ORIGINAL ARTICLE

Separable spatial modeling of spillovers and disturbances

R. Kelley Pace · Shuang Zhu

Abstract The single spatial parameter in the spatial autoregressive model affects both the estimation of spillovers and the estimation of spatial disturbances. Consequently, the spatial autoregressive model has the undesirable property that if the degree of spatial dependence in the disturbances differs from that in the spillovers, neither may be estimated correctly. We show theoretically that the dependence structure for the spillovers and disturbances can differ and conduct a Monte Carlo experiment that verifies these findings. In contrast, estimates from a simple separable model show little bias in all the scenarios. We also show differences between the spatial autoregressive model and the separable model on five empirical examples.

Keywords Spatial autoregressive model · Separable · Spatial autoregressive lag · Matrix exponential

JEL Classification C01 · C13 · C18 · C21

1 Introduction

In terms of applications, the spatial error model (autoregressive disturbances) and the spatial autoregressive model (autoregressive regressand) have been the most

R. K. Pace (✉)
LREC Endowed Chair of Real Estate, Department of Finance,
E.J. Ourso College of Business Administration, Louisiana State University,
Baton Rouge, LA 70803-6308, USA
e-mail: kelley@spatial.us

S. Zhu
Department of Finance, E.J. Ourso College of Business Administration,
Louisiana State University, Baton Rouge, LA 70803-6308, USA

widely used spatial econometric models. Nonetheless, both models have desirable and undesirable features.

The spatial error model (SEM) possesses the desirable feature that incorrect specification of spatial dependence of the disturbances does not create bias in the estimation of the regression parameters. However, the traditional form of the SEM does not provide any information about positive or negative spatial externalities (spillovers). To the degree that quantifying spillovers are necessary when dealing with spatial data and often are the objective of interest (Elhorst 2010), the inability of the traditional form of the spatial error model to handle spillovers reduces its appeal.

In contrast, the spatial autoregressive model (SAR) natively produces estimates of spillovers, but the one parameter in the model affects both the estimation of spillovers and the estimation of spatial disturbances. Consequently the spatial autoregressive model has the undesirable property that if the degree of spatial dependence in the disturbances differs from that in the spillovers, neither may be estimated correctly. This has implications concerning the estimated magnitude of spillovers from the SAR model. For example, disturbances that exhibit a high degree of spatial dependence could lead to the SAR model overestimating the magnitude of positive or negative externalities. Cressie (1988, p. 443) in his discussion of the spatial autoregressive model succinctly summarized this problem as "confounding large- and small-scale effects."

In this manuscript, we provide a theoretical development of why spatial dependence may differ in the spillovers and the disturbances. One mechanism leading to such a schism is that the spatial autoregressive model DGP emerges from the long-run equilibrium of a spatiotemporal process where the underlying innovations at each location are the same over time (persistent). However, an assumption of independent innovations over time leads to a different long-run equilibrium with a different specification of the disturbances. This specification will typically result in lower estimates of spatial dependence in the disturbances.

As another consideration, omitted variables provide a commonly used justification for spatial dependence in the disturbances. If the omitted variable is not correlated with included variables and the omitted variable is spatially dependent, the disturbances will also be spatially dependent. As shown later with 2010 Census data, many economic explanatory variables exhibit high levels of autocorrelation with autoregressive parameter estimates between 0.85 and 0.95. If such variables were omitted, this could lead to a higher level of spatial dependence in the disturbances relative to those in the spillovers.

Putting the spatiotemporal development together with omitted variables leads to a long-run equilibrium where the disturbances can show either more or less spatial dependence than the spillovers. We conduct a Monte Carlo experiment that verifies these findings and shows that the spatial autoregressive model can perform badly in various scenarios where the spatial dependence in the disturbances differs from the spatial dependence in the spillovers. In contrast, estimates of a simple separable model (a model where $E(y)$ and the covariance involve totally separate parameters) show little bias in all the scenarios.

We also examine five empirical examples and show that the estimates of the autoregressive parameter from the spatial autoregressive model can materially differ

from the estimates of the autoregressive parameter associated with the spillovers provided by the separable model. Specifically, in three out of the five examples, the estimated autoregressive parameter differed materially between the SAR and separable estimators. However, in two examples, the estimates of the parameter governing the spillover and the estimates of the spatial dependence of the disturbances almost matched. However, in every example, the estimated spillovers from the separable model were less than the spatial autoregressive model. This has implications for applied practice.

Finally, we examine some possible alternatives to using the spatial autoregressive model. In particular, we look at the spatial Durbin error model (SDEM) as proposed by LeSage and Pace (2009, pp. 41–42). The SDEM has the robustness of the spatial error model to incorrect specification of the disturbances, a simple and correct means of measuring direct effects (own partial derivatives) and indirect effects (cross-derivatives or spillovers), as well as the ability to have direct and indirect effects with either the same or the different signs. We also discuss the spatial Durbin model as well as using various lag specifications such as the Koyck, matrix exponential, Shiller, and Almon in spatial models.

