Spatial panel models and common factors

J. Paul Elhorst
University of Groningen, Department of Economics, Econometrics and Finance
Groningen, the Netherlands
E-mail: j.p.elhorst@rug.nl

Elhorst J.P. (2019) Spatial Panel Models and Common Factors. In: Fischer M., Nijkamp P. (eds) *Handbook of Regional Science*. Springer, Berlin, Heidelberg. Improved version of submitted manuscript.

Abstract This chapter provides a survey of the existing literature on spatial panel data models. Both static, dynamic, and dynamic models with common factors will be considered. Common factors are modeled by time-period fixed effects, cross-sectional averages, or principal components. It is demonstrated 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 long term. Direct effects can be used to test the hypothesis as to whether a particular variable has a significant effect on the own dependent variable, and indirect effects to test the hypothesis whether spatial spillovers affect the dependent variable of other units. To illustrate these models, their effects estimates, and the impact of the type of common factors, 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 spillovers, common factors, estimation

JEL Classification C21, C23, C51

1 Introduction

Spatial econometrics deals with interaction effects among geographical units, such as zip codes, neighborhoods, municipalities, counties, regions, states, or countries. 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 are also used to explain the behavior of economic agents other than geographical units, such as individuals, firms, or governments. The best example is a hedonic price equation in which the price of each house is explained by the price and characteristics of other houses that have sold prior and in the neighborhood of that house (Kelejian and Piras 2017, p.13).

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 (Handbook Chapter 76 Cross-section spatial regression models). 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, to dynamic specifications, 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, as well as common factors. In section 5, we provide so-called "effects estimates" (Handbook Chapter 77 Interpreting spatial econometric models), 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 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 $N\times 1$ vector consisting of one observation on the dependent variable for every unit in the sample (i=1,…,N), ι_N is an $N\times 1$ vector of ones associated with the constant term parameter α, X denotes an $N\times K$ matrix of exogenous explanatory variables, with the associated parameters β contained in a $K\times 1$ vector, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$ is a vector of disturbance terms, 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 that nests a series of simpler models, which would ideally represent all 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,

$$\text{Dependent variable } y \text{ of unit } A \leftrightarrow \text{Dependent variable } y \text{ of unit } B \tag{2}$$

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.

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 (3)

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).

iii. Interaction effects among the error terms

Error term u of unit A ↔ Error term u of unit B (4)

Interaction effects among the error terms do not require a theoretical model for a spatial or social interaction process, but instead, are consistent with a situation where determinants of the dependent variable omitted from the model are spatially autocorrelated, and with a situation where unobserved shocks 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, known as the general nesting spatial (GNS) model, takes the form

$$Y = \rho WY + \alpha \iota_N + X\beta + WX\theta + u, \tag{5a}$$

$$u = \lambda Wu + \varepsilon, \tag{5b}$$

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×1 vector of fixed but unknown parameters. W is a nonnegative N×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), LeSage and Pace (2009), Kelejian and Piras (2017), Handbook Chapter 78 Maximum likelihood estimation, Handbook Chapter 79 Bayesian MCMC estimation, and Handbook Chapter 80 Instrumental variables/methods of moments estimation.

Technically, there are no obstacles to estimating the GNS 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 due to overparameterization (Burridge et al. 2016). This is one of the reasons why the spatial econometrics literature distinguishes six simpler types of spatial econometric models. The differences between these models depend on the type and number of spatial interaction effects that are included. The first two columns of Table 1 provide an overview, including their designations and abbreviations. The last column will be further discussed in section 5.

Table 1 Spatial econometric models with different combinations of spatial interaction effects and their flexibility regarding spatial spillovers

Type of model	Interaction(s)	Flexibility spatial spillovers
SAR, Spatial autoregressive model*	*WY*	Constant ratios
SEM, Spatial error model	*Wu*	Zero by construction
SLX, Spatial lag of *X* model	*WX*	Fully flexible
SAC, Spatial autoregressive combined model**	*WY*, *Wu*	Constant ratios
SDM, Spatial Durbin model	*WY*, *WX*	Fully flexible
SDEM, Spatial Durbin error model	*WX*, *Wu*	Fully flexible
GNS, General nesting spatial model	*WY*, *WX*, *Wu*	Fully flexible

* Also known as the spatial lag model, ** Also known as the SARAR model

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 Eq. (5), for a cross-section of N observations 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

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

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

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 data points observed prior or after that date might be significantly different from one another.

