in economics[1]. Recently, there is a growing interest in extending the unit root and cointegration phenomena in the time series to the spatial setting. This is because the SAR model has sometimes been regarded as a generalization of an autoregressive model in the time series to the spatial setting. Fingleton (1999) has investigated implications of the unit root in the SAR model and has detected spurious regressions via Monte Carlo simulations. In order to generate unit root spatial data, he introduces an unconnected central unit such that the spatial weights matrix has a zero row. In that setting, unit root can be present and the system is an equilibrium one. This setting is analogous to having an initial observation in the time series autoregressive model. Mur and Trivez (2003) follow up Fingleton (1999) by allowing a deterministic intercept, which can take large values in the data generating process (DGP). By arguing that the intercept may generate a spatial trend variable in the reduced form equation, they investigate possible spurious regression features by Monte Carlo simulations. They find that the generated deterministic trend can also cause spurious regression. Lauridsen and Kosfeld (2006, 2007) develop a two step LM test to distinguish the possible two sources of spurious regression: one is the unit root in the DGPs of the regressant and/or regressors, and the other is the spatial error in the regression.

However, the assumption of an unconnected unit might be too strong and have limited applications in empirical studies. In empirical work, it is common to have a row-normalized spatial weights matrix with a few exceptions. In this paper, we investigate the spurious regression in the spatial setting with attention to a weights matrix with row-normalization, where there is no unconnected central unit. While the spatial weights matrix in Fingleton (1999) would not be row-normalized, the row-normalization of the weights matrix is allowed in the studies of Lauridsen and Kosfeld (2006, 2007). As pointed out by Mur and Trivez (2003), the possible spurious regression phenomenon in Fingleton (1999) might likely be generated by large variances of the process instead of non-circularity, as the introduced unconnected central unit might not generate accumulated noise disturbances for other spatial units. Hence, the spurious regression phenomenon could be generated by large variances in the process instead of non-circularity. Because of the popularity of the row-normalized weights matrix, we intend to investigate possible spurious regression in this spatial setting. For a spatial weights matrix being row-normalized, the spatial effect cannot be equal to one (Ord 1975). Therefore, we consider the case where the spatial effect is close to one. In the time series, the near unit root feature turns out to be similar to that of the exact unit root case in terms of asymptotic analysis (see Phillips (1987)). In the spatial setting, if the true spatial effect is near unity, the spatial dependence across units will be strong and the variances of the dependent variables will become large even though the spatial weights matrix is row and column sum bounded. Lee and Yu (2007) derive the asymptotic properties of the instrumental variable (IV) and quasi-maximum likelihood (QML) estimators when the true spatial effect can go to unity at any rate. In this paper, we investigate the spurious regression when we have possible nonstationary components in the DGP of the regressant and/or the regressors, where the nonstationarity is caused by near unit roots.

In the time series, Granger and Newbold (1974) report the spurious regression by Monte Carlo. Phillips (1986) provides analytical results of linear regressions involving general integrated random processes, including the spurious regression of Granger and Newbold (1974) and the cointegration regression of Granger and Engle (1985). With a row-normalized weights matrix and a near unit root spatial effect, this paper investigates analytically whether spurious regression phenomena could occur or not in the spatial setting. We find that the possible spurious regression phenomena in the spatial setting are much weaker than those of the integrated time series. The least square estimates of the coefficients in a spurious regression might or might not converge to 0. They might be divergent only when the regressant has a near unit root much closer to unity than those of the regressor. Even so, the corresponding t-statistic is asymptotically normal and does not diverge. The Moran I or LM test for spatial correlation in the least squares residual will show a significant correlation. However, the determinant of coefficient R^2 will always converge to 0 in probability. This implies that when high values of t or F were observed, low values of R^2 would indicate the spurious regression in the spatial setting. The analytical results are given and simulation results for finite samples are provided to support the theoretical implications.

The rest of this paper is organized as follows. Section 2 specifies the DGP of the regressant and regressors. Section 3 investigates the possible consequences of spurious regression in terms of the least squares estimates, the associated t or F statistics, R^2, and Moran I and LM test statistics (to test the spatial effect in the regression residuals). Simulation results are reported in Section 4 and conclusions are made in Section 5. Several lemmas are provided in the Appendices.

2 The DGP

Consider the cross sectional (first order) SAR models

$$\begin{aligned} Y_{1n} &= \lambda_{1n} W_{1n} Y_{1n} + Z_{1n}\gamma_1 + \epsilon_{1n}, \\ Y_{2n} &= \lambda_{2n} W_{2n} Y_{2n} + Z_{2n}\gamma_2 + \epsilon_{2n}, \\ &\vdots \\ Y_{mn} &= \lambda_{mn} W_{mn} Y_{mn} + Z_{mn}\gamma_m + \epsilon_{mn}, \end{aligned} \tag{2.1}$$

where, for $j = 1, ..., m$, Y_{jn} is $n \times 1$ vector of dependent variables, Z_{jn} is $n \times k_j$ nonstochastic exogenous variables, W_{jn} is a nonstochastic spatial weights matrix, and the disturbance of the n-dimensional vector ϵ_{jn} is *i.i.d.* $(0, \sigma_j^2)$. In spatial econometric literature, $W_{jn}Y_{jn}$ is usually referred to as a spatial lag of Y_{jn}.

Suppose that W_{jn} is diagonalizable with eigenvalues d_{ij} such that either $d_{ij} = 1$ or $|d_{ij}| < 1$[2] for $i = 1, ..., n$. Furthermore, suppose that there are m_j unit eigenvalues, and the remaining $(n - m_j)$ eigenvalues are all uniformly bounded away from 1 in the absolute value for all n. For any value λ in its parameter space, λ can be reparameterized as $\lambda = 1 - \frac{1}{\psi}$. The near unit root case refers to the situation that, for the true spatial effect parameter $\lambda_{jn} = 1 - \frac{1}{\psi_{jn}}$, ψ_{jn} goes to infinity as n goes to infinity. Thus, as n goes to infinity, λ_{jn} approaches 1. The ψ_{jn} specifies, in its general form, how fast λ_{jn} approaches unity as the sample size n increases[3]. Note that because the SAR model is specified to be an equilibrium model, it is not meaningful to assume that $\lambda_{jn} = 1$. With our specification of the near unit root case, as $\lambda_{jn} < 1$, this model is still an equilibrium model for any finite n. This is because $(I_n - \lambda_{jn} W_{jn})$ is invertible. With W_{jn} being diagonalizable, let R_{jn} be the eigenvector matrix such that

$$W_{jn} = R_{jn} D_{jn} R_{jn}^{-1}, \tag{2.2}$$

where $D_{jn} = diag\{\mathbf{1}_{m_j}, d_{m_j+1}, \cdots, d_{nj}\}$ is the diagonal eigenvalue matrix, $\mathbf{1}_{m_j}$ is $1 \times m_j$ vector of ones and $diag\{a_n\}$ is a diagonal matrix formed by elements of a row vector a_n. This implies that $(I_n - \lambda_{jn} W_{jn}) = R_{jn}(I_n - \lambda_{jn} D_{jn}) R_{jn}^{-1}$. As $|\lambda_{jn}| < 1$ and $|d_{ij}| \leq 1$ for all i, $(I_n - \lambda_{jn} D_{jn})$ is invertible and, so is $(I_n - \lambda_{jn} W_{jn})$. Thus, with our near unit root specification, for any finite n, this model is still a well-defined equilibrium model.

For the eigenvalue matrix D_{jn}, it can be decomposed into two parts as $D_{jn} = J_{jn} + \tilde{D}_{jn}$ where $J_{jn} = diag\{\mathbf{1}_{m_j}, 0, \cdots, 0\}$ and $\tilde{D}_{jn} = diag\{0, \cdots, 0, d_{m_j+1,j}, \cdots, d_{nj}\}$. The J_{jn} consists of all the m_j unit eigenvalues and $\tilde{D}_{jn}$ consists of all the eigenvalues with their absolute values less than one. Accordingly, W_{jn} can also be decomposed into two parts: $W_{jn} = W_{jn}^u + \tilde{W}_{jn}$, where $W_{jn}^u = R_{jn} J_{jn} R_{jn}^{-1}$ and $\tilde{W}_n = R_{jn} \tilde{D}_{jn} R_{jn}^{-1}$. Denote $S_{jn} = I_n - \lambda_{jn} W_{jn}$, from (2.1), the equilibrium vector Y_{jn} is $Y_{jn} = S_{jn}^{-1}(Z_{jn}\gamma_j + \epsilon_{jn})$. For $S_{jn} = I_n - \lambda_{jn} W_{jn} = R_{jn}(I_n - \lambda_{jn} D_{jn}) R_{jn}^{-1}$, we have $S_{jn}^{-1} = R_{jn}(I_n - \lambda_{jn} D_{jn})^{-1} R_{jn}^{-1}$. Because $(I_n - \lambda_{jn} D_{jn})$ is a diagonal matrix, $(I_n - \lambda_{jn} D_{jn})^{-1} = \psi_{jn} \lambda_{jn} J_{jn} + (I_n - \lambda_{jn} \tilde{D}_{jn})^{-1}$ as is derived in Lee and Yu (2007). Hence, denoting $G_{jn} = W_{jn} S_{jn}^{-1}$, we have

$$\begin{aligned} S_{jn}^{-1} &= \psi_{jn} \lambda_{jn} W_{jn}^u + (I_n - \lambda_{jn} \tilde{W}_{jn})^{-1}, \\ G_{jn} &= \psi_{jn} \lambda_{jn} W_{jn}^u + W_{jn}(I_n - \lambda_{jn} \tilde{W}_{jn})^{-1}, \end{aligned} \tag{2.3}$$

because $W_{jn} W_{jn}^u = R_{jn} D_{jn} J_{jn} R_{jn}^{-1} = W_{jn}^u$. Therefore, from (2.3), we have the following representation for any value of λ_{jn} in terms of ψ_{jn},

$$Y_{jn} = \psi_{jn} Y_{jn}^u + \tilde{Y}_{jn}, \tag{2.4}$$

where

$$Y_{jn}^u = \lambda_{jn} W_{jn}^u (Z_{jn}\gamma_j + \epsilon_{jn}),$$

and

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

Equation (2.3) is revealing in that, when λ_{jn} is near unity, S_{jn} is ill conditioned and its inverse has the large factor ψ_{jn}, which implies that the first term of Y_{jn} on the right hand side of (2.4) may present an unstable component with increasing variances as ψ_{jn} approaches infinity. Regardless whether λ_{jn} is close to one or not, the second term of Y_{jn} in (2.4) is a stable one. If λ_{jn} is close to one, the implied variance of Y_{jn} can be large because the first component has the factor ψ_{jn}, which can be explosive when λ_{jn} is near the unit root. Also, Y_{jn} and $W_{jn}Y_{jn}$ have the same unstable component because $W_{jn}Y_{jn}^u = Y_{jn}^u$, and, consequently, $(I_n - W_{jn})Y_{jn}$ becomes stable[4]. Thus, Y_{jn} and $W_{jn}Y_{jn}$ feature spatial nonstationarity in the cross sectional setting when λ_{jn} is near unity.

To analyze the model, we make the following assumptions.

