Spatial panel models

J. Paul Elhorst
University of Groningen, Department of Economics, Econometrics and Finance
P.O. Box 800, 9700 AV Groningen, the Netherlands
Phone: +31 50 3633893, Fax: +31 50 3637337, E-mail: j.p.elhorst@rug.nl

September 2011

Abstract This paper provides a survey of the existing literature on spatial panel data models. Both static and dynamic models will be considered. The paper also demonstrates that spatial econometric models that include lags of the dependent variable and of the independent variables in both space and time provide a useful tool to quantify the magnitude of direct and indirect effects, both in the short term and in long term. Direct effects can be used to test the hypothesis as to whether a particular variable has a significant effect on the dependent variable in its own economy, and indirect effects to test the hypothesis whether spatial spillovers exist. To illustrate these models and their effects estimates, a demand model for cigarettes is estimated based on panel data from 46 U.S. states over the period 1963 to 1992.

Keywords Spatial panels, dynamic effects, spatial spillover effects, identification, estimation methods

JEL Classification C21, C23, C51

1 Introduction

Spatial econometrics deals with interaction effects among geographical units.[1] Examples are economic growth rates of OECD countries over T years, monthly unemployment rates of EU regions in the last decade, and annual tax rate changes of all jurisdictions in a country since the last election. Spatial econometric models can also be used the explain the behavior of economic agents other than geographical units, such as individuals, firms, or governments, but this type of research is still in its infancy. Examples are research productivity of N universities located in a particular country, and consumption of the representative consumer in each zip code of the trade area of a commercial firm.

In modeling terms, three different types of interaction effects can be distinguished: endogenous interaction effects among the dependent variable (Y), exogenous interaction effects among the independent variables (X), and interaction effects among the error terms (ε). Originally, the central focus of spatial econometrics has been on one type of interaction effect in a single equation cross-section setting. Usually, the point estimate of the coefficient of this interaction effect was used to test the hypothesis as to whether spatial spillover effects exist. Most of the work was inspired by research questions arising in regional science and economic geography, where the units of observations are geographically determined and the structure of the dependence among these units can somehow be related to location and distance. However, more recently, the focus has shifted to models with more than one type of interaction effects, to panel data, and to the marginal effects of the explanatory variables in the model rather than the point estimates of the interaction effects.

In this chapter, we review and organize this recent literature. In section 2, we present the linear regression model with spatial interaction effects for cross-section data and, in section 3, its extension to panel data. In section 4, the latter model is further extended to include dynamic effects in both space and time. In section 5, we provide so-called "effects estimates" (after LeSage and Pace, 2009), which are required for making correct inferences regarding the effect of independent variables on the dependent variable. In section 6, we estimate a demand model for cigarettes based on panel data from 46 U.S. states over the period 1963 to 1992 to empirically illustrate the different models. This data set is taken from Baltagi (2005) and has been used for illustration purposes in other studies too. Finally, we

[1] Zip codes, neighborhoods, municipalities, counties, regions, states, countries, etc.

conclude this chapter with a number of important implications for econometric modeling of relationships based on spatial panel data.

2 Linear spatial dependence models for cross-section data

The standard approach in most empirical work is to start with a non-spatial linear regression model and then to test whether or not the model needs to be extended with spatial interaction effects. This approach is known as the specific-to-general approach. The non-spatial linear regression model takes the form

$$Y = \alpha\iota_N + X\beta + \varepsilon, \tag{1}$$

where Y denotes an N×1 vector consisting of one observation on the dependent variable for every unit in the sample (i=1,…,N), ι_N is an N×1 vector of ones associated with the constant term parameter α, X denotes an N×K matrix of exogenous explanatory variables, with the associated parameters β contained in a K×1 vector, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)^T$ is a vector of disturbance terms,[2] where ε_i are independently and identically distributed error terms for all i with zero mean and variance σ^2. Since the linear regression model is commonly estimated by Ordinary Least Squares (OLS), it is often labeled the OLS model. Furthermore, even though the OLS model in most studies focusing on spatial interaction effects is rejected in favor of a more general model, its results often serve as a benchmark.

The opposite approach is to start with a more general model containing, nested within it as special cases, a series of simpler models that ideally should represent all the alternative economic hypotheses requiring consideration. Generally, three different types of interaction effects may explain why an observation associated with a specific location may be dependent on observations at other locations:

i. Endogenous interaction effects, where the decision of a particular unit A (or its economic decision makers) to behave in some way depends on the decision taken by other units, among which, say, unit B,

   Dependent variable y of unit A ↔ Dependent variable y of unit B (1)

[2] The superscript T denotes the transpose of a vector or matrix.

