J Geogr Syst (2012) 14:5–28
DOI 10.1007/s10109-011-0158-4

ORIGINAL ARTICLE

Dynamic spatial panels: models, methods, and inferences

J. Paul Elhorst

Received: 1 February 2011 / Accepted: 26 July 2011 / Published online: 24 September 2011

Abstract This paper provides a survey of the existing literature on the specification and estimation of dynamic spatial panel data models, a collection of models for spatial panels extended to include one or more of the following variables and/or error terms: a dependent variable lagged in time, a dependent variable lagged in space, a dependent variable lagged in both space and time, independent variables lagged in time, independent variables lagged in space, serial error autocorrelation, spatial error autocorrelation, spatial-specific and time-period-specific effects. The survey also examines the reasoning behind different model specifications and the purposes for which they can be used, which should be useful for practitioners.

Keywords Dynamic effects · Spatial spillover effects · Identification · Estimation methods · Stationarity conditions

JEL Classification C21 · C23 · C51

1 Introduction

In the last decade, the spatial econometrics literature has exhibited a growing interest in the specification and estimation of dynamic regression equations based on spatial panels. This interest can be explained by the fact that panel data offer researchers extended modeling possibilities when compared to the single equation cross-sectional setting, which was the primary focus of the spatial econometrics literature for a long time. Ideally, a dynamic model in space and time should be able to deal with (1) serial dependence between the observations on each spatial unit

J. P. Elhorst (✉)
Department of Economics and Econometrics, University of Groningen,
P.O. Box 800, 9700 AV Groningen, The Netherlands
e-mail: j.p.elhorst@rug.nl

over time, (2) spatial dependence between the observations at each point in time, (3) unobservable spatial and/or time-period-specific effects, and (4) endogeneity of one or more of the regressors other than dependent variables lagged in space and/or time. The first problem is the domain of the voluminous time-series econometrics literature (Hamilton 1994; Enders 1995; Hendry 1995), the second problem of the spatial econometrics literature (Anselin 1988; Anselin et al. 2008; LeSage and Pace 2009), and the last two problems of the panel data econometrics literature (Hsiao 2003; Arrelano 2003; Baltagi 2005), to mention just a few well-known textbooks in these fields.

This paper provides a survey of the existing literature on dynamic spatial panel data models.[1] Ten years ago, there was no straightforward estimation procedure for dynamic spatial panel data models. This was because methods developed for dynamic but non-spatial and for spatial but non-dynamic panel data models produced biased estimates when these methods/models were put together. The literature reviewed in this paper includes the main methodological studies that have attempted to solve this shortcoming.

Ten years after the first paper of Elhorst (2001) dealing with dynamic models in space and time, it is also time to ask the question for which purposes the different model specifications that have been put forward in the literature can be used. To answer that question, we deal with short-term effects, long-term effects, direct effects, and indirect or spatial spillover effects.

2 A generalized dynamic model in space and time

In this section, we initially focus on a dynamic model in space and time that generalizes several simpler models that have been considered in the literature. We hasten to stress that this generalized model suffers from identification problems and thus is not useful for empirical research. However, when these econometric models are arranged in a framework and their mutual relationships exemplified, it may help us to identify which models are the most likely candidates to study space–time data, dependent on the purpose of a particular empirical study.

The most general model when written in vector form for a cross-section of observations at time t reads as

$$\begin{aligned} Y_t = \tau Y_{t-1} + \delta W Y_t + \eta W Y_{t-1} + X_t \beta_1 \\ + W X_t \beta_2 + X_{t-1} \beta_3 + W X_{t-1} \beta_4 + Z_t \theta + v_t, \end{aligned} \quad (1a)$$

$$v_t = \gamma v_{t-1} + \rho W v_t + \mu + \lambda_t \iota_N + \varepsilon_t, \quad (1b)$$

$$\mu = \kappa W \mu + \xi, \quad (1c)$$

where Y_t denotes an $N \times 1$ vector consisting of one observation of the dependent variable for every spatial unit ($i = 1,\ldots,N$) in the sample at time t ($t = 1,\ldots,T$), X_t is

[1] Recent surveys on static spatial panel data models are provided by Elhorst (2010a) and Lee and Yu (2010a). Lee and Yu's (2010a) overview also deals with dynamic spatial panel data models. We will come back to this paper several times.

an $N \times K$ matrix of exogenous explanatory variables, and Z_t is an $N \times L$ matrix of endogenous explanatory variables. A vector or a matrix with subscript $t-1$ denotes its serially lagged value, while a vector or a matrix premultiplied by W denotes its spatially lagged value. The $N \times N$ matrix W is a nonnegative matrix of known constants describing the spatial arrangement of the units in the sample. Its diagonal elements are set to zero by assumption, since no spatial unit can be viewed as its own neighbor. The scalars τ, δ, and η are the response parameters of successively the dependent variable lagged in time, Y_{t-1}, the dependent variable lagged in space, WY_t, and the dependent variable lagged in both space and time, WY_{t-1}. The restrictions that need to be imposed on these parameters and on W to obtain a stationary model are set out in the next section. The $K \times 1$ vectors β_1, β_2, β_3, and β_4 contain the response parameters of the exogenous explanatory variables and the $L \times 1$ vector θ of the endogenous explanatory variables in the model.

The $N \times 1$ vector v_t reflects the error term specification of the model, which is assumed to be serially correlated and to be spatially correlated; γ is the serial autocorrelation coefficient and ρ is the spatial autocorrelation coefficient. In contrast to Eq. 1a, the error term lagged in both space and time, Wv_{t-1}, is not included, since we did not come across this term in the literature. The $N \times 1$ vector $\mu = (\mu_1, \ldots, \mu_N)^T$ contains spatial-specific effects, μ_i, and are meant to control for all spatial-specific, time-invariant variables whose omission could bias the estimates in a typical cross-sectional study (Baltagi 2005). Similarly, λ_t ($t = 1,\ldots T$) denote time-period-specific effects, where ι_N is a $N \times 1$ vector of ones, meant to control for all time-specific, unit-invariant variables whose omission could bias the estimates in a typical time-series study. These spatial- and time-period-specific effects may be treated as fixed or as random effects. In addition to this, the spatial-specific effects are assumed to be spatially autocorrelated with spatial autocorrelation coefficient κ. Finally, $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^T$ and ξ are vectors of i.i.d. disturbance terms, whose elements have zero mean and finite variance σ^2 and σ_ξ^2, respectively.

2.1 Stationarity

To achieve stationarity in a dynamic spatial panel data model, restrictions need to be imposed on the parameters of the model and on the spatial weights matrix W.

Lee (2004) shows that the matrix $(I_N - \kappa W)$ in a cross-sectional equation like (1c) should be non-singular and that the characteristic roots of the matrix $(I_N - \kappa W)^{-1}$ should lie in the unit circle, where I_N represents the identity matrix of order N. For a symmetric W, this condition is satisfied as long as κ is in the interior of $(1/\omega_{\min},1/\omega_{\max})$, where $\omega_{\min}$ denotes the smallest (i.e., most negative) and $\omega_{\max}$ the largest real characteristic root of W. If W is row-normalized subsequently, the latter interval takes the form $(1/\omega_{\min},1)$, since the largest characteristic root of W equals unity in this situation. If W is an asymmetric matrix before it is row-normalized, it may have complex characteristic roots. LeSage and Pace (2009, pp. 88–89) demonstrate that in that case κ is restricted to the interval $(1/r_{\min},1)$, where $r_{\min}$ equals the most negative purely real characteristic root of W after this matrix has been row-normalized. Finally, one of the following two conditions

