MODELS FOR DYNAMIC PANELS IN SPACE AND TIME
AN APPLICATION TO REGIONAL UNEMPLOYMENT IN THE EU

J.Paul Elhorst
University of Groningen
Department of Economics
P.O. Box 800
9700 AV Groningen
the Netherlands
Phone 31.50.3633893
Fax 31.50.3637337
Email J.P.Elhorst@eco.rug.nl

One of the central research questions in modelling space-time data is the right econometric model. Three potential problems one must deal with are serial dependence between the observations on each spatial unit over time, spatial dependence between the observations on the spatial units at each point in time, and unobservable spatial and time period specific effects. As we have no a priori reasons to believe that one problem is more important than another, this paper presents a general model assembling time-series, spatial cross-section and panel data econometrics within one framework. We also express an economic explanation for this model, derive the unconditional maximum likelihood function, present the conditions under which the model is stationary, and give an overview of parameter restrictions that lead to simpler models. As an application, the relationship between the unemployment rate, the labor force participation rate and the employment growth rate is investigated based on space-time data of 113 regions across 9 countries of the EU over the period 1989-1996.

Paper prepared for 45th meetings of the European Regional Science Association in Amsterdam, 23-27 August 2005.

1. INTRODUCTION

Econometric modeling of space-time data calls for quite complex stochastic specifications. Three potential problems one must deal with are serial dependence between the observations on each spatial unit over time, spatial dependence between the observations on the spatial units at each point in time, and unobservable spatial and time period specific effects. The first problem is the domain of the time-series econometrics literature (Hamilton, 1994; Hendry, 1995), the second problem of the spatial cross-section econometrics literature (Anselin, 1988) and the third problem of the panel data econometrics literature (Hsiao, 1986; Baltagi, 2001). As we have no a priori reason to believe that one problem is more important than another, the best we can do is to start with a general model assembling time-series, spatial cross-section and panel data econometrics within one framework and then let the data speak. For this purpose we consider a first-order autoregressive distributed lag model in both space and time with spatial and time period fixed effects. Along with this model, a set of simpler econometric models is presented that are subsumed by this model, some of which are standard and frequently used in applied research.

The panel data literature has extensively discussed the dynamic but non-spatial panel data model (Hsiao, 1986, Ch.4; Baltagi, 2001, Ch.8; among others). The most serious estimation problem caused by the introduction of a serially lagged dependent variable (Y_{t-1}) is that the least-squres dummy variables (LSDV) estimator of the coefficient of the lagged dependent variable and the coefficients of the explanatory variables is inconsistent if the number of observations in the time domain (T) is fixed, regardless of the number of observation in the cross-section domain (N). Two procedures to remove this inconsistency are being intensely discussed in the panel data literature.

The first procedure considers the unconditional likelihood function of the model formulated in levels. Regression equations that include variables lagged one period in time are often estimated conditional upon the first observations. When estimating these models by ML it is also possible to obtain unconditional results by taking into account the density function of the first observation of each time-series of observations. This so-called unconditional likelihood function has shown to exist when applying this procedure to a standard linear regression model without exogenous explanatory variables (Hamilton, 1994; Johnston and Dinardo, 1997, pp.229-230), and on a random effects model without exogenous explanatory variables (Hsiao, Pesaran and Thamiscioglu, 2002). Unfortunately, the unconditional likelihood function does not exist when applying this procedure on 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 left-hand side variable Y_t and the right-hand side variables Y_{t-1} and X also does not work, because this technique creates a correlation of order (1/T) between Y_{t-1} and the demeaned error terms

(Nickell, 1981; Hsiao, 1986, pp.73-76), as a result of which the coefficient of Y_{t-1} cannot be estimated consistently. Only when T tends to infinity, does this inconsistency disappear.

The second procedure first differences the model to eliminate the fixed effects, and then applies GMM (generalized method-of-moments) using a set of appropriate instruments.[1] This is the most popular estimation method of dynamic panel models at the moment. Recently, Hsiao, Pesaran and Thamiscioglu (2002) have suggested a third procedure that combines the preceding two. This procedure first differences the model to eliminate the fixed effects and then considers the unconditional likelihood function of the first-differenced model taking into account the density function of the first-differenced observations on each spatial unit. Hsiao, Pesaran and Thamiscioglu (2002) prove that this procedure yields a consistent estimator of response parameters when the cross-sectional dimension N tends to infinity, regardless of the size T. They also show that the ML estimator is asymptotically more efficient that the GMM estimator. The advantage of the last procedure is that it also opens the possibility to estimate a fixed effects dynamic panel model extended to include a spatially lagged dependent variable (WY_t),[2] which is necessary to meet the objective of this paper. The choice to utilize the ML estimator and not the GMM estimator is that the ML estimator of models with a spatially lagged dependent variable (or a spatially autocorrelated error term) tends to be more accurate than the GMM estimator.[3] This is because the coefficient τ of the serially lagged dependent variable (Y_{t-1}), the coefficient δ of the spatially lagged dependent variable (WY_t), and the coefficient η of the serially/spatially lagged dependent variable (WY_{t-1}) are bounded from below and above using ML, whereas they are unbounded using GMM; the transformation of the estimation model from the error term to the variable to explain contains a Jacobian term, which the ML approach takes into account but the GMM approach does not. In its simplest form, this Jacobian term leads to the standard condition known from the time-series literature that $|\tau|<1$, and to the standard condition known from the spatial econometrics literature that $1/\omega_{min} < \delta < 1/\omega_{max}$, where ω_{min} and ω_{max} denote the minimum and maximum eigenvalue of the matrix W describing the spatial arrangement of the units in the sample. We find that the simultaneous modeling of spatial and temporal effects leads to constraints on the coefficients that go beyond these standard conditions.