Endogenous interaction effects are typically considered as the formal specification for the equilibrium outcome of a spatial or social interaction process, in which the value of the dependent variable for one agent is jointly determined with that of the neighboring agents. In the empirical literature on strategic interaction among local governments, for example, endogenous interaction effects are theoretically consistent with the situation where taxation and expenditures on public services interact with taxation and expenditures on public services in nearby jurisdictions (Brueckner 2003).

ii. Exogenous interaction effects, where the decision of a particular unit to behave in some way depends on independent explanatory variables of the decision taken by other units

Independent variable x of unit B → Dependent variable y of unit A (2)

Consider, for example, the savings rate. According to standard economic theory, saving and investment are always equal. People cannot save without investing their money somewhere, and they cannot invest without using somebody's savings. This is true for the world as a whole, but it is not true for individual economies. Capital can flow across borders; hence the amount an individual economy saves does not have to be the same as the amount it invests. In other words, per capita income in one economy also depends on the savings rates of neighboring economies. It should be stressed that, if the number of independent explanatory variables in a linear regression model is K, the number of exogenous interaction effects might also be K, provided that the intercept is considered as a separate variable**.** In other words, not only the savings rate but also other explanatory variables may affect per capita income in neighboring economies. It is for this reason that economic growth in both the theoretical and the empirical literature on economic growth and convergence among countries or regions is not only taken to depend on the initial income level and the rates of saving, population growth, technological change and depreciation in the own economy, but also on those of neighboring economies (Ertur and Koch 2007; Elhorst et al. 2010).

iii. Interaction effects among the error terms

Error term u of unit A ↔ Error term u of unit B (3)

Interaction effects among the error terms do not require a theoretical model for a spatial or social interaction process, but, instead, is consistent with a situation where determinants of the dependent variable omitted from the model are spatially autocorrelated, and with a situation where unobserved shock follow a spatial pattern. Interaction effects among the error terms may also be interpreted to reflect a mechanism to correct rent-seeking politicians for unanticipated fiscal policy changes (Allers and Elhorst 2005).

A full model with all types of interaction effects takes the form

$$Y = \rho WY + \alpha\iota_N + X\beta + WX\theta + u\,, \tag{4a}$$

$$u = \lambda Wu + \varepsilon\,, \tag{4b}$$

where the variable WY denotes the endogenous interaction effects among the dependent variables, WX the exogenous interaction effects among the independent variables, and Wu the interaction effects among the disturbance terms of the different units. ρ is called the spatial autoregressive coefficient, λ the spatial autocorrelation coefficient, while θ, just as β, represents a $K\times 1$ vector of fixed but unknown parameters. W is a nonnegative $N\times N$ matrix of known constants describing the arrangement of the units in the sample. Its diagonal elements are set to zero by assumption, since no unit can be viewed as its own neighbor.

Three methods have been developed to estimate models that include interaction effects. One is based on maximum likelihood (ML) or quasi maximum likelihood (QML), one on instrumental variables or generalized method of moments (IV/GMM), and one on the Bayesian Markov Chain Monte Carlo (MCMC) approach. QML and IV/GMM estimators are different in that they do not rely on the assumption of normality of the disturbances. Detailed descriptions of these estimation techniques can be found in Anselin (1988), Lee (2004), Kelejian and Prucha (1998), and LeSage and Pace (2009).

Technically, there are no obstacles to estimating a model with interaction effects among the dependent variable, the independent variables and the disturbance terms. Often, however, the parameters cannot be interpreted in a meaningful way since the different types

of interaction effects cannot be distinguished from each other. Lee et al. (2010) prove that there is at least one spatial weights matrix so that all parameters are identified. They consider G groups, each consisting of N_g cross-sectional units, and assume that the elements of the spatial weights matrix measuring the interaction effects are $w_{ij}=1/(N_g-1)$ if units i and j belong to the same group (except if i=j), and zero otherwise. Starting with this spatial weight matrix, it is shown that the parameters are identified either if both N and N_g tend to infinity, with at least two units in each group, or if the number of units in each group does not tend to infinity faster than or equal to the number of groups. Whether the parameters are also identified for other specifications of the spatial weights matrix still needs to be investigated. If not, the best option is to exclude the spatially autocorrelated error term and to consider a model with endogenous and exogenous interaction effects 'only'. This model is known as the spatial Durbin model. Only is put in quotation marks, because this model covers K+1 of the K+2 potential interaction effects. According to LeSage and Pace (2009, pp. 155-158), the cost of ignoring spatial dependence in the dependent variable and/or in the independent variables is relatively high since the econometrics literature has pointed out that if one or more relevant explanatory variable are omitted from a regression equation, the estimator of the coefficients for the remaining variables is biased and inconsistent. In contrast, ignoring spatial dependence in the disturbances, if present, will only cause a loss of efficiency.

3 Linear spatial dependence models for panel data

In recent years, the spatial econometrics literature has exhibited a growing interest in the specification and estimation of econometric relationships based on spatial panels. This interest can be partly explained by the increased availability of more data sets in which a number of spatial units are followed over time, and partly by the fact that panel data offer researchers extended modeling possibilities as compared to the single equation cross-sectional setting. Panel data are generally more informative, and they contain more variation and less collinearity among the variables. The use of panel data results in a greater availability of degrees of freedom, and hence increases efficiency in the estimation. Panel data also allow for the specification of more complicated behavioral hypotheses, including effects that cannot be addressed using pure cross-sectional data (see Baltagi 2005, and the references therein).