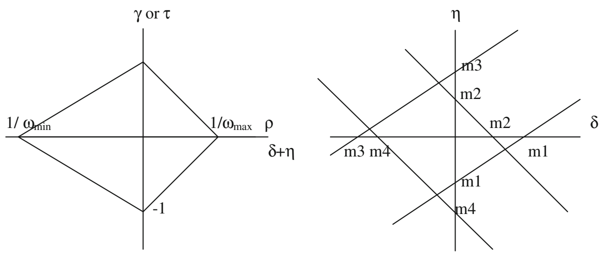

Fig. 1 Stationary regions of different model equations. **a** Stationarity region of $\delta + \eta$ and τ in Eq. (1a), and ρ and γ in Eq. (1b). **b** Stationarity region of δ and η, given τ, in Eq. (1a). *m1*, *m2 m3*, and *m4* denote points of intersection with the horizontal or vertical axes, where $m1 = \frac{1+\tau}{\omega_{\max}} > 0$, $m2 = \frac{1-\tau}{\omega_{\max}} > 0$, $m3 = \frac{1+\tau}{\omega_{\min}} < 0$, and $m4 = \frac{1-\tau}{\omega_{\min}} < 0$

should be satisfied: (a) the row and column sums of the matrices W and $(I - \kappa W)^{-1}$ before W is row-normalized should be uniformly bounded in absolute value as N goes to infinity or (b) the row and column sums of W before W is row-normalized should not diverge to infinity at a rate equal to or faster than the rate of the sample size N. Condition (a) is originated by Kelejian and Prucha (1998, 1999), and condition (b) by Lee (2004). Both conditions limit the cross-sectional correlation to a manageable degree, i.e., the correlation between two spatial units should converge to zero as the distance separating them increases to infinity.

In a similar fashion, Elhorst (2008a) demonstrates that the characteristic roots of the matrix $\gamma(I_N - \rho W)^{-1}$ in a space–time equation like (1b) should lie within the unit circle. Since the smallest and largest characteristic roots of this matrix take the form $\gamma/(1 - \rho\omega_{\min})$ and $\gamma/(1 - \rho\omega_{\max})$, or vice versa (dependent on whether γ is positive or negative), stationarity in time requires the conditions

$$|\gamma| < 1 - \rho\omega_{\max} \quad \text{if } \rho \geq 0, \tag{2a}$$

$$|\gamma| < 1 - \rho\omega_{\min} \quad \text{if } \rho < 0. \tag{2b}$$

These stationarity conditions are graphed in Fig. 1a and show that there exists a trade-off between the serial and spatial autocorrelation coefficients.

Finally, Elhorst (2001) derives that the characteristic roots of the matrix $(I - \delta W)^{-1}$ $(\tau I + \eta W)$ in a space–time equation like (1a) should lie within the unit circle (re-derived in Parent and LeSage 2011), which is the case when

$$\tau < 1 - (\delta + \eta)\omega_{\max} \quad \text{if } \delta + \eta \geq 0, \tag{3a}$$

$$\tau < 1 - (\delta + \eta)\omega_{\min} \quad \text{if } \delta + \eta < 0, \tag{3b}$$

$$-1 + (\delta - \eta)\omega_{\max} < \tau \quad \text{if } \delta - \eta \geq 0, \tag{3c}$$

$$-1 + (\delta - \eta)\omega_{\min} < \tau \quad \text{if } \delta - \eta < 0. \tag{3d}$$

If ρ is replaced by $\delta + \eta$ and γ by τ, then the stationarity condition between $\delta + \eta$ on the one hand and τ on the other are similar to those graphed for ρ and γ in Fig. 1a. This implies that there exists a trade-off between the serial autoregressive coefficient and the sum of the two spatial autoregressive coefficients. The stationarity conditions between δ and η, given τ, are graphed in Fig. 1b. This figure shows that the stationarity region of the two spatial autoregressive coefficients takes the form of a rhombus. The location and the size of this rhombus depend on τ and the smallest and largest characteristic roots of the spatial weights matrix.

The graphs in Fig. 1 make clear that the stationarity region implied by the restriction $|\tau| + |\delta| + |\eta| < 1$, put forward in Yu et al. (2008), is too restrictive. For example, whereas the combination of values $(\tau, \delta, \eta) = (0.1, 0.9, -0.1)$ based on the restriction $|\tau| + |\delta| + |\eta| < 1$ should be rejected, it is not based on the results presented in this paper. This is because the largest characteristic root of the matrix $(I - \delta W)^{-1}(\tau I + \eta W)$ is smaller than one for this combination of values. By contrast, the stationarity region implied by the restriction $\tau + \delta + \eta < 1$, put forward in Lee and Yu (2010a), is not restrictive enough. For example, whereas the combination of values $(\tau, \delta, \eta) = (1.1, -0.2, 0.0)$ is permitted by the restriction $\tau + \delta + \eta < 1$, it should be rejected based on the results presented in this paper. This is because the largest characteristic root of the matrix $(I - \delta W)^{-1}(\tau I + \eta W)$ is greater than one for this combination of values, indicating that the model would explode under these circumstances.

If a model appears to be unstable, that is, if the parameter estimates do not satisfy one of the stationarity conditions, Lee and Yu (2010a) propose to take every variable in Eq. 1 in deviation of its spatially lagged value. Mathematically, this is equivalent with multiplying Eq. 1 by the matrix $(I_N - W)$. The largest characteristic root $\omega_{\max}$, part of the stationarity conditions (3a) and (3c), may then be replaced by $\omega_{\max-1}$, the second largest characteristic root of the spatial weights matrix W. Since these newly obtained restrictions are less restrictive than the original ones, the spatial first-differenced model might be stable as a result. In a study on financial liberalization of 62 countries over the period 1976–2005, Elhorst et al. (2010a) estimate both a dynamic spatial panel data model in levels and a model reformulated in spatial first-differences. They find that the coefficient estimates of the first specification point to unstable components in the dependent variable, whereas the coefficient estimates of the second specification do not. Up to now, this is one of the few empirical studies that have found that taking spatial first-differences is an effective tool to obtain a stable model.

In conclusion, we can say that the stationarity conditions on the spatial and temporal parameters shown in (2) and (3) go beyond the standard condition $|\tau| < 1$ in serial models and the standard condition $1/\omega_{\min} < \delta < 1/\omega_{\max}$ in spatial models, and that they are considerably more difficult to work with.

The stationarity conditions that need to be imposed on the $N \times N$ spatial weights matrix W in a panel data setting are set forth in Yu et al. (2008). The matrix $I_N - pW$ for $p = \delta, \rho$ should be non-singular, and the row and column sums of the matrices W and $(I_N - pW)^{-1}$ should be uniformly bounded in absolute value as N goes to infinity. In addition, $\sum_{h=1}^{\infty} \text{abs}\{[(I_N - \delta W)^{-1}(\tau I_N + \eta W)]^h\}$ should be uniformly bounded. Let ω_i $(i = 1,\ldots,N)$ denote the characteristic roots of W and R_N the corresponding $N \times N$ matrix of normalized characteristic vectors, then this formula may be rewritten as $\sum_{h=1}^{\infty} \text{abs}\{[R_N D_N R_N^{-1}]^h\}$, where D_N is a diagonal matrix whose diagonal elements are $(\tau + \eta\omega_i)/(1 - \delta\omega_i)$ $(i = 1,\ldots,N)$. This expression represents the stationarity region of the parameters τ, η, and δ shown in Fig. 1. The assumption that the row and column sums of W before row-normalization should not diverge to infinity at a rate equal to or faster than the rate of the sample size N, which is made in the cross-sectional setting, is not explicitly made in a panel data setting. In this respect, a couple of studies have paid attention to a spatial weights matrix with equal weights, that is, a matrix where all non-diagonal elements are defined as $1/(N-1)$. Lee (2004), Kelejian and Prucha (2002), and Elhorst (2010c) demonstrate that this matrix leads to inconsistent parameter estimates in a cross-sectional setting, since the ratio between the row sums of this matrix before normalization and the sample size, $(N-1)/N$, converges to one instead of zero as N goes to infinity. By contrast, in a panel data setting, this spatial weights matrix causes no problems, provided that time-period effects are not included (see Kelejian and Prucha 2002; Kelejian et al. 2006). However, if time-period fixed effects are also considered, the estimators to be discussed in this paper become inconsistent again. In sum, a spatial weights matrix with equal weights is an extremely interesting matrix from a theoretical point of view because it may help to understand the stationarity conditions. In empirical research, however, this matrix is hardly relevant.