Assumption 1 *W_{jn} is diagonalizable, row sum and column sum bounded[5] (for short, UB), and has m_j unit roots, and the remaining eigenvalues d_{ij}, $i = m_j + 1, \cdots, n$, are uniformly bounded away from 1 in the absolute value for all n and i.*

Assumption 2 *$\lambda_{jn} = 1 - \frac{1}{\psi_{jn}}$ and $|\lambda_{jn}| < 1$. ψ_{jn} is positive, which might remain finite or approach infinity as n goes to infinity.*

Assumption 3 *Elements of $n \times 1$ disturbances $\{\epsilon_{jn}\}$ of the jth equation in (2.1) are i.i.d with zero mean, variance σ_j^2 and its higher than fourth moment exists. Also, ϵ_{in} and ϵ_{jn} are independent for $i \neq j$.*

Assumption 4 *$(I_n - \lambda_{jn}\tilde{W}_{jn})^{-1}$ and W^u_{jn} are UB.*

Assumption 1 says that W_{jn} is diagonalizable so that we can decompose the process into a stable part and a possible unstable part. As an example, a weights matrix row-normalized from a symmetric matrix is diagonalizable; and all its eigenvalues are real, less than or equal to one in the absolute value and its largest eigenvalue is always 1 (see Ord (1975)). Assumption 2 specifies that the true spatial effect can be near unity and how close it is near unity when ψ_{jn} approaches infinity. It also allows the stable case where ψ_{jn} remains finite[6]. Assumption 3 is a standard assumption. In this paper, it is to control the stochastic boundedness of ϵ_{jn}, so that the (possible) large variance nonstationary behavior of Y_{jn} comes only through the near unit λ_{jn}. Assumption 4 is to guarantee that the stable and (possible) unstable parts of S^{-1}_{jn} in (2.3) are UB after being rescaled. In the standard case with the true λ_0 being strictly less than one, an important assumption due to Kelejian and Prucha (1998) is that $S^{-1}_n(\lambda_0)$ is UB. For the (possible) near unit root case, $S^{-1}_{jn}(\lambda_{jn})$ will not be uniformly bounded in either row or column sum norms due to the explosive factor ψ_{jn} for the unstable component in (2.3). By taking out the ψ_{jn} factor, the magnitude of the remaining matrices will not grow with n. For our subsequent analysis, we shall focus on the setting that the unstable component Y^u_{jn} does not vanish asymptotically. This would be the case if W^u_{jn} were not dominated asymptotically by $\tilde{W}_{jn}$ in the sense that the share of unit eigenvalues, m_j/n, would not vanish as n approaches infinity for the near unit root process.

3 Spurious Regression

To study the possible spurious regression, we will focus on the case where there are no exogenous variables included in (2.1) as in Fingleton (1999). Hence, the DGP of the variables of interest are

$$\begin{aligned} Y_{1n} &= \lambda_{1n} W_{1n} Y_{1n} + \epsilon_{1n}, \qquad (3.1) \\ &\vdots \\ Y_{mn} &= \lambda_{mn} W_{mn} Y_{mn} + \epsilon_{mn}, \end{aligned}$$

where ϵ_{jn} is *i.i.d.*, and independent of ϵ_{in} for $i \neq j$. Denote $\mathbf{Y}_{-1,n} = (Y_{2n}, \cdots, Y_{mn})$, $X_n = [l_n, \mathbf{Y}_{-1,n}]$ where l_n is the $n \times 1$ column vector of ones, α is a scalar and $\beta = (\beta_2, \cdots, \beta_m)'$ is an $(m-1) \times 1$ vector. We will investigate the OLS estimate of $\delta \equiv (\alpha, \beta')'$ in the regression of Y_{1n} on l_n and $\mathbf{Y}_{-1,n}$ as if

$$Y_{1n} = \alpha l_n + \mathbf{Y}_{-1,n}\beta + V_n = X_n\delta + V_n, \tag{3.2}$$

where V_n is the residual vector.

The following Lemma 3.1 is useful for the asymptotic properties of the OLS estimates of α and β and their related statistics. Denote $Y_{jn}^* = \frac{1}{\psi_{jn}} Y_{jn}$, so that $\mathbf{Y}_{-1,n}^* = (Y_{2n}^*, \cdots, Y_{mn}^*)$. Also, denote $\mathbb{S}_{jn} = \frac{1}{\psi_{jn}^2} S_{jn}^{-1} S_{jn}'^{-1}$ for $j = 1, ..., m$. Hence, as $S_{jn}^{-1} = \psi_{jn}\lambda_{jn}W_{jn}^u + (I_n - \lambda_{jn}\tilde{W}_{jn})^{-1}$ from (2.3) where W_{jn}^u and $(I_n - \lambda_{jn}\tilde{W}_{jn})^{-1}$ are UB by Assumption 4, $\mathbb{S}_{jn}$ is UB regardless of whether ψ_{jn} is finite or large.

Lemma 3.1 *Let $X_n^* = [l_n, \mathbf{Y}_{-1,n}^*]$. Under Assumptions 1, 2, 3 and 4, for $i, j = 1, ..., m$,*

$$\frac{1}{n} Y_{in}^{*\prime} Y_{jn}^* = \begin{cases} \sigma_j^2 \frac{1}{n} tr(\mathbb{S}_{jn}) + O_p(\frac{1}{\sqrt{n}}) & \text{if } i = j \\ O_p(\frac{1}{\sqrt{n}}) & \text{if } i \neq j \end{cases}, \tag{3.3}$$

and

$$\frac{1}{\sqrt{n}} X_n^{*\prime} Y_{1n}^* = \frac{1}{\sqrt{n}} \begin{pmatrix} l_n' Y_{1n}^* \\ Y_{2n}^{*\prime} Y_{1n}^* \\ \vdots \\ Y_{mn}^{*\prime} Y_{1n}^* \end{pmatrix} \xrightarrow{d} N(0, \Sigma_m^*), \tag{3.4}$$

where $\Sigma_{m,ij}^* = \begin{cases} \sigma_1^2 \lim_{n\to\infty} \frac{1}{n} l_n' \mathbb{S}_{1n} l_n & \text{if } i = j = 1 \\ \sigma_j^2 \sigma_1^2 \lim_{n\to\infty} \frac{1}{n} tr(\mathbb{S}_{jn}\mathbb{S}_{1n}) & \text{if } i = j > 1 \\ 0 & \text{otherwise} \end{cases}$, *under the assumption that those limits were well defined.*

Proof. *See Appendix B.* ■

We note that the variance matrix Σ_m^* is a diagonal matrix because the different SAR processes in (3.2) are mutually independent.

3.1 Estimates

The OLS estimates for α and β in (3.2) are $\begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix} = (X_n' X_n)^{-1} (X_n Y_{1n})$. For the DGPs in (3.1), because the variables of the different processes are independent, one would expect that the linear regression of Y_{1n} on $X_n = [l_n, Y_{2n}, \cdots, Y_{mn}]$ should yield insignificant coefficients $\hat{\beta}_n$; otherwise, we might have some degree of spurious regressions. Denote

$$\Upsilon_m = \begin{pmatrix} 1 & 0 \\ 0 & \Upsilon_{2m} \end{pmatrix} \text{ where } \Upsilon_{2m} = \begin{pmatrix} \psi_{2n} & 0 & \cdots & 0 \\ 0 & \psi_{3n} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_{mn} \end{pmatrix},$$

we have

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix} = \psi_{1n} \Upsilon_m^{-1} \left(\frac{1}{n} X_n^{*\prime} X_n^* \right)^{-1} \left(\frac{1}{\sqrt{n}} X_n^{*\prime} Y_{1n}^* \right). \tag{3.5}$$

From Lemma 3.1, we have $\frac{1}{n} X_n^{*\prime} X_n^* = D_{n,xx}^* + O_p(\frac{1}{\sqrt{n}})$ where

$$D_{n,xx}^* \equiv diag\{1, \sigma_2^2 \frac{1}{n} tr(\mathbb{S}_{2n}), ..., \sigma_m^2 \frac{1}{n} tr(\mathbb{S}_{mn})\},$$

and $\frac{1}{\sqrt{n}} X_n^{*\prime} Y_{1n}^*$ will be asymptotically normally distributed with the limiting variance matrix

$$\Sigma_m^* = \sigma_1^2 \cdot diag\{\lim_{n\to\infty} \frac{1}{n} l_n' \mathbb{S}_{1n} l_n, \sigma_2^2 \lim_{n\to\infty} \frac{1}{n} tr(\mathbb{S}_{2n}\mathbb{S}_{1n}), ..., \sigma_m^2 \lim_{n\to\infty} \frac{1}{n} tr(\mathbb{S}_{mn}\mathbb{S}_{1n})\}.$$

Assumption 5 $\lim_{n\to\infty} \frac{1}{n} tr(\mathbb{S}_{jn}) \neq 0$ *for* $j = 1, 2, ..., m$.

Denote $D_{xx}^* \equiv \lim_{n\to\infty} D_{n,xx}^* = diag\{1, \sigma_2^2 \lim_{n\to\infty} \frac{1}{n} tr(\mathbb{S}_{2n}), ..., \sigma_m^2 \lim_{n\to\infty} \frac{1}{n} tr(\mathbb{S}_{mn})\}$. As $\frac{1}{n} X_n^{*\prime} X_n^* \xrightarrow{p} D_{xx}^*$, Assumption 5 specifies that the rescaled $\frac{1}{n} X_n^{*\prime} X_n^*$ is of full rank in the limit. Hence,

$$\frac{\sqrt{n}}{\psi_{1n}} \Upsilon_m (\hat{\alpha}_n, \hat{\beta}_n')' \xrightarrow{d} N(0, [D_{xx}^*]^{-1} \Sigma_m^* [D_{xx}^*]^{-1}). \tag{3.6}$$

Therefore, whether $\hat{\beta}_{jn}$, where $j = 2, ...m$, converges in probability to 0 or not will depend on the factors in $\frac{\sqrt{n}}{\psi_{1n}} \Upsilon_m$. The situations can be divided into three cases:

Case (1): when $\frac{\psi_{1n}}{\psi_{jn}\sqrt{n}} \to 0$, $\hat{\beta}_{jn}$ is $\frac{\psi_{jn}\sqrt{n}}{\psi_{1n}}$ consistent and $\frac{\psi_{jn}\sqrt{n}}{\psi_{1n}} \hat{\beta}_{jn}$ will be asymptotically normally distributed.

Case (2): when $\frac{\psi_{1n}}{\psi_{jn}\sqrt{n}} \to c$ where c is a positive finite constant, $\hat{\beta}_{jn}$ is asymptotically normally distributed. Because its limiting distribution is not degenerate, the estimate $\hat{\beta}_{jn}$ does not converge to 0.

Case (3): when $\frac{\psi_{1n}}{\psi_{jn}\sqrt{n}} \to \infty$, $\hat{\beta}_{jn}$ will diverge to infinity, i.e., it is not stochastically bounded.

Thus, $\hat{\beta}_{jn}$ will not converge to 0 when $\lim_{n\to\infty} \frac{\psi_{1n}}{\psi_{jn}\sqrt{n}} \neq 0$. In the spurious regression of integrated time series, the least squares estimated coefficients have nondegenerate distributions (Phillips, 1986). For our spatial situation, this feature may also appear. We note that this spurious feature on $\hat{\beta}_{jn}$ $j = 2, ..., m$, appears only when λ_{1n} approaches one faster than λ_{jn} in the sense that $\frac{\psi_{1n}}{\psi_{jn}\sqrt{n}}$ converges to a positive constant or diverges to infinity. Intuitively, this means that λ_{1n} is much closer to one than λ_{jn} for $j \neq 1$. If both λ_{1n} and λ_{jn} approach one with a similar rate, case (1) will apply and $\hat{\beta}_{jn}$ will converge to 0 and $\sqrt{n}\hat{\beta}_{jn}$ is asymptotically normal with zero mean and a finite limiting variance; also, the spurious regression feature of nondegenerate limiting distribution for $\hat{\beta}_{jn}$ will not occur. For $\hat{\alpha}_n$, we can see from (3.6) that it is divergent as long as λ_{1n} approaches one.

For the estimate $\hat{\sigma}_n^2$ of the variance of V_n, denote e_n as the regression residual $e_n = (I_n - P_n)Y_{1n}$ where $P_n = X_n(X_n'X_n)^{-1}X_n'$. We see that $\frac{1}{n}e_n'e_n = \frac{1}{n}Y_{1n}'Y_{1n} - \frac{1}{n}Y_{1n}'P_nY_{1n}$. The following lemma is useful to get the order of $\hat{\sigma}_n^2$.

Lemma 3.2 *Under Assumptions 1, 2, 3 and 4, for any nonstochastic UB square matrix B_n,*

$$\frac{1}{n}Y_{1n}^{*\prime}B_nP_nY_{1n}^* = O_p(\frac{1}{n}). \tag{3.7}$$

Proof. *See Appendix B.* ■

From Lemma 3.2, we have $\frac{1}{\psi_{1n}^2}\frac{1}{n}Y_{1n}'P_nY_{1n} = \frac{1}{n}Y_{1n}^{*\prime}P_nY_{1n}^* = O_p(\frac{1}{n})$; also, from Lemma 3.1, we have $\frac{1}{n}Y_{1n}^{*\prime}Y_{1n}^* = \sigma_1^2\frac{1}{n}tr(\mathbb{S}_{1n}) + O_p(\frac{1}{\sqrt{n}})$. Hence,

$$\frac{1}{\psi_{1n}^2}\hat{\sigma}_n^2 = \frac{1}{\psi_{1n}^2}\left(\frac{e_n'e_n}{n-m}\right) = \sigma_1^2\frac{1}{n}tr(\mathbb{S}_{1n}) + O_p(\frac{1}{\sqrt{n}}). \tag{3.8}$$

Thus, $\hat{\sigma}_n^2$ will be divergent at the ψ_{1n}^2 rate when ψ_{1n} goes to infinity. The divergence of $\hat{\sigma}_n^2$ is expected simply due to the large variance of the regressant Y_{1n}.

