J.Paul Elhorst, University of Groningen, the Netherlands
Seminar, St. Andrews, January 18, 2010

- What are spatial interaction effects and what is a spatial econometric model?
- How to estimate a spatial econometric model?
- A theoretical model that is tested using a spatial econometric model.

Seminar is based on three recent papers:
- J.P.Elhorst (2009) Spatial Panel Data Models. In: M.M. Fischer and A. Getis (eds.), Handbook of Applied Spatial Analysis. Berlin, Springer.
- Elhorst J.P., Fréret S. (2009) Evidence of Political Yardstick Competition in France Using a Two-regime Spatial Durbin Model with Fixed Effects. *Journal of Regional Science* 49: 931-951.
- J.P.Elhorst (2010) Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis, Forthcoming.

see www.regroningen.nl/elhorst (spatial econometrics)

J.P.Elhorst (2010) Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis, Forthcoming.

The idea that cross-sectional units interact with others has recently received considerable attention, as evidenced in the development of theoretical frameworks to explain social phenomena such as social norms, peer influence, neighborhood effects, network effects, contagion, epidemics, social interactions, interdependent preferences, etc.

Interaction effect = the average behavior in some group influences the behavior of the individuals that comprise the group (Manski, 1993).

$$Y = \rho WY + \alpha\iota_N + X\beta + WX\theta + u$$ Vector Notation

$$u = \lambda Wu + \varepsilon$$ Cross-section data

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 $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^T$ 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.

Space-time data: Add subscript t to Y, X, u and ε.

Endogenous interaction effects (WY) = the propensity of an individual to behave in some way varies with the behavior of the group.

Consider the choice of high school after primary school: the individual choice tends to vary with the choice made by friends.

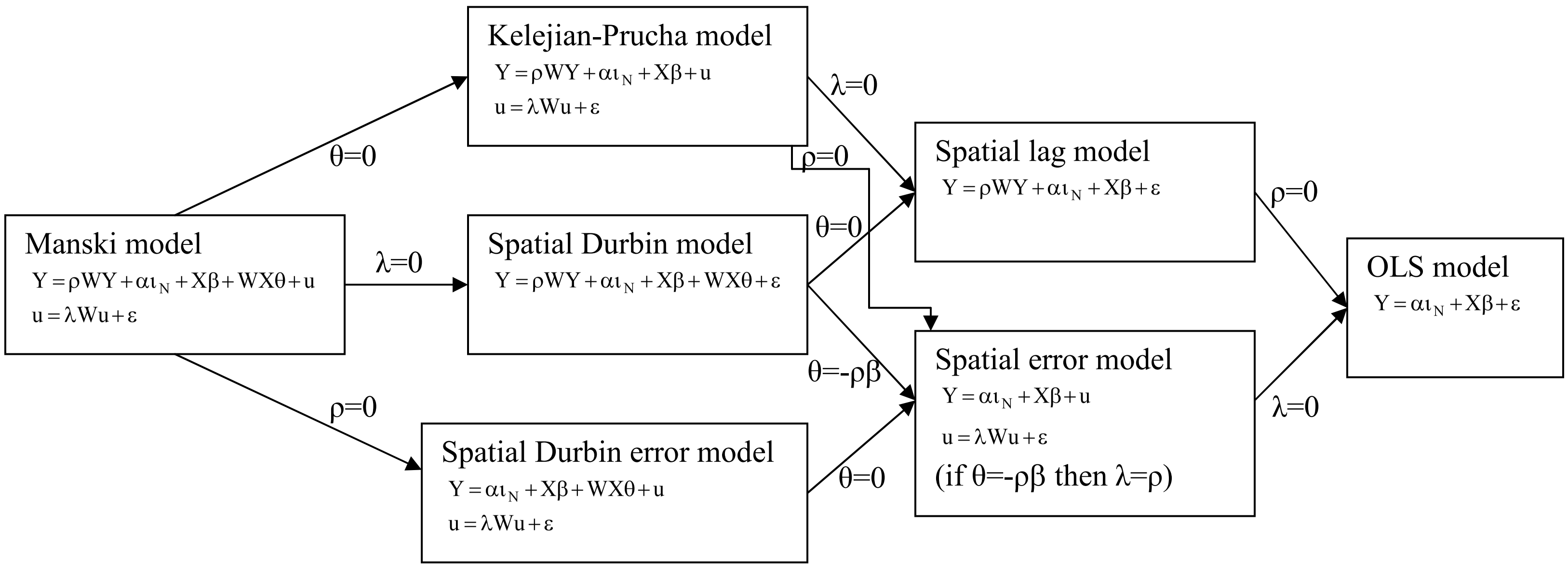
Exogenous interaction effects (WX) = the propensity of an individual to behave in some way varies with the exogenous characteristics of the group (mostly there are K exogenous explanatory variables, and thus K exogenous interaction effects).

