

Omitted Variable Biases of OLS and Spatial Lag Models

R. Kelley Pace and James P. LeSage

1 Introduction

Numerous authors have suggested that omitted variables affect spatial regression methods less than ordinary least-squares (OLS; Dubin 1988; Brasington and Hite 2005, Cressie 1993). To explore these conjectures, we derive an expression for OLS omitted variable bias in a univariate model with spatial dependence in the disturbances and explanatory variables. There are a number of motivations for making this set of assumptions regarding the disturbances and explanatory variables. First, in spatial regression models each observation represents a region or point located in space, for example, census tracts, counties or individual houses. Sample data used as explanatory variables in these models typically consists of socioeconomic, census and other characteristics of the regional or point locations associated with each observation. Therefore, spatial dependence in the explanatory variables seems likely, motivating our choice of this assumption. Note, the literature rarely examines the spatial character of the explanatory variables, but this can affect the relative performance of OLS as shown below. Second, application of OLS models to regional data samples frequently leads to spatial dependence in the regression disturbances, providing a justification for this assumption. Finally, there are a host of latent unobservable and frequently unmeasurable influences that are likely to impact spatial regression relationships. For example, factors such as location and other types of amenities, highway accessibility, school quality or neighborhood prestige may exert an influence on the dependent variable in hedonic house price models. It is unlikely that explanatory variables are readily available to capture all of these types of latent influences. This type of reasoning motivates our focus on the impact of omitted explanatory variables. Since the omitted and included explanatory variables are both likely to exhibit spatial dependence based on the same spatial connectivity structure, it seems likely that omitted and included variables will exhibit

R.K. Pace (✉)

E. J. Ourso College of Business Administration, Department of Finance, Louisiana State University, Baton Rouge, LA 70803-6308, USA,
e-mail: kelley@pace.am

non-zero covariance. The expression we derive for OLS bias in these circumstances shows that positive dependence in the disturbances and explanatory variables when omitted variables are correlated with included explanatory variables magnifies the magnitude of conventional least-squares omitted variables bias.

We extend the considerations above to also include models where the dependent variable exhibits spatial dependence, following a spatial autoregressive process. LeSage and Pace (2009) provide a number of different motivations for how spatial dependence in the dependent variable arises in spatial regression relationships. It is well-known that spatial dependence in the dependent variable leads to bias in OLS estimates (Anselin 1988). We show that this type of spatial dependence in the presence of omitted variables exacerbates the usual bias that arises when applying OLS to this type of sample data. In particular, the bias is magnified, with the magnitude of bias depending on the strength of spatial dependence in: the disturbances, the dependent variable, and the explanatory variable included in the model.

Our derivation shows that the combination of an omitted variable, spatial dependence in the disturbances, dependent and explanatory variables leads to an implied model specification that includes spatial lags of both the dependent and explanatory variables. This type of model has been labeled a spatial Durbin model (SDM) in the literature (Anselin 1988). Estimates based on the SDM specification which matches the implied DGP in this set of circumstances shrinks the bias relative to OLS.

In the following section, we consider the implications of omitted variables in the presence of spatial dependence for OLS estimates. Next we demonstrate that the SDM model specification matches a reparameterization of the DGP that results from various assumptions on omitted variables and spatial dependence. We consider an expression for the omitted variables bias that arises when the SDM model is used to produce estimates, and compare this to the bias expression for OLS estimates. We show that the magnitude of omitted variable bias for the SDM model does not exhibit the magnification of OLS and it no longer depends on the magnitude of spatial dependence in the disturbances, dependent, or independent variables. These desirable properties of the SDM model provide a strong motivation for use of this model specification in applied practice.

2 Spatial Dependencies and OLS Bias

We begin with a frequently used spatial econometric model specification shown in (1) and (2). Equation (1) represents a spatial autoregressive process governing the dependent variable and (2) adds the assumption of spatial autoregressive disturbances. This model has been labeled SAC by LeSage (1999) and a spatial autoregressive model with autoregressive disturbances by Kelejian and Prucha (1998). It should be noted that we will work with a model involving simple univariate explanatory and omitted variable vectors for simplicity. There is no reason to believe that the results we derive here would not extend to the more general case involving matrices of explanatory variables in place of the univariate vector.