2.2 Fixed versus random effects

The spatial-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, while in the random effects model, μ_i is treated as a random variable that is independently and identically distributed with zero mean and variance σ_μ^2. Furthermore, it is assumed that the random variables μ_i and ε_{it} are independent of each other. A similar distinction can be made for time-period-specific effects.

Whether the random effects model is an appropriate specification for a given set of adjacent spatial units, such as all counties of a state or all regions in a country, remains controversial. If the data happen to be a random sample of the population, unconditional inference about the population necessitates estimation with random effects. If, however, the objective is limited to making conditional inferences about the sample, then fixed effects should be specified. Beenstock and Felsenstein (2007) point out that the random effects model should be the default option in principle, since researchers are usually interested in making unconditional inferences about the population and the fixed effects model would lead to an enormous loss of degrees of freedom. However, "if the sample happens to be the population" (Beenstock and Felsenstein 2007, p. 178), or if the population may be said "to be sampled exhaustively" (Nerlove and Balestra 1996, p. 4),[2] specific effects should be fixed because each spatial unit represents itself and has not been sampled randomly. Similar observations have been made by Anselin (1988, p. 51) "the individual spatial units have characteristics that actually set them apart from a larger population," and Beck (2001, p. 272), "the critical issue is that the spatial units be fixed and not sampled, and that inference be conditional on the observed units" [see also Hsaio (2003, p. 43) for a more general explanation]. In addition, the traditional

[2] This remark through Balestra and Nerlove is striking especially since they are the devisers of the random effects model (Balestra and Nerlove 1966).

assumption of zero correlation between the random effects μ_i and the explanatory variables, which also needs to be made, is particularly restrictive.

Many studies that derive test statistics for spatial effects or that develop estimation methods for the parameters in dynamic panel data models with random effects can be criticized for not paying (sufficient) attention to the reasoning behind the random effects specification. The motivation to consider random effects rather than fixed effects in one of the first studies that derived a Lagrange Multiplier (LM) test for spatial interaction effects when combining time-series cross-section data consists of all-in-all one single sentence (Baltagi et al. 2003, p. 124): "Heterogeneity across the cross-sectional units is usually modeled with an error component model."[3] Experience shows that spatial econometricians tend to work with space–time data of adjacent spatial units located in unbroken study areas; otherwise, the spatial weights matrix cannot be defined. Consequently, the study area often takes a form similar to all counties of a state or all regions in a country. Under these circumstances the fixed effects model is more appropriate than the random effects model. For example, to explain cigarette demand using a panel of 46 US states over the period 1963–1992, Yang et al. (2006) adopt a dynamic spatial panel data model with random effects. However, since these states cover almost the whole United States, a fixed effects model would have been a better choice (cf. Elhorst 2005).

This position does not mean that the fixed effects model is free of problems. First, the spatial fixed effects can only be estimated consistently when T is sufficiently large, because the number of observations available for the estimation of each μ_i is T. Sampling more observations in the cross-sectional domain is no solution for insufficient observations in the time domain, since the number of unknown parameters increases as N increases, a situation known as the incidental parameters problem. Fortunately, the inconsistency of μ_i is not transmitted to the estimator of the slope coefficients β, since it is not a function of the estimated μ_i. Consequently, the incidental parameters problem does not matter when β are the coefficients of interest and the spatial fixed effects μ_i are not, which is the case in many empirical studies.

Second, any variable that does not change over time or only varies a little cannot be estimated when controlling for spatial fixed effects. This is the main reason for many studies not to control for spatial fixed effects, for example, because such variables are the main focus of the analysis. On the other hand, 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 (Greene 2005). This also holds true for spatial fixed effects and is known as the omitted regressor bias.

[3] The loss of degrees of freedom if N is large may be one of the main reasons to adopt the random effects model (Cressie and Wikle 2011). However, the number of observations in the cross-sectional domain of most empirical studies in spatial economics and economic geography is smaller than 1,000. Furthermore, methodological studies considering further extensions of the random effects models with spatial interaction effects also tend to present Monte Carlo simulation experiments for relatively small values of N ($N < 500$).

3 Feasible models

Figure 2 presents two regressions equations that have extensively been discussed in the econometric literature: the dynamic model without spatial interaction effects and the spatial model without dynamic effects. We will first briefly discuss these two models and then consider mixtures of both models.

3.1 Dynamic non-spatial panel data models

A panel data model without spatial interaction effects and without dynamic effects (i.e., a static panel data model) can be estimated by the least-squares dummy variables (LSDV) estimator if the spatial-specific effects μ are treated as fixed effects, and by the generalized least-squares (GLS) estimator if the spatial-specific effects μ are treated as random effects (Hsiao 2003, Ch. 3; Baltagi 2005, Ch. 2). The most serious estimation problem caused by the extension of this model with a dependent variable lagged in time, Y_{t-1}, is that these two estimators become inconsistent *if T is fixed, regardless of the size of N* (Hsiao 2003, Ch. 4; Arrelano 2003; Baltagi 2005, Ch. 8). This is because the right-hand variable Y_{t-1} is correlated with the spatial-specific effect μ, and uncorrelatedness of regressors and disturbances is one of the basic conditions that needs to be satisfied in regression analysis. Three procedures to remove this inconsistency if T is fixed have been developed.

The first and most popular procedure is generalized method-of-moments (GMM). By defining and solving a set of moment conditions that need to be satisfied at the true values of the parameters to be estimated, one obtains a set of exogenous variables correlated with Y_{t-1} but orthogonal to the errors, which as a results can be used to instrument Y_{t-1}. The Arrelano and Bond (1991) *difference* GMM estimator is based on moment conditions after taking first differences to eliminate the spatial-specific effects. Typically, this GMM estimator instruments ΔY_{t-1} by the variables Y_1 up to Y_{t-2} and X_1 up to X_{t-1} ($t \geq 3$). In practice, the difference GMM estimator has been shown to perform poorly on data with persistent series. The explanation is that, under these circumstances, the lagged levels of variables tend to have only weak correlation with the first-differenced lagged dependent variable. The Blundell

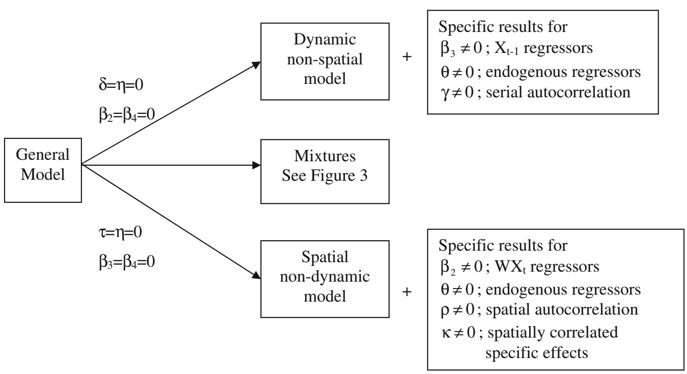

Fig. 2 Two feasible models that ignore either spatial or dynamic effects

and Bond (1998) *system* GMM estimator has been shown to offer much increased efficiency and less finite sample bias compared with the difference GMM estimator, since this estimator also utilizes lagged first differences for the equation in levels. Typically, this GMM estimator also instruments Y_{t-1} by the variables ΔY_1 up to ΔY_{t-2} and ΔX_2 up to ΔX_{t-1} ($t \geq 3$). See Baltagi (2005) for a summary of the main results and the main references, and Kukenova and Monteiro (2009) for a more detailed mathematical exposition.