There are exogenous effects if school choice tends to vary with the exogenous characteristics of other people, e.g. the opinion of the parents of friends.

Correlated effects (disturbance term Wu) = individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments (these may be unobserved).

There are correlated effects if children make the same choice because they have similar backgrounds (neighbourhood) or because they are taught by the same teachers.

The relationships between different spatial dependence models for cross-section data



Up to 2007 spatial econometricians were mainly interested in models containing one type of spatial interaction effect: the spatial lag model and the spatial error model.

Identification problem: One of the K+2 interaction effect should be dropped. Following LeSage and Pace (pp. 155-158), the best option is to exclude the spatially autocorrelated error term. 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.

**

W is an N×N matrix describing the spatial arrangement of the spatial units in the sample. Lee (2004) shows that W should be a nonnegative matrix of known constants. The diagonal elements are set to zero by assumption, since no spatial unit can be viewed as its own neighbour. The matrices I-ρW and I-λW should be non-singular, where I represents the identity matrix of order N. For a symmetric W, this condition is satisfied as long as ρ and λ are in the interior of $(1/\omega_{min},1/\omega_{max})$, where ω_{min} denotes the smallest (i.e., most negative) and ω_{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 (pp. 88-89) demonstrate that in that case ρ and λ are restricted to the interval $(1/r_{min},1)$, where r_{min} equals the most negative purely real characteristic root of W after this matrix is row-normalized.

1	2	3

Row-normalizing $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ gives $W = \begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$.

Finally, one of the following two conditions should be satisfied:

(a) the row and column sums of the matrices W, $(I-\rho W)^{-1}$ and $(I-\lambda W)^{-1}$ 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.

When the spatial weights matrix is a binary contiguity matrix, (a) is satisfied. Normally, no spatial unit is assumed to be a neighbour to more than a given number, say q, of other units.

By contrast, when the spatial weights matrix is an inverse distance matrix, (a) may not be satisfied. Consider an infinite number of spatial units that are linearly arranged. The distance of each spatial unit to its first left- and right-hand neighbour is d; to its second left- and right-hand neighbour, the distance is 2d; and so on. When W is an inverse distance matrix and the off-diagonal elements of W are of the form $1/d_{ij}$, where d_{ij} is the distance between two spatial units i and j, each row sum is $2\times(\frac{1}{d}+\frac{1}{2d}+\frac{1}{3d}+...)$, representing a series that is not finite. This is perhaps the reason why some empirical applications introduce a cut-off point d* such that $w_{ij}=0$ if $d_{ij}>d^*$. However, since the ratio $2\times(\frac{1}{d}+\frac{1}{2d}+\frac{1}{3d}+...)/N\to 0$ as N goes to infinity, condition (b) is satisfied, which implies that an inverse distance matrix without a cut-off point does not necessarily have to be excluded in an empirical study for reasons of consistency.

The opposite situation occurs when all cross-sectional units are assumed to be neighbours of each other and are given equal weights. In that case all off-diagonal elements of the spatial weights matrix are $w_{ij}=1$. Since the row and column sums are N-1, these sums diverge to infinity as N goes to infinity. In contrast to the previous case, however, $(N-1)/N \rightarrow 1$ instead of 0 as N goes to infinity. This implies that a spatial weights matrix that has equal weights and that is row-normalized subsequently, $w_{ij}=1/(N-1)$, must be excluded for reasons of consistency.

Note: some of the regularity conditions may change in a panel data setting (Yu et al. 2007).

**

J.P.Elhorst (2009) Spatial Panel Data Models. In: M.M. Fischer and A. Getis (eds.), Handbook of Applied Spatial Analysis. Berlin, Springer.