3.2 t, F Statistics and R^2

The t-statistic for each β_j where $j = 2, ..., m$, is $t_{\beta_j} = \frac{\hat{\beta}_{jn}}{\hat{\sigma}_n\sqrt{[(X_n'X_n)^{-1}]_{jj}}} = \frac{\sqrt{n}\frac{\psi_{jn}}{\psi_{1n}}\hat{\beta}_{jn}}{\frac{\hat{\sigma}_n}{\psi_{1n}}\sqrt{\left[(\frac{1}{n}X_n^{*\prime}X_n^*)^{-1}\right]_{jj}}}$. As $\frac{1}{n}X_n^{*\prime}X_n^* = D_{n,xx}^* + O_p(\frac{1}{\sqrt{n}})$, we have $(\frac{1}{n}X_n^{*\prime}X_n^*)^{-1} = [D_{n,xx}^*]^{-1} + O_p(\frac{1}{\sqrt{n}})$ because $D_{n,xx}^*$ is nonsingular in the limit from Assumption 5. Hence, from (3.5) and $\frac{1}{\sqrt{n}}X_n^{*\prime}Y_{1n}^*$ is $O_p(1)$ from Lemma 3.1, we have

$$t_{\beta_j}=\frac{\{[D_{n,xx}^*]^{-1}\}_{jj}\frac{1}{\sqrt{n}}Y_{jn}^{*\prime}Y_{1n}^*+O_p(\frac{1}{\sqrt{n}})}{\frac{\hat{\sigma}_n}{\psi_{1n}}(\{[D_{n,xx}^*]^{-1/2}\}_{jj}+O_p(\frac{1}{\sqrt{n}}))}.$$

As $\frac{\hat{\sigma}_n}{\psi_{1n}}=\sigma_1\sqrt{\frac{1}{n}tr(\mathbb{S}_{1n})}+O_p(\frac{1}{\sqrt{n}})$ from (3.8), we have

$$t_{\beta_j}=\frac{\{[D_{n,xx}^*]^{-1/2}\}_{jj}}{\sigma_1\sqrt{\frac{1}{n}tr(\mathbb{S}_{1n})}}\left(\frac{1}{\sqrt{n}}Y_{jn}^*Y_{1n}^*\right)+O_p(\frac{1}{\sqrt{n}})\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{n\cdot tr(\mathbb{S}_{jn}\mathbb{S}_{1n})}{tr(\mathbb{S}_{jn})\cdot tr(\mathbb{S}_{1n})}\right). \quad (3.9)$$

Hence, we see that the asymptotic distribution of the individual t-statistics is normal, but its variance is, in general, different from 1. For the special case that $W_{jn}=W_{1n}$ is a symmetric matrix and both λ_{jn} and λ_{1n} converge to 1, i.e., both are near unit roots, then one can easily show, using the property $\frac{1}{\psi_{jn}}S_{jn}$ is approximated by W_{jn}^u and $R_{jn}^{-1}=R_{jn}'$, that the limiting variance $\lim_{n\to\infty}\frac{n\cdot tr(\mathbb{S}_{jn}\mathbb{S}_{1n})}{tr(\mathbb{S}_{jn})\cdot tr(\mathbb{S}_{1n})}=\lim_{n\to\infty}\frac{n}{m_1}$. In this case, the variance is greater than one as the number of unit eigenvalues m_1 is less than n.[7]

Denote A_{22} as obtained from A with the first row and column deleted for any square matrix A. For the F test of the overall significance of regressors, i.e., $H_0:\beta_2=...=\beta_k=0$, the test statistic is $F=\frac{1}{m-1}\hat{\beta}_n'\{\hat{\sigma}_n^2[(X_n'X_n)^{-1}]_{22}\}^{-1}\hat{\beta}_n$. By using the rescaled variables,

$$F=\frac{1}{(m-1)\frac{\hat{\sigma}_n^2}{\psi_{1n}^2}}\left(\frac{\sqrt{n}}{\psi_{1n}}\Upsilon_{2m}\hat{\beta}_n\right)'\cdot[(\frac{1}{n}X_n^{*\prime}X_n^*)^{-1}]_{22}\cdot\left(\frac{\sqrt{n}}{\psi_{1n}}\Upsilon_{2m}\hat{\beta}_n\right).$$

We have $[(\frac{1}{n}X_n^{*\prime}X_n^*)^{-1}]_{22}=\{(\frac{1}{n}X_n^{*\prime}X_n^*)_{22}-(\frac{1}{n}X_n^{*\prime}X_n^*)_{21}\cdot[(\frac{1}{n}X_n^{*\prime}X_n^*)_{11}]^{-1}\cdot(\frac{1}{n}X_n^{*\prime}X_n^*)_{12}\}^{-1}$ from the inverse of a partitioned matrix, where $(\frac{1}{n}X_n^{*\prime}X_n^*)_{pq}$ is the corresponding partitioned matrix for $p=1,2$ and $q=1,2$. As $(\frac{1}{n}X_n^{*\prime}X_n^*)_{11}=1$, $(\frac{1}{n}X_n^{*\prime}X_n^*)_{12}=[(\frac{1}{n}X_n^{*\prime}X_n^*)_{21}]'=O_p(\frac{1}{\sqrt{n}})$ and $(\frac{1}{n}X_n^{*\prime}X_n^*)_{22}=[D_{n,xx}^*]_{22}+O_p(\frac{1}{\sqrt{n}})$ from Lemma 3.1, $[(\frac{1}{n}X_n^{*\prime}X_n^*)^{-1}]_{22}=\{[D_{n,xx}^*]_{22}\}^{-1}+O_p(\frac{1}{\sqrt{n}})$. Similarly, from (3.5), we have $\frac{\sqrt{n}}{\psi_{1n}}\Upsilon_{2m}\hat{\beta}_n=\{[D_{n,xx}^*]_{22}\}^{-1}\left(\frac{1}{\sqrt{n}}\mathbf{Y}_{-1,n}'^{*}Y_{1n}^*\right)+O_p(\frac{1}{\sqrt{n}})$. Hence,

$$F=\frac{1}{(m-1)}\frac{1}{\sigma_1^2\frac{1}{n}tr(\mathbb{S}_{1n})}\left(\frac{1}{\sqrt{n}}\mathbf{Y}_{-1,n}'^{*}Y_{1n}^*\right)\{[D_{n,xx}^*]_{22}\}^{-1}\left(\frac{1}{\sqrt{n}}\mathbf{Y}_{-1,n}'^{*}Y_{1n}^*\right)+O_p(\frac{1}{\sqrt{n}}).$$

Hence, from Lemma 3.1,

$$F\overset{d}{=}\frac{1}{(m-1)}U_m'U_m,$$

where U_m is an $(m-1)\times 1$ vector and its element $(u_{2m}, ..., u_{mm})$ are independent normal where for $j = 2, ..., m$,

$$u_{jm} \sim N\left(0, \lim_{n\to\infty} \frac{n \cdot tr(\mathbb{S}_{jn}\mathbb{S}_{1n})}{tr(\mathbb{S}_{jn}) \cdot tr(\mathbb{S}_{1n})}\right).$$

This implies that, the F-statistic is asymptotically the average of $m-1$ independent square of normal random variables, which might have different variance[8]. Hence, the F-statistic multiplied by $(m-1)$ is not asymptotically chi-square distributed with $(m-1)$ degrees of freedom in general. Additionally, we can see that

$$F \overset{d}{=} \frac{1}{(m-1)} \sum_{j=2}^{m} t_{\beta_j}^2. \tag{3.10}$$

For the coefficient of determination, we have $R^2 = 1 - \frac{e_n' e_n}{Y_{1n}' M_n^0 Y_{1n}}$ where $M_n^0 = I_n - \frac{1}{n} l_n l_n'$. Hence, $R^2 = 1 - \frac{e_n' e_n/(n\psi_{1n}^2)}{Y_{1n}' M_n^0 Y_{1n}/(n\psi_{1n}^2)} = \frac{\frac{1}{n} Y_{1n}^{*\prime} P_n Y_{1n}^* - (\frac{1}{n} l_n' Y_{1n}^*)^2}{\frac{1}{n} Y_{1n}^{*\prime} Y_{1n}^* - (\frac{1}{n} l_n' Y_{1n}^*)^2}$. As $\frac{1}{n} Y_{1n}^{*\prime} P_n Y_{1n}^* = O_p(\frac{1}{n})$ from Lemma 3.2, $\frac{1}{n} l_n' Y_{1n}^* = O_p(\frac{1}{\sqrt{n}})$ from Lemma 3.1, and $\frac{1}{n} Y_{1n}^{*\prime} Y_{1n}^* = \sigma_1^2 \frac{1}{n} tr(\mathbb{S}_{1n}) + O_p(\frac{1}{\sqrt{n}})$ with $\lim_{n\to\infty} \frac{1}{n} tr(\mathbb{S}_{1n}) \neq 0$ from Assumption 5, we have

$$R^2 = \frac{\frac{1}{n} Y_{1n}^{*\prime} P_n Y_{1n}^* - (\frac{1}{n} l_n' Y_{1n}^*)^2}{\frac{1}{n} Y_{1n}^{*\prime} Y_{1n}^* - (\frac{1}{n} l_n' Y_{1n}^*)^2} = O_p(\frac{1}{n}) \xrightarrow{p} 0. \tag{3.11}$$

These imply that the t-statistic and F-statistic may not be reliable to test the null of $\beta_j = 0$ for $j = 2, ..., m$ because their asymptotic distributions are not the usual ones for the conventional t and F statistics. However, from (3.11), we can see that R^2 is a good indicator of the insignificance of β_j.

3.3 Moran I and LM Tests of Spatial Error in OLS

In addition to the testing of β, it is of interest to test the spatial effect in the disturbances of the OLS regression. The Moran I test statistic is $I_{Moran} = \frac{n}{S_0} \frac{e_n' W_n e_n}{e_n' e_n}$ with $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{n,ij}$ and $e_n = (I_n - P_n) Y_{1n}$, where W_n is a nonstochastic UB spatial weights matrix. When W_n is row-normalized, $S_0 = n$ so that $I_{Moran} = \frac{e_n' W_n e_n}{e_n' e_n}$. Hence,

$$I_{Moran} = \frac{\frac{1}{n} e_n' W_n e_n}{\frac{1}{n} e_n' e_n} = \frac{\frac{1}{n} Y_{1n}' W_n Y_{1n} - \frac{1}{n} Y_{1n}' W_n P_n Y_{1n} - \frac{1}{n} Y_{1n}' P_n W_n Y_{1n} + \frac{1}{n} Y_{1n}' P_n W_n P_n Y_{1n}}{\frac{1}{n} Y_{1n}' Y_{1n} - \frac{1}{n} Y_{1n}' P_n Y_{1n}}.$$

Denote $S_{jn}^* = \psi_{jn} S_{jn}$. Hence, $S_{jn}^{*-1} = \psi_{jn}^{-1} S_{jn}^{-1}$ is UB by (2.3) and Assumption 4. For the numerator of I_{Moran}, from Lemma 3.1, we have $\frac{1}{\psi_{1n}^2} \frac{1}{n} Y_{1n}' W_n Y_{1n} = \sigma_1^2 \frac{1}{n} tr(S_{1n}^{*\prime -1} W_n S_{1n}^{*-1}) + O_p(\frac{1}{\sqrt{n}})$; also, from Lemma

3.2, the remaining three terms are $O_p(\frac{1}{n})$ after being rescaled by $\frac{1}{\psi_{1n}^2}$. Similarly for the denominator, we have $\frac{1}{\psi_{1n}^2}\frac{1}{n}Y_{1n}'Y_{1n} = \sigma_1^2\frac{1}{n}tr(S_{1n}^{*\prime-1}S_{1n}^{*-1}) + O_p(\frac{1}{\sqrt{n}})$ and $\frac{1}{n}Y_{1n}'P_nY_{1n} = O_p(\frac{1}{n})$. Hence,

$$I_{Moran} = \frac{tr(S_{1n}^{*\prime-1}W_nS_{1n}^{*-1})}{tr(S_{1n}^{*\prime-1}S_{1n}^{*-1})} + O_p(\frac{1}{\sqrt{n}}). \tag{3.12}$$

The case $W_n = W_{1n}$ is of special interest. Because $W_{jn}^uW_{jn}^u = W_{jn}^u$ and $\tilde{W}_{jn}W_{jn}^u = 0$, they imply that $tr(W_{jn}'^uW_{jn}W_{jn}^u) = tr(W_{jn}'^uW_{jn}^u)$. Hence, when $\psi_{1n} \to \infty$, the dominant term in S_{1n}^{*-1} is $\lambda_{1n}W_{jn}^u$ so that $I_{Moran} \xrightarrow{p} 1$.