This paper consists of one technical, one empirical and one concluding section. In the technical section we introduce the general model, its mathematics and its justification. In the empirical section, the relationship between the unemployment rate, the labor force participation rate and the employment growth rate is investigated based on space-time data of

[1] Although these instruments can be obtained from the moment conditions in principle, the number and kind of moment conditions, and therefore the number and kind of instruments involved, are in a state of flux (see for an overview Baltagi, 2001, Ch.8).
[2] The spatial weight matrix W is introduced below.
[3] See Das, Kelejian and Prucha (2003) for the loss of accuracy in models containing a spatially lagged dependent variable and/or a spatially autocorrelated error term.

157 regions across 10 countries of the EU over the period 1988-1997. The concluding section recapitulates our major findings.

2. A GENERAL MODEL

This paper focuses on a first-order serial and spatial autoregressive distributed lag model with spatial and time period fixed effects. The model is considered in vector form for a cross-section of observations at time t

$$Y_t = \tau Y_{t-1}+\delta WY_t+\eta WY_{t-1}+X_t\beta_1+X_{t-1}\beta_2+WX_t\beta_3+WX_{t-1}\beta_4+\mu+\lambda_t e_N+v_t, \quad (1)$$

where Y_t denotes a N×1 vector consisting of one observation for every spatial unit (i=1,...,N) of the dependent variable in the tth time period (t=1,...,T). X_t denotes a N×K matrix of the independent variables. A vector or matrix with subscript t-1 denotes its serially lagged value. It is assumed that the vector Y_0 and matrix X_0 of initial observations are observable. W represents an N×N non-negative spatial weight matrix with zeros on the diagonal describing the spatial arrangement of the spatial units. A vector or matrix premultiplied by W denotes its spatially lagged value. The scalars τ, δ, η and the vectors β_1, β_2, β_3 and β_4 represent the response parameters of the model. The disturbance term consists of $\mu = (\mu_1,...,\mu_N)'$, $\lambda_t e_N$ and $v_t = (v_{1t},...,v_{Nt})'$, where μ_i denotes a spatial-specific effect, λ_t denotes a time-specific effect, which is multiplied with a N×1 vector of unit elements, and v_{it} are indepedently and identically distributed error terms for all i and t with zero mean and variance σ^2. The standard reasoning behind spatial- and time-specific effects is that they control for all time-invariant variables and all spatial-invariant variables whose omission could bias the estimates in a typical cross-section or time-series study.

In section 2.2 we give an explanation for introducing lags in both space and time and in section 2.3 for introducing fixed effects. In section 2.4 we derive the unconditional maximum likelihood function of the general model and in section 2.5 we present the conditions under which this model is stationary. In section 2.6 we present parameter restrictions that lead to simpler models.

2.2 THE EXPLANATION OF LAGS IN SPACE AND TIME

There are a number of reasons why serial lags appear in econometric equations. First, a household may not change its consumption level and labor supply immediately in response to a change in prices or its income. Similarly, a firm may react with some delay to changes in costs and to changes in the demand for its product. Second, lags can arise because of imperfect information. Economic agents require time to gather relevant information, and this delays the making of decisions. There are also occasions when institutional factors can result

in lags. Households may be contractually obliged to supply a certain level of labor hours, even though other conditions would indicate a reduction or increase in labor supply.

Similarly, there are a number of reasons why spatial lags appear in econometric equations. The main reason that one observation associated with a location depends on observations at other locations is that distance affects household and firm behavior. Each household may change its location, consumption and labor supply decisions, and each firm may change its location, input demand and output supply decisions, depending on the market conditions in the home region compared to other regions and on the distance or travel time to these regions. These notions have been formulated in regional science theory that relies on notions of spatial interaction and diffusion effects, hierarchies of place and spatial spillovers. Brueckner (2003) gives an overview of empirical studies analyzing local government behavior.

Another reason to expect both serial and spatial lags is that data collection of observations associated with spatial units over time might reflect measurement error. This would occur if the administrative boundaries for collecting information ─ the arbitrary delineation of space into different units (countries, states, provinces, countries, tracts or zip codes), and of time into different units (years, quarters or months) ─ do not accurately reflect the nature of the underlying process generating the sample data. As an example, consider the relationship between the unemployment rate and the labor force participation rate. Since laborers may travel up and down daily from one spatial unit to another, unemployment and labor force participation rates measured on the basis of where people live could exhibit spatial dependence. Similarly, as unemployed people may find a job that starts the next time period, unemployment and labor force participation rates measured on the basis of people's labor market status at a particular point in time could exhibit serial dependence.