In recent years, the spatial econometrics literature has exhibited a growing interest in the specification and estimation of econometric relationships based on spatial panels. Spatial panels typically refer to data containing time series observations of a number of spatial units (zip codes, municipalities, regions, states, jurisdictions, countries, etc.). This interest can be explained by the fact that panel data offer researchers extended modeling possibilities as compared to the single equation cross-sectional setting, which was the primary focus of the spatial econometrics literature for a long time.

Spatial lag model (Endogenous interaction effects only)

$$y_{it} = \delta \sum_{j=1}^{N} w_{ij} y_{jt} + x_{it}\beta + \mu_i + \varepsilon_{it},$$

δ is called the spatial autoregressive coefficient and w_{ij} is an element of a spatial weights matrix W describing the spatial arrangement of the units in the sample.

Note: In this paper, I do not use the vector form notation, but a notation in individual observations. Furthermore, I use δ instead of ρ as the coefficient of endogenous interaction effects (WY). Finally, exogenous interaction effects can be included by replacing x_{it} by $\bar{x}_{it} = [x_{it} \quad \sum_{j=1}^{N} w_{ij} x_{jt}]$ (spatial Durbin model).

Reason to consider fixed effects/random effects models.

The standard reasoning behind spatial specific effects is that they control for all space-specific time-invariant variables whose omission could bias the estimates in a typical cross-sectional study (Baltagi, 2005).

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. Furthermore, it is assumed that the random variables μ_i and ε_{it} are independent of each other.

I first consider the fixed effects model and assume that the data are sorted first by time and then by spatial units, whereas the classic panel data literature tends to sort the data first by spatial units and then by time.

According to Anselin et al. (2006), the extension of the fixed effects model with a spatially lagged dependent variable raises two complications. First, the endogeneity of $\Sigma_j w_{ij} y_{jt}$ violates the assumption of the standard regression model that $E[(\Sigma_j w_{ij} y_{jt})\varepsilon_{it}]=0$. In model estimation, this simultaneity must be accounted for. Second, the spatial dependence among the observations at each point in time may affect the estimation of the fixed effects.

The log-likelihood function of the spatial lag model with fixed effects is

$$\text{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + T\log|I_N - \delta W| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \delta\sum_{j=1}^{N} w_{ij} y_{jt} - x_{it}\beta - \mu_i)^2,$$

where the second term on the right-hand side represents the Jacobian term of the transformation from ε to y taking into account the endogeneity of $\Sigma_j w_{ij} y_{jt}$ (Anselin 1988, p. 63).

Computational problems of the Jacobian term.

$$\mathrm{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + T\log|I_N - \delta W| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \delta\sum_{j=1}^{N} w_{ij} y_{jt} - x_{it}\beta - \mu_i)^2,$$

The partial derivatives of the log-likelihood with respect to μ_i are

$$\frac{\partial \mathrm{LogL}}{\partial \mu_i} = \frac{1}{\sigma^2}\sum_{t=1}^{T}(y_{it} - \delta\sum_{j=1}^{N} w_{ij} y_{jt} - x_{it}\beta - \mu_i) = 0, \quad i=1,\ldots,N.$$

When solving for μ_i, one obtains

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \delta\sum_{j=1}^{N} w_{ij} y_{jt} - x_{it}\beta), \quad i=1,\ldots,N.$$

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \delta\sum_{j=1}^{N} w_{ij} y_{jt} - x_{it}\beta), \quad i=1,\ldots,N.$$

This equation shows that the standard formula for calculating the spatial fixed effects applies to the fixed effects spatial lag model in a straightforward manner. Corrections for the spatial dependence among the observations at each point in time, other than the addition of the spatially lagged dependent variable to these formulas, are not necessary.

