2. SPATIALLY LAGGED DEPENDENT VARIABLES

In this chapter, we describe a statistical model that incorporates spatial dependence explicitly by adding a "spatially lagged" dependent variable *y* on the right-hand side of the regression equation. This model goes by many different names. Anselin (1988) calls this the *spatial autoregressive* model, but this terminology is potentially confusing since the term *autoregressive* is used to denote quite different spatial models in the geostatistical literature. For simplicity, we will here call it the spatially lagged *y* model, since its main feature is the presence of a spatially lagged dependent variable among the covariates.

The spatially lagged *y* model is appropriate when we believe that the values of *y* in one unit *i* are directly influenced by the values of *y* found in *i*'s "neighbors." This influence is above and beyond other covariates specific to *i*. If we believe that *y* is not influenced directly by the value of *y* as such among neighbors but rather that there is some spatially clustered feature that influences the value of *y* for *i* and its neighbors but is omitted from the specification, we may consider an alternative model with spatially correlated errors, which we discuss subsequently. For the spatially lagged *y* model to be appropriate, the dependent variable *y* must be considered as a continuous variable. In this book, we do not examine the generally more complicated case of binary dependent variables. These are more complicated since they often do not have a closed-form solution and must be estimated with iterative techniques outside the range of this book (see Ward & Gleditsch, 2002).

2.1 Regression With Spatially Lagged Dependent Variables

To motivate and illustrate the spatially lagged *y* model, we return to our example of the distribution of democracy around the world. We have seen that the distribution of democracy displays spatial clustering in the sense that countries are more likely to have higher values on the POLITY democracy score if they are surrounded by countries that also have high levels of democracy. Although some of the clustering in democracy obviously could be due to spatial clustering in GDP per capita, which in turn is positively related to democracy, we have shown that the spatial clustering in the democracy data does not completely disappear when we condition on a country's level of GDP per capita. The assumption that the errors ϵ_i of a model treating democracy as a function of GDP per capita are independent

can easily be examined by testing for possible spatial dependence in the residuals from the regression; that is, $\hat{\epsilon}_i = (\hat{y}_i - y)$, using the Moran $\mathcal{I}$ correlation coefficient and our specified pattern of connectivities in the matrix **C**, where states in this case are defined as connected if they are within a 200-km distance threshold of one another. In this instance, we found strong evidence of residual spatial correlation. The Moran $\mathcal{I}$ statistic for these residuals is 0.40, which has an associated z score of approximately 8.[1] This is far from what we would expect if the null hypothesis of spatial independence were true. Stated differently, this implies that there is considerable positive association between the democracy level of a country and that of its geographical neighbors, above and beyond what we would expect from their levels of GDP per capita. This result is fairly typical, and it will often be the case that including spatially clustered covariates alone will not completely remove spatial clustering in the outcome of interest.

Given that the distribution of democracy still displays spatial clustering after conditioning on a country's GDP per capita, we should look for possible ways to incorporate this spatial dependence in our previous regression model. As in the case of serial clustering over time, we can think of spatial autocorrelation either as nuisance or substance. Spatial dependence leads to problems with the regression estimate $\hat{\beta}$ for the effect of GDP per capita and its standard errors, since the errors cannot be considered to be independent among connected units. These problems in estimating the effect of GDP per capita on democracy can, in principle, be addressed through alternative estimators that take into account the spatial correlation of the errors, that is, the residual variation not captured by GDP per capita alone. This approach is often known as the spatial error model, an approach we discuss subsequently.

However, our broader interest here is in what influences democracy, not just estimating the association between a country's GDP per capita on its prospects for democracy. If a country's level of democracy appears to be associated with its neighbors' level of democracy, this tells us something important about the distribution of democracy itself and provides an opportunity for learning something about possible influences from spatial dependence on prospects and constraints on democracy. As such, a more plausible and interesting approach is to consider the spatial association as a substantive feature of democracy rather than as a statistical nuisance.