The second procedure applies maximum likelihood (ML) based on the *unconditional* likelihood function of the model. Regression equations that include variables lagged one period in time are often estimated conditional upon the first observations. Nerlove (1999, p. 139), however, points out that conditioning on those initial values is an undesirable feature, especially when the time dimension of the panel is short. If the process generating the data in the sample period is stationary, the initial values convey a great deal of information about this process since they reflect how it has operated in the past. By taking account of the density function of the first observation of each time-series of observations, the unconditional likelihood function is obtained. This procedure has been applied successfully to random effects dynamic panel data models formulated in levels (Bhargava and Sargan 1983). Unfortunately, the unconditional likelihood function does not exist when applying this procedure to the fixed effects model, even without exogenous explanatory variables. The reason is that the coefficients of the fixed effects cannot be estimated consistently, since the number of these coefficients increases as N increases. The standard solution to eliminate these fixed effects from the regression equation by demeaning the Y and X variables also does not work, because this technique creates a correlation of order ($1/T$) between the serial lagged dependent variable and the demeaned error terms, known as the Nickell (1981) bias, as a result of which the common parameter τ cannot be estimated consistently. Only when T tends to infinity, does this inconsistency disappear. More recently, Hsiao et al. (2002) have suggested an alternative procedure for the fixed effects dynamic panel data model. This procedure first-differences the model to eliminate the spatial fixed effects and then considers the unconditional likelihood function of the first-differenced model taking into account the density function of the first first-differenced observations on each cross-sectional unit. They find that this likelihood function is well defined, depends on a fixed number of parameters, and satisfies the usual regularity conditions. Thereupon, they conclude that the ML estimator is consistent and asymptotically normally distributed when N tends to infinity, regardless of the size of T. They also find that the ML estimator is asymptotically more efficient than the GMM estimator.

The third procedure is to bias-correct the LSDV estimator. Kiviet (1995), Hahn and Kuersteiner (2002), and Bun and Carree (2005) develop bias correction procedures when both the number of cross-sectional units (N) and the number of time points (T) in the sample go to infinity such that the limit of the ratio of N and T exists and is bounded between zero and infinity ($0 < \lim(N/T) < \infty$). The only problem is that in most empirical studies based on space–time data, the most relevant asymptotics are believed to be N tends to infinity and T is fixed. When T is fixed, the spatial-specific effects must be eliminated by first-differencing, whereas first-differencing is not necessary when T tends to infinity.

Specific problems occur when the dynamic panel data model is also extended to include a serially autocorrelated error term, $\gamma \neq 0$, regressors lagged in time, X_{t-1}, or endogenous regressors, Z_t. The consistency of the difference and system GMM estimators relies on the assumptions that there is no first-order serial autocorrelation in the errors of the level equation, and that the instruments are truly exogenous and therefore valid to define the moment conditions. The Arrelano and Bond (1991) test for serial autocorrelation tests the hypothesis that there is no second-order serial correlation in the first-differenced residuals, which in turn implies that the errors from the level equation are serially uncorrelated. However, correlation coefficients of observations on variables made in single spatial units 1 year, 2 years, up to $T-1$ years apart tend to be large and to diminish only slightly over time (Elhorst 2008b). Consequently, the null hypothesis of no serial autocorrelation of the error terms must often be rejected. One remedial reaction could be to re-estimate the model using methods that assume that the errors are generated by a first-order serial autoregressive process, but this approach has been severely criticized. Rather than improving an initial model when it appears to be unsatisfactory, Hendry (1995, Ch. 7) argues that it is better to start with a more general model containing a series of simpler models nested within it as special cases. The general model Hendry recommends as a generalization of the first-order serial autocorrelation model for time-series data is the first-order serial autoregressive distributed lag model, a linear dynamic regression model in which the dependent variable Y_t is regressed on Y_{t-1} and the explanatory variables X_t and X_{t-1}. For this reason, dynamic panel data models extended to include explanatory variables X_{t-1} are more popular than dynamic panel data models extended to include serial autocorrelation. Another reason is that the econometric literature has paid much attention to estimators of the covariance matrix that are robust to serial autocorrelation and heteroskedasticity, affecting inferences regarding the statistical significance of the explanatory variables in the model (Newey and West 1987; Greene 2005).

If one or more of the explanatory variables are endogenous (Z_t), they need to be instrumented too. Since the GMM estimator already instruments Y_{t-1}, this estimator can easily be extended to include additional endogenous explanatory variables. See Kukenova and Monteiro (2009) how to adjust the GMM estimator when having both endogenous and exogenous explanatory variables (Z_t and X_t).

3.2 Spatial non-dynamic panel data models

When imposing the parameter restrictions $\tau = \eta = 0$ and $\beta_3 = \beta_4 = 0$ on Eq. 1a, as shown in Fig. 2, one obtains the spatial Durbin model, which reads as

$$Y_t = \delta W Y_t + X_t \beta_1 + W X_t \beta_2 + v_t. \tag{4}$$

The spatial Durbin model further simplifies to the spatial lag model when imposing the restriction $\beta_2 = 0$ and to the spatial error model when imposing the restriction $\beta_2 + \delta\beta_1 = 0$.

The estimation of static spatial panel data models is extensively discussed in Elhorst (2003, 2010a) and Lee and Yu (2010a, b). Elhorst (2003, 2010a) presents the ML estimator of the fixed effects spatial lag model and of the random effects spatial lag model. It should be noted 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 spatial lag model can be estimated by concentrating out the fixed effects first, called demeaning. In a model with spatial fixed effects but no time-period fixed effects, this procedure takes the form $D_t - \frac{1}{T}\sum_t D_t$ for $D_t = Y_t$, WY_t, X_t, WX_t (see Baltagi 2005; and Elhorst 2010a for more 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. Similarly, the random effects spatial lag model can be estimated by quasi-demeaning each variable in the model, $D_t - (1-\psi)\frac{1}{T}\Sigma_t D_t$, where ψ is an additional parameter to be estimated.

Lee and Yu (2010b) show that the ML estimator of the spatial lag 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 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 random effects spatial error model. Recently, these routines have been updated for the bias correction procedure of Lee and Yu (2010b).

An important development in the spatial econometrics literature is the increasing attention for direct, indirect, and spatial spillover effects of the independent variables. This applies to cross-sectional as well as to spatial panel data models. By rewriting the spatial Durbin model as (leaving the subscript t aside for notational simplicity)

$$Y = (I - \delta W)^{-1}(X\beta_1 + WX\beta_2) + (I - \delta W)^{-1}v, \tag{5}$$

the matrix of partial derivatives of Y with respect to the kth explanatory variable of X in unit 1 up to unit N (say x_{ik} for $i = 1,\ldots,N$, respectively) can be seen to be

$$\begin{aligned}\begin{bmatrix}\frac{\partial Y}{\partial x_{1k}} & \cdots & \frac{\partial Y}{\partial x_{Nk}}\end{bmatrix} &= \begin{bmatrix}\frac{\partial y_1}{\partial x_{1k}} & \cdots & \frac{\partial y_1}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_N}{\partial x_{1k}} & \cdots & \frac{\partial y_N}{\partial x_{Nk}}\end{bmatrix} \\ &= (I_N - \delta W)^{-1}\begin{bmatrix}\beta_{1k} & w_{12}\beta_{2k} & \cdots & w_{1N}\beta_{2k} \\ w_{21}\beta_{2k} & \beta_{1k} & \cdots & w_{2N}\beta_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\beta_{2k} & w_{N2}\beta_{2k} & \cdots & \beta_{1k}\end{bmatrix} \\ &= (I - \delta W)^{-1}[\beta_{1k}I_N + \beta_{2k}W],\end{aligned} \tag{6}$$

where w_{ij} is the (i, j)th element of W, β_{1k} is the kth element of the vector β_1, and β_{2k} is the kth element of the vector β_2. To obtain the last term in Eq. 6, we made use of the property that the diagonal elements of W are zero.