The extension of the spatial econometric model, presented in equation (2), for a cross-section of N observation to a space-time model for a panel of N observations over T time periods is obtained by adding a subscript t, which runs from 1 to T, to the variables and the error terms of that model

$$Y_t = \rho W Y_t + \alpha \iota_N + X_t \beta + W X_t \theta + u_t, \tag{5a}$$

$$u_t = \lambda W u_t + \varepsilon_t. \tag{5b}$$

This model can be estimated along the same lines as the cross-sectional model, provided that all notations are adjusted from one cross-section to T cross-sections of N observations.

However, the main objection to pooling the data like this is that the resulting model does not account for spatial and temporal heterogeneity. Spatial units are likely to differ in their background variables, which are usually space-specific time-invariant variables that do affect the dependent variable, but which are difficult to measure or hard to obtain. Examples of such variables abound: one spatial unit is located at the seaside, the other just at the border; one spatial unit is a rural area located in the periphery of a country, the other an urban area located in the center; norms and values regarding labor, crime and religion in one spatial unit might differ substantially from those in another unit, etc. Failing to account for these variables increases the risk of obtaining biased estimation results. One remedy is to introduce a variable intercept μ_i representing the effect of the omitted variables that are peculiar to each spatial unit considered. In sum, spatial specific effects control for all time-invariant variables whose omission could bias the estimates in a typical cross-sectional study. Similarly, the justification for adding time-period specific effects is that they control for all spatial-invariant variables whose omission could bias the estimates in a typical time-series study (Baltagi, 2005). Examples of such variables also exist: one year is marked by economic recession, the other by a boom; changes in legislation or government policy can significantly affect the functioning of an economy as from the date of implementation, as a result of which before and after observations might be significantly different from one another.

The space-time model in (5) extended with spatial specific and time-period specific effects reads as

$$Y_t = \rho W Y_t + \alpha \iota_N + X_t \beta + W X_t \theta + \mu + \xi_t \iota_N + u_t, \tag{6a}$$

$$u_t = \lambda W u_t + \varepsilon_t, \tag{6b}$$

where $\mu = (\mu_1, \ldots, \mu_N)^T$. The spatial and time-period specific effects may be treated as fixed

effects or as random effects. In the fixed effects model, a dummy variable is introduced for each spatial unit and for each time period (except one to avoid perfect multicollinearity), while in the random effects model, μ_i and ξ_t are treated as random variables that are independently and identically distributed with zero mean and variance σ_μ^2 and σ_ξ^2, respectively. Furthermore, it is assumed that the random variables μ_i, ξ_t and ε_{it} are independent of each other.

The estimation of static spatial panel data models is extensively discussed in Elhorst (2003, 2010a) and Lee and Yu (2010a). Elhorst (2003, 2010a) presents the ML estimator of the spatial lag model and of the error model extended to include fixed effect or random effects. Further note that the spatial Durbin model can be estimated as a spatial lag model with explanatory variables [X WX] instead of X. The response parameters of the fixed effects models can be estimated by concentrating out the fixed effects first, called demeaning (see Baltagi 2005; and Elhorst 2010a for mathematical details). The resulting equation can then be estimated by the ML estimation procedure developed by Anselin (1988) for the spatial lag model, provided that this procedure is generalized from one single cross-section of N observations to T cross-sections of N observations. The estimation of the random effects model is somewhat more complicated (see Elhorst, 2010a for details).

Lee and Yu (2010a) show that the ML estimator of the spatial lag and of the spatial error model with spatial fixed effects, as set out in Elhorst (2003, 2010a), will yield an inconsistent parameter estimate of the variance parameter (σ^2) if N is large and T is small, and inconsistent estimates of all parameters of the spatial lag and of the spatial error model with spatial and time-period fixed effects if both N and T are large. To correct for this, they propose a simple bias correction procedure based on the parameter estimates of the uncorrected approach. Elhorst (2010b) provides Matlab routines at his website www.regroningen.nl/elhorst for both the fixed effects and the random effects spatial lag model, as well as the fixed effects and the random effects spatial error model. Recently, these routines have been updated for the bias correction procedure of Lee and Yu (2010a).

4 Dynamic linear spatial dependence models for panel data

To make the spatial panel data model, presented in equation (6), dynamic, one might add time lags of the variables Y_t and WY_t, to get

$$Y_t = \tau Y_{t-1} + \rho W Y_t + \eta W Y_{t-1} + \alpha \iota_N + X_t \beta + W X_t \theta + \mu + \xi_t \iota_N + u_t . \quad (5)$$

This model is known as a dynamic spatial Durbin model (Debarsy et al., 2011). Similarly, one might consider time lags of the variables X_t and WX_t, and of the error terms u_t and Wu_t. According to Anselin et al. (2008), however, the parameters of such a model will not be identified. More than that, they argue that the parameters of the dynamic spatial panel data model in (8) are not identified either. Elhorst (2012) gives an overview of the main restrictions that have been considered in the literature to get rid of this identification problem: (i) θ=0 to exclude exogenous interaction effects (WX_t), (ii) ρ=0 to exclude contemporaneous endogenous interaction effects (WY_t); (iii) η=0 to exclude lagged endogenous interaction effects (WY_{t-1}), and (iv) η=-$\tau\rho$. The latter restriction implies that η, the parameter associated with lagged endogenous interaction effects, is equal to $-\tau\rho$, the two parameters of the dependent variables respectively lagged in time and lagged in space. In sum, each of these restrictions eliminates one type of interaction effects.