The space-time model in Eq. (6) extended with spatial specific and time-period specific effects reads as

$$Y_t = \rho W Y_t + \alpha \iota_N + X_t \beta + W X_t \theta + \mu + \xi_t \iota_N + u_t, \tag{7a}$$

$$u_t = \lambda W u_t + \varepsilon_t. \tag{7b}$$

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_t are independent of each other.

The estimation of static spatial panel data models is extensively discussed in Elhorst (2014), and Lee and Yu (2015). Elhorst (2014) presents the ML estimator of the spatial lag model and of the spatial error model extended to include fixed effect or random effects. Halleck Vega and Elhorst (2015) further note that the spatial Durbin model can be estimated as a spatial lag model and the spatial Durbin error model as a spatial error 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 2014 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 2014 for details).

Lee and Yu (2015) show that the ML estimator of the spatial lag and of the spatial error model with spatial and time-period fixed effects effects, as set out in Elhorst (2014), will yield inconsistent parameter estimates 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. If time-period fixed effects are not included or if time period fixed effects are included and *T* is fixed, only the variance of the error term, σ^2, needs to be bias-corrected. Elhorst (2014) provides Matlab routines at his website www.spatial-panels.com 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, accounting for this bias correction procedure.

4 Dynamic linear spatial dependence models for panel data

To make the spatial panel data model, presented in Eq. (7), dynamic, one might add time lags of the variables Y_t and WY_t, to get

$$Y_t = \tau Y_{t-1} + \rho W Y_t + \eta W Y_{t-1} + \alpha \iota_N + X_t \beta + W X_t \theta + \mu + \xi_t \iota_N + u_t. \tag{8}$$

This model is known as the dynamic spatial Durbin model (Debarsy et al. 2012). In addition, 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), the parameters of the model in Eq. (8) and of models with these extensions are not identified, but Lee and Yu (2016) provide an econometric-theoretical proof showing the opposite.

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. Detailed descriptions of these estimation techniques can be found in Lee and Yu (2015), and Parent and LeSage (2010). In addition, Elhorst (2010) and Parent and LeSage (2010) treat the first period cross-section as endogenous, and find that when applying respectively the ML and the Bayesian MCMC estimators, the correct treatment of the initial observations (endogenous instead of exogenous) is important, especially in cases when T is small.

One objection to the model in Eq. (8) is that each time dummy ξ_t has the same homogenous impact on all observations in period t, while it is likely that, for example, business cycle effects may hit one unit harder than others. Similarly, each spatial fixed effect μ_i has the same homogenous impact on all observations in unit i. An alternative is to replace the spatial and time-period fixed effects by common factors that have a different heterogenous impact on the observations

$$Y_t = \tau Y_{t-1} + \rho W Y_t + \eta W Y_{t-1} + \alpha \iota_N + X_t \beta + W X_t \theta + \sum_r \Gamma_r^T f_{rt} + u_t, \quad (9)$$

where f_{rt} denotes the r^{th} $N\times 1$ common factor observed at time t and Γ_r is the corresponding $N\times 1$ vector of coefficient estimates. If two factors are considered, $f_{1t} = (1, \dots, 1)^T$ and $f_{2t} = (\xi_1, \dots, \xi_T)^T$, and the parameter restrictions $\Gamma_1^T = (\mu_1, \dots, \mu_N)$ and $\Gamma_2^T = (1, \dots, 1)$ are imposed, the model in Eq. (8) is obtained. This implies that controls for spatial and time-period fixed effects can be considered as two common factors. Formally, the spatial fixed effects represent one common factor (f_{1t}) which is constant over time but with heterogenous coefficients (Γ_1). The time-period fixed effects represent another common factor of length T (f_{2t}) which changes over time but which has homogenous coefficients (Γ_2). The total number of common factor parameters to be estimated in this setting amounts to N+T.