The spatial association observed here suggests that we have dependence among observations such that the expected value of democracy for a country i differs notably depending on the level of democracy in neighboring states j. Instead of letting expected democracy for a country i depend just on GDP per capita, we devise a model where democracy is a function of

both its own GDP per capita and the level of democracy among neighbors, defined by $w_i.y_i$ where the entries of the connectivity vector $w_i.$ (i.e., row i from matrix $\mathbf{W}$) acquire nonzero values for all states j that are defined as connected to i. Recall again that the $\mathbf{W}$ connectivity matrix is row standardized so that each row in $\mathbf{W}$ sums to 1.

This reasoning suggests a spatially lagged dependent variable model of the form

$$y_i = \beta_0 + \beta_1 x_i + \rho w_i.y_i + \epsilon_i, \qquad [2.1]$$

where a positive value for the parameter associated with the spatial lag (ρ) indicates that countries are expected to have higher democracy values if, on average, their neighbors have high democracy values.

One can think of the spatially lagged y model as analogous to an autoregressive time series model where temporal serial correlation is addressed by including a lagged dependent variable y_{t-1} on the right-hand side when we estimate the effects of other right-hand side covariates (say x_t) on y_t. The $\hat{\beta}_1$ coefficient in the spatially lagged y model differs from the coefficient calculated via the OLS regression model in that we are now assessing the effect of GDP per capita on the democracy level of a country, while controlling for spatial dependence in y, or the extent to which variation in a country i's level of democracy can be accounted for by the value of y in other connected countries j. Hence, we will need to take into account the spatial ramifications when assessing the effect of changes in x.

Tables 2.1 and 2.2 provide the estimates from an OLS regression on the level of democracy on the natural log of per capita GDP in 158 countries in 2002 with and without a spatial lag of y. We observe a large positive coefficient for the log of GDP per capita of 1.68 in the OLS without the spatially lagged y. In contrast, in the spatially lagged y model, the estimated coefficient for the log of GDP per capita is 0.76, less than half of its original size,

TABLE 2.1
OLS Without Spatial Lag

OLS	$\hat{\beta}$	$SE(\hat{\beta})$	t Value
Intercept	–9.69	2.43	–3.99
ln GDP per capita	1.68	0.31	5.36

$N = 158$
Log likelihood ($df = 3$) = –513.62
$F = 28.77$ ($df_1 = 1$, $df_2 = 156$)

TABLE 2.2
OLS With Spatial Lag

OLS	$\hat{\beta}$	$SE(\hat{\beta})$	t Value
Intercept	–4.98	2.07	–2.40
ln GDP per capita	0.76	0.28	2.72
ρ	0.76	0.088	8.65

$N = 158$
Log likelihood ($df = 4$) = –482.48
$F = 58.64$ ($df_1 = 2$, $df_2 = 155$)

although it continues to be far away from 0 by the conventional standards for significance tests.

The estimate for the spatially lagged y term is large and positive (0.76) and highly statistically significant by standard criteria. This provides support for the conjecture that a country's level of democracy covaries with the level of democracy among its geographical neighbors. In substantive terms, the model implies that a country's expected level of democracy would be –7.6 points lower if its neighbors had an average democracy score at the minimum possible score (i.e., –10) compared with a neighbor average of 0, which is close to the historical average POLITY score since 1945. Conversely, a country with the maximum neighbor average democracy score of 10 would be expected to be 7.6 points "more democratic" relative to a country with a neighbor average of 0. These estimates reflect the clustering of democracy illustrated previously. Although most democracies tend to have higher GDP per capita, we do observe clusters of democracy in 2002 in areas where GDP per capita is not particularly high, as in Latin America, and clustering of autocracies in areas with high average GDP, as in the Gulf states.

Comparing the measures of the overall fit for the model assuming independent observations in Table 2.1 and the model with the spatially lagged y in Table 2.2 indicates that the model with the spatially lagged y term fits the data notably better. It has a higher F statistic and a higher log likelihood than the model assuming independent observations. This, in turn, reinforces our belief that the spatial lag of y adds something important to specifying the distribution of democracy, beyond what we would expect from a country's GDP per capita. Model heuristics alone do not provide the compelling reason for using the spatial approach, however. The spatial approach is better not because of the heuristics it produces alone, but because it specifies a plausible form of the feedback or dependency among observations.