The expression in (6) illustrates that the partial derivatives of Y with respect to the kth explanatory variable in the spatial Durbin model have the following properties. First, if a particular explanatory variable in a particular unit changes, not only will the dependent variable in that unit itself change but also the dependent variables in other units. The first is called a *direct effect* and the second an *indirect effect*. It should be noted that indirect effects do not occur if both $\delta = 0$ and $\beta_{2k} = 0$. Second, indirect effects that occur if $\beta_{2k} \neq 0$ are known as *local effects*, as opposed to indirect effects that occur if $\delta \neq 0$ and that are known as *global effects*. Local effects got their name because they arise only from a unit's neighborhood set; if the element w_{ij} of the spatial weights matrix is zero (non-zero), then the effect of x_{jk} on y_i is also zero (non-zero). Global effects got their name because they also arise from units that do not belong to a unit's neighborhood set. This follows from the fact that the matrix $(I_N - \delta W)^{-1}$, in contrast to W, does not contain zero elements (provided that $\delta \neq 0$).

Since the direct and the indirect effects are different for different units in the sample, LeSage and Pace (2009) propose to report one direct effect measured by the average of the diagonal elements, and one indirect effect measured by the average of the row sums of the non-diagonal elements of that matrix. The results reported for model 0 in Table 1 give an overview of the type of effects that can be calculated from the static spatial Durbin model, where the superscript $\bar{d}$ denotes the operator that calculates the mean diagonal element of a matrix and the superscript $\overline{\text{rsum}}$ denotes the operator that calculates the mean row sum of the non-diagonal elements.

One of the strengths of the spatial Durbin model is that it does not impose prior restrictions on the magnitude of the indirect effects. These indirect effects, also known as *spatial spillover* effects, are often the main focus of an empirical study using spatial econometric techniques. In a non-spatial model or in a spatial error model, these spatial spillover effects are set to zero by construction, while in a spatial lag model, the magnitude of these spatial spillover effects in relation to the direct effects is the same for every explanatory variable (Elhorst 2010c). These restrictions make clear that these models are less appropriate in empirical research than the spatial Durbin model. In contrast, the main shortcoming of a static spatial Durbin model is that it cannot be used to calculate short-term effects of the explanatory variables (see the last column of Table 1).

Specific problems occur when the static spatial panel data model is also extended to include a spatially autocorrelated error term, $\rho \neq 0$, or endogenous regressors, Z_t. Moran's I of observations made on variables in different spatial units at one particular point in time often support the view that neighboring values are more similar than those further apart (Elhorst 2008b). Consequently, the null hypothesis of no spatial autocorrelation among the error terms must often be rejected if such variables are related to each other. One remedial reaction could be to re-estimate the model using methods that assume that the errors are generated by a first-order spatial autoregressive process, but this approach has been criticized. Rather than improving an initial model when it appears to be unsatisfactory, Burridge (1981) argues that it is better to start with a more general model containing a series of simpler models nested within it as special cases. The general model Burridge

Table 1 Short-term, long-term, direct and indirect (spatial spillover) effects of different models

Type of model	Short-term direct effect	Short-term indirect effect	Long-term direct effect	Long-term indirect effect	Shortcoming
0. Static spatial Durbin model	–	–	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	No short-term effects
1. Error terms lagged in space and/or in time	–	–	β_{1k}	–	No short-term effects No indirect effects
2. Dynamic model + spatial error corr.	β_{1k}	–	$\beta_{1k}/(1-\tau)$	–	No indirect effects
3. Dynamic spatial Durbin model	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	$[((1-\tau)I-(\delta+\eta)W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[((1-\tau)I-(\delta+\eta)W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	Parameters not identified
4. $\beta_2=0$	$[(I-\delta W)^{-1}(\beta_{1k}I_N)]^{\bar{d}}$	$[(I-\delta W)^{-1}(\beta_{1k}I_N)]^{\overline{\mathrm{rsum}}}$	$[((1-\tau)I-(\delta+\eta)W)^{-1}(\beta_{1k}I_N)]^{\bar{d}}$	$[((1-\tau)I-(\delta+\eta)W)^{-1}(\beta_{1k}I_N)]^{\overline{\mathrm{rsum}}}$	Ratio ind./dir effects the same for every X
5. $\delta=0$	$[(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	$[((1-\tau)I-\eta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[((1-\tau)I-\eta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	No short-term global indirect effects
6. $\eta=-\tau\delta$	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	$\left[\frac{1}{1-\tau}(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)\right]^{\bar{d}}$	$[\frac{1}{1-\tau}(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	Ratio ind./dir effects constant over time
7. $\eta=0$	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	$[((1-\tau)I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}}$	$[((1-\tau)I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}$	–

recommends as a generalization of the first-order spatial autocorrelation model for spatial data is the first-order spatial autoregressive distributed lag model, a linear dynamic regression model in which the dependent variable Y is regressed on WY and the explanatory variables X and WX. Today, this model is known as the spatial Durbin model. Another reason to adopt the spatial Durbin model rather than the spatial error model is put forward by 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 (Greene 2005, pp. 133–134). In contrast, ignoring spatial dependence in the disturbances, if present, will only cause a loss of efficiency. Furthermore, the spatial Durbin model produces unbiased coefficient estimates, also if the true data-generation process is a spatial lag or a spatial error model. A final reason is that the spatial econometric literature has started to develop estimators of the covariance matrix that are robust to spatial autocorrelation (Driscoll and Kraay 1998) and spatial autocorrelation and heteroskedasticity (Kelejian and Prucha 2010), affecting inferences regarding the statistical significance of the explanatory variables in the model.

If one or more of the explanatory variables are endogenous (Z_t), they need to be instrumented. In applied econometrics work, the presence of endogenous variables on the right-hand side is a common occurrence, as endogeneity may be the result of measurement errors on explanatory variables, of omitted variables correlated with included explanatory variables or of the existence of an unknown set of simultaneous structural equations. For the time being, the best estimation method under these circumstances is the IV/GMM estimator developed by Fingleton and LeGallo (2008; see also Elhorst 2010c for more discussion). The main argument of applying GMM estimators rather than traditional spatial maximum likelihood estimators is that the former can also be used to instrument endogenous explanatory variables (other than the variables Y_{t-1} and WY_t), while the latter cannot. ML estimators of models with a spatial lag and additional endogenous variables do not feature in the spatial econometrics literature and would be difficult, if not impossible, to derive.

3.3 A taxonomy of dynamic models in space and time

Figure 3 presents six different models that have mixed dynamics in both space and time.

A first set of studies (model 1 in Fig. 3) have mixed space and time in the error term specification. The parameters in Eqs. 1b and 1c that are allowed to vary and those that have been not been included in these studies are reported below in parentheses. Baltagi et al. (2003) consider the testing of spatial error correlation in a model with spatial random effects (ρ, μ; γ, λ, κ not included). Baltagi et al. (2006) extend this study to include serial autocorrelation (γ, ρ, μ; λ, κ not included). Elhorst (2008a) considers ML estimation of a model with serial and spatial autocorrelation (γ, ρ; μ, λ, κ not included). Kapoor et al. (2007) consider GMM estimation of a