Three methods have been developed in the literature to estimate models that have mixed dynamics in both space and time. One method is to bias-correct the maximum likelihood (ML) or quasi-maximum likelihood (QML) estimator, one method is based on instrumental variables or generalized method of moments (IV/GMM), and one method utilizes the Bayesian Markov Chain Monte Carlo (MCMC) approach.

Yu et al. (2008) construct a bias corrected estimator for a dynamic model (Y_{t-1}, WY_t and WY_{t-1}) with spatial fixed effects. Lee and Yu (2010b) extend this study to include time-period fixed effects. They first estimate the model by the ML estimator developed by Elhorst (2003, 2010a) for the spatial lag model with spatial (and time-period) fixed effects. This estimator is called the LSDV estimator and is based on the conditional log-likelihood function of the model, i.e., conditional upon the first observation of every spatial unit in the sample due to the regressors Y_{t-1} and WY_{t-1}. Next, they provide a rigorous asymptotic theory for the LSDV estimator and their bias corrected LSDV (BCLSDV) estimator when both the number of spatial units (N) and the number of time points (T) in the sample go to infinity such that the limit between N and T exists and is bounded between zero and infinity ($0<\lim(N/T)<\infty$). In the words of Lee and Yu (2010c, p. 2), this condition implies that '$T \to \infty$ where T cannot be too small relative to N'. The bias correction is derived for both normally distributed error terms (ML) and for error terms that do not rely on the normality assumption (QML). In the latter case the first four moments are required. Finally, it is to be noted that this BCLSDV

estimator can also be used when either the variable Y_{t-1} or the variable WY_{t-1} is eliminated from the model.

Elhorst (2010c) investigates the small sample properties of this BCLSDV estimator. For this purpose, he extends the unconditional ML estimator proposed by Hsiao et al. (2002) with the variable WY_t. To determine the expected value and the variance of the first first-differenced observations in the sample, needed to obtain the unconditional log-likelihood function, he applies the Bhargava and Sargan (1983) approximation. One of his conclusions is that the parameter estimate ρ of the variable WY_t is still considerably biased when using this unconditional ML estimator. However, if the parameter estimate ρ is based on the BCLSDV estimator and the other parameters, given ρ, on the unconditional ML estimator, then this so-called mixed ML/BCLSDV estimator outperforms the BCLSDV estimator of Yu et al. (2008) for small values of T (T=5).

A couple of studies have considered IV/GMM estimators, building on previous work of Arrelano and Bond (1991) and Blundell and Bond (1998). Elhorst (2010c) extends the Arrelano and Bond difference GMM estimator to include endogenous interaction effects and finds that this estimator can still be severely biased, especially with respect to the parameter estimate ρ of the variable WY_t. He notes a bias of 0.061. The explanation for this can be found in Lee and Yu (2010c). They find that a 2SLS estimator like the Arrelano and Bond GMM estimator which is based on lagged values of Y_{t-1}, WY_{t-1}, X_t and WX_t is not consistent due to too many moments, and that the dominant bias is caused by the endogeneity of the variable WY_t rather than the variable Y_{t-1}. To avoid these problems, they propose an optimal GMM estimator based on linear moment conditions, which are standard, and quadratic moment conditions, which are implied by the variable WY_t, and therefore not standard in dynamic panel data models. They prove that this GMM estimator is consistent, also when T is small relative to N.

Parent and LeSage (2010, 2011) point out that the Bayesian MCMC approach considers conditional distributions of each parameter of interest conditional on the others, which leads to some computational simplification. Just as Elhorst (2010c), they treat the first period cross-section as endogenous, using the Bhargava and Sargan (1983) approximation, and find that the correct treatment of the initial observations (endogenous instead of exogenous) is important, especially in cases when T is small.

5 Direct, indirect, and spatial spillover effects

Many empirical studies use point estimates of one or more spatial regression model specifications to test the hypothesis as to whether or not spatial spillovers exist. One of the key contributions of LeSage and Pace's book (2009, p. 74) is the observation that this may lead to erroneous conclusions, and that a partial derivative interpretation of the impact from changes to the variables of different model specifications represents a more valid basis for testing this hypothesis.