In terms of the organization of the paper, in Sect. 2, we review the common spatial autoregressive and spatial error models to help motivate a simple separable spatial model introduced in Sect. 3 Moreover, in Sect. 3, we develop some plausible data generating processes that yield different levels of spatial dependence in the spillovers and disturbances. In Sect. 4, we provide Monte Carlo and empirical evidence showing the need for separable spatial models. In Sect. 5, we set forth alternative separable spatial specifications, and Sect. 6 summarizes some key findings and discusses some implications of this research.

2 Conventional spatial models

A wide variety of models have been used in spatial econometrics, but the two most commonly employed are the spatial autoregressive model (SAR) and the spatial error model (SEM). This section briefly reviews these conventional models in Sects. 2.1 and 2.2 to set up the separable models introduced in Sect. 3.

2.1 Spatial autoregressive dependent variable model

We begin with the SAR model which has the estimation form given in (1) and the data generating process (DGP) form given in (2) and (3). In these equations, y represents a vector containing n observations on the regressand, X represents a matrix containing n observations on k exogenous explanatory variables (typically a column contains a constant vector and the other p columns contain non-constant vectors), the vector ε contains n normal *iid* variates, and the scalar parameter ρ captures spatial dependence associated with neighboring observations as specified by the n by n spatial weight matrix W. The spatial weight matrix W contains positive elements in W_{ij} if observation j affects observation i (j is a neighbor to i) and 0 otherwise. By convention, observations do not directly affect themselves and so $W_{ii} = 0$. All the elements of W are exogenous

non-negative scalars. For simplicity, we assume that W is symmetric ($W_{ij} = W_{ji}$) and doubly stochastic so that each row and column sum to one. In this case, W will have all real eigenvalues with the principal eigenvalue of one. We assume, as a sufficient condition, that $\rho \in (-1, 1)$ to ensure that $(I_n - \rho W)^{-1}$ exists.

$$y = X\beta + \rho W y + \varepsilon \tag{1}$$

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1} \varepsilon \tag{2}$$

$$\varepsilon \sim N(0, \sigma^2 I_n) \tag{3}$$

The expected value of y for the SAR model appears in (4) while the covariance of the SAR disturbances appears in (5).

$$E(y) = (I_n - \rho W)^{-1} X\beta \tag{4}$$

$$\Omega_{\text{SAR}} = \sigma^2 (I_n - \rho W)^{-2} \tag{5}$$

A feature of the SAR model is that a change in x_{ir} for any non-constant variable $r = 2 \ldots k$ for any observation $i = 1 \ldots n$ affects all the elements of y. Specifically,

$$\frac{\partial E(y)}{\partial X_r'} = S_r(W) = (I_n - \rho W)^{-1} \beta_r \tag{6}$$

The own partials are found on the diagonal of $S_r(W)$ and the cross-partials are found on the off-diagonals of $S_r(W)$. The cross-partials are the marginal effect of changing the jth observation's value of the rth variable x_{jr} on $E(y_i)$ for $i \neq j$. These represent the spillovers and estimating these is of great interest in many economic contexts (Elhorst 2010). For example, how does reducing crime at business j affect the insurance costs for business i? If reducing crime through mitigation effects such as security lighting and cameras for a business leads to positive spillovers, the full benefits of such improvements may not be realized by this business and such efforts will be underutilized relative to a scenario where a business could fully capture the positive externalities stemming from their investments.

As another perspective on the SAR model, (4) can be rewritten as a varying parameter model (where α is an intercept parameter and ι_n is a n by 1 vector of ones). As shown in (7), variables do not have a constant impact over space, but instead have a more local interpretation (Pace and LeSage 2010). Thus, in a global model $S_r(W) \propto I_n$, but in (7) $S_r(W) \propto (I - \rho W)^{-1}$ and this is not constant on the diagonal.

$$E(y) = \alpha \iota_n + S_1(W)X_1 + S_2(W)X_2 + \cdots + S_k(W)X_k \tag{7}$$

An undesirable aspect of the SAR model, however, is that $E(y)$ in (4) is a function of ρ and the covariance in (5) is also a function of ρ. Therefore, misspecification of either part potentially contaminates the estimation of the other part. In other words, estimation of $E(y)$ and the covariances are not separable.

To see this in more detail, assume the DGP in (8) and (9) where ρ_1 governs the spatial dependence in the spillovers and ρ_2 governs the spatial dependence in the disturbances. The process in (10) leads to the estimation model in (8) and (11).

$$y = (I_n - \rho_1 W)^{-1} X\beta + (I_n - \rho_2 W)^{-1} \varepsilon \tag{8}$$

$$\varepsilon \sim N(0, \sigma^2 I_n) \tag{9}$$

$$y = AX\beta + \rho_2 Wy + \varepsilon \tag{10}$$

$$A = (I_n - \rho_2 W)(I_n - \rho_1 W)^{-1} = \sum_{q=0}^{\infty} \pi_q W^q \tag{11}$$

The matrix A will only equal I_n and the estimation model will only be the same as SAR when ρ_1 is the same as ρ_2. When ρ_1 diverges from ρ_2, use of the SAR estimation model omits the various powers of W^j ($q = 1 \ldots \infty$) implied by the series expansion of A (where π_q are the coefficients associated with the expansion) and this misspecification leads to bias in the estimated parameters.