A standard ordinary least squares regression has the following form:

$$y_i = \mathbf{x_i}\beta + \varepsilon_i.$$

If ε_i is decomposed into a spatially lagged term for the dependent variable—which is correlated with the dependent variable—and an independent error term, $\varepsilon_i = \rho\mathbf{w}_i y_i + \epsilon_i$, this leads to the formulation for spatially lagged dependent variables:

$$y_i = \mathbf{x_i}\beta + \rho\mathbf{w}_i y_i + \epsilon_i.$$

If, however, we specify this differently, $\varepsilon_i = \lambda\mathbf{w}_i \xi_i + \epsilon_i$, we get

$$y_i = \mathbf{x_i}\beta + \lambda\mathbf{w}_i \xi_i + \epsilon_i,$$

which is a spatial error formulation.

We turn next to an examination of the spatially lagged dependent variable model; the spatial error model is addressed in Chapter 3.[2]

It is tempting to interpret the coefficient estimate for GDP per capita in the model with the spatially lagged y in Table 2.2 and compare this directly with Table 2.1, suggesting a seemingly larger effect of GDP per capita. However, this interpretation is not correct. The coefficient estimates have different interpretations, as the model with the spatially lagged y in Equation 2.1 is an autoregressive specification, so that the coefficient for the impact of x now reflects the short-run impact of x_i on y_i rather than the net effect, as is the case of the coefficient for x in the OLS model without the spatially lagged y. Since the value of y_i will influence the level of democracy in other states y_j and these y_j, in turn, feed back on to y_i, we need to take into account the additional effects that the short impact of x_i exerts on y_i through its impact on the level of democracy in other states.

This is analogous to the interpretation of the coefficient β for a covariate x_t in a time series model where we have a temporal lag of the dependent variable y_{t-1} on the right-hand side, for example,

$$y_t = \beta x_t + \phi y_{t-1} + \epsilon_t.$$

In this equation, β indicates the immediate effect of x_t on y_t. But this will, in turn, affect y_{t-1} in the following time period and the long-run effect of x_t must thus take into account the part of the net effect that works through the autoregressive part or the estimated coefficient for the impact of the lag y_{t-1}. The long-run effect of x_t will be $\beta / (1-\phi)$. In a situation where ϕ is large, the long-run effect $\beta / (1-\phi)$ can be substantially larger than β.

Continuing this analogy, imagine if we could increase the log of GDP per capita by one unit in only a single country i, which would have an immediate impact on that country's level of democracy of β_1. However, the model in Equation 2.1 implies spatial dynamics with a feedback effect between countries, where country i's level of democracy is also held to have an effect on its neighbors' level of democracy. Hence, an increase in democracy that affects i's level of democracy will then influence democracy in the neighbors of j. Contemporaneously, in turn, the neighbors' neighbors will also be affected, throughout all connected countries. In general, all countries will have some neighbors so that eventually the influence of all countries will be affected. But note that Equation 2.1 includes democracy for all countries in the system y, so if the democracy level of other countries connected to i increases so will the level of democracy in i. In this way, an exogenous shock to one observation, such as our thought experiment, will have a reverberating effect throughout the system with feedback among observations and flow through the system as a series of adjustments until it settles on some new stable equilibrium (Cressie, 1993; Lin, Wu, & Lee, 2006).

Rather than focusing on the coefficient estimates for x_i alone in a spatially lagged y model, it is important to consider the equilibrium effects. Unfortunately, the long-run effect for the spatially lagged y cannot be stated in a form as simple as in the case of the long-run effect in the presence of a temporally lagged y. We will return later for how to characterize and estimate the equilibrium effect of covariates in a spatially lagged y model. First, however, we turn to problems posed by the presence of the spatial lag of y on the right-hand side of the equation and the implied endogeneity problems related to consistent estimation of the model in the ordinary least squares setup.

The following section relies on matrix algebra and focuses on issues of estimation and why a maximum likelihood estimator (MLE) may be preferable to estimating the spatially lagged y model. Since using the MLE by itself does not require an understanding of all the details in this section, readers who are not interested in issues of estimation may skip the details in this section and proceed immediately to the next section.

2.2   Estimating the Spatially Lagged y Model

In a time series model with a temporally lagged y_{t-1} on the right-hand side, the presence of the temporal lag y_{t-1} does not create problems for estimation with OLS, provided there is no serial correlation in the residuals of the regression model. More precisely, OLS with a lagged dependent