Substituting the solution for μ_i into the log-likelihood function, and after rearranging terms, the concentrated log-likelihood function with respect to β, δ and σ^2 is obtained

$$\text{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + T\log|I_N - \delta W| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}^* - \delta[\sum_{j=1}^{N} w_{ij} y_{jt}]^* - x_{it}^*\beta)^2,$$

where the asterisk denotes the demeaning procedure

$$y_{it}^* = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}, \sum_{j=1}^{N} w_{ij} y_{jt}^* = \sum_{j=1}^{N} w_{ij} y_{jt} - \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N} w_{ij} y_{jt} \text{ and } x_{it}^* = x_{it} - \frac{1}{T}\sum_{t=1}^{T} x_{it}.$$

$$\mathrm{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + T\log|I_N - \delta W| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}^* - \delta[\sum_{j=1}^{N} w_{ij} y_{jt}]^* - x_{it}^*\beta)^2,$$

1. Regress y_{it}, $\Sigma w_{ij}y_{jt}$ on x_{it} (+*) by OLS → b_0 and b_1
2. Compute residuals
3. Substitute residuals into log-likelihood function, concentrate it with respect to σ^2, and maximize it with respect to δ
4. $\beta = b_0 - \delta b_1$ and σ^2
5. Determine variance-covariance matrix (see paper)
6. Recover fixed effects

Go to <u>http://www.regroningen.nl/elhorst</u> (and click on software) for software to estimate spatial panels.

**

Whether the random effects model is an appropriate specification in spatial research remains controversial. When the random effects model is implemented, the units of observation should be representative of a larger population, and the number of units should potentially be able to go to infinity.

There are two types of asymptotics that are commonly used in the context of spatial observations: (a) The 'infill' asymptotic structure, where the sampling region remains bounded as $N \to \infty$. In this case more units of information come from observations taken from between those already observed; and (b) The 'increasing domain' asymptotic structure, where the sampling region grows as $N \to \infty$. In this case there is a minimum distance separating any two spatial units for all N.

According to Lahiri (2003), there are also two types of sampling designs: (a) The stochastic design where the spatial units are randomly drawn; and (b) The fixed design where the spatial units lie on a nonrandom field, possibly irregularly spaced.

The spatial econometric literature mainly focuses on increasing domain asymptotics under the fixed sample design (Cressie 1993, p. 100; Griffith and Lagona 1998; Lahiri 2003).

Although the number of spatial units under the fixed sample design can potentially go to infinity, it is questionable whether they are representative of a larger population. For a given set of regions, such as all counties of a state or all regions in a country, the population may be said

- ‘to be sampled exhaustively’ (Nerlove and Balestra 1996, p. 4),
- ‘the individual spatial units have characteristics that actually set them apart from a larger population’ (Anselin 1988, p. 51).
- ‘the critical issue is that the spatial units be fixed and not sampled, and that inference be conditional on the observed units’ Beck (2001, p. 272).

In addition, the traditional assumption of zero correlation between μ_i in the random effects model and the explanatory variables, which also needs to be made, is particularly restrictive. **

Interpretation Coefficients

If the spatial Durbin model is taken as point of departure and rewritten as

$$Y=(I-\rho W)^{-1}\alpha\iota_N+(I-\rho W)^{-1}(X\beta+WX\theta)+(I-\rho W)^{-1}\varepsilon,$$

the matrix of partial derivatives of Y with respect to the k^{th} explanatory variable of X in unit 1 up to unit N (say x_{ik} for i=1,…,N, respectively) is relatively easy to obtain

$$\begin{bmatrix}\frac{\partial Y}{\partial x_{1k}} & . & \frac{\partial Y}{\partial x_{Nk}}\end{bmatrix}=\begin{bmatrix}\frac{\partial y_1}{\partial x_{1k}} & . & \frac{\partial y_1}{\partial x_{Nk}}\\ . & . & .\\ \frac{\partial y_N}{\partial x_{1k}} & . & \frac{\partial y_N}{\partial x_{Nk}}\end{bmatrix}=(I-\rho W)^{-1}\begin{bmatrix}\beta_k & w_{12}\theta_k & . & w_{1N}\theta_k\\ w_{21}\theta_k & \beta_k & . & w_{2N}\theta_k\\ . & . & . & .\\ w_{N1}\theta_k & w_{N2}\theta_k & . & \beta_k\end{bmatrix},$$

where w_{ij} is the (i,j)th element of W.