Instead of the Moran I test statistic, the associated LM test is

$$LM = \frac{e_n'W_ne_n}{\hat{\sigma}_n^2 \cdot tr^{1/2}(W_n^2 + W_n'W_n)} = \frac{\sqrt{n}}{[\frac{1}{n}tr(W_n^2 + W_n'W_n)]^{1/2}} \cdot \frac{n}{n-m}I_{Moran}, \tag{3.13}$$

which diverges to infinity at the $\sqrt{n}$ rate as $n \to \infty$, regardless of whether ψ_{jn} is large or not as long as I_{Moran} does not converge to 0.

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

When there are constant terms included in the DGP (3.1) so that Y_{jn}'s have nonzero means, we show that they do not change the estimates of β_{jn}. This is so even the constant terms may take on large values (or divergent) as in Mur and Trivez (2003). For notational purposes, we use Y_{jn}^c as the counterpart of Y_{jn} so that

$$\begin{aligned} Y_{1n}^c &= \lambda_{1n}W_{1n}Y_{1n}^c + c_{1n}l_n + \epsilon_{1n}, \\ &\vdots \\ Y_{mn}^c &= \lambda_{mn}W_{mn}Y_{mn}^c + c_{mn}l_n + \epsilon_{mn}, \end{aligned} \tag{3.14}$$

where c_{jn} is a sequence of nonstochastic scalars, for $j = 1, ..., m$. Denote $\mathbf{Y}_{-1,n}^c = (Y_{2n}^c, \cdots, Y_{mn}^c)$, α^c as a scalar and $\beta^c = (\beta_2^c, \cdots, \beta_m^c)'$ an $(m-1) \times 1$ vector. Consider the OLS regression of Y_{1n}^c on l_n and $\mathbf{Y}_{-1,n}^c$ as if

$$Y_{1n}^c = \alpha^c l_n + \mathbf{Y}_{-1,n}^c\beta^c + V_n^c, \tag{3.15}$$

where V_n^c is the residual vector. Rather than developing a lemma similar to Lemma 3.1 where c_{jn} could possibly diverge, we can compare (3.15) with (3.2). When W_{jn} is row-normalized, we have $Y_{jn}^c = S_{jn}^{-1}(c_{jn}l_n + \epsilon_{jn}) = \psi_{jn}c_{jn}l_n + S_{jn}^{-1}\epsilon_{jn} = \psi_{jn}c_{jn}l_n + Y_{jn}$. Hence, (3.15) can be re-written as

$$Y_{1n} = (\alpha^c - \psi_{1n}c_{1n} + \textstyle\sum_{j=2}^m \psi_{jn}c_{jn}\beta_{jn}^c)l_n + \mathbf{Y}_{-1,n}\beta^c + V_n^c.$$

Compared with (3.2), we can see that $\alpha^c - \psi_{1n}c_{1n} + \sum_{j=2}^m \psi_{jn}c_{jn}\beta_{jn}$ is reparameterized as α and $\beta^c = \beta$, where V_n^c is the same as V_n after this reparameterization. This implies that we will have the same estimates of β as in (3.2) even though a constant term (possibly large) is included in the DGP of Y_{jn}. Consequently, the t-statistic and F-statistic involving estimated β^c would be the same as those for the estimate β in (3.2).

Note that $\alpha^c = \alpha + \psi_{1n}c_{1n} - \sum_{j=2}^m \psi_{jn}c_{jn}\beta_{jn}$. As $\frac{\psi_{jn}\sqrt{n}}{\psi_{1n}}\hat{\beta}_{jn} = O_p(1)$ and $\frac{\sqrt{n}}{\psi_{1n}}\hat{\alpha}_n = O_p(1)$, we have $\frac{\sqrt{n}}{\psi_{1n}}\hat{\alpha}_n^c = O_p(\max(1, \sqrt{n}c_{1n}, c_{2n}, ..., c_{mn}))$. However, inference about the constant term in the OLS regression might not be of much interest.

3.5 Intuitions

It might be helpful to see the intuition behind the results we have derived. For illustration, consider the simple OLS regression of Y_{1n}^* on l_n and Y_{2n}^*, via Y_{1n} on l_n and Y_{2n} (so that $m = 2$) as done in the previous sections. These two problems are related because

$$\begin{aligned}
& \min_{\alpha,\beta}(Y_{1n} - \alpha l_n - Y_{2n}\beta)'(Y_{1n} - \alpha l_n - Y_{2n}\beta) \\
= \quad & \psi_{1n}^2 \min_{\alpha,\beta}(\frac{Y_{1n}}{\psi_{1n}} - \frac{\alpha}{\psi_{1n}}l_n - \frac{Y_{2n}}{\psi_{2n}}\frac{\psi_{2n}\beta}{\psi_{1n}})'(\frac{Y_{1n}}{\psi_{1n}} - \frac{\alpha}{\psi_{1n}}l_n - \frac{Y_{2n}}{\psi_{2n}}\frac{\psi_{2n}\beta}{\psi_{1n}}) \\
= \quad & \psi_{1n}^2 \min_{\alpha^*,\beta^*}(Y_{1n}^* - \alpha^* l_n - Y_{2n}^*\beta^*)'(Y_{1n}^* - \alpha^* l_n - Y_{2n}^*\beta^*),
\end{aligned}$$

where $\alpha^* = \frac{\alpha}{\psi_{1n}}$ and $\beta^* = \frac{\psi_{2n}\beta}{\psi_{1n}}$. Hence, $\hat{\alpha}_n = \psi_{1n}\hat{\alpha}_n^*$ and $\hat{\beta}_n = \frac{\psi_{1n}}{\psi_{2n}}\hat{\beta}_n^*$. Similarly,

$$\begin{aligned}
\hat{\sigma}_n^2 \quad &= \quad \frac{1}{n-2}(Y_{1n} - \hat{\alpha}_n l_n - Y_{2n}\hat{\beta}_n)'(Y_{1n} - \hat{\alpha}_n l_n - Y_{2n}\hat{\beta}_n) \\
&= \quad \psi_{1n}^2 \frac{1}{n-2}(Y_{1n}^* - \hat{\alpha}_n^* l_n - Y_{2n}^*\hat{\beta}_n)'(Y_{1n}^* - \hat{\alpha}_n^* l_n - Y_{2n}^*\hat{\beta}_n^*) = \psi_{1n}^2\hat{\sigma}_n^{*2},
\end{aligned}$$

where $\hat{\sigma}_n^{*2} = \frac{1}{n-2}(Y_{1n}^* - \hat{\alpha}_n^* l_n - Y_{2n}^* \hat{\beta}_n)'(Y_{1n}^* - \hat{\alpha}_n^* l_n - Y_{2n}^* \hat{\beta}_n)$. Hence, as the regression of Y_{1n}^* on l_n and Y_{2n}^* is a standard OLS regression with the regular $O(1)$ order, we can get the implied orders of $\hat{\alpha}_n$, $\hat{\beta}_n$ and $\hat{\sigma}_n^2$.

Additionally, it follows that

$$\begin{aligned} R^2 &= 1 - (n-2)\hat{\sigma}_n^2/(Y_{1n} - \frac{l_n' Y_{1n}}{n} l_n)'(Y_{1n} - \frac{l_n' Y_{1n}}{n} l_n) \\ &= 1 - \psi_{1n}^2 (n-2)\hat{\sigma}_n^{*2}/[\psi_{1n}^2 (Y_{1n}^* - \frac{l_n' Y_{1n}^*}{n} l_n)'(Y_{1n}^* - \frac{l_n' Y_{1n}^*}{n} l_n)] \\ &= 1 - (n-2)\hat{\sigma}_n^{*2}/[(Y_{1n}^* - \frac{l_n' Y_{1n}^*}{n} l_n)'(Y_{1n}^* - \frac{l_n' Y_{1n}^*}{n} l_n)] = R^{*2}, \end{aligned}$$

and

$$\begin{aligned} t_\beta &= [(Y_{2n} - \frac{l_n' Y_{2n}}{n} l_n)'(Y_{2n} - \frac{l_n' Y_{2n}}{n} l_n)]^{1/2} \frac{\hat{\beta}_n}{\hat{\sigma}_n} \\ &= [(\frac{Y_{2n}}{\psi_{2n}} - \frac{l_n'}{n}\frac{Y_{2n}}{\psi_{2n}} l_n)'(\frac{Y_{2n}}{\psi_{2n}} - \frac{l_n'}{n}\frac{Y_{2n}}{\psi_{2n}} l_n)]^{1/2} \frac{\frac{\psi_{2n}}{\psi_{1n}}\hat{\beta}_n}{\frac{\hat{\sigma}_n}{\psi_{1n}}} \\ &= [(Y_{2n}^* - \frac{l_n' Y_{2n}^*}{n} l_n)'(Y_{2n}^* - \frac{l_n' Y_{2n}^*}{n} l_n)]^{1/2} \frac{\hat{\beta}_n^*}{\hat{\sigma}_n^*} = t_{\beta^*}. \end{aligned}$$

We would expect R^{*2} to converge to 0 as Y_{1n}^* and Y_{2n}^* are independent. t_{β^*} would not diverge as Y_{1n}^* and Y_{2n}^* have the regular $O_p(1)$ order. This explains why the R^2 is close to 0 and t_β is $O_p(1)$. The asymptotic distribution of t_{β^*} would not necessarily have the conventional asymptotic $N(0,1)$ distribution because elements of Y_{1n}^* might not be independent with a homoskedastic variance. The proper asymptotic distributions of β^*, t_{β^*}, etc., are derived as in previous sections.

4 Monte Carlo

We run a small Monte Carlo to check possible spurious regression in the spatial setting. The DGPs are

$$\begin{aligned} Y_{1n} &= \lambda_{1n} W_n Y_{1n} + \epsilon_{1n}, \\ Y_{2n} &= \lambda_{2n} W_n Y_{2n} + \epsilon_{2n}, \end{aligned}$$

where ϵ_{1n} and ϵ_{2n} are independent $N(0, I_n)$, and we have the same weights matrix W_n. Both λ_{1n} and λ_{2n} are near unity, but might have different values. We choose 0.95 and 0.999 as possible values so that $\psi_{jn} = 20$ or

1000. For W_n, we first generate a (row-normalized) 4×4 queen matrix and then construct a block diagonal matrix. Hence, we have $\frac{m_j}{n} = 1/4$ so that we have a significant portion of the unstable component. We use 125 blocks so that we have $n = 500$. The regression equation is specified as

$$Y_{1n} = \alpha_n l_n + \beta Y_{2n} + V_n, \tag{4.1}$$

where the repetition is 1000. After we run the regression of (4.1), we calculate the t-statistics for β. As Y_{1n} and Y_{2n} are independent, conventional cases for regression should have low frequency of rejecting the null of $\beta = 0$ so that most of the t-statistics should lie within the range $(-2, 2)$ with approximately 5% level of significance for a two-sided test. If we have a very high frequency of rejecting the null hypothesis as the t-statistics are large in the absolute value, we might have a spurious regression in the spatial setting.