Instead of spatial and time-period fixed effects, common factors can also be measured by time-specific cross-sectional averages of the observable variables in the model or by principal components. A detailed summary of the econometric literature in this field can be found in Pesaran (2015a, section 29.4). The idea to replace time-period effects by cross-sectional averages dates back to Pesaran (2006). In view of Eq. (9), he suggests to use one or more variables from the set of cross-sectional averages, $\bar{Y}_t = \frac{1}{N}\sum_{i=1}^{N} Y_{it}$, $\bar{Y}_{t-1} = \frac{1}{N}\sum_{i=1}^{N} Y_{it-1}$, and $\bar{X}_{kt} = \frac{1}{N}\sum_{i=1}^{N} X_{ikt}$ (k=1,..,K), as common factors such that each common factor has N spatial specific estimates, one for each unit in the sample. Together with the remaining N spatial fixed effects, the total number of common factor parameters to be estimated, if all cross-sectional avarages are included, then increases to N+(2+K)N. Just as time-period fixed effects, these cross-sectional averages may be treated as exogenous explanatory variables based on the assumption that the contribution of each unit to the cross-sectional averages at a particular point in time goes to zero if N goes to infinity (Pesaran 2006, assumption 5 and remark 3). Since the numbers of parameters to be estimated increases rapidly with the number of common factors, most empirical studies try to keep the number of cross-sectional averages to a minimum. Cicarelli and Elhorst (2018) find that, using cigarette demand data of 69 Italian regions over the period 1877-1913, controlling for $\bar{Y}_t$ and $\bar{Y}_{t-1}$ only already effectively filters out the common time trends in the data. The authors of this paper used Matlab routines to estimate their model. However, it is also possible to estimate the parameters in Eq. (9) using the xlsme command in Stata (Belotti et al. 2017), illustrated in the textbox below.

The command in Stata running a dynamic spatial panel data model with spatial fixed effects and common factor $\bar{Y}_t$ reads as

xsmle Y X $\bar{Y}_{1t}$ … $\bar{Y}_{Nt}$ wmat(W) model(sdm) durbin(X) dlag(3) fe type(ind) effects nsim(1000)

The option type(ind) controls for spatial fixed effects and the variable list $\bar{Y}_{1t} \ldots \bar{Y}_{Nt}$ controls for the cross-sectional average of the dependent variable with N unit-specific coefficients. Time dummies should not be included to avoid (near) perfect multicollinearity. The data structure to read in the cross-sectional averages takes the following form:

Unit	Time	Unit 1	Unit 2	…	Unit N
1	1	$\bar{Y}_1$	0	…	0
2	1	0	$\bar{Y}_1$	…	0
⋮	⋮	⋮	⋮		⋮
N	1	0	0		$\bar{Y}_1$
⋮	⋮	⋮	⋮		⋮
1	T	$\bar{Y}_T$	0	…	0
2	T	0	$\bar{Y}_T$	…	0
⋮	⋮	⋮	⋮		⋮
N	T	0	0	…	$\bar{Y}_T$

The second possibility is to approach the unobservable common factors by one or more principal components. In that case the Γ parameters represent the factor loadings of the principal components. Shi and Lee (2017) developed a QML estimator for the model in Eq. (9), including a spatially autocorrelated error term $u_t = \lambda W u_t + \varepsilon_t$. This estimator does not require any specification of the distribution function of the disturbance term. The coefficients estimates are bias-corrected for the Nickell bias and the impact of this bias on the other coefficients in the equation. For this purpose, a Matlab routine called SFactors has been developed, which the first author (Shi) made available at his web site www.w-shi.net. A potential disadvantage of principal components is that they are often difficult to interpret, especially if they are compared with cross-sectional averages. Every principal component requires the estimation of 2N parameters.

To find out which set of common factors is able to filter out common factors most effectively, the cross-sectional dependence (CD) test developed by Pesaran (2015b) may be used. This test is based on the correlation coefficients between the time-series observations of each pair of units with respect to a particular variable, in this case the residuals of Eq. (9), resulting in $N(N-1)$ correlations. Denoting these estimated correlation coefficients between the time-series for units i and j as κ_{ij}, the test statistic is defined as $CD = \sqrt{2T/(N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\kappa_{ij}$. The $N(N-1)$ mutual correlation coefficients, each calculated over T observations, explain respectively the division by $N(N-1)$ and the multiplication by T in the expression $\sqrt{2T/N(N-1)}$. The number two is added since the correlation matrix is symmetric. Consequently, it is sufficient to calculate the CD statistic over the upper triangular elements of the correlation matrix only and to multiply the outcome by two so as to also represent the impact of the lower triangular elements. The null hypothesis of the CD-test assumes the existence of weak cross-sectional dependence in the data (Pesaran, 2015b, Theorem 3), better known as local spatial dependence, while the alternative hypothesis reflects strong cross-sectional dependence, among which common factors. The CD statistic is a two-sided test statistic whose limiting distribution converges to the standard normal distribution, and thus −1.96 and 1.96 as critical values at the 5% significance level, provided that N goes to infinity faster than T or when T is fixed, reflecting the case in most spatial econometric studies.