By rewriting the spatial econometric model with dynamic effects in space and time in (5) as

$$Y_t = (I-\rho W)^{-1}(\tau I+\eta W)Y_{t-1} + (I-\rho W)^{-1}(X_t\beta + WX_t\theta) + R\,, \tag{6}$$

where R is a rest term containing the intercept and the error terms, the matrix of partial derivatives of Y with respect to the kth explanatory variable of X in unit 1 up to unit N at a particular point in time can be seen to be

$$\begin{bmatrix} \frac{\partial Y}{\partial x_{1k}} & \cdots & \frac{\partial Y}{\partial x_{Nk}} \end{bmatrix}_t = (I-\rho W)^{-1}[\beta_k I_N + \theta_k W]. \tag{7}$$

These partial derivatives denote the effect of a change of a particular explanatory variable in a particular spatial unit on the dependent variable of all other units in the *short term*. Similarly, the *long-term* effects can be seen to be

$$\begin{bmatrix} \frac{\partial Y}{\partial x_{1k}} & \cdots & \frac{\partial Y}{\partial x_{Nk}} \end{bmatrix} = [(1-\tau)I-(\rho+\eta)W]^{-1}[\beta_k I_N + \theta_k W]. \tag{8}$$

LeSage and Pace (2009) and Debarsy et al. (2011) define the direct effect as the average of the diagonal elements of the matrix on the right-hand side of (7) or (8), and the indirect effect as the average of either the row sums or the column sums of the non-diagonal elements of these matrices (since the numerical magnitudes of these two calculations of the indirect effect are the same, it does not matter which one is used). The outcomes are independent from the time index; this explains why the right-hand sides of these equations do not contain the

symbol t. The expressions in (7) and (8) also show that short-term indirect effects do not occur if both $\rho=0$ and $\theta_k=0$, while long-term indirect effects do not occur if both $\rho=-\eta$ and $\theta_k=0$.

Using the expressions in (7) and (8), it is also possible to indicate the disadvantages of certain parameter restrictions put forward in the previous section needed for identification. The disdvantage of imposing the restriction $\theta=0$ is that the ratio between the indirect effect and the direct effect becomes the same for very explanatory variables. In other words, if this ratio happens to be p percent for one variable, it is also p percent for any other variable. The disadvantage of imposing the restriction $\rho=0$ is that the matrix $(I-\rho W)^{-1}$ degenerates to the identity matrix, as a result of which the short-term indirect effects depend on θ only. This loss of flexibility makes the model less suitable for empirical research focusing on short-term effects. The disadvantage of imposing the restriction is $\eta=-\tau\rho$ is that the ratio between the indirect effect and the direct effect of a particular explanatory variable remains constant over time. In other words, if this ratio happens to be p percent for one variable in the short term, it is also p percent in the long term. By contrast, if the restriction $\eta=0$ is imposed, no prior restrictions are imposed on the effects estimates, even though still some flexibility of the model gets lost.

6 Empirical illustration

Baltagi and Li (2004) estimate a demand model for cigarettes based on a panel from 46 U.S. states in which real per capita sales of cigarettes by persons of smoking age (14 years and older) measured in packs of cigarettes per capita (C_{it}) is regressed on the average retail price of a pack of cigarettes measured in real terms (P_{it}) and on real per capita disposable income (Y_{it}). Moreover, all variables are taken in logs. Whereas Baltagi and Li (2004) use the first 25 years for estimation to reserve data for out of sample forecasts, we use the full data set covering the period 1963-1992.[3] More details, as well as reasons to include state-specific effects (μ_i) and time-specific effects (ξ_t), are given in Baltagi (2005). The spatial weights matrix is specified as a row-normalized binary contiguity matrix whose elements are one if two states share a common border, and zero otherwise.

Column (1) of Table 1 reports the estimation results when adopting a non-dynamic spatial Durbin model without spatial and time-period fixed effects, and column (2) when including these effects. To investigate whether or not the fixed effects are jointly significant, one may test the

[3] The dataset can be downloaded freely from www.wiley.co.uk/baltagi/. An adapted version of this dataset is available at www.regroningen.nl/elhorst.

hypothesis H_0: $\mu_1=\ldots=\mu_N=\xi_1=\ldots=\xi_T=\alpha$, where α is the intercept of the model without fixed effects. To test this (null) hypothesis, one may perform a likelihood ratio (LR) test, which is based on the log-likelihood function values of both models. The number of degrees of freedom is equal to the number of restrictions that needs to be imposed on the fixed effects to get one overall intercept, which in this particular case is N+T-1. The outcome of this test (2×(1691.4-475.5)=2431.8 with N+T-1=46+30-1=75 df) justifies the extension of the model with spatial and time-period effects.[4] It is to be noted that the coefficient of any variable that does not change over time or only a little cannot be estimated when controlling for spatial fixed effects. Similarly, the coefficient of any variable that does not change across space or only a little cannot be estimated when controlling for time-period fixed effects. For many empirical studies this is a reason not to control for fixed effects, for example, because such time-invariant or space-invariant variables are the main focus of the analysis. However, if one or more relevant explanatory variables are omitted from the regression equation, when they should be included, the estimator of the coefficients of the remaining variables is biased and inconsistent. This also holds true for fixed effects and is known as the omitted regressor bias.

Instead of fixed effects we can also treat μ and ξ as random effects. Hausman's specification test can then be used to test the random effects model against the fixed effects model. However, whether the random effects model is an appropriate specification if the population may be said to be sampled exhaustively, such as all counties of a state or all regions in a country, remains controversial. A detailed discussion of this issue can be found in Elhorst (2012).