If $W = \begin{bmatrix} 0 & 1 & 0 \\ w_{21} & 0 & w_{23} \\ 0 & 1 & 0 \end{bmatrix}$ and $(I-\rho W)^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1-w_{23}\rho^2 & \rho & \rho^2 w_{23} \\ \rho w_{21} & 1 & \rho w_{23} \\ \rho^2 w_{21} & \rho & 1-w_{21}\rho^2 \end{bmatrix}$, we get

$$\begin{bmatrix} \frac{\partial Y}{\partial x_{1k}} & \frac{\partial Y}{\partial x_{2k}} & \frac{\partial Y}{\partial x_{3k}} \end{bmatrix} =$$

$$\frac{1}{1-\rho^2}\begin{bmatrix} (1-w_{23}\rho^2)\beta_k + (w_{21}\rho)\theta_k & \rho\beta_k + \theta_k & (w_{23}\rho^2)\beta_k + (\rho w_{23})\theta_k \\ (w_{21}\rho)\beta_k + w_{21}\theta_k & \beta_k + \rho\theta_k & (w_{23}\rho)\beta_k + w_{23}\theta_k \\ (w_{21}\rho^2)\beta_k + (w_{21}\rho)\theta_k & \rho\beta_k + \theta_k & (1-w_{21}\rho^2)\beta_k + (w_{23}\rho)\theta_k \end{bmatrix}.$$

Direct effect: Mean diagonal element

Indirect effect: Mean row sum of non-diagonal elements.

Table 1. Direct and indirect effects of different model specifications [N=3, W as in (4)]

Type of model	Direct effect	Indirect effect
Spatial Durbin model Manski model	$\frac{(3-\rho^2)}{3(1-\rho^2)}\beta_k + \frac{2\rho}{3(1-\rho^2)}\theta_k$	$\frac{3\rho+\rho^2}{3(1-\rho^2)}\beta_k + \frac{3+\rho}{3(1-\rho^2)}\theta_k$
Spatial lag model Kelejian-Prucha model	$\frac{(3-\rho^2)}{3(1-\rho^2)}\beta_k$	$\frac{3\rho+\rho^2}{3(1-\rho^2)}\beta_k$
Spatial Durbin error model	β_k	θ_k
OLS model Spatial error model	β_k	0

- OLS model: indirect effects are zero by construction.
- Spatial Durbin error model: it can still be seen from the coefficient estimates and the corresponding standard errors or t-values (derived from the variance-covariance matrix) whether indirect effects are significant.
- Spatial lag model: limitation is that the ratio between the indirect and direct effects in the spatial lag model is the same for every explanatory variable.
- Spatial Durbin model: no prior restrictions are imposed on the magnitude of both the direct and indirect effects and thus that the ratio between the indirect and the direct effect may be different for different explanatory variables.

**

Elhorst J.P., Fréret S. (2009) Evidence of Political Yardstick Competition in France Using a Two-regime Spatial Durbin Model with Fixed Effects. *Journal of Regional Science* 49: 931-951.

Strategic interaction among governments (municipalities, regions or states) has become a major focus of theoretical and empirical work in public economics. Many studies have found that an increase in the tax burden of neighboring jurisdictions of one euro/dollar is matched by an increase of 18 to 66 cents per unit of tax in a jurisdiction's own tax burden. A related literature focuses on expenditure interdependence and has found similar figures.

One explanation is yardstick competition: voters use information from other jurisdictions to judge the performance of their own governors. The reason for this behavior is asymmetric information; voters do not know what level of services can be provided relative to a certain tax level.

Taxes cover the minimal production cost of public goods plus any extra resources lost to waste or rent seeking. These lost resources cannot be observed by voters. Since tax rates and expenditure levels in nearby jurisdictions are more easily observed, they can serve as a benchmark and used in elections to discipline and select the type of governor.

However, if voters consider relative performance, rational governors will do the same and (partly) mimic the tax rates and expenditure levels of their neighbors. This is called yardstick competition.

Preferences of representative resident of jurisdiction i and individual budget constraint are given by

$U(y_i - T_i, z_i; X_i)$

y_i=per capita income, T_i=tax payment per capita, z_i=level of a public good, X_i=characteristics of jurisdiction i other than income.

Let z_i/T_i denote the minimum level of public good provision relative to taxes that must be delivered for jurisdiction's i's government to remain in office. This required level depends on observed public good levels observed in other jurisdictions: $z_i/T_i = \varphi([z/T]_{-i})$.