$$y = x\beta + \alpha Wy + \varepsilon \quad (1)$$

$$\varepsilon = \rho W\varepsilon + \xi \quad (2)$$

$$\xi = xy + u \quad (3)$$

$$x = \phi Wx + v \quad (4)$$

In (1) through (4), the n by 1 vector y represents observations on the dependent variable, x represents an n by 1 vector of observations on a non-constant explanatory variable, ε, ξ, u , and v represent various types of n by 1 disturbance vectors, $\alpha, \beta, \rho, \phi$, and γ represent scalar parameters, and W is an n by n non-negative symmetric spatial weight matrix with zeros on the diagonal. We assume that u is distributed $N(0, \sigma_u^2 I_n)$, v is distributed $N(0, \sigma_v^2 I_n)$, and u is independent of v . For simplicity, we exclude the intercept term from the model.

We extend the conventional SAC model specification using (3) that adds the assumption of an omitted variable correlated with the explanatory variable x . The strength of correlation is determined by the parameter γ and the variance of the noise vector u , σ_u^2 . The last equation, (4) adds the assumption of a spatial dependence in the explanatory variable x , which is governed by a spatial autoregressive process with dependence parameter ϕ . We focus on non-negative spatial dependence, by assuming $\alpha, \phi, \rho \in [0, 1)$. We note that in the case where $\gamma = 0$, there is no covariance between the included explanatory variable x and the omitted variable ξ . In the case where $\phi = 0$, the explanatory variable does not exhibit spatial dependence, and when $\alpha = 0$, the dependent variable does not exhibit spatial dependence. Similarly $\rho = 0$ eliminates spatial dependence in the disturbances.

The weight matrix has positive elements W_{ij} when observations i and j are neighbors, and we assume each observation has at least one neighbor. The symmetry of W contrasts with the usual lag operator matrix L from time series, since L is strictly triangular containing zeros on the diagonal. Powers of L are also strictly triangular with zeros on the diagonal, so that L^2 specifies a two-period time lag whereas L creates a single period time lag. It is never the case that produces observations that point back to include the present time period. In contrast, W^2 specifies neighbors to the neighbors identified by the matrix W , and since the neighbor of the neighbor to an observation includes the observation itself due to symmetry, W^2 has positive elements on the diagonal. This results in a form of simultaneous dependence among spatial observations that does not occur in time series analysis, making spatial regression models distinct from time series regressions. We use the same spatial weight matrix W to specify the pattern of spatial dependence in the explanatory variable, which seems reasonable since this matrix reflects the spatial configuration of both the dependent and independent variable observations.

We assume that W is symmetric and real, so the n by 1 vector of eigenvalues λ_W is real. For simplicity, we assume the eigenvalues are unique, the principal eigenvalue of W equals 1, and this is the maximum eigenvalue as well. This is not a restrictive assumption since dividing any candidate weight matrix by its principal eigenvalue would yield a weight matrix with a principal eigenvalue of 1.

Since the sum of the eigenvalues ($tr(W)$) equals 0, the minimum eigenvalue is negative, but the minimum eigenvalue is not the principal eigenvalue, and $\min(\lambda_W) > -1$. Therefore, for some real scalar $\theta \in (\min(\lambda_W)^{-1}, 1)$, $I_n - \theta W$ will be symmetric, positive definite, and thus $(I_n - \theta W)^{-1}$ exists. Clearly, $\theta \in [0, 1)$ is sufficient for positive definite $I_n - \theta W$. Finally, since the maximum eigenvalue equals 1, $tr(W^{2j}) \geq 1$ (all eigenvalues of even powered matrices are non-negative and the largest eigenvalue equals 1) and 0 for any integer $j > 0$ (traces of non-negative matrices are non-negative).

We rewrite (1) to solve for y using $F(\alpha) = (I_n - \alpha W)^{-1}$ and this yields (5). We rewrite (2) to solve for ε using $G(\rho) = (I_n - \rho W)^{-1}$ and substitute $x\gamma + u$ for ξ via (3) to yield (6). Similarly, we rewrite (4) to isolate x using $H(\phi) = (I_n - \phi W)^{-1}$ to produce (7). Equation (8) summarizes the definitions.