The main shortcoming of a non-dynamic spatial Durbin model is that it cannot be used to calculate short-term effect estimates of the explanatory variables. This is made clear in Table 2, which reports the corresponding effects estimates of the models presented in Table 1; since a non-dynamic model only produces long-term effects estimates, the cells reporting short-term effects estimates are left empty.

The direct effects estimates of the two explanatory variables reported in column (2) of Table 2 are significantly different from zero and have the expected signs. Higher prices restrain people from smoking, while higher income levels have a positive effect on cigarette demand. The price elasticity amounts to -1.01 and the income elasticity to 0.594. Note that these direct effects estimates are different from the coefficient estimates of -1.001 and 0.603 reported in column (2) of Table 1 due to feedback effects that arise as a result of impacts passing through neighboring states and back to the states themselves.

[4] Note that one may also separately test for the inclusion of spatial fixed effects and time-period fixed effects.

The spatial spillover effects (indirect effects estimates) of both variables are negative and significant. Own-state price increases will restrain people not only from buying cigarettes in their own state, but to a limited extent also from buying cigarettes in neighboring states (elasticity -0.22). By contrast, whereas an income increase has a positive effects on cigarette consumption in the own state, it has a negative effect in neighboring states. We come back to this result below. Further note that the non-dynamic spatial Durbin model without spatial and time-period effects indicates a positive rather than a negative spatial spillover effect of price increases, and that only the latter result would be consistent with Baltagi and Levin (1992), who found that price increases in a particular state —due to tax increases meant to reduce cigarette smoking and to limit the exposure of non-smokers to cigarette smoke— encourage consumers in that state to search for cheaper cigarettes in neighboring states. However, there are two reasons why this comparison is invalid. First, whereas Baltagi and Levin's (1992) model is dynamic, it is not spatial.[5] Second, whereas our model contains spatial interaction effects, it is not (yet) dynamic. For these reasons it is interesting to consider the estimation results of our dynamic spatial panel data model.

Column (3) of Table 2 reports the direct and indirect effects of the dynamic model, both in the short term and in the long term. Consistent with microeconomic theory, the short-term direct effects appear to substantially smaller than the long-term direct effects; -0.262 versus -1.931 for the price variable and 0.099 versus 0.770 for the income variable. This is because it takes time before price and income changes fully settle. The long-term direct effects in the dynamic spatial Durbin model, on their turn, appear to be greater (in absolute value) than their counterparts in the non-dynamic spatial Durbin model; -1.931 versus -1.013 for the price variable and 0.770 versus 0.594 for the income variable. Apparently, the non-dynamic model underestimates the long-term effects. The short-term spatial spillover effect of a price increase turns out to be positive; the elasiticty amounts to 0.16 and is highly significant (t-value 3.49). This finding is in line with the original finding of Baltagi and Levin (1992) in that a price increase in one state encourages consumers to search for cheaper cigarettes in neighboring states. The positive spatial spillover effect of a price increase we found earlier for the non-dynamic spatial Durbin model demonstrates that a non-dynamic approach falls short here. Although greater and again positive, we do not find empirical evidence that the long-term spatial spillover effect is also significant. A similar result is found by Debarsy et al. (2011).

[5] They consider the price of cigarettes in neighboring states, but not any other spatial interaction effects.

The long-term effect spatial spillover effect of the income variable derived from the dynamic spatial panel data model appears to be positive, which suggests that an income increase in a particular state has a positive effect on smoking not only in that state itself, but also in neighboring states. Furthermore, the spatial spillover effect is smaller than the direct effect, which makes sense since the impact of a change will most likely be larger in the place that instigated the change. However, the spatial spillover effect of an income increase is not significant. A similar result is found by Debarsy et al. (2011). Interestingly, the spatial spillover effect of the income variable in the non-dynamic spatial panel data model appeared to be negative and significant. Apparently, the decision whether to adopt a dynamic or a non-dynamic model represents an important issue. Some researchers prefer simpler models to more complex ones (Occam's razor). One problem of complex models is overfitting, the fact that excessively complex models are affected by statistical noise, whereas simpler models may capture the underlying process better and may thus have better predictive performance. However, if one can trade simplicity for increased explanatory power, the complex model is more likely to be the correct one.

To investigate whether the extension of the non-dynamic model to the dynamic spatial panel data model increases the explanatory power of the model, one may test whether the coefficients of the variables Y_{t-1} and WY_{t-1} are jointly significant using an LR-test. The outcome of this test ($2\times(2623.3-1691.4)=1863.8$ with 2 df) evidently justifies the extension of the model with dynamic effects.