Tables 1-3 are the empirical densities of relevant statistics when $n = 500$. Table 1 presents the empirical densities of β_n and the corresponding t-statistic. Table 2 presents the empirical densities of $\hat{\sigma}_n^2$, R^2, and Table 3 presents the empirical densities of I_{Moran} and LM test statistics. To have a more-detailed look at the empirical density of the t-statistics, we also report the frequencies of the t-statistics in Table 7. From Tables 1-3, we can see that

(1) For the estimates of β, they will diverge only when $\psi_{1n} = 1000$ and $\psi_{2n} = 20$, while the t-statistics do not diverge, with a fat tail compared to the standard normal distribution[9].

(2) For $\hat{\sigma}_n^2$, they diverge, and will have huge values when $\psi_{1n} = 1000$; for R^2, they are close to 0 for all cases.

(3) For Moran I test statistics, it is close to 1[10]; and the LM test statistics are of the order $\sqrt{n}$.

All the results (1)-(3) are consistent with the theoretical prediction[11], i.e., value of $\hat{\beta}_n$ is in (3.5) such that $\hat{\beta}_n = O_p(\frac{\psi_{1n}}{\psi_{2n}\sqrt{n}})$ with zero mean, the t-statistic is in (3.9) which are asymptotically normal but not necessarily $N(0,1)$, the value of $\hat{\sigma}_n^2$ is in (3.8) with $\hat{\sigma}_n^2 = O_p(\psi_{1n}^2)$, and R^2, Moran I and LM test statistics are in (3.11), (3.12) and (3.13). In our simulation result, the t-statistic has a fat tail compared to the $N(0,1)$ distribution. This is so, because in our simulation, $\frac{n \cdot tr(\mathbb{S}_{2n}\mathbb{S}_{1n})}{tr(\mathbb{S}_{2n}) \cdot tr(\mathbb{S}_{1n})}$ is about[12] 4, which implies that the t_β has approximately the standard deviation of 2. This is consistent with the empirical density in Table 1.

We run additional simulations with a smaller n to see the finite sample behavior of the estimates and the statistics of interest. The frequency of t-statistics for $n = 100$ is reported in Table 7[13].

Finally, we run the simulation where the number of unit roots in W_n is small. We choose the 49×49 weights matrix used in Anselin (1988) so that $\frac{m_j}{n} = 1/49$. Tables 4-6 are the counterparts of Tables 1-3 where we use 10 blocks of the Anselin (1988)'s matrix so that $n = 490$. With Anselin (1988)'s weights matrix, we have $\frac{n\cdot tr(\mathbb{S}_{2n}\mathbb{S}_{1n})}{tr(\mathbb{S}_{2n})\cdot tr(\mathbb{S}_{1n})} = 48.9754$ when $\lambda_{1n} = \lambda_{2n} = 0.999$ (when $\lambda_{1n} = 0.95$ and $\lambda_{2n} = 0.999$, or vice versa, $\frac{n\cdot tr(\mathbb{S}_{2n}\mathbb{S}_{1n})}{tr(\mathbb{S}_{2n})\cdot tr(\mathbb{S}_{1n})} = 25.4925$). This explains that we have a larger variance for the t_β compared to the cases using the queen matrix.

Tables 1-6 here

Table 7 here

5 Conclusion

This paper investigates possible spurious regression phenomena in the spatial setting where the regressant and/or regressors are generated from possible nonstationary SAR processes. With a row-normalized spatial weights matrix (rather than one with an unconnected unit), it is shown that the possible spurious regression phenomena in the spatial setting are relatively weaker than those in the nonstationary time series scenario. The OLS estimates of the regression coefficients might or might not converge to 0. The divergence cases occur only when the near unit root of the regressant is much closer to 1 than those of the regressors. For the t and F statistics, there could be over-rejection of the null of uncorrelatedness under certain situations, but they do not diverge. The Moran I test may reject the null of no spatial dependence with the least squares residuals. However, the coefficient of determination R^2 converges to 0, which provides strong evidence on a spurious regression. Simulation results about different statistics are in line with the theoretical results we derive in this paper.

In the time series, Granger and Newbold (1974) state that low value of the Durbin-Watson statistic might imply the spurious regression. Phillips (1986) derives the asymptotic distribution of relevant statistics. It is shown that the Durbin-Watson statistic converges in probability to 0, while the regression R^2 has a non-degenerate limiting distribution as $T \to \infty$. Also, the t and F statistics will diverge. In the spatial setting for the near unit root case where the spatial weights matrix is row-normalized, we have the following observations: (a) The t statistic will not diverge. However, it is not asymptotically a standard normal statistic because the asymptotic variance of the statistic is different from 1. Hence, even though we might have some degree of spurious regression in terms of the t and F statistics, they are not as strong as the time series counterpart where the t and F statistics diverge at $T^{1/2}$ and T rate respectively; (b) The $R^2 \xrightarrow{p} 0$, which is different from the time series literature. This implies that low R^2 with high t value can provide a good indicator for the spurious regression in the spatial setting; (c) The Moran I test is $O_p(1)$ and will converge to 1 when $\psi_{1n} \to \infty$. The Moran I test statistic (which is the Durbin-Watson statistic in the time series) might be used as an indicator for the misspecification of a regression relation with homogenous disturbances in the spatial setting.

Our results are different from Fingleton (1999), which has a special specification of the spatial weights matrix. In Fingleton (1999) where the spatial effect can be 1, an unconnected central unit is introduced. In our specification, we use a row-normalized spatial weights matrix, which is used in empirical applications and has a long history for the SAR model (Ord 1975). With the nonstationarity of the spatial data coming from the near unit root specification, we find that the spurious regression is weaker than the one in Fingleton (1999).

In a spatial setting of a spurious regression, a significant Moran I or LM test statistic (to test the spatial error in the OLS regression) may be caused by the regressant being a spatial correlation process. Under the specification with an unconnected central unit in the spatial weights matrix as in Fingleton (1999), the high value of the Moran I or LM test statistic can be caused by the presence of a unit root in the DGP of variables in the regression. Similarly, in our specification of row-normalized weights matrix and nonstationarity, the high value of the Moran I or LM test statistic can also be caused by the near unit root in the DGP of the regressant. Based on the results we derived, we can combine the Moran I or LM test with the R^2 measure to have better indicators to detect possible spurious regression in the spatial setting where the spatial weights matrix is row-normalized. Lauridsen and Kosfeld (2006, 2007) develop a two step LM test to distinguish two possible sources of spurious regression: one is the unit root in the DGPs of the regressant, and the other is the spatial error process in the spurious regression. Their procedure is essentially testing whether the spatial process of the regressant is stable or nonstable. Hence, their procedure may also be applicable to our case of near unit root. However, the detailed analysis of that test procedure is beyond the scope of this paper and shall be left for future investigation.

Notes

[1]The estimation and testing for spatial dependence in cross sectional data can be found in Anselin (1988, 1992), Kelejian and Robinson (1993), Cressie (1993), Anselin and Florax (1995), Anselin and Rey (1997), Anselin and Bera (1998), Kelejian and Prucha (1998, 2001, 2007) and Lee (2003, 2004, 2007), among others.

[2]When W_{jn} is row normalized from a symmetric matrix, W_{jn} is diagonalizable. See Lemma A.1 in Yu et al. (2007). A weights matrix row normalized has real eigenvalues, with all its eigenvalues less than or equal to one in the absolute value and its largest eigenvalue always 1 (see Ord (1975)). Hence, W_{jn} being diagonalizable with the specified eigenvalues, is a slight generalization of W_{jn} being row-normalized from a symmetric matrix.

[3]In a time series near unit root model, the deviation from the unit root is measured through a noncentrality parameter, where the AR(1) coefficient is usually specified as $\exp(c/n)$ or $1 - \frac{c}{n}$ with c being the noncentrality parameter (see Phillips (1987)). For the near unit root in the SAR model, ψ_n can take a general form as long as it is increasing in n, which does not need to be specified in empirical applications.

[4]This property might be useful for the estimation. However, we do not explore the use of this spatial difference operator, $I_n - W_{jn}$, in this paper.

[5]We say a (sequence of $n \times n$) matrix P_n is uniformly bounded in row and column sum norms if $\sup_{n \geq 1} \|P_n\|_\infty < \infty$ and $\sup_{n \geq 1} \|P_n\|_1 < \infty$, where $\|P_n\|_\infty \equiv \sup_{1 \leq i \leq n} \sum_{j=1}^n |p_{ij,n}|$ is the row sum norm and $\|P_n\|_1 = \sup_{1 \leq j \leq n} \sum_{i=1}^n |p_{ij,n}|$ is the column sum norm.

[6]As our asymptotic analyses below can allow both the near unit root and stable cases, this generality provides a unified framework for our study.

[7]When both λ_{1n} and λ_{jn} are near unit root, one can easily evaluate the variance of t_{β_j} in (3.9) via W_{1n}^u and W_{jn}^u, and the t-statistic can be adjusted to be asymptotically $N(0,1)$ distributed. For a general case, one needs to estimate λ values in order to adjust such a variance.

[8]Equivalently, it is asymptotically a weighted sum of $(m-1)$ independent $\chi^2(1)$ random variables.

[9]For the estimates of α, the values are large, and they diverge when $\psi_{1n} = 1000$, while the t-statistics do not diverge, but with a fat tail compared to the standard normal distribution. As the inference of α is not of much interest, we do not report the relevant statistics.

[10]In an additional simulation where we set $\psi_{1n} = 2$ so that λ_{1n} is not close to 1, I_{Moran} is smaller with the mean value 0.4831 for $\psi_{2n} = 2$ and 0.4828 for $\psi_{2n} = 1000$. Hence, the Moran I statistic is close to 1 only when ψ_{1n} is large, which is implied by (3.12).

[11]Using the 4 by 4 queen matrix, we can increase the number of blocks from 11 to 125 to see the effect of the sample size. From the results (due to the space limit, the tables are not presented), the effect of the sample size on $\hat{\beta}_n$, t and Moran I statistics are not apparent. However, $\hat{\sigma}_n^2$ and the LM statistic are increasing in n and R^2 is decreasing in n.

[12]The values of $\frac{n \cdot tr(\mathbb{S}_{2n}\mathbb{S}_{1n})}{tr(\mathbb{S}_{2n}) \cdot tr(\mathbb{S}_{1n})}$ are 3.9656, 3.9656 and 4 respectively when $\lambda_{1n} = 0.95$ and $\lambda_{2n} = 0.999$, $\lambda_{1n} = 0.999$ and $\lambda_{2n} = 0.95$, and $\lambda_{1n} = \lambda_{2n} = 0.999$.

[13]The other results are similar. Due to the space limit, we do not report them in the tables.

Appendices

A Some Lemmas

Lemma A.1 *Denote* $B_n = [b_{n,gh}]_{g,h=1}^n$ *an* $n \times n$ *nonstochastic UB matrix. Under Assumption 3, for* $i, j = 1, ..., m$,

$$\frac{1}{n}\epsilon'_{in}B_n\epsilon_{jn} = \begin{cases} \sigma_j^2 \frac{1}{n}tr(B_n) + O_p(\frac{1}{\sqrt{n}}) & \text{if } i = j \\ O_p(\frac{1}{\sqrt{n}}) & \text{if } i \neq j \end{cases}.$$

Lemma A.2 *Denote* $B_{jn} = [b_{jn,gh}]_{g,h=1}^n$ *an* $n \times n$ *nonstochastic UB matrix for* $j = 2, ..., m$, *and denote* $C_{jn} = [c_{jn,g}]_{g=1}^n$ *an* $n \times 1$ *constant vector. Under Assumption 3,*

$$\Sigma_{Q,n}^{-1/2} \cdot \frac{1}{\sqrt{n}} \sum_{j=2}^m (\epsilon'_{jn}B_{jn} + C'_{jn})\epsilon_{1n} \xrightarrow{d} N(0,1),$$

where $\Sigma_{Q,n} = \frac{1}{n}\sigma_1^2 \sum_{j=2}^m \sigma_j^2 tr(B'_{jn}B_{jn}) + \frac{1}{n}\sigma_1^2 \sum_{j=2}^m C'_{jn}C_{jn}$.