$$y = F(\alpha)x\beta + F(\alpha)\varepsilon \quad (5)$$

$$\varepsilon = G(\rho)(x\gamma + u) \quad (6)$$

$$x = H(\phi)v \quad (7)$$

$$F(\alpha) = (I_n - \alpha W)^{-1} \quad (8)$$

$$G(\rho) = (I_n - \rho W)^{-1}$$

$$H(\phi) = (I_n - \phi W)^{-1}$$

Taken together, (5), (6), and (7) lead to the DGP shown in (9).

$$y = F(\alpha)H(\phi)v\beta + F(\alpha)G(\rho)H(\phi)v\gamma + F(\alpha)G(\rho)u \quad (9)$$

Given the assumptions made concerning the matrix W , the matrix inverses: $F(\alpha)$, $G(\rho)$, $H(\phi)$ exist. We refer to (9) as a DGP since this expression could be used with vectors v, u of random deviates to generate a dependent variable vector y from the model and assumptions set forth. Given the structure of the model set forth in (1)–(4), the parameters $\alpha, \rho, \phi, \gamma$ allow us to generate dependent variable vectors that reflect varying combinations of our assumptions. For example, setting $\gamma = 0$ and maintaining positive values for α, ρ, ϕ would produce a vector y reflecting no covariance between the included and omitted variable vectors x and ξ . Similarly, setting $\phi = 0$ while maintaining positive values for the other parameters (α, γ, ρ) would produce a vector y from a model having no spatial dependence in the explanatory variable x .

OLS estimates $\hat{\beta}_o = (x'x)^{-1}x'y$ represent “best linear unbiased” estimates when the DGP matches that of the ordinary regression model: $y = x\beta + \varepsilon$, and the Gauss–Markov assumptions. These require the vector x to be fixed in repeated sampling and the disturbances to have constant variance and zero covariance.

However, suppose that the true DGP is (9) and we apply the least-squares expressions to produce the estimates in (10). That is, we apply least-squares in the circumstances considered here involving spatial dependence in the dependent variable, disturbances and the model contains an omitted variable that is correlated with the spatially dependent included variable.

$$\hat{\beta}_o = \frac{v'H(\phi)F(\alpha)H(\phi)v}{v'H(\phi)^2v}\beta + \frac{v'H(\phi)F(\alpha)G(\rho)H(\phi)v}{v'H(\phi)^2v}\gamma + \frac{v'H(\phi)F(\alpha)G(\rho)u}{v'H(\phi)^2v} \quad (10)$$

This expression can be further simplified. To do so, we turn to some additional results. We begin by defining (11),

$$R(A) = \frac{d'Ad}{d'd} \quad (11)$$

$$Q(A) = \frac{d'Ar}{d'd}$$

where d , r are distributed $N(0, \sigma_d^2 I_n)$, $N(0, \sigma_r^2 I_n)$ with r independent of d , and A is a n by n symmetric real matrix. Using different techniques, both Barry and Pace (1999), and Girard (1989) show that:

$$E(R(A)) = \frac{tr(A)}{n} \quad (12)$$

$$\sigma_{R(A)}^2 = \frac{2\sigma_{\lambda(A)}^2}{n}$$

$$E(Q(A)) = 0$$

where tr denotes the trace operator and $\sigma_{\lambda(A)}^2$ is the variance of the eigenvalues of matrix A . Obviously, $E(d'Ar) = 0$ due to the independence of r and d , while $d'd > 0$ so that $E(Q(A)) = 0$.

Consider a variation of (11) involving n by n symmetric real matrices A and B .

$$R(A/B) = \frac{d'Ad}{d'Bd} = \frac{(d'd)^{-1}d'Ad}{(d'd)^{-1}d'Bd} \quad (13)$$

From (12), the expectation of the numerator of (13) equals $tr(A)/n$, and the expectation of the denominator of (13) equals $tr(B)/n$. Also, an implication of (12) is that as $n \rightarrow \infty$, the variance of the numerator and denominator go to 0. Therefore,

$$plim_{n \rightarrow \infty} R(A/B) = \frac{tr(A)}{tr(B)} \quad (14)$$

Applying these results to expression (10), results in the third term of (10) vanishing asymptotically via (12). Applying result (14) to the first two terms of (10), and using the cyclical properties of the trace, produces expression (15) and its equivalent abbreviated form in (16).