If the levels of public good provision relative to taxes in other jurisdictions, $[z/T]_{-i}$, increases, government i is forced to raise z_i/T_i to remain in office. Since $z_i=T_i\ \varphi([z/T]_{-i})$, we have (and using median voter theorem)

$$U(y_i-T_i,z_i;X_i)= U(y_i-T_i,\ T_i\ \varphi([z/T]_{-i});X_i) \equiv V(z_i,z_{-i};X_i).$$

Note: Instead of one consumer, we may have different consumers in every jurisdiction with preferences ranging along a spectrum on most public services. The median voter theorem states that, if preferences are single-peaked and government policy is decided by representatives elected by a majority vote, government policy will reflect the preferences of the median voter.

First-order condition

$$\frac{\partial V}{\partial z_i} \equiv V_{z_i} = 0 \;\Rightarrow\; z_i = R(z_{-i}; X_i),$$

where R represents a reaction function to the choices of other jurisdictions. The slope of the reaction function with respect to z_{-i} can be positive or negative. A test of the null hypothesis that the reaction function's slope is zero is effectively a test for the existence of spillovers. Furthermore, interaction may expected to be more pronounced if governors are politically sensitive to fiscal policy changes in neighboring jurisdictions. In this paper: Departments governed by a small political majority mimic neighboring expenditures on welfare to a greater extent than do Departments governed by a large political majority. **

$$y_{it} = \delta_1 d_{it} \sum_{j=1}^{N} w_{ij} y_{jt} + \delta_2 (1 - d_{it}) \sum_{j=1}^{N} w_{ij} y_{jt} + \alpha + x_{it}\beta + \sum_{j=1}^{N} w_{ij} x_{jt}\theta + \mu_i + \lambda_t + \varepsilon_{it},$$

Reason to consider two regimes: the theoretical and empirical literature on public economics offers two alternative explanations for the existence of tax and expenditure interaction effects, which have the same reaction function. They may also be the result of spillover effects, for example, because expenditures on local public services may have beneficial or detrimental effects on nearby jurisdictions (see Case et al., 1993 for a theoretical explanation), or be the result of tax or welfare competition.

$$y_{it} = \delta_1 d_{it} \sum_{j=1}^{N} w_{ij} y_{jt} + \delta_2 (1 - d_{it}) \sum_{j=1}^{N} w_{ij} y_{jt} + \alpha + x_{it}\beta + \sum_{j=1}^{N} w_{ij} x_{jt}\theta + \mu_i + \lambda_t + \varepsilon_{it},$$

One reason to add spatially lagged independent variables is taken from Boarnet and Glazer (2002). They argue that a negative grant spillover effect can also be interpreted as a form of yardstick competition. A voter who sees that a neighboring jurisdiction received a grant which his jurisdiction did not receive, may think poorly of the ability of the local governors and, therefore, reduce his demand for local spending. In terms of modeling, if spending on public services is taken to depend on grants received by the jurisdiction, the spatial average of grants received by neighboring jurisdictions also affects spending on public services.

LeSage and Pace (2008) offer another reason to add spatially lagged independent variables, namely, an omitted variables motivation to include the variables in a regression relationship that seeks to explore interaction effects in a spatial context. If unobserved or unknown but relevant variables following a first-order spatial autoregressive process do not appear in the model, and these variables happen to be correlated with independent variables not omitted from the model, a spatial lag model extended to include spatially lagged independent variables will produce unbiased coefficient estimates, whereas a spatial lag model without these variables cannot.

Reason to consider time-period fixed effects (and not spatially autocorrelated error terms): Identification problem Manski (1993). Time-period fixed effects correct for spatial interaction effects among the error terms, such as unobserved shocks following a spatial pattern or variables that increase or decrease together in different jurisdictions along the same (business) cycle over time.

The mathematical explanation is that time-period fixed effects are identical to a spatially autocorrelated error term with a spatial weights matrix whose elements are all equal to 1/*N*, including the diagonal elements. When this spatial weights matrix would be adopted, one obtains e.g. $y_{it} - \sum_{j=1}^{N} w_{ij} y_{jt} = y_{it} - \frac{1}{N}\sum_{j=1}^{N} y_{jt}$ which is equivalent to the demeaning procedure of Eq. (6) but then for fixed effects in time.

$$y_{it} = \delta_1 d_{it} \sum_{j=1}^{N} w_{ij} y_{jt} + \delta_2 (1 - d_{it}) \sum_{j=1}^{N} w_{ij} y_{jt} + \alpha + x_{it}\beta + \sum_{j=1}^{N} w_{ij} x_{jt}\theta + \mu_i + \lambda_t + \varepsilon_{it},$$