Proof for Lemma A.1: Denote μ_{j4} as the fourth moment of elements of ϵ_{jn}. For $i = j$, as $E\frac{1}{n}\epsilon'_{jn}B_n\epsilon_{jn} = \sigma_j^2 \frac{1}{n}tr(B_n)$ and $var(\epsilon'_{jn}B_n\epsilon_{jn}) = (\mu_{j4} - 3\sigma_j^4)\sum_{g=1}^n b_{n,gg}^2 + \sigma_j^4(trB_n(B_n + B'_n)) = O(n)$, we have $\frac{1}{n}\epsilon'_{jn}B_n\epsilon_{jn} = \sigma_j^2 \frac{1}{n}tr(B_n) + O_p(\frac{1}{\sqrt{n}})$. For $i \neq j$, as $E\frac{1}{n}\epsilon'_{in}B_n\epsilon_{jn} = 0$ and $var(\epsilon'_{in}B_n\epsilon_{jn}) = \sigma_i^2\sigma_j^2 tr(B'_nB_n) = O(n)$, we have $\frac{1}{n}\epsilon'_{in}B_n\epsilon_{jn} = O_p(\frac{1}{\sqrt{n}})$.

Proof for Lemma A.2: Kelejian and Prucha (2007) extend the central limit theorem in Kelejian and Prucha (2001) to a system of quadratic forms. For $\sum_{j=2}^m \epsilon'_{jn}B_{jn}\epsilon_{1n}$, denote

$$\zeta_n = (\epsilon_{1n,1}, \epsilon_{2n,1}, ..., \epsilon_{mn,1}, ..., \epsilon_{1n,n}, \epsilon_{2n,n}, ..., \epsilon_{mn,n})' = (\zeta'_{n,1}, ..., \zeta'_{n,n})',$$

where $\zeta_{n,i} = (\epsilon_{1n,i}, \epsilon_{2n,i}, ..., \epsilon_{mn,i})'$. Thus, $\sum_{j=2}^m \epsilon'_{jn}B_{jn}\epsilon_{1n} = \zeta'_n\mathbf{A}_n\zeta_n$ where $\mathbf{A}_n$ is an $(n \times m) \times (n \times m)$ matrix such that $\mathbf{A}_n = [a_{n,gh}]_{g,h=1}^n$ where $a_{n,gh} = \begin{pmatrix} 0 & b_{2n,gh} & \cdots & b_{mn,gh} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$ is an $m \times m$ square matrix. Also, $\sum_{j=2}^m C'_{jn}\epsilon_{1n} = \boldsymbol{\alpha}'_n\zeta_n$ where $\boldsymbol{\alpha}_n = (c_{1n,1}, \mathbf{0}_{1\times(m-1)}, ..., c_{mn,n}, \mathbf{0}_{1\times(m-1)})'$ is an $(n \times m) \times 1$ vector. Hence,

$$\sum_{j=2}^m (\epsilon'_{jn}B_{jn} + C'_{jn})\epsilon_{1n} = \zeta'_n\mathbf{A}_n\zeta_n + \boldsymbol{\alpha}'_n\zeta_n,$$

where $var(\frac{1}{\sqrt{n}}\zeta_n'\mathbf{A}_n\zeta_n + \frac{1}{\sqrt{n}}\boldsymbol{\alpha}_n'\zeta_n) = \Sigma_{Q,n}$. As $\mathbf{A}_n$ is UB because m is finite, and elements of $\boldsymbol{\alpha}_n$ is bounded, we have the result from Theorem A.1 in Kelejian and Prucha (2007).

B Proof for Lemma 3.1 and 3.2

Proof for Lemma 3.1: From (2.4), $Y_{jn} = \psi_{jn}Y_{jn}^u + \tilde{Y}_{jn}$ where $Y_{jn}^u = \lambda_{jn}W_{jn}^u\epsilon_{jn}$ and $\tilde{Y}_{jn} = (I_n - \lambda_{jn}\tilde{W}_{jn})^{-1}\epsilon_{jn}$. Hence,

$$Y_{jn}^* = \lambda_{jn}W_{jn}^u\epsilon_{jn} + \frac{1}{\psi_{jn}}\tilde{Y}_{jn}.$$

When $\psi_{jn} \to \infty$, the dominant term of Y_{jn}^* is $\lambda_{jn}W_{jn}^u\epsilon_{jn}$. When ψ_{jn} is a finite constant (or a convergent sequence), Y_{jn}^* is just $\frac{1}{\psi_{jn}}S_{jn}^{-1}\epsilon_{jn}$. By Assumption 4 and Lemma A.1, $\frac{1}{n}Y_{jn}^{*\prime}Y_{jn}^* = \sigma_j^2\frac{1}{n}tr(\mathbb{S}_{jn}) + O_p(\frac{1}{\sqrt{n}})$ and $\frac{1}{n}Y_{in}^{*\prime}Y_{jn}^* = O_p(\frac{1}{\sqrt{n}})$ for $i \neq j$. This proves the first part.

For the second part, $\frac{1}{\sqrt{n}}l_n'Y_{1n}^* = \frac{1}{\sqrt{n}}l_n'\frac{1}{\psi_{1n}}S_{1n}^{-1}\epsilon_{1n}$ and $\frac{1}{\sqrt{n}}Y_{jn}^{*\prime}Y_{1n}^* = \frac{1}{\sqrt{n}}\epsilon_{jn}'\frac{1}{\psi_{jn}}S_{jn}'^{-1}\frac{1}{\psi_{1n}}S_{1n}^{-1}\epsilon_{1n}$. As $\frac{1}{\psi_{jn}}S_{jn}^{-1}$ is UB from Assumption 4, we have the result from Lemma A.2.

Proof for Lemma 3.2: As $P_n = X_n(X_n'X_n)^{-1}X_n'$, we have $\frac{1}{n}Y_{1n}^{*\prime}B_nP_nY_{1n}^* = \frac{1}{n}Y_{1n}^{*\prime}B_nX_n(X_n'X_n)^{-1}X_n'Y_{1n}^* = \frac{1}{n}Y_{1n}^{*\prime}B_nX_n^*(X_n^{*\prime}X_n^*)^{-1}X_n^{*\prime}Y_{1n}^* = \frac{1}{n}Y_{1n}^{*\prime}B_nX_n^*(\frac{1}{n}X_n^{*\prime}X_n^*)^{-1}\frac{1}{n}X_n^{*\prime}Y_{1n}^*$. For $\frac{1}{n}Y_{1n}^{*\prime}B_nX_n^*$, as $X_n^* = [l_n, \mathbf{Y}_{-1,n}^*]$ which does not include Y_{1n}^*, we have $\frac{1}{n}Y_{1n}^{*\prime}B_nX_n^* = O_p(\frac{1}{\sqrt{n}})$ as $E(\frac{1}{n}Y_{1n}^{*\prime}B_nX_n^*) = 0$ and $var(\frac{1}{n}Y_{1n}^{*\prime}B_nX_n^*) = O(\frac{1}{n})$. From Lemma 3.1, $\frac{1}{n}X_n^{*\prime}X_n^* = D_{n,xx}^* + O_p(\frac{1}{\sqrt{n}})$ and $\frac{1}{n}X_n^{*\prime}Y_{1n}^* = O_p(\frac{1}{\sqrt{n}})$ where $D_{n,xx}^*$ is $O(1)$. Hence, $\frac{1}{n}Y_{1n}^{*\prime}B_nP_nY_{1n}^* = \frac{1}{n}Y_{1n}^{*\prime}B_nX_n^*(\frac{1}{n}X_n^{*\prime}X_n^*)^{-1}\frac{1}{n}X_n^{*\prime}Y_{1n}^* = O_p(\frac{1}{n})$.

References

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

Anselin, L. (1992) Space and applied econometrics, Anselin (ed.), Special Issue, *Regional Science and Urban Economics,* 22.

Anselin, L. and A.K. Bera (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics, in A. Ullah and D.E.A. Giles (ed.) *Handbook of Applied Economics Statistics*. New York: Marcel Dekker, 237-289.

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

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

Cliff, A.D. and J.K. Ord (1973) *Spatial Autocorrelation*, London,Pion Ltd.

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

Fingleton, B. (1999) Spurious spatial regression: some Monte Carlo results with a spatial unit root and spatial cointegration, *Journal of Regional Science,* 39, 1-19.

Granger, C.W.J. and P. Newbold (1974) Spurious regressions in econometrics, *Journal of Econometrics* 2, 111-120.

Granger, C.W.J. and R.F. Engle (1985) Dynamic model specification with equilibrium constraints: Cointegration and error correction, Discussion paper no. 85-18, UCSD.

Kelejian, H.H. and I.R. Prucha (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance, *Journal of Real Estate Finance and Economics,* 17:1, 99-121.

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

Kelejian H.H. and I.R. Prucha (2007) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Working Paper, University of Maryland.

Kelejian, H.H. and D. Robinson (1993) A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model, *Papers in Regional Science,* 72, 297-312.

Lauridsen, J. and R. Kosfeld (2006) A test for spurious spatial regression, spatial nonstationarity, and spatial cointegration, *Papers in Regional Science,* 85, 363-377.

Lauridsen, J. and R. Kosfeld (2007) Spatial cointegration and heteroscedasticity, *Journal of Geography Systems,* 9, 253-265.

Lee, L.F. (2003) Best spatial two-stage least squares estimator for a spatial autoregressive model with autoregressive disturbances, *Econometric Reviews,* 22, 307-335.

Lee, L.F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models, *Econometrica,* 72, 1899-1925.

Lee, L.F. (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics,* 137, 489-514.

Lee, L.F., J. Yu (2007) Near unit root in the spatial autoregressive model, Working Paper, The Ohio State University.

Mur, J. and F.J. Trivez (2003) Unit roots and deterministic trends in spatial econometric models, *International Regional Science Review,* 26, 289-312.

Ord, J.K. (1975) Estimation methods for models of spatial interaction, *Journal of the American Statistical Association,* 70, 120-297.

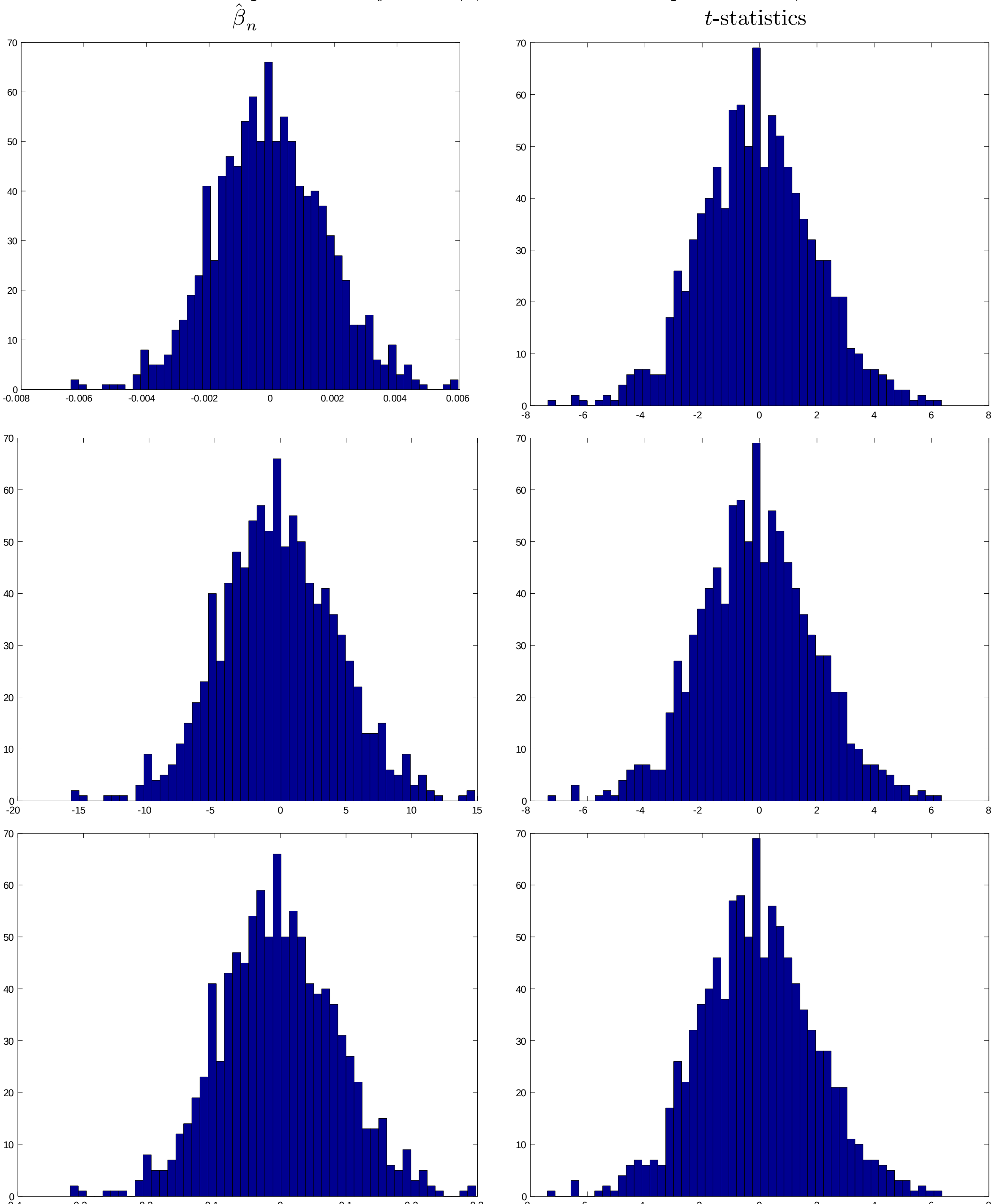
Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics, *Journal of Econometrics,* 33, 311-340.