$$\begin{aligned}
p \lim_{n \rightarrow \infty} \hat{\beta}_o &= \frac{tr [H(\phi)^2 F(\alpha)]}{tr [H(\phi)^2]} \beta + \frac{tr [H(\phi)^2 F(\alpha) G(\rho)]}{tr [H(\phi)^2]} \gamma \\
&= T_\beta(\phi, \alpha) \beta + T_\gamma(\phi, \alpha, \rho) \gamma
\end{aligned} \tag{15}$$

$$\begin{aligned}
T_\beta(\phi, \alpha) &= \frac{tr [H(\phi)^2 F(\alpha)]}{tr [H(\phi)^2]} \\
&= T_\gamma(\phi, \alpha, \rho) = \frac{tr [H(\phi)^2 F(\alpha) G(\rho)]}{tr [H(\phi)^2]}
\end{aligned} \tag{16}$$

Naturally, as the factors $T_\beta(\phi, \alpha)$ and $T_\gamma(\phi, \alpha, \rho)$ rise above 1, the bias of using OLS to produce estimates for a model with a dependent variable y generated using our spatial DGP from (9) can increase. This will be especially true when β and γ have the same signs. We will show that $T_\beta(\phi, \alpha) > 1$ for $\alpha > 0$ and $T_\gamma(\phi, \alpha, \rho) > 1$ when spatial dependence in the dependent variable y or disturbances exists ($\alpha > 0$ or $\rho > 0$), and that spatial dependence in the explanatory variable $\phi > 0$ amplifies these factors. Our strategy involves showing that when no spatial dependence in the explanatory variable exists ($\phi > 0$), $T_\beta(0, \alpha) > 1$ when $\alpha > 0$ and $T_\gamma(0, \alpha, \rho) > 1$ when $\alpha > 0$ or $\rho > 0$. We then show that $T_\beta(\phi, \alpha) > T_\beta(0, \alpha)$ and that $T_\gamma(\phi, \alpha, \rho) > T_\gamma(0, \alpha, \rho)$ when $\phi > 0$.

We begin by showing that $T_\beta(\phi, \alpha) > 1$ when $\alpha > 0$, $\phi > 0$ and $T_\gamma(\phi, \alpha, \rho) > 1$ when α or ρ are positive and $\phi = 0$ (no spatial dependence in the explanatory variable). To see the first assertion, let θ_m represent some positive scalar parameter. Since $(I_n - \theta_m W)^{-1} = I_n + \theta_m^2 W^2 + \dots$, and since $tr(I_n) = n$, $tr(W^{2j}) \geq 1$, and $tr(W^{2j-1}) \geq 0$, $tr(I_n - \theta_m W)^{-1} > n$. To generalize this, let $\theta_1 > 0$ or $\theta_2 > 0$ and consider $P(\theta_1, \theta_2) = (I_n - \theta_1 W)^{-1} (I_n - \theta_2 W)^{-1} = I_n + \pi_1 W + \pi_2 W^2 + \dots$ where $\pi > 0$. Since products and sums of positive parameters (θ_1, θ_2) are positive, $tr[P(\theta_1, \theta_2)] > n$ because $tr(I_n) = n$, $tr(W^{2j}) \geq 1$, and $tr(W^{2j-1}) \geq 0$. When $\phi = 0$, this describes the numerator of both $T_\beta(0, \alpha)$ and $T_\gamma(0, \alpha, \rho)$ and the denominator of both terms is n . Consequently, $T_\beta(0, \alpha) > 1$ and $T_\gamma(0, \alpha, \rho) > 1$.

We now turn to the effect of positive spatial dependence in the explanatory variable ($\phi > 0$) on $T_\beta(\phi, \alpha)$ and $T_\gamma(\phi, \alpha, \rho)$. We show that $T_\beta(\phi, \alpha) > T_\beta(0, \alpha)$ and $T_\gamma(\phi, \alpha, \rho) > T_\gamma(0, \alpha, \rho)$ for $\phi > 0$. Let Ω and Ψ be monotonic functions of similar symmetric matrices so that both Ω and Ψ are symmetric positive definite and are not proportional to an identity matrix. Given these assumptions, consider the assertion in (17). Multiplying both sides by the positive scalar $tr(\Psi)/n$ does not change the direction of the inequality and this leads to (18). Since Ψ and Ω are based upon the same eigenvalues (similar matrices) and are monotonic functions of these eigenvalues, the eigenvalues of Ψ and Ω have the same ordering. Moreover, the eigenvalues of $\Psi\Omega$ are the product of these ordered eigenvalues as shown in equation (19).