One potential objection to the dynamic spatial Durbin model, however, is that its parameters are not identified (Anselin et al., 2008). To investigate this, we carried out a Monte Carlo simulation experiment. The basic idea is to randomly draw (e.g., 1,000 times) the error terms based on σ^2 of the estimated equation, to generate the dependent variable given this error term and the independent variables and their coefficient estimates reported in column (3) of Table 1, and then to re-estimate the model. On average, these results should be similar to those of the "original" parameter estimates. Column (1) of Table 3 reports the results. These results show that the biases in the coefficient estimates, based on 1,000 replications, in this particular case are not very large. The largest bias is found in the coefficient of the spatial lag of the price variable (W*Log(P)); its original coefficient is 0.170, while its simulated coefficient is 0.182, which represents a bias of 0.012 or 7.1% of the original parameter value. For this reason, we have tested which of the four restrictions presented in the two previous sections is acceptable on the data. For this purpose we have used a Wald test, which has two degrees of freedom when considering the restriction $\theta=0$, and

one degree of freedom when considering any of the other restrictions. The following results were obtained: (i) $\theta=0$ has a test statistic of 15.04 with p-value smaller than 0.01 (K=2 degrees of freedom, since we have two explanatory variables), (ii) $\rho=0$ has a test statistic of 3.97 with p-value 0.046, (iii) $\eta=0$ has a test statistic of 0.15 with p-value 0.701, and (iv) $\eta=-\tau\rho$ has a test statistic of 5.66 with p-value 0.017. These results indicate that only the restriction $\eta=0$ is acceptable on the data. Remarkably, Debarsy et al. (2011) found that the restriction $\eta=-\tau\rho$ cannot be rejected, but they estimated the parameters of the model by Bayesian MCMC developed by Parent and LeSage (2010, 2011), whereas we used the BCLSDV estimator developed by Yu and Lee (2010).

Since the restriction $\eta=0$ is acceptable on the data, we re-estimated the dynamic spatial Durbin model, excluding the variable WY_{t-1}. The parameter estimates are reported in column (4) of Table 1, and the corresponding effects estimates in column (4) of Table 2. These results do not change much, mainly because the coefficient estimate of the variable WY_{t-1} reported in column (3) of Table 1 was not different from zero significant statistically. Column (2) of Table 3 reports the Monte Carlo simulation results of this model. These results confirm that identification problems of the parameters of this model are not longer at issue.

7 Conclusion

Spatial econometric models that include lags of the dependent variable and of the independent variables in both space and time provide a useful tool to quantify the magnitude of direct and indirect effects, both in the short term and in long term. A demand model for cigarettes based on panel data from 46 U.S. states over the period 1963 to 1992 is used to empirically illustrate this. Direct effects should be used to test the hypothesis as to whether a particular variable has a significant effect on the dependent variable in its own economy rather than the coefficient estimate of that variable. Similarly, indirect effects should be used to test whether or not spatial spillovers exist rather than the coefficient estimate of the spatially lagged dependent variable and/or the coefficients estimates of the spatially lagged independent variables.

One difficulty is that it cannot be seen from the coefficient estimates and the corresponding standard errors or t-values (derived from the variance-covariance matrix) whether the direct and indirect effects in a spatial econometric model are significant. This is because these effects are composed of different coefficient estimates according to complex mathematical formulas and the dispersion of these effects depends on the dispersion of all coefficient estimates involved. Fortunately, some individual researchers have made software

available at their Web sites programmed in Matlab or R that calculates these effects and their corresponding standard errors or t-values. Nevertheless, the availability of easier accessible packages such as Stata would probably encourage much more applied researchers to use these kind of models and to report direct and indirect effects estimates in addition to the point estimates of the parameters of the model. This is important since eventually only these effects estimates should be used to draw inferences regarding the relationships we are modeling.

References

Allers MA, Elhorst JP (2005) Tax mimicking and yardstick competition among governments in the Netherlands. International Tax and Public Finance 12(4):493-513

Anselin L (1988) Spatial econometrics: Methods and models. Kluwer, Dordrecht

Anselin L, Le Gallo J, Jayet H (2008) Spatial panel econometrics. In: Mátyás L, Sevestre P (eds) The econometrics of panel data, fundamentals and recent developments in theory and practice, 3rd edition. Kluwer, Dordrecht, pp 627-662

Arrelano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58(2):277-297

Baltagi BH (2005) Econometric analysis of panel data, 3rd edition. Wiley, Chichester

Baltagi BH, Levin D (1992) Cigarette taxation: raising revenues and reducing consumption. Structural Change and Economic Dynamics 3(2): 321-335.

Baltagi BH, Li D. (2004) Prediction in the panel data model with spatial autocorrelation. In Anselin L, Florax RJGM, Rey SJ (eds) Advances in spatial econometrics: methodology, tools, and applications. Springer, Berlin, pp 283-295

Bhargava A, Sargan JD (1983) Estimating dynamic random effects models from panel data covering short time periods. Econometrica 51(6):1635-1659

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87(1):115-143

Brueckner JK (2003) Strategic interaction among local governments: An overview of empirical studies. International Regional Science Review 26(2):175-188

Debarsy N, Ertur C, LeSage JP (2011) Interpreting dynamic space-time panel data models. Statistical Methodology doi:10.1016/j.stamet.2011.02.002

Elhorst JP (2003) Specification and estimation of spatial panel data models. International Regional Science Review 26(3):244-268

Elhorst JP (2010a) Spatial panel data models. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, Heidelberg and New York, pp 377-407

Elhorst JP (2010b) Matlab software for spatial panels. Paper presented at 4th World Conference of the Spatial Econometric Association, Chicago

Elhorst JP (2010c) Dynamic panels with endogenous interaction effects when T is small. Regional Science and Urban Economics 40(5):272-282

Elhorst JP (2012) Dynamic spatial panels: Models, methods and inferences. Journal of Geographical Systems, forthcoming.