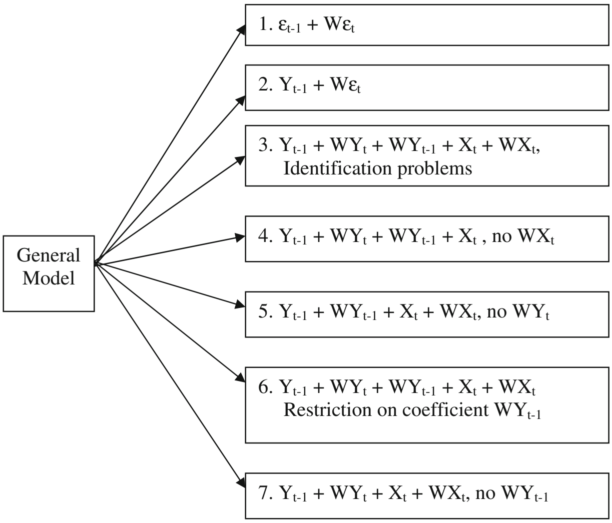

Fig. 3 Dynamic spatial panel data models that have been considered in the literature

spatial error model with time-period random effects (ρ, λ; γ, μ, κ not included). Finally, Baltagi et al. (2007) consider the testing of spatial autocorrelation in both the remainder error term and the spatial random effects (ρ, μ, κ; γ, λ not included). This short overview shows that not every model combination has been considered yet. It is questionable, however, whether more research is needed in this direction. First, the fixed effects model is often more appropriate than the random effects model when modeling spatial panel data (see the discussion in Sect. 2.2). Second, Lee and Yu (2010a) argue that the fixed effects model is robust to and also computationally simpler than the random effects model. Equation 1c can be rewritten as $\xi = (I_N - \kappa W)^{-1}\mu$. Consequently, if μ is treated as a vector of fixed effects for every spatial unit in the sample, so can ξ without having to estimate the parameter κ. Likewise, if μ is treated as a vector of random effects for every spatial unit in the sample, a vector of fixed effects for every spatial unit in the sample ξ can replace μ without having to estimate the parameter κ. In other words, by controlling for spatial fixed effects, spatial autocorrelation among the spatial-specific effects is automatically accounted for, no matter whether these effects are fixed or random and without having to estimate the magnitude of this form of spatial error autocorrelation. Third, spatial interaction effects among the dependent variable Y and/or the independent variables X are more important than spatial interaction effects among the error terms; when ignoring WY and/or WX variables, the estimator of the remaining parameter estimates will lose its property of being consistent. By contrast, when ignoring spatial interaction effects among the error term, $W\varepsilon_t$, the estimator of the remaining parameter estimates will "only" lose its property of being efficient. Fourth, these types of models cannot be used to determine short-term effects and indirect (spatial spillover) effects (see model 1 in Table 1), which are often the main purpose of the analysis.

Perhaps more important is the development of estimators of the covariance matrix that are robust to serial autocorrelation, spatial autocorrelation, and heteroskedasticity, as already pointed out in Sects. 3.1 and 3.2. Newey and West (1987) derive a consistent estimator of the covariance matrix robust to serial autocorrelation and heteroskedasticity. Similarly, Kelejian and Prucha (2010) derive a consistent estimator of the covariance matrix robust to spatial autocorrelation and heteroskedasticity. Whether these two estimators can be combined and be used in a panel data setting still needs to be investigated. Pesaran and Tosetti (2011) are among the first to consider such an estimator. This study, however, estimates one equation for every spatial unit in the sample, which requires T to be large while T in most space–time studies tends to be small, and does not consider WY_t and WX_t variables in the deterministic regression equation. Whether this approach is still practicable and whether the parameters are identified if this model is extended to include WY_t and WX_t variables is an interesting topic for further research.

A second set of studies (model 2 in Fig. 3) have mixed space and time by specifying the deterministic regression equation as a dynamic panel data model and the stochastic error term specification as a spatial error model. Elhorst (2005) considers ML estimation of this model extended to include spatial and time-period fixed effects, and Yang et al. (2006) extended to include spatial random effects (but no time-specific effects). Practice has shown that this setup of separating deterministic dynamic effects in time and stochastic interaction effects among different units across space is beneficial. First, it offers the opportunity to control for independent variables lagged in time, X_{t-1}. When interaction effects among different units across space are taken up in the regression equation rather than the error term specification, identification of the parameters requires the elimination of the variables X_{t-1} (Anselin et al. 2008). Besides, it also offers the opportunity to adjust the error specification such that endogeneous Z_t variables can be controlled for, also when the model is estimated by ML (see Elhorst 2008b). Second, the forecast performance of these models tends to be much better than that of dynamic panel data model that do not control for spatial autocorrelation (Elhorst 2005; Kholodilin et al. 2008). The disadvantage of this type of models, however, is that they cannot be used to determine indirect (spatial spillover) effects (see model 2 in Table 1).

A third set of studies (model 3 in Fig. 3) have considered a spatial Durbin model extended to include dynamic effects. These studies mainly deal with growth and convergence among countries or regions (Ertur and Koch 2007; Elhorst et al. 2010b). Typically, these studies regress economic growth on economic growth in neighboring economies, on the initial income level in the own and in neighboring economies, and on the rates of saving, population growth, technological change, and depreciation in the own and in neighboring economies. Elhorst et al. (2010b) demonstrate that this economic growth model can be represented by the dynamic regression equation

$$Y_t = \tau Y_{t-1} + \delta WY_t + \eta WY_{t-1} + X_t\beta_1 + WX_t\beta_2 + v_t, \tag{7}$$

which may be labeled a dynamic spatial Durbin model. By rewriting this model as

$$Y_t = (I - \delta W)^{-1}(\tau I + \eta W)Y_{t-1} + (I - \delta W)^{-1}(X_t\beta_1 + WX_t\beta_2) + (I - \delta W)^{-1}v_t, \quad (8)$$

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

$$\left[\frac{\partial Y}{\partial x_{1k}} \quad \cdots \quad \frac{\partial Y}{\partial x_{Nk}}\right]_t = (I - \delta W)^{-1}[\beta_{1k}I_N + \beta_{2k}W]. \quad (9)$$

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

$$\left[\frac{\partial Y}{\partial x_{1k}} \quad \cdots \quad \frac{\partial Y}{\partial x_{Nk}}\right] = [(1 - \tau)I - (\delta + \eta)W]^{-1}[\beta_{1k}I_N + \beta_{2k}W]. \quad (10)$$

The expressions in (9) and (10) show that short-term indirect effects do not occur if both $\delta = 0$ and $\beta_{2k} = 0$, while long-term indirect effects do not occur if both $\delta = -\eta$ and $\beta_{2k} = 0$. It should be noted that similar expressions have recently been derived by Debarsy et al. (2011). By simulating the effects of shocks in the error term, v_t, it is also possible to find the path along which an economy moves to its long-term equilibrium (Blanchard and Katz 1992; De Groot and Elhorst 2010).

The results reported in Table 1 show that this dynamic spatial Durbin model (model 3) can be used to determine short-term and long-term direct effects, and short-term and long-term indirect (spatial spillover) effects. In this respect, it seems to be the ideal model. However, three remarks are in place here. First, Anselin et al. (2008) criticize this model because it might suffer from identification problems. By continuous substitution of Y_{t-1} up to $Y_{t-(t-1)}$ into Eq. 8 and rearranging terms, this equation can be rewritten as

$$Y_t = (I - \delta W)^{-T}(\tau I + \eta W)^T Y_{t-T} + \sum_{p=1}^{T}(I - \delta W)^{-p}(\tau I + \eta W)^{p-1}(X_{t-(p-1)}\beta_1 + WX_{t-(p-1)}\beta_2 + v_{t-(p-1)}). \quad (11)$$

This expression shows that two global spatial multiplier matrices are at work at the same time, $(I - \rho W)^{-p}$ and $(\tau I + \eta W)^{p-1}$, in conjunction with one process that produces local spatial spillover effects, $WX_{t-(p-1)}\beta_2$. That is one process too much.[4,5] We come back to this below. Second, more empirical research is needed to find out whether the above-mentioned short-term and long-term direct and indirect

[4] To investigate this analytically, one should consider the expected value of the log-likelihood function and examine whether it is flat or almost flat. Alternatively, the algorithm that is used to find the optimum should be improved.

[5] As pointed out in Sects. 3.1 and 3.2, Hendry (1995) recommends to regress Y_t on Y_{t-1}, X_t and X_{t-1} as a generalization of the first-order serial autocorrelation model for time-series data, while Burridge (1981) recommends to regress Y_t on WY_t, X_t and WX_t as a generalization of the first-order spatial autocorrelation model for cross-section data. Elhorst (2001) combines these two recommendations and suggests to regress Y_t on Y_{t-1}, WY_t, WY_{t-1}, X_t, WX_t, X_{t-1} and WX_{t-1} when having data in space and time. This extension of Eq. 7, however, worsens the identification problem. See also Beenstock and Felsenstein (2007) but then presented in the form of a spatial VAR model.

effects make sense. Third, 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 avoid identification problems, one of the following four restrictions taken from the spatial econometrics literature on dynamic spatial panel data models might be imposed on the parameters in Eq. 7.