$$\frac{tr(\Omega)}{n} < \frac{tr(\Psi\Omega)}{tr(\Psi)} \quad (17)$$

$$\frac{tr(\Omega)}{n} \frac{tr(\Psi)}{n} < \frac{tr(\Psi\Omega)}{n} \quad (18)$$

$$\left[n^{-1} \sum_{i=1}^n \lambda(\Psi)_i \right] \left[n^{-1} \sum_{i=1}^n \lambda(\Omega)_i \right] < \left[n^{-1} \sum_{i=1}^n \lambda(\Psi)_i \lambda(\Omega)_i \right] \quad (19)$$

In fact, (19) is a restatement of the Chebyshev sum inequality from Gradshteyn and Ryzhik (1980). Expression (19) holds true as a strict inequality, since the eigenvalues are not all the same (because Ω and Ψ are not proportional to the identity matrix). Substitution of $\Psi = H(\phi)^2$ and $\Omega = F(\alpha)$ or $\Omega = F(\alpha)G(\rho)$ proves the assertion that $T_\beta(\phi, \alpha) > T_\beta(0, \alpha)$ and $T_\gamma(\phi, \alpha, \rho) > T_\gamma(0, \alpha, \rho)$ for $\phi > 0$, where the strict inequality arises because the eigenvalues of W and the monotonic functions of the eigenvalues of W are not similar to the identity matrix.

As already indicated, our expression (9) for the DGP allows us to consider special cases that arise from various settings of the control parameters $\alpha, \rho, \phi, \gamma$. We enumerate how some of these special cases impact omitted variables bias in various applied situations using our results applied to the expressions from (15).

1. Spatial dependence in the disturbances and explanatory variable, but no covariance between the explanatory variable and omitted variable. This results from setting the parameters ($p > 0, \phi > 0, \alpha = 0, \gamma = 0$). In this case, $plim_{n \rightarrow \infty} \hat{\beta}_o = \beta$, and there is no asymptotic bias.
2. Spatial dependence in the explanatory variable in the presence of an omitted variable that is correlated with the included explanatory variable but no spatial dependence in the dependent variable or disturbances. This results from setting the parameters ($\phi > 0, \gamma > 0, \alpha = 0, \rho = 0$). In this case, $plim_{n \rightarrow \infty} \hat{\beta}_o = \beta + \gamma$, and we have the standard omitted variable bias that would arise in the least-squares model.
3. Spatial dependence in y and the explanatory variable with no correlation between the explanatory and omitted. This results from setting the parameters ($\alpha > 0, \phi > 0, \gamma = 0$). In this case, $plim_{n \rightarrow \infty} \hat{\beta}_o = T_\beta(\phi > 0, \alpha > 0)\beta$, and OLS has asymptotic bias amplified by the parameter α reflecting the strength of spatial dependence in y and by ϕ representing the strength of dependence in the explanatory variable.
4. No spatial dependence in y , spatial dependence in the disturbances and the explanatory variables with an omitted variable that is correlated with the included explanatory variable. This results from setting the parameters ($\rho > 0, \gamma > 0, \phi > 0, \alpha = 0$). In this case, $plim_{n \rightarrow \infty} \hat{\beta}_o = \beta + T_\gamma(\phi > 0, \alpha = 0, \rho > 0)\gamma$, and OLS has omitted variables bias amplified by the spatial dependence in the disturbances and in the explanatory variable reflected by the magnitudes of the scalar parameters ρ and ϕ .

The first result is well-known, and the second is a minor extension of the conventional omitted variables case for least-squares. The third result shows the bias from

applying OLS when the true DGP produces spatial dependence in the dependent variable y , and there is spatial dependence in the included explanatory variable. The bias for this case exceeds that shown in Anselin (1988) due to the spatial dependence in the explanatory variable. The fourth case shows that the usual result that spatial dependence in the disturbances does not lead to bias in OLS estimates does not hold in the presence of an omitted variable (that is correlated with the included explanatory variable). We find that spatial dependence in the disturbances (and/or in the explanatory variable) in the presence of omitted variables leads to a magnification of the conventional omitted variables bias.