Phillips, P.C.B. (1987) Towards a unified asymptotic theory for autoregressions, *Biometrika,* 74, 535-547.

Phillips, P.C.B. (1988) Regression theory for near-integrated time series, *Econometrica,* 56, 1021-1043.

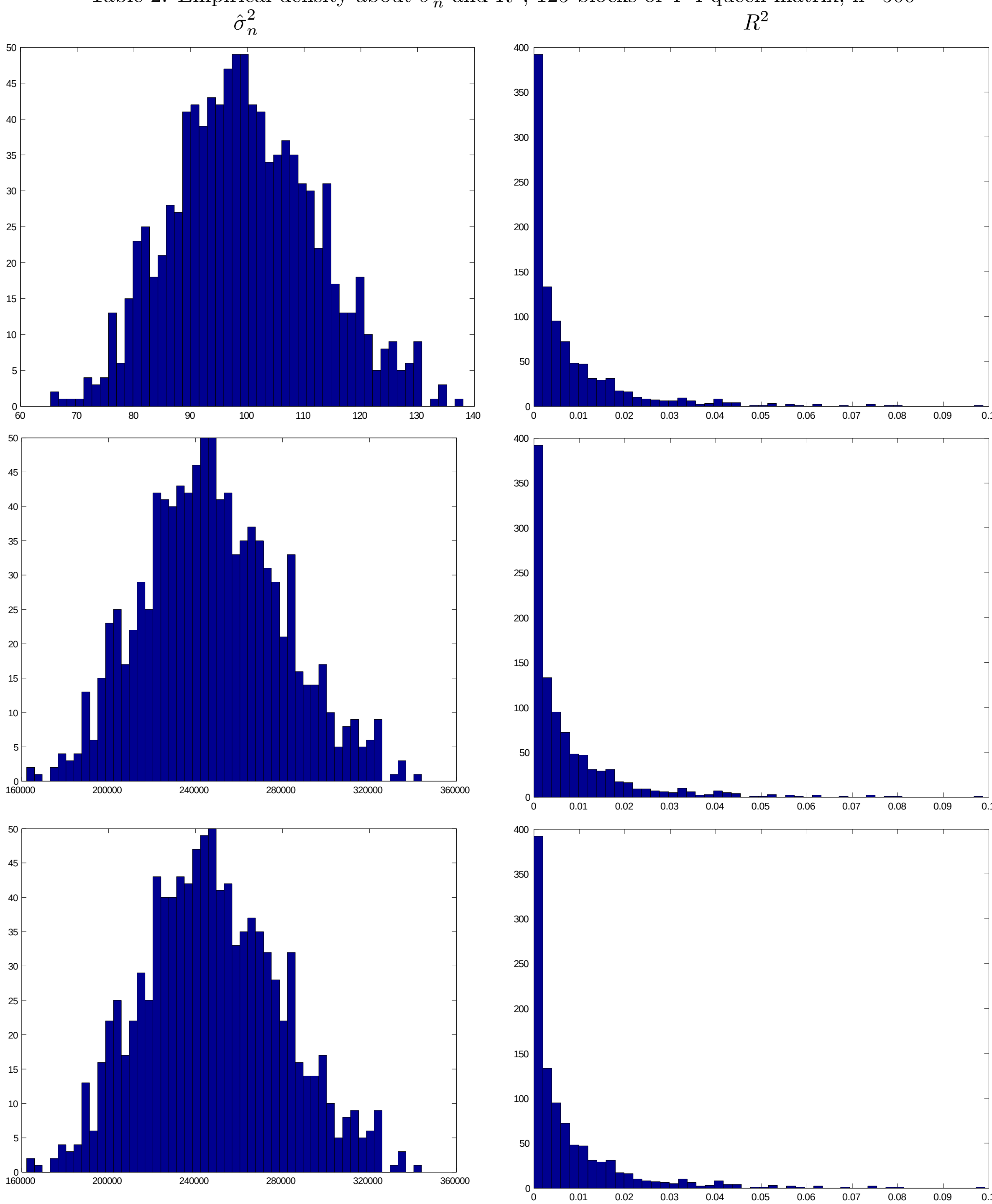
Yu, J., R. de Jong and L.F. Lee (2007) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: a nonstationary case, Working Paper, The Ohio State University.

Table 1: Empirical density about $\hat{\beta}$, 125 blocks of 4*4 queen matrix, n=500



From first to third row are λ_1= 0.95 and λ_2= 0.999, λ_1= 0.999 and λ_2= 0.95, and λ_1= λ_2= 0.999.

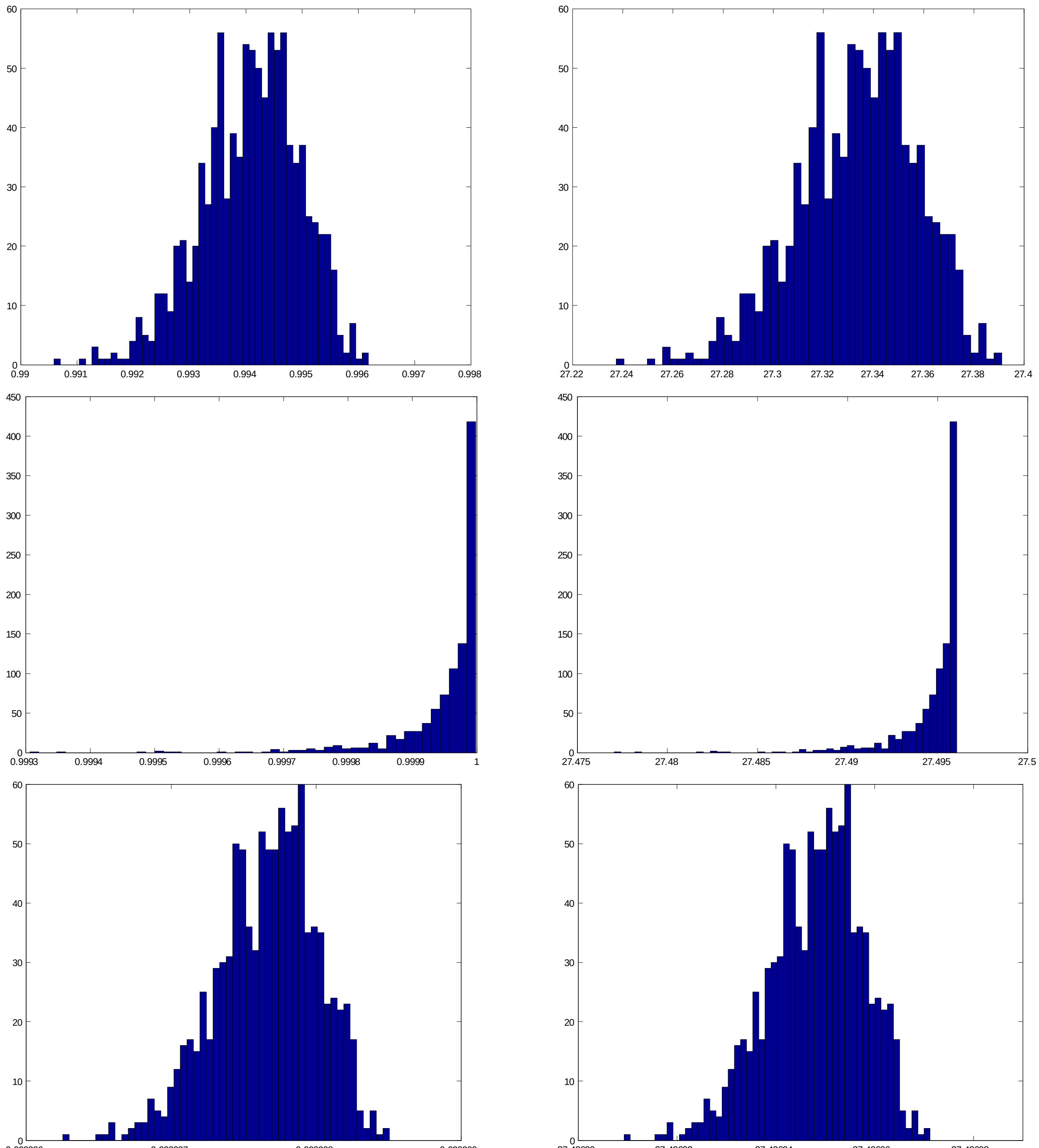
Table 2: Empirical density about $\hat{\sigma}_n^2$ and R^2, 125 blocks of 4*4 queen matrix, n=500



From first to third row are $\lambda_1 = 0.95$ and $\lambda_2 = 0.999$, $\lambda_1 = 0.999$ and $\lambda_2 = 0.95$, and $\lambda_1 = \lambda_2 = 0.999$.

Table 3: Empirical density about Moran I and LM, 125 blocks of 4*4 queen matrix, n=500

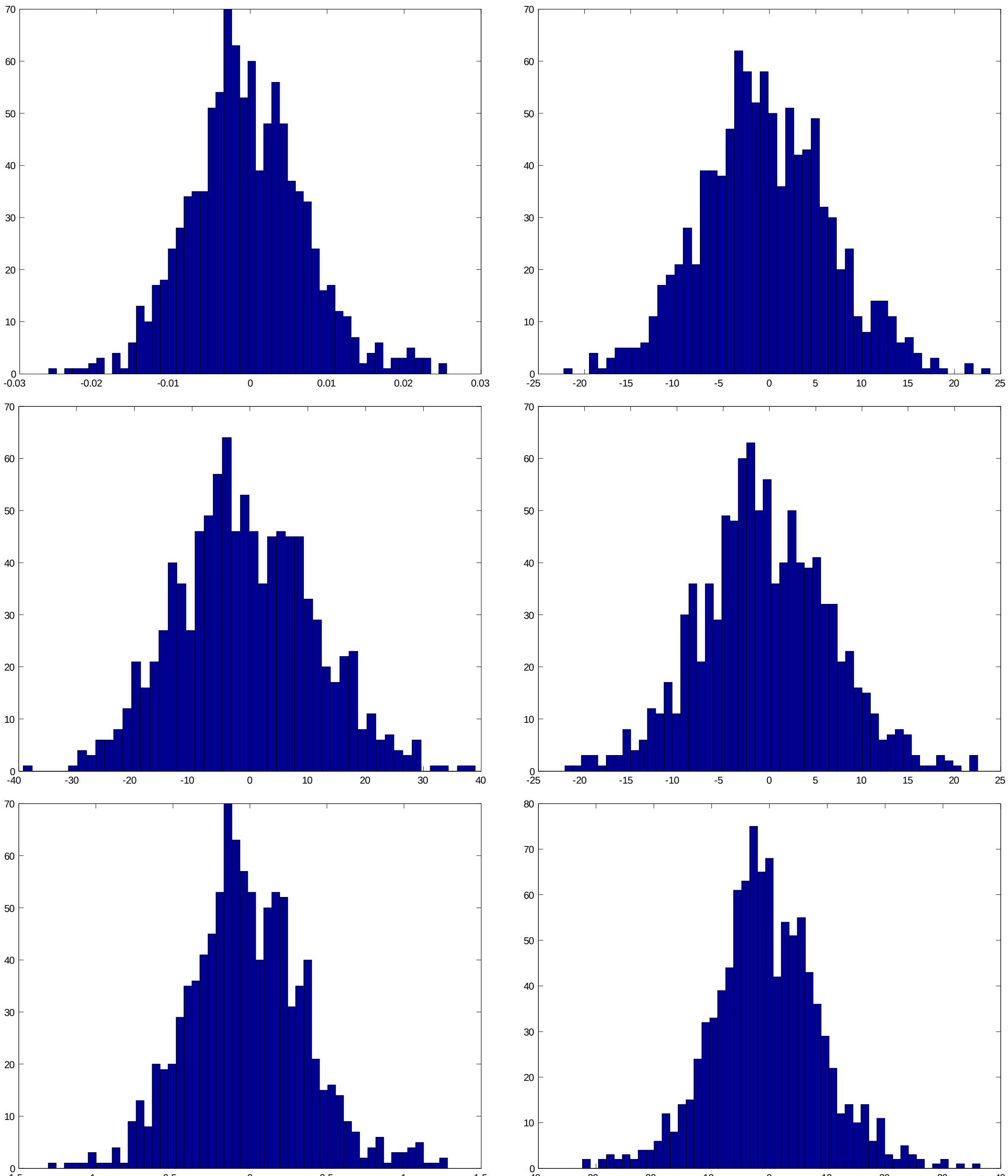
Moran I | LM



From first to third row are $\lambda_1 = 0.95$ and $\lambda_2 = 0.999$, $\lambda_1 = 0.999$ and $\lambda_2 = 0.95$, and $\lambda_1 = \lambda_2 = 0.999$.