The first restriction is $\beta_2 = 0$ (model 4 in Fig. 3 and Table 1). This model is considered in Yu et al. (2008) and Lee and Yu (2010c). However, the price that needs to be paid for identification is relatively high. Due to this restriction, the local indirect (spatial spillover) effects are set to zero by construction, as a result of which the indirect effects in relation to the direct effects become the same for every explanatory variable, both in the short term and in the long term. If this ratio happens to be p percent for one variable, it is also p percent for any other variable. This is because β_{1k} in the numerator and β_{1k} in the denominator of this ratio cancel each other out. For example, the ratio for the kth explanatory variable in the short term takes the form

$$\begin{aligned}&[(I-\delta W)^{-1}(\beta_{1k}I_N)]^{\overline{\mathrm{rsum}}}/[(I-\delta W)^{-1}(\beta_{1k}I_N)]^{\bar{d}}\\&=[(I-\delta W)^{-1}]^{\overline{\mathrm{rsum}}}/[(I-\delta W)^{-1}]^{\bar{d}},\end{aligned}\tag{12}$$

which shows that it is independent of β_{1k} and thus the same for every explanatory variable. A similar result is obtained when considering this ratio in the long term.

The second restriction that might be imposed is $\delta = 0$ (model 5 in Fig. 3 and Table 1). This model is considered in LeSage and Pace (2009, Ch.7) and Korniotis (2010). The disadvantage of this restriction is that the matrix $(I - \rho W)^{-1}$ degenerates to the identity matrix and thus the global short-term indirect (spatial spillover) effect of every explanatory variable to zero. In other words, this model is less suitable if the analysis focuses on spatial spillover effects in the short term.

The third restriction that might be imposed is $\eta = -\tau\delta$ (model 6 in Fig. 3 and Table 1). This restriction is put forward in Parent and LeSage (2010, 2011). The advantage of this restriction is that the impact of a change in one of the explanatory variables on the dependent variable can be decomposed into a spatial effect and a time effect; the impact over space falls by the factor δW for every higher-order neighbor, and over time by the factor τ for every next time period (see Elhorst 2010c for a mathematical derivation). The disadvantage is that the indirect (spatial spillover) effects in relation to the direct effects remain constant over time for every explanatory variable. The ratio of the kth explanatory variable takes the form

$$[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\overline{\mathrm{rsum}}}/[(I-\delta W)^{-1}(\beta_{1k}I_N+\beta_{2k}W)]^{\bar{d}},\tag{13}$$

both in the short term and the long term. In other words, if it is p percent for one variable in the short term, it is also p percent for that variable in the long term.

The fourth restriction that might be imposed is $\eta = 0$ (model 7 in Fig. 3; Table 1). This model restriction is considered in Franzese and Hays (2007), Kukenova and Monteiro (2009), Elhorst (2010d), Jacobs et al. (2009), and Brady (2011). Although this model also limits the flexibility of the ratio between indirect and direct effects, it seems to be the least restrictive model. More empirical research is needed to find out whether this is really the case.

3.4 Methods of estimation

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. These methods are (partly) based on previous studies discussed in Sects. 3.1 and 3.2.

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 (2010d) 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. In the latter case, the first four moments are required (QML). 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 (2010d) 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. This extension builds on previous work of Elhorst (2005) in which the Bhargava and Sargan approximation is compared with that of Nerlove and Balestra (1996), but then for a dynamic panel data model with Y_{t-1} and $W\varepsilon_t$. One of his conclusions is that the parameter estimate δ of the variable WY_t is still considerably biased when using this unconditional ML estimator. However, if the parameter estimate δ is based on the BCLSDV estimator and the other parameters, given δ, 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 ($T = 5$).

Korniotis (2010) constructs a bias-corrected LSDV estimator for a dynamic panel data model (Y_{t-1}, WY_{t-1}) with spatial fixed effects, also assuming $0 < \lim(N/T) < \infty$. The bias correction in this study is different from that in Yu et al. (2008), since the LSDV estimator does not have to account for endogenous interaction effects WY_t.

A couple of studies have considered IV/GMM estimators, building on previous work of Arrelano and Bond (1991), and Blundell and Bond (1998). Elhorst (2010d) 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 δ 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.

Both Kukenova and Monteiro (2009) and Jacobs et al. (2009) consider a dynamic panel data model (Y_{t-1}, WY_t) and extend the system GMM estimator of Blundell and Bond (1998) to account for endogenous interaction effects (WY_t). The former study also considers endogenous explanatory variables Z_t, and the latter spatially autocorrelated error terms $W\varepsilon_t$. The main argument of applying GMM estimators rather than traditional spatial maximum likelihood estimators is that the former can also be used to instrument endogenous explanatory variables (other than the variables Y_{t-1} and WY_t).

Both studies find that the system GMM estimator substantially reduces the bias in the parameter estimate of the WY_t variable, and that the system GMM estimator outperforms the Arrelano and Bond difference GMM estimator. The main message of these studies seems to be that the bias Lee and Yu (2010c) have recently found to occur in theory may reduce so strongly that they become acceptable in practice. In Jacobs et al. (2009), the bias in δ of the variable WY_t amounts to 0.50% of the true parameter value, on average. On the other hand, Monte Carlo simulation experiments can only cover a limited number of situations and therefore do not prove that these results hold in general. Kukenova and Monteiro (2009), for example, only consider positive values for the spatial autoregressive coefficient τ of the variable WY_t. Furthermore, in some cases, both studies also find biases that are rather large. For $T = 10$, $N = 50$, and $\delta = 0.3$, for example, Kukenova and Monteiro (2009, appendix 6.C) find a bias of -0.0219, or 7.3% of the true parameter value. Comparably, Jacobs et al. (2009, Table A.1) find an increasing bias, up to 6.1% of the true parameter value, in the spatial autoregressive coefficient τ of the variable WY_t, provided that spatial autocorrelation in the error terms is not accounted for. When correcting for spatial error correlation, this bias considerably diminishes, but then the bias in the spatial autocorrelation coefficient ρ increases.

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 in Elhorst (2001, 2005, 2010d), they treat the first period cross-section as endogenous, using the Bhargava and Sargan (1983) approximation. They find that the correct treatment of the initial observations (endogenous instead of exogenous) is important, especially in cases when T is small. Since Yu et al. (2008) and Elhorst (2010d) find that maximizing the log-likelihood function leads to biased estimates of the spatial autoregressive parameter δ of the variable WY_t, the former when considering the log-likelihood conditional upon the first cross-section of observations and the latter when considering the unconditional log-likelihood, the question arises whether the Bayesian MCMC estimator is not also subject to a bias. For $T = 5$, $N = 50$, and $\delta = 0.7$, for example, Parent and LeSage (2011, Table 3) find a bias of 0.0149, or 2.13% of the true parameter value.

4 Conclusion

Ten years ago, there was no straightforward estimation procedure for dynamic spatial panel data models. Today, they can be estimated by bias-corrected ML or QML, IV/GMM, and Bayesian MCMC methods. However, many problems remain. First, not every method is able to tackle the potential bias in the coefficient δ of the variable WY_t adequately. Second, some estimators underperform when T is small; treating the initial observations endogenous instead of exogenous may be beneficial under these circumstances. Third, not every estimator is able to deal with endogenous explanatory variables other than the dependent variables lagged in space and/or time. Fourth, the stationarity conditions that need to be imposed on the parameters of the model are not always implemented correctly. A final problem is the treatment of spatial-specific effects; many studies adopt a random effects specification, where a fixed effects specification might be more appropriate.

A dynamic panel data model can take several forms. In this paper, we presented the most popular ones. Each form appeared to have certain shortcomings. Dependent on the purpose of a particular empirical study and the structure of the data (e.g., large or small N, and small or large T), it is the researcher to determine which form is most appropriate and which estimator to use.