To obtain some feel for the magnitudes of these biases, we conducted a small Monte Carlo experiment. In the computations, we simulated a square random set of 1,000 locations and used these locations to compute a contiguity-based matrix W . The resulting 1,000 by symmetric spatial weight matrix W was standardized to be stochastic (doubly stochastic). We set $\beta = 0.75$ and $\gamma = 0.25$ for all trials. The setting for γ reflects a relatively low level of correlation between the included and omitted variables. Given W and a value for α , ρ , and ϕ we used the DGP to simulate y for 1,000 trials. For each trial we calculated the estimate $\hat{\beta}_o$ and recorded the average of the estimates. We did this for 27 combinations of α , ρ , and ϕ . For each of these 27 cases we also computed the theoretical $E(\hat{\beta}_o)$. Table 1 shows the empirical average of the estimates and the theoretically expected estimates for the 27 cases. The theoretical and empirical results show close agreement, and the table documents that serious bias can occur when omitted variables combine with spatial dependence in the disturbance process. This is especially true if there is spatial dependence in the regressors, a realistic prospect in applied use of spatial regression models that seems to have been overlooked in the literature. For example, OLS estimates yield an empirical average of 3.9984 (expectation of 4.0221) when ρ , α , and Ψ equal 0.8, even though $\beta = 0.75$ and $\gamma = 0.25$. That is, we have a fivefold bias in the OLS estimates.

3 A Comparison with Spatial Lag Models

We consider the contrast between the above results for least-squares estimates and those for estimates from spatial lag models that match the DGP arising from the presence of omitted variables in the face of spatial dependence.

We begin with the DGP (9) which we repeat in (20). In (21) we substitute in x for $H(\phi)v$ as we condition upon x in this analysis. We introduce the identity $G(\rho)G^{-1}(\rho)$ in (22), rearrange terms in (23) using the linearity of $G^{-1}(\rho) = I_n - \rho W$, and arrive at the final expression in (24).

$$y = F(\alpha)H(\phi)v\beta + F(\alpha)G(\rho)H(\phi)v\gamma + F(\alpha)G(\rho)u \quad (20)$$

$$y = F(\alpha)x\beta + F(\alpha)G(\rho)xy + F(\alpha)G(\rho)u \quad (21)$$

Table 1 Mean $\hat{\beta}_o$ and $E(\hat{\beta}_o)$ as function of spatial dependence ($\beta = 0.75, \gamma = 0.25$)

Case	Parameter			Empirical	Theoretical
	ϕ	ρ	α	Mean $\hat{\beta}_o$	$E(\hat{\beta}_o)$
1	0.0000	0.0000	0.0000	1.0020	1.0000
2	0.4000	0.0000	0.0000	0.9999	1.0000
3	0.8000	0.0000	0.0000	0.9993	1.0000
4	0.0000	0.0000	0.4000	1.0382	1.0376
5	0.4000	0.0000	0.4000	1.1366	1.1363
6	0.8000	0.0000	0.4000	1.3401	1.3438
7	0.0000	0.0000	0.8000	1.2531	1.2540
8	0.4000	0.0000	0.8000	1.6176	1.6161
9	0.8000	0.0000	0.8000	2.5559	2.5666
10	0.0000	0.4000	0.0000	1.0076	1.0094
11	0.4000	0.4000	0.0000	1.0361	1.0341
12	0.8000	0.4000	0.0000	1.0862	1.0860
13	0.0000	0.4000	0.4000	1.0604	1.0592
14	0.4000	0.4000	0.4000	1.1918	1.1915
15	0.8000	0.4000	0.4000	1.4722	1.4771
16	0.0000	0.4000	0.8000	1.3121	1.3080
17	0.4000	0.4000	0.8000	1.7348	1.7361
18	0.8000	0.4000	0.8000	2.8559	2.8723
19	0.0000	0.8000	0.0000	1.0638	1.0635
20	0.4000	0.8000	0.0000	1.1570	1.1540
21	0.8000	0.8000	0.0000	1.3897	1.3917
22	0.0000	0.8000	0.4000	1.1450	1.1458
23	0.4000	0.8000	0.4000	1.3759	1.3762
24	0.8000	0.8000	0.4000	1.9511	1.9552
25	0.0000	0.8000	0.8000	1.4952	1.5006
26	0.4000	0.8000	0.8000	2.1363	2.1475
27	0.8000	0.8000	0.8000	3.9984	4.0221

$$y = F(\alpha)G(\rho)G^{-1}(\rho)x\beta + F(\alpha)G(\rho)xy + F(\alpha)G(\rho)u \tag{22}$$

$$y = F(\alpha)G(\rho)x\beta + F(\alpha)G(\rho)Wx(-\rho\beta) + F(\alpha)G(\rho)xy + F(\alpha)G(\rho)u \tag{23}$$

$$y = F(\alpha)G(\rho)x[\beta + \gamma] + F(\alpha)G(\rho)Wx[-\rho\beta] + F(\alpha)G(\rho)u \tag{24}$$