2.2 Spatial error model

A version of the spatial error model appears in (12) with the disturbances specified in a more general way such that the disturbances involves a matrix function $F(W, \lambda)$ of the spatial weight matrix and a scalar parameter λ. The most common autoregressive error model comes from the assumption that $F(W) = (I_n - \lambda W)^{-1}$ where λ is a scalar parameter in (as a sufficient condition) the interval $(-1, 1)$. For any specification of $F(W, \lambda)$, the expected value of y equals $X\beta$ as shown in (13),

$$y = X\beta + F(W, \lambda)\varepsilon \tag{12}$$

$$E(y) = X\beta \tag{13}$$

This desirable insensitivity of the error model expectation to specification of the error dependence continues even when $F(W, \lambda) = I_n$, which leads to OLS. Other examples of $F(W, \lambda)$ include the matrix exponential ($F(W, \lambda) = e^{\lambda W}$) or moving average ($F(W, \lambda) = I_n + \lambda W$). Pace and LeSage (2008) used the unbiasedness of the error model in devising a spatial Hausman test for misspecification as all error models specifications should have similar estimates for β in the presence of a correct specification of the regression part of the model. Statistically significant deviations in the estimated regression parameters point to misspecification.

An aspect of the error model is that it does not natively allow for modeling of spillovers and is a global model where a change in an explanatory variable has the same impact at each location. In other words, the derivatives for each variable are the same globally as shown in (14) and (15).

$$\frac{\partial y_i}{\partial x_{ir}} = \beta_r \quad \text{global} \tag{14}$$

$$\frac{\partial y_i}{\partial x_{jr}} = 0 (i \neq j) \quad \text{no spillovers} \tag{15}$$

3 Separable models

Is there a way to obtain the spatial autoregressive model's spillover estimation along with the spatial error model's robustness to specification of the disturbances?

An obvious generalization of the two models to obtain both spillover and robustness properties would be to fit (16) using maximum likelihood or Bayesian methods,

$$y = (I_n - \rho W)^{-1} X\beta + F(W, \lambda)\varepsilon \tag{16}$$

where λ represents a scalar parameter. As long as ρ is not a parameter in $F(W, \lambda)$, this model has separable modeling of the spillovers and the disturbances.

The simplest separable specification would be to use $F(W, \lambda) = I_n$ which would lead to (17), a single parameter model that could be fit using non-linear least squares.

$$y = (I_n - \rho W)^{-1} X\beta + \varepsilon \tag{17}$$

If the disturbances were spatially dependent, estimation of (17) could be quite inefficient and that by itself could lead to indeterminant inferences. However, this might not be a factor in large sample sizes. In addition, misspecification of the error covariance could lead to incorrect standard errors.

Naturally, other models for spatial dependence such as the autoregressive, moving average, or matrix exponential could be used for $F(W, \lambda)$ and these should show greater efficiency and less biased inference. In addition, we discuss in Sect. 5 some more elaborate separable models. However, the focus here is not on estimation efficiency, but on the biases that can occur due to incorrect specification of the covariance structure in non-separable models. In this regard, the simple *iid* error model in (17) provides the simplest model that illustrates the benefits of separable modeling. Since $F(W, \lambda) = I_n$ is a dramatic misspecification of the covariance matrices in the various Monte Carlo and empirical examples, the ability of (17) to perform well demonstrates the advantages of separable modeling.

Having introduced the idea of separable modeling, we now turn to possible motivations of such separable models. Viewing spatial dependence in cross-sectional data as the equilibrium of a spatiotemporal process seems the most natural way to motivate simultaneous spatial dependence. Section 3.1 develops some possible data generating processes that stem from a spatiotemporal perspective.

Another natural way of motivating spatial dependence in cross-sectional data comes from the omission of variables in models. Section 3.2 introduces an omitted variable in the spatiotemporal process and shows additional plausible DGPs.

3.1 Spatiotemporal aspects of the DGP

Simultaneous spatial dependence arises whenever for some variable location i affects location j and vice versa. Although this could happen instantaneously, it seems more natural to assume that the interaction between the i and j locations happens over time. The temporal autoregressive specification is the most widely used model in time series analysis. The spatiotemporal equivalent in (18) and (19) allows the current value of a variable to depend on past values of the variable at its own location as well as neighboring locations. In (18) and (19), y_t represents a vector containing n observations on the regressand at period t, X represents the matrix containing n observations on k exogenous that are constant over time, the

vector of innovations ε_t contains n normal *iid* variates, the scalar parameter τ captures own observation time dependence, and the scalar parameter ρ_{st} captures time dependence associated with neighboring observations as specified by the n by n spatial weight matrix W. Despite the presence of t in the subscript, the parameter ρ_{st} does not vary over time, but serves as a reminder that it is a parameter in the spatiotemporal DGP.

$$y_t = Gy_{t-1} + X\beta + \varepsilon_t \tag{18}$$

$$G = \tau I_n + \rho_{\mathrm{st}} W \tag{19}$$

Naturally, one could make this spatiotemporal DGP more flexible. For example, X could vary over time (X_t). However, the additional complexity would not change the fundamental point that different assumptions about the spatiotemporal process can lead to a divergence in the forms associated with the spillovers and the disturbances in the resulting long-run, cross-sectional equilibrium.