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 Eq. (9) as

$$Y_t = (I_N - \rho W)^{-1}(\tau + \eta W)Y_{t-1} + (I_N - \rho W)^{-1}(X_t\beta + WX_t\theta) + R, \tag{10}$$

where R is a rest term covering all the remaining terms (intercept, spatial and time-period fixed effects, common factors, and/or the error terms), the matrix of partial derivatives of the expectation of Y_t, $E(Y_t)$, with respect to the k^{th} explanatory variable of X_t in unit 1 up to unit N can be seen to be

$$\left[\frac{\partial E(Y_t)}{\partial X_{1kt}} \cdots \frac{\partial E(Y_t)}{\partial X_{Nkt}}\right] = (I_N - \rho W)^{-1}(I_N\beta_k + W\theta_k). \tag{11}$$

This $N \times N$ matrix of partial derivatives denotes the effects of a change of a particular explanatory variable in a particular unit on the dependent variable of all other units in the *short term*. Note that this $N \times N$ matrix is actually the product of two $N \times N$ matrices. The elements of the first matrix, the inverse of $(I - \rho W)$, better known as the spatial multiplier matrix, are not worked out further since a simple analytical expression for this inverse does not exist. Similarly, the *long-term* effects can be seen to be

$$\left[\frac{\partial E(Y_t)}{\partial X_{1kt}} \cdots \frac{\partial E(Y_t)}{\partial X_{Nkt}}\right] = ((1-\tau)I_N - (\rho + \eta)W)^{-1}(I_N\beta_k + W\theta_k). \tag{12}$$

LeSage and Pace (2009) and Debarsy et al. (2012) define the direct effect as the average of the diagonal elements of the matrix on the right-hand side of Eqs. (11) or (12), and the indirect effect as the average of either the row sums or the column sums of the off-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). If the spatial weight matrix W does not change over time, the outcomes are independent from the time index; this explains why the right-hand sides of these equations do not contain the symbol t. A more appealing synonym for the indirect effect is spatial spillover effect, a term that therefore will be used in the remainder of this chapter.

Using the expressions in Eqs. (11) and (12), it is also possible to indicate the disadvantages of certain parameter restrictions adopted in previous studies (see Elhorst 2014 for an overview). Especially the SAR, SEM and SAC models are of limited use in empirical research due to initial restrictions on the spillover effects they can potentially produce. In the SAR (θ=λ=0) and the SAC (θ=0) models the ratio between the spillover effect and the direct effect is the same for every explanatory variable; if this ratio happens to be p percent for one variable, it is also p percent for any other variable, both in the short and in the long term. In the SEM model the spillover effects are even set to zero by construction; if both ρ=0 and θ_k=0, short-term spillover effects do not occur, while long-term spillover effects do not occur if both ρ=−η and θ_k=0. This loss of flexibility makes these models less suitable for empirical research focusing on spillover effects. Only in the SLX, SDEM, SDM and GNS models can they take any value. These findings are summarized in the last column of Table 1. Finally, the disadvantage of imposing the restriction η=−$\tau\rho$ is that the ratio between the indirect effect and the direct effect of a particular explanatory variable remains constant over time; if this ratio happens to be p percent for one variable in the short term, it is also p percent in the long term.

6 Empirical illustration

Baltagi (2005) estimates a demand model for cigarettes based on a panel from 46 U.S. states over the period 1963-1992 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. The dataset can be downloaded freely from www.wiley.co.uk/baltagi/, while an adapted version of this dataset is available at www.spatial-panels.com. 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 (before row-normalization) if two states share a common border, and zero otherwise.

Column (1) of Table 2 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 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 $N+T-1=46+30-1=75$ degrees of freedom justifies the extension of the model with spatial and time-period effects. Note that one may also separately test for the inclusion of spatial fixed effects and time-period fixed effects.