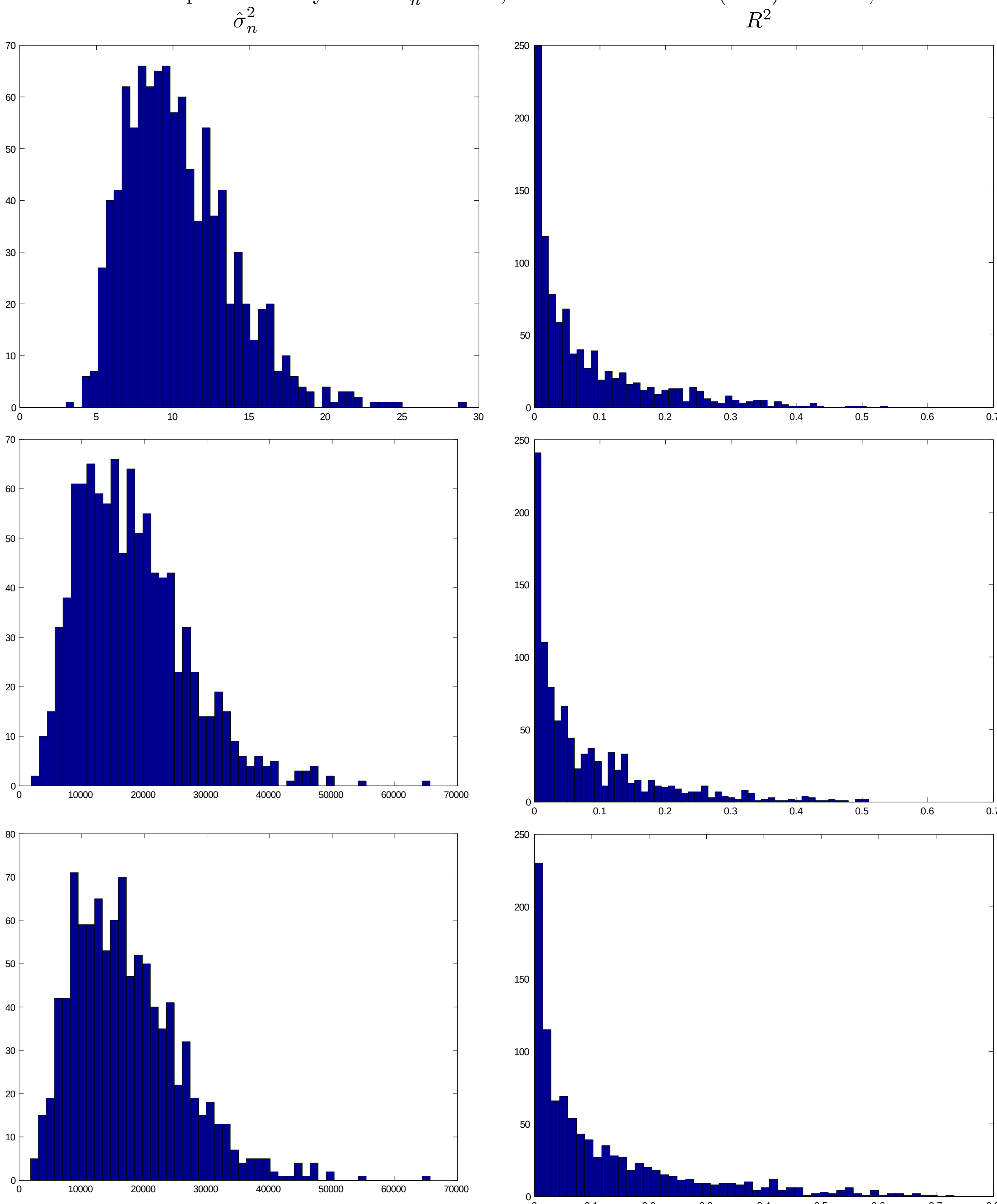
Table 4: Empirical density about $\hat{\beta}$, 10 blocks of Anselin (1988)'s matrix, n=490

$\hat{\beta}_n$ t-statistics



From first to third row are $\lambda_1 = 0.95$ and $\lambda_2 = 0.999$, $\lambda_1 = 0.999$ and $\lambda_2 = 0.95$, and $\lambda_1 = \lambda_2 = 0.999$.

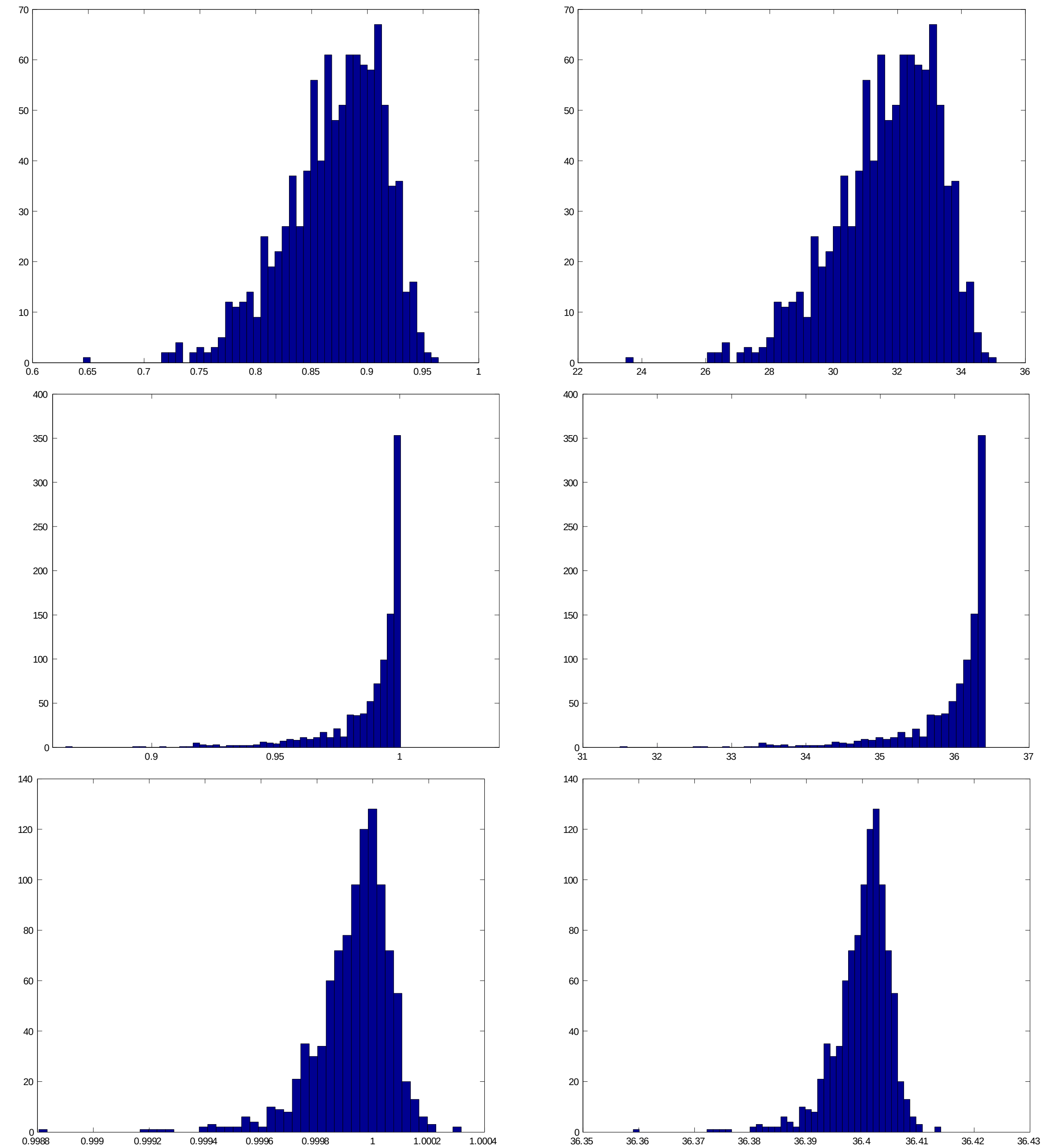
Table 5: Empirical density about $\hat{\sigma}_n^2$ and R^2, 10 blocks of Anselin (1988)'s matrix, n=490



From first to third row are $\lambda_1 = 0.95$ and $\lambda_2 = 0.999$, $\lambda_1 = 0.999$ and $\lambda_2 = 0.95$, and $\lambda_1 = \lambda_2 = 0.999$.

Table 6: Empirical density about Moran I and LM, 10 blocks of Anselin (1988)'s matrix, n=490

Moran I | LM



From first to third row are λ_1= 0.95 and λ_2= 0.999, λ_1= 0.999 and λ_2= 0.95, and λ_1= λ_2= 0.999.

Table 7: t-statistics for $\hat{\beta}_n$, rep=1000

	$n = 100$ from the queen matrix			$n = 98$ from Anselin (1988)		
	λ_1= 0.95 λ_2= 0.999	λ_1= 0.999 λ_2= 0.95	λ_1= 0.999 λ_2= 0.999	λ_1= 0.95 λ_2= 0.999	λ_1= 0.999 λ_2= 0.95	λ_1= 0.999 λ_2= 0.999
below -30	0	0	0	6	8	461
-12 to -30	0	0	0	134	148	29
-12 to -10	0	0	0	38	40	4
-10 to -8	0	0	0	47	40	3
-8 to -6	1	1	1	68	55	3
-6 to -4	25	25	25	64	71	3
-4 to -2	168	169	172	80	76	2
-2 to 0	335	334	331	72	66	1
0 to 2	293	293	291	63	72	5
2 to 4	141	141	143	73	87	2
4 to 6	37	37	37	63	64	1
6 to 8	0	0	0	54	53	2
8 to 10	0	0	0	60	47	2
10 to 12	0	0	0	49	40	2
12 to 30	0			118	129	28
30 above	0	0	0	11	4	452
	$n = 500$ from the queen matrix			$n = 490$ from Anselin (1988)		
	λ_1= 0.95 λ_2= 0.999	λ_1= 0.999 λ_2= 0.95	λ_1= 0.999 λ_2= 0.999	λ_1= 0.95 λ_2= 0.999	λ_1= 0.999 λ_2= 0.95	λ_1= 0.999 λ_2= 0.999
below -30	0	0	0	0	0	2
-12 to -30	0	0	0	41	49	94
-12 to -10	0	0	0	40	33	48
-10 to -8	0	0	0	51	63	50
-8 to -6	4	4	4	76	63	61
-6 to -4	25	25	25	92	94	86
-4 to -2	129	129	130	126	127	108
-2 to 0	374	374	373	123	120	101
0 to 2	320	320	320	86	90	72
2 to 4	123	123	123	107	95	82
4 to 6	24	24	23	89	89	71
6 to 8	1	1	2	58	59	59
8 to 10	0	0	0	39	50	52
10 to 12	0	0	0	24	27	30
12 to 30	0	0	0	48	41	80
30 above	0	0	0	0	0	4