We can transform the DGP in (24) to arrive at an estimation model in (26) containing spatial lags of the dependent and independent variables, which we label the spatial lag model (SLM).

$$G^{-1}(\rho)F^{-1}(\alpha)y = x\beta + Wx\Psi + v \tag{25}$$

$$(I_n - \rho W)(I_n - \alpha W)y = x\beta + Wx\Psi + v \tag{26}$$

$$y = x\beta + Wx\Psi + (\alpha + \rho)Wy - \alpha\rho W^2y + u \tag{27}$$

For the case where there is no spatial dependence in the disturbances so that $\rho = 0$, we have the SDM in (28).

$$\begin{aligned}(I_n - \alpha W) y &= x\beta + Wx\Psi + v \\ y &= x\beta + Wx\Psi + \alpha Wy + u\end{aligned}\tag{28}$$

The SLM model result in (27) points to a potential problem that has been discussed in the spatial econometrics literature. This model specification could lead to what is known as a *label switching identification problem*, if we do not impose the theoretically implied restriction on the estimated parameters α and ρ . In part, this potential for identification problems arises from our use the same spatial weight matrix W in the specification for dependence in both y and x as well as the disturbances for purposes of simplicity. Kelejian and Prucha (2007) show that in the absence of omitted variables the model is identified when using the same spatial weight matrix W for the dependent variable and disturbances, provided that the parameter $\beta = 0$. However, the absence of omitted variables in their consideration results in a simpler model that does not include the two expressions containing spatial lags of the dependent variable, $(\alpha + \rho) Wy$, and $-\alpha\rho W^2y$.

We proceed by working with the SLM model, but assume that the restrictions are used to avoid the potential identification problem. Unlike many restrictions, the restrictions on label switching will not affect the value of the likelihood. Assuming consistency of maximum likelihood estimates for the spatial lag model parameters β , Ψ , ρ , and α , these estimates from the SLM model will equal the underlying structural parameters from the DGP in large samples (Kelejian and Prucha 1998; Lee 2004; Mardia and Marshall 1984). In other words, the asymptotic expected values equal the corresponding parameters in the reparameterized DGP (27), so that $E(\tilde{\beta}) = \beta + \gamma$, $E(\tilde{\Psi}) = -\rho\beta$, $E(\tilde{\rho}) = \rho$, and $E(\tilde{\alpha}) = \alpha$. There is no asymptotic bias in estimates of α and ρ for the SLM model that arise from omitted variables. (This would also be true for the SDM model that would arise in cases where $\rho = 0$.)

However, the asymptotic bias in this model's estimates for β that arise from an omitted variable is $E(\tilde{\beta}) - \beta = \gamma$. Unlike the results for OLS in (15), the bias for the SLM does not depend on x , eliminating the influence of the parameter ϕ reflecting spatial dependence in the included variable x , nor does it depend on spatial dependence in the disturbances reflected by the parameter ρ . Instead, the SLM has a constant bias that depends only upon the strength of relation between the included and omitted explanatory variables reflected by γ . (The same holds true for the SDM model which arises in the case where $\rho = 0$.)

4 Conclusion

The nature of omitted variables bias arising in OLS estimates versus spatial lag model estimates was explored. We assumed that the DGP reflected a situation where spatial dependence existed in the disturbances, the dependent variable, and

the explanatory variables, and we assumed that the omitted variables were correlated with the included explanatory variable. We established that spatial dependence in the explanatory variable exacerbates the usual bias that arises when using OLS to estimate a model relationship generated by a typical spatial econometric model specification that includes dependence in both the disturbances as well as the dependent variable.