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

Table 2 Estimation results of cigarette demand using different model specifications

Determinants	(1)		(2)		(3)		(4)		(5)	
	Non-dynamic spatial Durbin model no fixed effects		Non-dynamic spatial Durbin model with fixed effects		Dynamic spatial Durbin model with fixed effects		Dynamic spatial Durbin model with cross-sectional averages*		Dynamic spatial Durbin model with principal components**	
Intercept	2.631	(15.82)								
Log(C)$_{-1}$					0.865	(67.14)	0.831	(61.54)	0.677	(40.04)
WLog(C)	0.337	(11.09)	0.264	(8.24)	0.076	(2.03)	0.001	(12.55)	0.097	(2.73)
WLog(C)$_{-1}$					−0.015	(−0.39)	0.077	(1.98)	-0.027	(-0.87)
Log(P)	−1.251	(−21.80)	−1.001	(−24.36)	−0.266	(−11.50)	−0.307	(−13.26)	−0.326	(−15.82)
Log(Y)	0.554	(14.96)	0.603	(10.27)	0.100	(3.35)	0.103	(1.34)	0.280	(9.64)
WLog(P)	0.780	(11.15)	0.093	(1.13)	0.170	(3.88)	0.292	(10.53)	0.247	(8.12)
WLog(Y)	−0.444	(11.09)	−0.314	(−3.93)	−0.022	(−0.57)	0.059	(1.67)	0.028	(0.72)
R^2	0.435		0.902		0.977		0.935		0.983	
LogL	475.5		1691.4		2623.3		3146.4		3293.9	
CD-test	25.25		−2.28		0.29		−2.82		−2.82	

Notes: t-values in parentheses; * cross-sectional averages at time t and t-1 of the dependent variables; ** two principal components 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 (2014, section 3.4).

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 by Table 3, which reports the corresponding effects estimates of the models presented in Table 2; 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 3 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 −1.013 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 2 due to feedback effects that arise as a result of impacts passing through neighboring states and back to the states themselves.

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 −0.220). By contrast, whereas an income increase has a positive effect 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 (column(1) of Table 3) 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. They consider the

Table 3 Effects estimates of cigarette demand using different model specifications

Determinants	(1)		(2)		(3)		(4)		(5)	
	Non-dynamic spatial Durbin model no fixed effects		Non-dynamic spatial Durbin model with fixed effects		Dynamic spatial Durbin model with fixed effects		Dynamic spatial Durbin model with cross-sectional averages*		Dynamic spatial Durbin model with principal omponents**	
Short-term direct effect Log(*P*)					−0.264	(−11.10)	−0.307	(−13.32)	−0.320	(−15.94)
Short-term indirect effect Log(*P*)					0.160	(3.39)	0.292	(10.92)	0.233	(7.66)
Short-term direct effect Log(*Y*)					0.101	(3.39)	0.010	(1.34)	0.281	(9.89)
Short-term indirect effect Log(*Y*)					−0.017	(−0.44)	0.060	(1.72)	0.058	(1.52)
Long-term direct effect Log(*P*)	−1.217	(−22.96)	−1.012	(−24.95)	−1.940	(−8.03)	−1.716	(−4.67)	−0.976	(−16.12)
Long-term indirect effect Log(*P*)	0.509	(7.38)	−0.220	(−2.30)	0.561	(0.47)	1.459	(1.04)	0.668	(6.37)
Long-term direct effect Log(*Y*)	0.529	(15.12)	0.591	(10.44)	0.768	(1.87)	0.109	(0.45)	0.883	(10.32)
Long-term indirect effect Log(*Y*)	−0.363	(−7.70)	−0.198	(−2.20)	0.297	(0.14)	0.676	(0.76)	0.332	(3.81)

Notes: t-values in parentheses; * cross-sectional averages at time t and t-1 of the dependent variables; ** two principal components price of cigarettes in neighboring states, but not any other spatial interaction effects. Second, whereas this model contains spatial interaction effects, it is not (yet) dynamic. For these reasons it is interesting to consider the estimation results of the dynamic spatial panel data models.

Columns (3), (4) and (5) of Table 3 reports the direct and indirect effects of the dynamic model, both in the short term and long term, based on the corresponding parameter estimates reported in columns (3), (4) and (5) of Table 2. Column (3) represents a classical dynamic spatial panel data model with spatial and time-period fixed effects. In column (4) the time-period fixed effects are replaced with cross-sectional averages of the dependent variable observed at time t and $t-1$, $\bar{Y}_t = \frac{1}{N}\sum_{i=1}^{N} Y_{it}$ and $\bar{Y}_{t-1} = \frac{1}{N}\sum_{i=1}^{N} Y_{it-1}$. In column (5), both the spatial and time-period effects are replaced by two principal components. A spatially lagged error term is not included.