Recursive substitutions of lagged values of (18) (i.e., $y_{t-1} = Gy_{t-2} + X\beta + \varepsilon_{t-1}$) as in Elhorst (2001) as well as LeSage and Pace (2009) lead to the state of the dynamic system after t periods in (20) and (21).

$$y_t = \left(I_n + G + G^2 + \cdots + G^{t-1}\right)X\beta + G^t y_0 + v \tag{20}$$

$$v = \varepsilon_t + G\varepsilon_{t-1} + G^2\varepsilon_{t-2} + \cdots + G^{t-1}\varepsilon_1 \tag{21}$$

If t is sufficiently far in the future to ensure convergence of the spatiotemporal process, $G^t y_0 \approx 0$ and $(I_n - G)^{-1} \approx I_n + G + G^2 + \cdots + G^{t-1}$. In this case, (22) will describe the cross-sectional values at y_t. To avoid tedium, we will treat the approximation from now on as an equality. Taking the expectation of (22) yields (23) which shows the expected value of the cross-sectional values at y_t.

$$y_t = (I_n - G)^{-1} X\beta + v \tag{22}$$

$$E(y_t) = (I_n - G)^{-1} X\beta \tag{23}$$

Expressing (23) in terms of the original temporal parameter τ and spatiotemporal parameter ρ_{st} yields (24) where the cross-sectional spatial parameter ρ_s is defined in (25). Equation (25) is important because it describes the relation between the spatial dependence in the spatiotemporal process ρ_{st} relative to ρ_s, the cross-sectional spatial dependence in the long-term equilibrium of the spatiotemporal process. Subject to the stability restrictions, it shows that a high level of spatial dependence in a cross-sectional sample can come from either small amounts of spatiotemporal dependence (low ρ_{st}) in the presence of strong temporal dependence (high τ) or large amounts of spatiotemporal dependence (high ρ_{st}) in the presence of weak temporal dependence (low τ). This means, for example, that running a spatiotemporal regression and finding small ρ_{st} and high τ is perfectly compatible with running a spatial cross-sectional regression and finding large levels of spatial dependence (ρ_s).

$$E(y_t) = (I_n - \rho_s W)^{-1} X \left[\frac{\beta}{1-\tau}\right] \tag{24}$$

$$\rho_s = \frac{\rho_{st}}{1 - \tau} \tag{25}$$

Turning attention to the distribution of the disturbances v from (21) leads to at least two possibilities. First, suppose that the innovations were persistent or permanent so that $\varepsilon_t = \varepsilon$ for all t. In that case, $v = (I_n - G)^{-1}\varepsilon$ and the disturbances would be distributed as $N(0, \Omega_{SAR})$ as in (26).

$$\Omega_{SAR} = E(vv') = \sigma^2(I_n - G)^{-2} \tag{26}$$

The importance of (26) is that it provides a rationale for using the SAR model since (22) coupled with (26) describe the DGP for the SAR model (c.f., (4)–(5)).

An alternative assumption is that the innovations ε_t are *iid* over space and over time so that $E(\varepsilon_{it}\varepsilon_{jq}) = 0$ as long as the subscripts it and jq are not the same. In that case, as LeSage and Pace (2009, p. 198) show that the disturbances would be distributed as $N(0, \Omega_{iid})$ as in (26).

$$\Omega_{iid} = E(vv') = \sigma^2(I_n - G)^{-1}(I_n + G)^{-1} \tag{27}$$

The importance of the alternative specification in (27) is that the DGP defined by (22) coupled with (27) does not equal the SAR model and using the SAR model in conjunction with this DGP could result in inconsistent estimates. Both are plausible, but mutually exclusive, DGPs. Note, the cancelation of some of the spatial terms makes Ω_{iid} less singular than Ω_{SAR}.

3.2 Omitted variables and the DGP

Omitted variables constitute another mechanism that can produce spatial dependence. Suppose a variable z_t follows the autoregressive spatiotemporal process in (28) and (29) where δ is the level of temporal dependence and ϕ_{st} is the level of spatiotemporal dependence. Also, suppose that s in (28), which is persistent or permanent, is independent of ε_t.[1] Following repeated recursive substitutions and simplifications yields (30).

$$z_t = Hz_{t-1} + s \tag{28}$$

$$H = \delta I_n + \phi_{st}W \tag{29}$$

$$z_t = (I_n - H)^{-1}s \tag{30}$$

This leads to a spatial equilibrium for $E(z_t)$ in (31) and (32).

$$E(z_t) = (I_n - \phi_s W)^{-1}s\left[\frac{\beta}{1 - \delta}\right] \tag{31}$$

$$\phi_s = \frac{\phi_{st}}{1 - \delta} \tag{32}$$

What are levels of ϕ_s found for typical explanatory variables? Fitting a SAR model with an intercept, but no other regressors to some common explanatory

---

[1] To maintain simplicity, we keep s as permanent and do not explore letting s be independent over time and space as that would create more scenarios to examine. However, this could certainly be done.

Table 1 Spatial autocorrelation of selected variables from census 2010

Variable	Census tract	Block group
White	0.919	0.946
Black (%)	0.910	0.943
Hispanic (%)	0.894	0.902
Age 18+	0.872	0.865
Land	0.953	0.950
Occupied unit (%)	0.906	0.916
n	1,112	3,360

variables at the tract and block group geography out of the 2010 Census for Louisiana yields the results shown in Table 1.[2]

Table 1 documents that many explanatory variables manifest high levels of spatial dependence (ϕ_s between 0.85 and 0.95). Consequently, missing variables could display high levels of dependence as well.