Unlike the standard least-squares result for the case of omitted variables, the presence of spatial dependence magnifies conventional omitted variables bias in OLS estimates. We derived expressions for the amplification in bias showing that this depends on the strength of spatial dependence in the disturbances, dependent variable, and explanatory variables. In contrast, we show that using spatial econometric model specifications containing spatial lags of both the dependent and explanatory variables (that we labeled SDM and SLM) produces estimates whose bias matches the conventional omitted variables case. Our results provide a strong econometric motivation for using spatial econometric model specification such as the SDM and SLM in applied situations where the presence of omitted variables are suspected. The theoretical results presented here also confirm conjectures made by number of authors that omitted variables affect spatial regression methods less than OLS (Brasington and Hite 2005; Dubin 1988; Cressie 1993).

To summarize our findings from the standpoint of a practitioner, we make the following observations. If only the disturbances and explanatory variables exhibit spatial dependence and there is no omitted variable that is correlated with the included explanatory variable, OLS and spatial models should both yield similar regression parameter estimates for large data sets (Pace 1997). This theoretical result is interesting in light of empirical studies that continue to uncover examples where the spatial and OLS estimates differ materially in large samples. The differential sensitivity to omitted variable bias set forth here may account for these observed differences between least-squares and spatial regression estimates reported in applied work. For example, Lee and Pace (2005) examined retail sales and found that OLS estimates for the impact of store size on sales had a significant, negative effect while the spatial model produced a positive significant estimate. In addition, they found that spatial estimates reversed the sign of a number of other counterintuitive OLS parameter estimates. Similarly, Brasington and Hite (2005) in a model of demand for environmental quality found that OLS produced positive and insignificant estimates for the price of environmental quality, whereas a spatial lag model resulted in negative and significant estimates.

Finally, the method used here may aid in understanding other spatial model specifications such as the matrix exponential, conditional autoregressions, and moving average autoregressions in the presence of omitted variables and spatial dependence (LeSage and Pace 2007; LeSage and Pace 2009). Related work considers the issue of omitted variables in a spatial context using a combination of GMM and HAC estimation procedure applied to models involving right-hand-side endogenous variables (Fingelton and Le Gallo 2009).

Acknowledgements Kelley Pace would like to acknowledge support from NSF SES-0729259 and from the Louisiana Sea grant program, while James LeSage is grateful for support from NSF SES-0729264 and the Texas Sea grant program. Both authors would like to thank Shuang Zhu and David Brasington for helpful comments on earlier versions of this chapter.

References

- Anselin L (1988) *Spatial econometrics: methods and models*. Kluwer, Dordrecht
- Barry R, Pace RK (1999) A Monte Carlo estimator of the log determinant of large sparse matrices. *Lin Algebra Appl* 289:41–54
- Brasington DM, Hite D (2005) Demand for environmental quality: a spatial hedonic analysis. *Reg Sci Urban Econ* 35:57–82
- Cressie N (1993) *Statistics for spatial data*, Revised edition. Wiley, New York
- Dubin R (1988) Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Rev Econ Stat* 70:466–474
- Fingelton B, Le Gallo J (2009) Endogeneity in a spatial context: properties of estimators. In: Páez A, Le Gallo J, Buliung R, Dall’Erba S (eds) *Progress in spatial analysis: methods and applications*. Springer, Berlin: 59–73
- Girard DA (1989) A fast Monte Carlo cross-validation procedure for large least squares problems with noisy data. *Numer Math* 56:1–23
- Gradshteyn IS, Ryzhik IM (1980) *Table of integrals, series, and products*. Corrected and enlarged edition, Academic, Orlando
- Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J R Estate Finance Econ* 17:99–121
- Kelejian HH, Prucha IR (2007) The relative efficiencies of various predictors in spatial econometric models containing spatial lags. *Reg Sci Urban Econ* 37:363–374
- Lee LF (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* 72:1899–1926
- Lee ML, Pace RK (2005) Spatial distribution of retail sales. *J R Estate Finance Econ* 31:53–69
- LeSage JP (1999) *Spatial econometrics*. Unpublished manuscript available at www.spatial-econometrics.com. Sept. 2009
- LeSage JP, Pace RK (2007) A matrix exponential spatial specification. *J Econom* 140:190–214
- LeSage JP, Pace RK (2009) *Introduction to spatial econometrics*. Taylor & Francis, New York, FL
- Mardia KV, Marshall RJ (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71:135–146
- Pace RK (1997) Performing large-scale spatial autoregressions. *Econ Lett* 54:283–291