Acknowledgments The author would like to thank Jan Jacobs, three anonymous referees, the editor of this journal, and the participants of the 5th World Conference of the Spatial Econometrics Association in Toulouse 2011 for valuable comments on a previous version of this paper.

References

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 edn. Kluwer, Dordrecht, pp 627–662

Arrelano M (2003) Panel data econometrics. Oxford University Press, Oxford
Arrelano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev Econ Stud 58(2):277–297
Balestra P, Nerlove M (1966) Pooling cross-section and time-series data in the estimation of a dynamic model: the demand for natural gas. Econometrica 34(3):585–612
Baltagi BH (2005) Econometric analysis of panel data, 3rd edn. Wiley, Chichester
Baltagi BH, Song SH, Koh W (2003) Testing panel data models with spatial error correlation. J Econom 117(1):123–150
Baltagi BH, Song SH, Jung BC, Koh W (2006) Testing for serial correlation, spatial autocorrelation and random effects using panel data. J Econom 140(1):5–51
Baltagi BH, Egger P, Pfaffermayr M (2007) A generalized spatial panel data model with random effects. http://ideas.repec.org/p/max/cprwps/113.html
Beck N (2001) Time-series-cross-section data: what have we learned in the past few years? Ann Rev Polit Sci 4:271–293
Beenstock M, Felsenstein D (2007) Spatial vector autoregressions. Spatial Econ Anal 2(2):167–196
Bhargava A, Sargan JD (1983) Estimating dynamic random effects models from panel data covering short time periods. Econometrica 51(6):1635–1659
Blanchard OJ, Katz LF (1992) Regional evolutions. Brookings Pap Econ Act 23(1):1–75
Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. J Econom 87(1):115–143
Brady RR (2011) Measuring the diffusion of housing prices across space and time. J Appl Econom 26(2):213–231
Bun M, Carree M (2005) Bias-corrected estimation in dynamic panel data models. J Bus Econ Stat 3(2):200–211
Burridge P (1981) Testing for a common factor in a spatial autoregression model. Environ Plan A 13(7):795–800
Cressie N, Wikle CK (2011) Statistics for spatial-temporal data. Wiley, Blackwell
Debarsy N, Ertur C, LeSage JP (2011) Interpreting dynamic space-time panel data models. Stat Methodol. doi:10.1016/j.stamet.2011.02.002
De Groot AJ, Elhorst JP (2010). Labour market effects of flexicurity from a regional perspective. Tijdschrift voor Economische en Sociale Geografie [J Econ Soc Geogr] 101(4):392–408
Driscoll JC, Kraay AC (1998) Consistent covariance matrix estimation with spatially dependent data. Rev Econ Stat 80(4):549–560
Elhorst JP (2001) Dynamic models in space and time. Geogr Anal 33(2):119–140
Elhorst JP (2003) Specification and estimation of spatial panel data models. Int Reg Sci Rev 26(3):244–268
Elhorst JP (2005) Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. Geogr Anal 37(1):62–83
Elhorst JP (2008a) Serial and spatial autocorrelation. Econ Lett 100(3):422–424
Elhorst JP (2008b) A spatiotemporal analysis of aggregate labour force behaviour by sex and age across the European Union. J Geogr Syst 10(2):167–190
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) Applied spatial econometrics: raising the bar. Spatial Econ Anal 5(1):9–28
Elhorst JP (2010d) Dynamic panels with endogenous interaction effects when T is small. Reg Sci Urban Econ 40(5):272–282
Elhorst JP, Piras G, Arbia G (2010a) Growth and convergence in a multi-regional model with space-time dynamics. Geogr Anal 42(3):338–355
Elhorst JP, Zandberg E, De Haan J (2010b) The impact of interaction effects among neighboring countries on financial reform: a dynamic spatial panel data approach. Mimeo, Groningen
Enders W (1995) Applied econometric time series. Wiley, New York
Ertur C, Koch W (2007) Growth, technological interdependence and spatial externalities: theory and evidence. J Appl Econom 22(6):1033–1062
Fingleton B, Le Gallo J (2008) Estimating spatial models with endogenous variables, a spatial lag en spatially dependent disturbances: finite sample properties. Pap Reg Sci 87(3):319–339

Franzese RJ Jr, Hays JC (2007) Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. Political Anal 15(2):140–164
Greene WH (2005) Econometric analysis, 6th edn. Pearson Prentice Hall, New Jersey
Hahn J, Kuersteiner G (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica 70(4):1639–1657
Hamilton JD (1994) Time series analysis. Princeton University Press, New Jersey
Hendry DF (1995) Dynamic econometrics. Oxford University Press, Oxford
Hsiao C (2003) Analysis of panel data, 2nd edn. Cambridge University Press, Cambridge
Hsiao C, Pesaran MH, Tahmiscioglu AK (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J Econom 109(1):107–150
Jacobs JPAM, Ligthart JE, Vrijburg H (2009) Dynamic panel data models featuring endogenous interaction and spatially correlated errors. http://ideas.repec.org/p/ays/ispwps/paper0915.html
Kapoor M, Kelejian HH, Prucha IR (2007) Panel data models with spatially correlated error components. J Econom 140(1):97–130
Kelejian HH, Prucha IR (1998) A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Estate Finance Econ 17(1):99–121
Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40(2):509–533
Kelejian HH, Prucha IR (2002) 2SLS and OLS in a spatial autoregressive model with equal spatial weights. Reg Sci Urban Econ 32(6):691–707
Kelejian HH, Prucha IR (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J Econom 157(1):53–67
Kelejian HH, Prucha IR, Yuzefovich Y (2006) Estimation problems in models with spatial weighting matrices which have blocks of equal elements. J Reg Sci 46(3):507–515
Kholodilin KA, Siliverstovs B, Kooths S (2008) A dynamic panel data approach to the forecasting of the GDP of German Länder. Spatial Econ Anal 3(2):195–207
Kiviet JF (1995) On bias, inconsistency and efficiency of some estimators in dynamic panel data models. J Econom 68(1):53–78
Korniotis GM (2010) Estimating panel models with internal and external habit formation. J Bus Econ Stat 28(1):145–158
Kukenova M, Monteiro JA (2009) Spatial dynamic panel model and system GMM: a Monte Carlo investigation. http://ideas.repec.org/p/pra/mprapa/11569.html
Lee LF (2004) Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6):1899–1925
Lee LF, Yu J (2010a) Some recent developments in spatial panel data models. Reg Sci Urban Econ 40(5):255–271
Lee LF, Yu J (2010b) Estimation of spatial autoregressive panel data models with fixed effects. J Econom 154(2):165–185
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, Yu J (2010d) A spatial dynamic panel data model with both time and individual fixed effects. Econom Theory 26(2):564–597
LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press Taylor & Francis Group, Boca Raton
Nerlove M (1999) Properties of alternative estimators of dynamic panel models: an empirical analysis of cross-country data for the study of economic growth. In: Hsiao C, Lahiri K, Lee LF, Pesaran MH (eds) Analysis of panels and limited dependent variable models. Cambridge University Press, Cambridge, pp 136–170
Nerlove M, Balestra P (1996) Formulation and estimation of econometric models for panel data. In: Mátyás L, Sevestre P (eds) The econometrics of panel data, 2nd revised edition. Dordrecht, Kluwer, pp 3–22
Newey W, West K (1987) A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55(3):703–708
Nickell S (1981) Biases in dynamic models with fixed effects. Econometrica 49(6):1417–1426
Parent O, LeSage JP (2010) A spatial dynamic panel model with random effects applied to commuting times. Transp Res Part B 44(5):633–645
Parent O, LeSage JP (2011) A space-time filter for panel data models containing random effects. Comput Stat Data Anal 55(1):475–490

Pesaran MH, Tosetti E (2011) Large panels with common factors and spatial correlation. J Econom 161(2):182–202

Yang Z, Li C, Tse YK (2006) Functional form and spatial dependence in spatial panels. Econ Lett 91(1):138–145

Yu J, De Jong R, Lee L (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. J Econom 146(1):118–134