Elhorst JP, Piras G, Arbia G (2010) Growth and convergence in a multi-regional model with space-time dynamics. Geographical Analysis 42(3):338-355

Ertur C, Koch W (2007) Growth, technological interdependence and spatial externalities: Theory and evidence. Journal of Applied Econometrics 22(6):1033-1062

Hsiao C, Pesaran MH, Tahmiscioglu AK (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics 109(1):107-150

Kelejian HH, Prucha IR (1998) A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17(1):99-121

Lee LF (2004) Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6):1899-1925

Lee LF, Yu J (2010a) Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154(2):165-185

Lee LF, Yu J (2010b) A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory 26(2):564-597

Lee LF, Yu J (2010c) Efficient GMM estimation of spatial dynamic panel data models with fixed effects. http://www.economics.smu.edu.sg/events/Paper/LungfeiLee.pdf.

Lee LF, Liu X, Lin X (2010) Specification and estimation of social interaction models with network structures. The Econometrics Journal 13(2): 145–176.

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press Taylor & Francis Group, Boca Raton

Parent O, LeSage JP (2010) A spatial dynamic panel model with random effects applied to commuting times. Transportation Research Part B 44(5):633-645

Parent O, LeSage JP (2011) A space-time filter for panel data models containing random effects. Computational Statistics and Data Analysis 55(1):475-490

Yu J, Jong R. de, Lee LF (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. Journal of Econometrics 146(1):118-134

Table 1. Estimation results of cigarette demand using different model specifications

Determinants	(1)		(2)		(3)		(4)	
	Non-dynamic spatial Durbin model no fixed effects		Non-dynamic spatial Durbin model with fixed effects		Dynamic spatial Durbin model with lag WY_{t-1}		Dynamic spatial Durbin model without lag WY_{t-1}	
Intercept	2.631	(15.82)						
$Log(C)_{-1}$					0.865	(65.04)	0.864	(65.16)
W*Log(C)	0.337	(11.09)	0.264	(8.25)	0.076	(2.00)	0.064	(3.23)
$W*Log(C)_{-1}$					-0.015	(-0.29)		
Log(P)	-1.251	(-21.80)	-1.001	(-24.36)	-0.266	(-13.19)	-0.266	(-13.19)
Log(Y)	0.554	(14.96)	0.603	(10.27)	0.100	(4.16)	0.100	(4.17)
W*Log(P)	0.780	(11.15)	0.093	(1.13)	0.170	(3.66)	0.172	(3.68)
W*Log(Y)	-0.444	(11.09)	-0.314	(-3.93)	-0.022	(-0.87)	-0.023	(-0.89)
R^2		0.435		0.902		0.977		0.977
LogL		475.5		1691.4		2623.3		2623.0

Notes: t-values in parentheses

Table 2. Effects estimates of cigarette demand using different model specifications

Determinants	(1)		(2)		(3)		(4)	
	Non-dynamic spatial Durbin model no fixed effects		Non-dynamic spatial Durbin model with fixed effects		Dynamic spatial Durbin model with lag WY_{t-1}		Dynamic spatial Durbin model without lag WY_{t-1}	
Short-term direct effect Log(P)					-0.262	(-11.48)	-0.263	(-11.36)
Short-term indirect effect Log(P)					0.160	(3.49)	0.163	(3.60)
Short-term direct effect Log(Y)					0.099	(3.36)	0.100	(3.43)
Short-term indirect effect Log(Y)					-0.018	(-0.45)	-0.019	(-0.50)
Long-term direct effect Log(P)	-1.216	(-23.39)	-1.013	(-24.73)	-1.931	(-9.59)	-1.912	(-9.51)
Long-term indirect effect Log(P)	0.508	(7.27)	-0.220	(-2.26)	0.610	(0.98)	0.581	(0.94)
Long-term direct effect Log(Y)	0.530	(15.48)	0.594	(10.45)	0.770	(3.55)	0.776	(3.64)
Long-term indirect effect Log(Y)	-0.366	(-7.47)	-0.197	(-2.15)	0.345	(0.48)	0.372	(0.62)

Notes: t-values in parentheses

Table 3. Identification of parameters

Determinant/Parameter	Dynamic spatial Durbin model with lag WY_{t-1}		Dynamic spatial Durbin model without lag WY_{t-1}	
	Original parameter value*	Simulated parameter value	Original parameter value**	Simulated parameter value
$Log(C)_{-1}$	0.865	0.864	0.864	0.865
W*Log(C)	0.076	0.074	0.064	0.064
$W*Log(C)_{-1}$	-0.015	-0.005	-	-
Log(P)	-0.266	-0.264	-0.266	-0.264
Log(Y)	0.100	0.100	0.100	0.100
W*Log(P)	0.170	0.182	0.172	0.173
W*Log(Y)	-0.022	-0.025	-0.023	-0.022

* Based on column (3) of Table 1, ** Based on column (4) of Table 1