Given some idea of the magnitude of spatial dependence for z_t from Table 1, suppose that z_t is truly part of the DGP for y_t as shown in (33),

$$y_t = Gy_{t-1} + X\beta + z_t + \varepsilon_t \tag{33}$$

where (33) is an extension of (18) and (19).

Recursive substitutions of lagged values of (33) (i.e., $y_{t-1} = Gy_{t-2} + X\beta + z_{t-1} + \varepsilon_{t-1}$) lead to the state of the dynamic system after t periods in (34), (35), and (36).

$$y_t = \left(I_n + G + G^2 + \cdots + G^{t-1}\right)X\beta + G^t y_0 + u + v \tag{34}$$

$$u = z_t + Gz_{t-1} + G^2 z_{t-2} + \cdots + G^{t-1} z_1 \tag{35}$$

$$v = \varepsilon_t + G\varepsilon_{t-1} + G^2 \varepsilon_{t-2} + \cdots + G^{t-1} \varepsilon_1 \tag{36}$$

Since u and v are independent of each other, $E((u+v)(u+v)') = E(uu') + E(vv')$. Previously, (26) and (27) gave the form for $E(vv')$ for the case of persistent and *iid* disturbances over time and $E(uu')$ appears in (37),

$$E(uu') = \sigma_z^2 (I_n - G)^{-2}(I_n - H)^{-2} \tag{37}$$

where the order of H and G does not matter here since both are functions of the same symmetric W. The scalar constant σ_z^2 is the variance of s from (28). Note, the covariance matrix for the disturbance term related to the omitted variable is more singular or spatially dependent than the covariance matrix for the disturbance term related to the other sources of error. This happens since the omitted variable term combines omitting a spatially dependent variable, and this is further amplified by the usual mechanism caused by a lagged dependent variable. This suggests that disturbance terms could have an even more complicated spatial structure than spillovers.

---

[2] We used Louisiana because it is one of the first states with data releases in the 2010 Census and because more effort than normal was spent in data collection due to concerns about population loss after Katrina.

Therefore, the overall disturbances follow either a $N(0, \Omega_{o\text{SAR}})$ or $N(0, \Omega_{oiid})$ distribution,

$$\Omega_{o\text{SAR}} = \sigma^2 (I_n - G)^{-2} + \sigma_z^2 (I_n - G)^{-2} (I_n - H)^{-2} \tag{38}$$

$$\Omega_{oiid} = \sigma^2 (I_n - G)^{-1} (I_n + G)^{-1} + \sigma_z^2 (I_n - G)^{-2} (I_n - H)^{-2} \tag{39}$$

and $E(y)$ is still given by (23).

In other words, whether a spatiotemporal DGP involves *iid* innovations (innovations over space and time) or permanent innovations (innovations over space, but not time) and various levels of omitted variables that are independent of included variables, the resulting DGP can have SAR type spillovers in the mean as given by (23) and disturbances that materially differ.

Various components of, for example, $\Omega_{o\text{SAR}}$ display lesser or greater singularity. Although one could define singularity in various ways, the norm of the variance–covariance matrix provides one measure of singularity. Specifically, a large norm would indicate a more singular variance–covariance matrix. In terms of norms, the maximum absolute row sum norm of the variance–covariance matrix is a convenient method when using a row-stochastic W. For example, this norm for $\sigma^2 (I_n - G)^{-2}$ is $\sigma^2 (I_n - \tau - \rho_{\text{st}})^{-2}$. If $\tau + \rho_{\text{st}}$ approach one, this norm becomes quite large. Similarly, the norm associated with $\sigma_z^2 (I_n - G)^{-2} (I_n - H)^{-2}$ is $\sigma_z^2 (I_n - \tau - \rho_{\text{st}})^{-2}$ $(I_n - \delta - \phi_{\text{st}})^{-2}$. The part affected by σ^2 displays less singularity than the part affected by σ_z^2. Therefore, the overall singularity the various variance–covariance matrices will likely depend on the relative magnitudes of σ^2 and σ_z^2.

Note, one could imagine that a number of omitted variables exist for any given problem and that these omitted variables may take on various magnitudes and levels of spatial dependence. Aggregation of such random variables could lead to a long memory process such as discussed by LeSage and Pace (2009) in conjunction with their fractional differencing estimator.

4 Monte Carlo and empirical support for separable specifications

Section 4.1 examines the effects of these various DGPs on the SAR and separable estimators via a Monte Carlo experiment while Sect. 4.2 estimates separable and conventional SAR models for five models that use different data and/or specifications and shows that in some situations the separable specification provides materially different results than the conventional models.