Previous studies on yardstick competition

- Besley and Case (1995), Revelli (2006)

Two-equations spatial lag model estimated by IV based on panel data
Objection: Coefficients of control variables not identical; results may also cover differences other than the political process

- Bordignon et al. (2003), Allers and Elhorst (2005)

Two regimes spatial lag/error model estimated by ML based on cross-sectional data. Objection: No control for spatial fixed effects

- Case (1993), Schaltegger and Küttel (2002) and Sollé-Ollé (2003)

Spatial lag model with cross-product variables estimated by IV based on panel data. Objection: Nonstationarity; Jacobian term not defined for all observations.

$$T * \ln | I_N - \delta W - \sum_{j=1}^{J} \eta_j \,\mathrm{diag}(P_j) W |$$

- One major shortcoming of all studies: No spatial Durbin model.

$$y_{it} = \delta_1 d_{it} \sum_{j=1}^{N} w_{ij} y_{jt} + \delta_2 (1-d_{it}) \sum_{j=1}^{N} w_{ij} y_{jt} + \alpha + x_{it}\beta + \sum_{j=1}^{N} w_{ij} x_{jt}\theta + \mu_i + \lambda_t + \varepsilon_{it},$$

ML estimation (IV would ignore the Jacobian term, instrumental variables?)

$$\text{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + \sum_{t=1}^{T}\log|I_N - \delta_1 D_t W - \delta_2 (I_N - D_t) W|$$
$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it} - \delta_1 d_{it}\sum_{j=1}^{N} w_{ij} y_{jt} - \delta_2 (1-d_{it})\sum_{j=1}^{N} w_{ij} y_{jt} - \alpha - x_{it}\beta - \sum_{j=1}^{N} w_{ij} x_{jt}\theta - \mu_i - \lambda_t]^2,$$

Solve for intercept and spatial and time-period fixed effects

$$\text{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + \sum_{t=1}^{T}\log|I_N - \delta_1 D_t W - \delta_2 (I_N - D_t) W|$$
$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it}^{*} - \delta_1 (d_{it}\sum_{j=1}^{N} w_{ij} y_{jt})^{*} - \delta_2 ((1-d_{it})\sum_{j=1}^{N} w_{ij} y_{jt})^{*} - x_{it}^{*}\beta - \sum_{j=1}^{N} w_{ij} x_{jt}^{*}\theta]^2,$$

$$\mathrm{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + \sum_{t=1}^{T}\log|I_N - \delta_1 D_t W - \delta_2 (I_N - D_t) W|$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}[y_{it}^* - \delta_1 (d_{it}\sum_{j=1}^{N} w_{ij} y_{jt})^* - \delta_2 ((1-d_{it})\sum_{j=1}^{N} w_{ij} y_{jt})^* - x_{it}^*\beta - \sum_{j=1}^{N} w_{ij} x_{jt}^*\theta]^2,$$

1. Regress y_{it}, $d_{it}\Sigma w_{ij}y_{it}$ and $(1\text{-}d_{it})\Sigma w_{ij}y_{jt}$ on x_{it} (+*) by OLS $\rightarrow$ b_0, b_1 and b_2
2. Compute residuals
3. Substitute residuals into log-likelihood function, concentrate it with respect to σ^2, and maximize it with respect to δ_1 and δ_2.
4. $\beta = b_0 - \delta_1 b_1 - \delta_2 b_2$ and σ^2.
5. Determine variance-covariance matrix. Not reported in the literature before.
6. Recover fixed effects

Table 2 Estimation results: Welfare spending by French Departments

Explanatory variables	Two-way spatial Durbin model (1)	One-way spatial Durbin model (2)	Two-way spatial Durbin model, two regimes* (3)	
Operational grant	0.000	0.031	0.001	
	(0.03)	(7.73)	(0.02)	
W*Operational grant	-0.035	0.031	-0.032	
	(-2.07)	(4.00)	(-1.88)	
δ	0.083	0.282	0.167	int.eff.=0.18 / 0.20 in studies with controls for spatial fixed effects, 0.20 / 0.66 in studies without controls for spatial fixed effects.
	(1.61)	(6.24)	0.034	
			(7.07)	
			(1.52)	
Spatial FE	yes	no	yes	
Time-period FE	yes	yes	yes	
Regime dummy			yes	
LogL	1305.83	638.22	1314.15	
R^2	0.941	0.703	0.942	

t-values in parenthesis, * Governors backed by majority less than or equal to 75% and greater than 75%, respectively