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 for the model in column (3) with the lowest log-likelihood function value of the three dynamic models amounts to 2(2623.3−1691.4)=1863.8 with 2 degree of freedom, and evidently justifies the extension of the model with dynamic effects.

Another way to look at this are the results of the CD-test statistic. When testing for cross-sectional dependence based on the consumption patterns of the 46 states over the period 1963-1992, this statistic amounts to 101.52 with an average pairwise correlation coefficient of 0.58. This outcome is highly statistically significant, indicating that cross-sectional dependence needs to be accounted for. When controlling for endogenous and exogenous interaction effects this statistic drops down to 25.25 (column (2) of Table 2), and when additionally controlling for fixed effects in space and time it drops further down to −2.28 (column (3) of Table 2). However, only when also controlling for habit persistence, i.e. the coefficient estimate of the variable Y_{t-1} amounts to 0.865 and appears to be highly significicant, does the CD test statistic take a value within the interval [−1.96,+1.96] (0.29, see column (3) of Table 2). It indicates that only this dynamic model is able to effectively factor out the observed cross-sectional dependence in the data.

It is tempting to conclude that the estimation results of the other two dynamic models are comparable, but when taking a closer look especially at the coefficient estimates of the spatial interaction effects in Table 2 we do see differences. Although the spatial autoregressive coefficient of WLog(C) is significant in all three models, its magnitude is almost negligible when controlling for cross-sectional averages. Whereas the spatiotemporal coefficient η of WLog(C)$_{-1}$ is positive and significant when controlling for cross-sectional averages, it is negative and insignificant when controlling for spatial and time-period fixed effects or principal components. Even though the spatial interaction coefficient of the price of cigarettes, WLog(P), is significant in all three models, its magnitude ranges from 0.170 to 0.292. Finally, whereas the spatial interaction coefficient of income, WLog(Y), is insignificant in all three models, its sign is sometimes negative and sometimes positive.

In line with this, also the short and long term direct and spatial spillover effects and their significance levels, reported in Table 3, are different. The explanation is that the short-term direct and spatial spillover effects depend on three parameters, among which two are spatial interaction effects (Eq. (11)), and their long term counterparts on five parameters, among which three spatial interaction effects (Eq. (12)). Any change in the magnitude and the significance levels of these coefficients, therefore works through in the effects estimates derived from them. The biggest difference concerns the long-term price and income spillover effects. Whereas these two effects are positive and significant when controlling for principal components, they are insignificant when controlling for spatial and time-period fixed effects or cross-sectional averages, even if the magnitude of these effects as in the model with spatial and time-period fixed effects are almost similar.

Since the dynamic model with spatial and time-period fixed effects appears to be the only model that effectively filtered out common time trends, i.e., only for this model does the CD-test applied to the residuals of Eq. (9) takes a value within the interval $[-1.96,+1.96]$, the focus below will be on the effects estimates produced by this model. It should be noted, however, that two other recent studies found different results. Cicarelli and Elhorst (2018) find that the dynamic model with cross-sectional averages outperforms its counterpart with spatial and time-period fixed effects. Similarly, Elhorst et al. (2018) find that the dynamic model with principal components outperforms its counterpart with spatial and time-period fixed effects. The conclusion must be that the best model to control for common time trends might differ from one empirical study to another. Since the log-likelihood function values of the models in Table 2 with cross-sectional averages and common factors exceed that of the model with spatial and time-period fixed effects, while their CD-test statistics still take values outside the interval $[-1.96,+1.96]$, another reading is that the spatial weight matrix is not correctly specified. Using the same data set, Halleck Vega and Elhorst (2015) parameterize this matrix and find that it is denser than the binary contiguity matrix, while Kelejian and Piras (2014) find that this matrix is endogenous. It indicates that the log-likelihood function values of the models with cross-sectional averages and common factors might increase even further when a parameterized and/or an endogeneous spatial weight matrix is adopted. This is an interesting issue for further research.