4.1 Monte Carlo performance

To obtain some idea of the performance of the separable estimator (17) versus the SAR estimator, we ran a simple Monte Carlo experiment. We generated y_t using (33) with assumptions on the various parameter values. Specifically, we set $\rho_{\text{st}} = 0.6$, $\tau = 0.25$, $\delta = 0$, and $\phi_s = 0.9$. The level of $\phi_s = 0.9$ is motivated by Table 1. This meant that the equilibrium cross-sectional spatial dependence

$\rho_s = 0.6/(1 - 0.25) = 0.8$. The equilibrium spatial dependence for z_t was 0.9. The explanatory variable matrix X had two columns. The first column was a constant and the second was an *iid* random uniform and each had a $\beta = 1$. The symmetric, doubly stochastic spatial weight matrix was based on contiguity. We simulated 500 periods each with 1,000,000 observations. Given the large sample size of 1,000,000 observations, the reported results are from one draw for each case. However, we repeated the experiment several times and found little variation in these results. The large number of periods guaranteed convergence (for these parameter values) and the large number of observations per cross-section reduce the effects of efficiency on estimation so that this exercise focuses on the demonstration of bias.

In terms of the simulation, we looked at eight scenarios differing by the level of noise in ε (σ_ε), the importance of the omitted variable σ_z, and whether disturbances over time were permanent (the same over time) or whether these were *iid* over time (*iid*).

We estimated the cross-sectional models using the separable approach (ρ_{SEP}) in (40) and the SAR model (ρ_{SAR}) in (41). For the SAR model, we employed the traditional maximum likelihood procedure. For the separable approach, we fit (40) using non-linear least squares so that the estimate $\hat{\rho}_{\text{SEP}}$ was the value of ρ_{SEP} that minimized the sum-of-squared residuals. Note, y is not transformed in the separable approach, and this avoids the need for a determinant term.

$$y = (I_n - \rho_{\text{SEP}} W)^{-1} X\beta + e_{\text{SEP}} \tag{40}$$

$$y = \rho_{\text{SAR}} Wy + X\beta + e_{\text{SAR}} \tag{41}$$

Table 2 shows the results of the experiment. In cases where the DGP was correct (cases 5 and 7), the SAR estimator returned 0.800 and 0.800 (true value 0.8). However, for some of the other scenarios, the SAR estimator performed poorly. Specifically, for cases 2 and 6 which had a large omitted variable component, the SAR estimator returned estimates of 0.976. On the other hand, for the *iid* errors with a large standard deviation (without the omitted variable), the SAR estimate of 0.391 was very low.

In contrast the separable estimator performed well in terms of estimating ρ with a maximum deviation of -0.006 across all eight cases, even without any modeling of the spatial disturbances (variance–covariance matrix of $\sigma^2 I_n$).

The problem for the SAR estimator is that it leans on correct specification of the disturbances to arrive at its estimates of dependence for both spillovers and disturbances and when this is misspecified, it can perform poorly. All else equal, *iid* disturbances over time have less singular spatial dependence than those from temporally persistent disturbances, and this tends to lead the SAR estimator to underestimate ρ. On the other hand, the high spatial dependence in the omitted variable leads the SAR model to overestimate ρ. Combining *iid* disturbances and omitted variables means that the SAR model has the potential for biases of unknown direction.

Table 2 Performance of the separable and SAR estimators

	σ_ε	σ_z	$\hat{\rho}_{\text{SEP}}$	$\tilde{\rho}_{\text{SAR}}$
iid	0.1000	0.0000	0.8000	0.7720
iid	0.1000	0.1000	0.7970	0.9760
iid	1.0000	0.0000	0.8010	0.3910
iid	1.0000	0.1000	0.7940	0.6530
Persistent	0.1000	0.0000	0.8000	0.8000
Persistent	0.1000	0.1000	0.7970	0.9760
Persistent	1.0000	0.0000	0.8030	0.8000
Persistent	1.0000	0.1000	0.8000	0.8350

4.2 Empirical examples

To see whether any difference exists between the estimated spillover parameter ρ_{SEP} from the separable model in (17) and the dependence parameter in the SAR model ρ_{SAR}, we estimated five models that use different data and/or specifications. First, we estimated the election model in Pace and Barry (1997) which uses 3,107 county level observations on voter participation as a function of number of voters, education, homeownership, and income. Second, we estimated tract level housing prices ($n = 62{,}226$) in year 2000 as a function of 1990 house age, number of households, median income, median years of education, and tract size. Third, we estimated the same housing model augmented with the level of house prices in 1990. Fourth, we estimated median incomes in year 2000 as a function of 1990 number of households, median years of education, and tract size. Fifth, we estimated the same income model augmented with the level of median incomes in 1990. All the variables were logged.

Table 3 shows the results from estimating the separable and the SAR model. In all five cases, the separable parameter estimate ρ_{SEP} was less than the parameter estimate of the SAR model ρ_{SAR}. Specifically, in three out of the five examples (election and housing examples), the separable parameter estimate ρ_{SEP} was materially less than the parameter estimate of the SAR model ρ_{SAR}. However, for the two income examples, the estimated parameters were similar in magnitude. Therefore, use of a separable model has the potential to make a difference in empirical work.