Table 3 Yardstick competition and voting margin

	55%	60%	65%	70%	75%	80%	85%
Number of observations	105	203	315	430	543	659	739
and δ_1 when margin is	0.102	0.114	0.127	0.142	0.167	0.168	0.206
less than or equal to .%	(3.61)	(3.50)	(5.48)	(5.35)	(7.07)	(3.73)	(2.35)
Number of observations	732	634	522	407	294	178	98
and δ_2 when margin is	-0.016	0.029	0.020	0.033	0.034	0.056	0.075
greater than .%	(0.27)	(0.62)	(0.70)	(1.29)	(1.52)	(1.92)	(2.50)
δ_1 - δ_2	0.118	0.085	0.108	0.110	0.134	0.112	0.130
T-value of difference	(1.78)	(1.51)	(3.01)	(3.00)	(4.26)	(2.10)	(1.40)
Difference intercepts	0.622	0.462	0.582	0.588	0.711	0.576	0.680
LogL	1309.23	1309.77	1311.54	1311.96	1314.15	1311.61	1309.05

Table 3. Bias and RMSE of δ_1 and δ_2 of different experimental parameter combinations when using the ML estimator and when using IV*

	ML estimator					Instrumental variables (IV)				
	Bias in δ_1					Bias in δ_1				
$\delta_1 \setminus \delta_2$	-0.066	-0.016	0.034	0.084	0.134	-0.066	-0.016	0.034	0.084	0.134
0.067	-0.024	-0.022	-0.003	-0.005	-0.002	0.082	0.026	0.004	-0.002	0.000
0.117	-0.012	-0.039	-0.029	-0.003	-0.002	0.005	0.116	0.000	0.004	0.001
0.167	-0.001	-0.010	-0.031	-0.010	-0.002	0.006	0.010	0.144	0.007	0.003
0.217	-0.002	-0.005	-0.018	-0.029	-0.007	0.001	0.001	0.005	0.130	0.009
0.267	0.001	0.000	-0.005	-0.017	-0.033	0.002	0.002	0.002	0.022	0.071
	RMSE of δ_1					RMSE of δ_1				
$\delta_1 \setminus \delta_2$	-0.066	-0.016	0.034	0.084	0.134	-0.066	-0.016	0.034	0.084	0.134
0.067	0.052	0.045	0.032	0.019	0.014	0.101	0.060	0.036	0.020	0.014
0.117	0.032	0.051	0.039	0.028	0.017	0.038	0.128	0.052	0.030	0.017
0.167	0.021	0.038	0.064	0.036	0.025	0.022	0.046	0.151	0.043	0.027
0.217	0.016	0.022	0.036	0.057	0.030	0.016	0.023	0.043	0.121	0.033
0.267	0.013	0.015	0.022	0.037	0.046	0.013	0.015	0.023	0.046	0.098

* Based on 100 repetitions

Conclusions

Strong evidence in favor of political yardstick competition: If Departments are governed by a political majority in the council less than or equal to 75%, they will change their spending on welfare by seventeen cents in reaction to a change in spending on welfare of one euro by neighboring Departments. By contrast, if Departments are governed by a political majority greater than 75%, this interaction effect decreases to three cents.

Both the interaction effect of seventeen cents and the difference between these two interaction effects of fourteen cents appeared to be significant. In contrast to previous studies, we can be sure that this significant difference does not stem from ignoring spatially lagged independent variables or a spatially correlated error term (Manski, 1993), since the former were explicitly taken into account, while the latter was covered by time-period fixed effects. Furthermore, we found no empirical evidence of any additional spatial patterns in the error terms of the final model.

Conclusions

The model developed in this paper is a two-regime spatial Durbin model with spatial and time-period fixed effects. We demonstrated the ML estimator of this model and found that this estimator performs as well as, if not better than, its counterpart based on instrumental variables. Since we expect that this model and the application of this estimation technique will also prove beneficial to other empirical applications, a Matlab routine has been developed which can be freely downloaded from the first author's website.

**