Consistent with microeconomic theory, the short-term direct effects appear to substantially smaller than the long-term direct effects; −0.264 versus −1.940 for the price variable and 0.101 versus 0.768 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: −1.940 versus −1.012 for the price variable and 0.768 versus 0.591 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 elasticity amounts to 0.160 and is highly significant (t-value 3.39). 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 price spatial spillover effect is also significant. A similar result is found by Debarsy et al. (2012). It is to be noted that they estimate the parameters of the model by the Bayesian MCMC estimator developed by Parent and LeSage (2010), whereas we use the consistent bias-corrected ML estimator developed by Lee and Yu (2015). Furthermore, the spatial weights matrix used in that study is based on lengths of state borders in common between each state and its neighboring states, whereas we use a binary contiguity matrix.

The long-term 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, it 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 income spatial spillover effect is not significant. A similar result is found by Debarsy et al. (2012). 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 again 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.

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 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. However, due to the availability of software written in Stata, R and Matlab, they are no major obstacles any more for applied researchers to report and discuss direct and indirect effects estimates in addition to the point estimates of the parameters of the model. In addition, there are no obstacles any more to control for common factors other than time-period fixed effects, including applied researchers familiar with Stata only.

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

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.

Belotti F, Hughes G, Mortari AP (2017) Spatial panel-data models using Stata. The Stata Journal 17(1):139-180

Burridge P, Elhorst JP, Zigova K (2017) Group interaction in research and the use of general nesting spatial models. In: Baltagi BH, LeSage JP, Pace RK (eds) Spatial Econometrics: Qualitative and Limited Dependent Variables (Advances in Econometrics, Volume 37). Emerald Bingley (UK), pp 223 – 258

Ciccarelli C, Elhorst JP (2018) A dynamic spatial econometric diffusion model with common factors: the rise and spread of cigarette consumption in Italy. Regional Science and Urban Economics 72(September):131-142

Debarsy N, Ertur C, LeSage JP (2012) Interpreting dynamic space-time panel data models. Statistical Methodology 9(1-2):158-171

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

Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels. Springer Heidelberg New York Dordrecht London

Elhorst, J.P., Madre, J.-L., Pirotte, A. (2018) Car traffic, habit persistence, cross-sectional dependence, and spatial heterogeneity: New Insights using French departmental data. University of Groningen, working paper under review.

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

Halleck Vega S, Elhorst JP (2015) The SLX model. Journal of Regional Science 55(3): 339-363

Kelejian HH, Piras G (2014) Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. Regional Science and Urban Economics 46(May):140-149

Kelejian HH, Piras G (2017) Spatial econometrics. Elsevier, London.

Lee LF, Yu J (2015) Spatial panel data models. In: Baltagi BH (ed) The Oxford Handbook of Panel Data. Oxford, University Press, pp 363 – 401

Lee LF, Yu J (2016) Identification of spatial Durbin models. Journal of Applied Econometrics 31(1): 133-162

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

Pesaran MH (2006) Estimation and inferences in large heterogenous panels with a multifactor structure. Econometrica 74(4): 967-1012

Pesaran MH (2015a) Time series and panel data econometrics. Oxford University Press

Pesaran MH (2015b) Testing weak cross-sectional dependence in large panels. Econometric Reviews 34(6-10): 1089-1116

Shi W, Lee LF (2017) Spatial dynamic panel data model with interactive fixed effects. Journal of Econometrics 197(2):323-347

Cross-references in the Handbook of Regional Science

LeGallo J (2014) Cross-section spatial regression model. In: Fischer MM, Nijkamp P (eds) Handbook of Regional Science. Springer Heidelberg, pp 1511-1533

LeSage JP, Pace RK (2014) Interpreting spatial econometric models. In: Fischer MM, Nijkamp P (eds) Handbook of Regional Science. Springer Heidelberg, pp 1535-1552

Mills JA, Parent O (2014) Bayesian MCMC estimation. In: Fischer MM, Nijkamp P (eds) Handbook of Regional Science. Springer Heidelberg, pp 1571-1595

Pace RK (2014) Maximum likelihood estimation. In: Fischer MM, Nijkamp P (eds) Handbook of Regional Science. Springer Heidelberg, pp 1553-1569

Prucha IR (2014) Instrumental variables/method of moments estimation. In: Fischer MM, Nijkamp P (eds) Handbook of Regional Science. Springer Heidelberg, pp 1597-1617