Table 3 Separable and SAR estimates on empirical data

	Empirical example	$\hat{\rho}_{\text{SEP}}$	$\tilde{\rho}_{\text{SAR}}$
Election	1	0.18	0.53
Housing sans temporal lag	2	0.66	0.78
Housing with temporal lag	3	0.31	0.54
Income sans temporal lag	4	0.75	0.78
Income with temporal lag	5	0.32	0.36

5 Alternative separable specifications

Although it is possible to use the separable estimator introduced earlier in (17), this estimator suffers from low efficiency in the presence of spatially dependent disturbances. Like the SAR model, it suffers from an implicit restriction that the direct effects (own partials) have the same signs as the indirect effects (cross-partials). This is an undesirable feature of the SAR model (Elhorst 2010). Also, it requires some care in the interpretation of the estimates as the β parameters are often incorrectly treated like those from OLS by many applied researchers (LeSage and Pace (2009).

For these reasons, the Spatial Durbin Error Model (SDEM), introduced by LeSage and Pace (2009, pp. 41–42), provides a simpler and more useful estimator that has separable spillover and disturbance modeling. The DGP for the SDEM appears in (42) and (43) while (44) provides a form suitable for estimation.

$$y = \iota_n \alpha + X\beta + WX\theta + (I_n - \lambda W)^{-1}\varepsilon \qquad (42)$$

$$\varepsilon \sim N(0, \sigma^2 I_n) \qquad (43)$$

$$(I_n - \lambda W)y = \iota_n \alpha^* + (I_n - \lambda W)[X\beta + WX\theta] + \varepsilon \qquad (44)$$

In (42), α and α^* are intercept parameters, ι_n is an n element column vector of ones, λ is the separable dependence parameter governing the spatial dependence among disturbances, and X contains only non-constant variables. The expected value of y appears in (45) and the covariance matrix of the disturbances Ω_{SDEM} appears in (46).

$$E(y) = \iota_n \alpha + X\beta + WX\theta \qquad (45)$$

$$\Omega_{\text{SDEM}} = \sigma^2 (I_n - \lambda W)^{-2} \qquad (46)$$

The SDEM has the very attractive property that β measures the direct effect and θ measures the indirect effect of a change in X_r for $r = 1 \ldots p$ (where p is the number of non-constant columns in X) as in (47) and (48). Also, β_r and θ_r can have the same or differing signs (unlike the SAR model). In fact, the SAR model not only has indirect effects that are the same signs as the direct effects, but the magnitude of the indirect effects to the direct effects are always a constant ratio across variables $(S_r(W) \propto (I_n - \rho W)^{-1})$. The SDEM does not share this inflexibility. However, the SDEM is a global model and does not have the varying parameter interpretation of the SAR model.

$$\frac{\partial E(y_i)}{\partial x_{ir}} = \beta_r \qquad (47)$$

$$\frac{\partial E(y_i)}{\partial x_{jr}} = W_{ij}\theta_r \quad (i \neq j) \qquad (48)$$

Estimation of the SDEM can proceed using the usual error model routines by maximum likelihood or Bayesian methods. Of course, it also has the attractive separable property and so mistakes in specification of the disturbances do not create bias in the estimation of the direct effect and indirect effect (spillovers). Naturally,

mistakes in the specification of the disturbances lower the efficiency of the estimator and create potential bias in the standard errors. However, with larger sample sizes neither of these drawbacks is as important as reducing bias in the estimation of the regression parameters β and θ.

Another possibility is to use the spatial Durbin model (SDM) where (49) and (50) define the data generating process, (51) is the expectation of the DGP, and (52) defines an estimation form.

$$y = \iota_n\alpha + (I_n - \rho W)^{-1}[X\beta + WX\theta] + (I_n - \rho W)^{-1}\varepsilon \qquad (49)$$

$$\varepsilon \sim N(0, \sigma^2 I_n) \qquad (50)$$

$$E(y) = \iota_n\alpha + (I_n - \rho W)^{-1}[X\beta + WX\theta] \qquad (51)$$

$$y = \iota_n\alpha^* + X\beta + WX\theta + \rho Wy + \varepsilon \qquad (52)$$

Although $E(y)$ depends on parameters that are used in modeling the disturbances, since the inverse of that form appears in the model through X and WX, the SDM can successfully handle some types of misspecification in the disturbances. For example, the SDM subsumes the spatial error model with spatial autoregressive disturbances. Thus, if the DGP is the SEM, then the SDM will produce (for large enough n) estimates of $\theta = -\rho\beta$ which will allow the model part to have the conditional mean of $X\beta$. Therefore, the SDM can handle the conditional mean part differently than the disturbance part of the model.

An extended SDM that includes higher order variables $W^q X$ for $q > 1$ can handle even more complicated spillovers without being affected by misspecification of the disturbances. LeSage and Pace (2009) discuss extensively some of the properties of the SDM and the extended SDM.

Another alternative separable specification would be to fit a model with a geometric lag (Koyck 1954) term $W(I_n - \rho_K W)^{-1}X$ as in (53) using any desired error specification for d.

$$y = X\beta + W(I_n - \rho_K W)^{-1}X\theta + d \qquad (53)$$

If $\theta = \beta$, the regression part of the model would match the SAR specification of the regression part of the model. However, unlike the SAR model, this allows for the variable and its lags to differ in signs and magnitudes and hence allows for the possibility of having direct effects (own partials) of opposite signs as the indirect effects (cross-partials).

Alternatively, one could implement a separable model with a matrix exponential lag term as in (54),

$$y = X\beta + We^{\rho_e W}X\theta + d \qquad (54)$$

where ρ_e is a scalar real parameter. LeSage and Pace (2007) as well as LeSage and Pace (2009) discussed the properties of matrix exponentials.

Naturally, one could look at other lag specifications such as those from Almon (1965) or Shiller (1973) within the context of an error model. The regression model part would focus on accurately estimating spillovers and the disturbance part would focus on maximizing efficiency and validity of inference.

As mentioned above, one could use maximum likelihood or Bayesian methods to estimate these models. However, in the case where there are multiple parameters in the concentrated log-likelihood, residual maximum likelihood (REML) which matches the marginal likelihood may perform better (Pace et al. 2010, pp. 29–30). Another alternative is for any model parameterized by λ to treat each level or choice of λ as an individual model and to perform Bayesian or frequentist model averaging over the individual models. See LeSage and Pace (2009, pp. 173–184) and LeSage and Pace (2007) for some examples of model averaging in a spatial context.

6 Conclusion

The two most commonly used spatial econometrics in the literature, the spatial error model and the spatial autoregressive model, have complementary strengths and weaknesses. Specifically, estimation of the regression parameters in the spatial error model is robust to misspecification in spatial dependence, but does not natively produce estimates of spillovers while the spatial autoregressive model natively produces estimates of spillovers but is sensitive to misspecification in the spatial dependence in the disturbances. We propose a separable model that could natively produce estimates of spillovers while maintaining robustness to misspecification of the spatial dependence in the disturbances.

We show that plausible data generating processes based on spatiotemporal processes and omission of spatially dependent regressors can lead to a divergence in the spatial dependence in the spillovers and spatial dependence in the disturbances. We use a Monte Carlo experiment to show that the spatial autoregressive model is sensitive to misspecification of the disturbances while the separable model performs well in this setting. We also look at five empirical examples where the separable model indicates that the spatial dependence in the spillovers and disturbances differs.

We discussed some possible separable specifications. Specifically, we discussed the spatial Durbin error model (SDEM), the extended spatial Durbin model, and spatial error models that use richer specifications of spillovers based on geometric, matrix exponential, Almon, and Shiller spatial lags.

There are some aspects of separable models that we did not explore here, but may help in other contexts. For example, a possible benefit of the separable approach is that computationally it may be easier to estimate a model with dependence in the disturbances than in the dependent variable in some cases. For example, the probit approach with spatially structured random effects (tantamount to a spatial error term) using the approach of Smith and LeSage (2004) may be easier to implement and better computationally than spatial probit with a lagged latent dependent variable as in LeSage and Pace (2009).

As another example, the multiple spatial weight matrix problem (i.e., $W = a_1 W_1 + \cdots a_q W_q$), such as discussed in Pace and LeSage (2002), may be easier to address by just estimating the multiple W as spillovers and using some W of convenience to help with the disturbances. Specification of the multiple W in the spillovers and using a separate W for the disturbances avoids the need to compute

the log-determinant which poses a problem in the traditional multiple W model. In many applications, the W_1 and W_2 are mutually exclusive (border counties, interior counties). A simple SDEM model $y = X\beta + W_1\theta + W_2\xi + \varepsilon$ has an easy interpretation as β measures the direct effects, θ measures the indirect effects or spillovers stemming from W_1 (border counties), and ξ measures the indirect effects or spillovers from W_2 (interior counties).

Finally, even though spatial dependence in the regressand emerges as a natural outcome of spatiotemporal processes, philosophically some researchers prefer not to use the spatial autoregressive model. However, many of the same researchers will use the spatial error model. The separable model is still an error model with specification in terms of the exogenous explanatory variables that captures spillovers and so may be more appealing to this group.

References

Almon S (1965) The distributed lag between capital appropriations and expenditures. Econometrica 33(1):178–196

Cressie N (1993) Statistics for spatial data, revised edition. Wiley, New York

Elhorst JP (2001) Dynamic models in space and time. Geogr Anal 33(2):119–140

Elhorst JP (2010) Applied spatial econometrics: raising the bar. Spatial Econ Anal 5(1):9–28

Koyck LM (1954) Distributed lags and investment analysis. North-Holland

LeSage JP, Pace RK (2007) A matrix exponential spatial specification. J Econ 140(1):190–214

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press, UK

Pace RK, Barry RP (1997) Quick computation of spatial autoregressive estimators. Geogr Anal 29(3):232–246

Pace RK, LeSage JP (2002) Semiparametric maximum likelihood estimates of spatial dependence. Geogr Anal 34(1):76–90

Pace RK, LeSage JP (2008) A spatial Hausman test. Econ Lett 101(3):282–284

Pace RK, LeSage JP (2010) Spatial econometrics. Handbook of spatial statistics. CRC Press, UK

Pace RK, LeSage JP, Zhu S (2010) Spatial dependence in regressors and its effect on estimator performance (manuscript)

Shiller RJ (1973) A distributed lag estimator derived from smoothness priors. Econometrica 41(4):775–788

Smith TE, LeSage JP (2004) A Bayesian probit model with spatial dependencies. In: LeSage JP, Pace RK (eds) Advances in econometrics: vol 18: spatial and spatiotemporal econometrics. Elsevier, Oxford, pp 127–160