One crucial practical question is the order of the lag structure that should be assigned to each variable. One may look to economic theory for guidance, but while economic theory is often quite helpful to entail long-run equilibrium relationships it often has little to say about the short-run dynamics of how this equilibrium is approached. In practice, dynamic factors can often only be uncovered by allowing actual data to determine the appropriate structure. This paper is restricted to lag structures of the first-order. As the emphasis in this paper is on models with both serial and spatial effects, higher order lag structures are not discussed.

2.3 THE EXPLANATION OF SPATIAL AND TIME PERIOD FIXED EFFECTS

A space-time model, even if it is dynamic, still assumes that the spatial units are completely homogeneous, differing only in their explanatory variables. A panel data approach would presume that spatial heterogeneity is a feature of the data and attempt to model that heterogeneity.

The need to account for spatial heterogeneity is that spatial units are likely to differ in their background variables, which are usually space-specific time-invariant variables that affect the

dependent variable, but are difficult to measure or hard to obtain. Omission of these variables leads to bias in the resulting estimates. One remedy is to introduce a variable intercept μ_i representing the effect of the omitted variables that are peculiar to each spatial unit considered (Baltagi 2001, ch.1). Conditional upon the specification of the variable intercept μ_i, the regression equation can be estimated as a fixed or a random effects model. In the fixed effects model, a dummy variable is introduced for each spatial unit as a measure of the variable intercept. In the random effects model, the variable intercept is treated as a random variable that is independently and identically distributed with zero mean and variance σ_μ^2.

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: (i) 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 (ii) the 'increasing domain' asymptotic structure, where the sampling region grows as $N \to \infty$ and the sample design is such that 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: (i) the stochastic design where the spatial units are randomly drawn; and (ii) 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 (Lahiri, 2003; Cressie, 1993, p.100; Griffith and Lagona, 1998). 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 spatial units, 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), and 'the individual spatial units have characteristics that actually set them apart from a larger population' (Anselin, 1988, p.51). According to 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'. In addition, the traditional assumption of zero correlation between μ_i in the random effects model and the explanatory variables is particularly restrictive. For these reasons the random effects model is often not employed.

A similar type of reasoning apply to time-specific effects. Time periods are likely to differ in their background variables, which are usually time-specific spatial-invariant variables that affect the dependent variable, but are difficult to measure or hard to obtain. Omission of these variables leads to bias in the resulting estimates. Generally, time period effects (λ_t) are assumed to be fixed and justified by events such as policy interventions, structural breaks, sudden shocks (war, climatic catastrophes), nonlinear time trends, et cetera.

2.3 UNCONDITIONAL MAXIMUM LIKELIHOOD ESTIMATION

To shorten the notation, we rewrite the general model as

$$BY_t = AY_{t-1} + \beta_{X_t} + \mu + v_t, \quad \text{with } B = I_N - \delta W,\ A = \tau I_N + \eta W, \text{ and}$$
$$\beta_{X_t} = \lambda_t e_N + X_t\beta_1 + X_{t-1}\beta_2 + WX_t\beta_3 + WX_{t-1}\beta_4. \tag{2}$$

Taking first differences of (2), the model changes into

$$B\Delta Y_t = A\Delta Y_{t-1} + \beta_{\Delta X_t} + \Delta v_t. \tag{3}$$

Note that first-differencing a regression equation formulated in levels to eliminate the fixed effects in the cross-sectional domain, does not eliminate the time period fixed effects. Furthermore, the structure of these first-differenced time period fixed effects is such that common time dummies can replace them. Only the first time dummy disappears. This implies that $\beta_{\Delta X_t}$ can be written as

$$\beta_{\Delta X_t} = \lambda_t e_N + \Delta X_t\beta_1 + \Delta X_{t-1}\beta_2 + W\Delta X_t\beta_3 + W\Delta X_{t-1}\beta_4. \tag{4}$$

ΔY_t is well defined for t=2,…,T, but not for ΔY_1 because ΔY_0 is not observed. To be able to specify the maximum likelihood function of the complete sample ΔY_t (t=1,…,T), the probability function of ΔY_1 must be derived first. Therefore, we repeatedly lag equation (2) by one period. For ΔY_{t-m} (m≥1) we get

$$B\Delta Y_{t-m} = A\Delta Y_{t-(m+1)} + \beta_{\Delta X_{t-m}} + \Delta v_{t-m}, \text{ or}$$
$$\Delta Y_{t-m} = AB^{-1}\Delta Y_{t-(m+1)} + B^{-1}\beta_{\Delta X_{t-m}} + B^{-1}\Delta v_{t-m}. \tag{5}$$

Then by substitution of ΔY_{t-1} into (2), next ΔY_{t-2} into (2) up to $\Delta Y_{t-(m-1)}$ into (2), we get

$$\begin{aligned} B\Delta Y_t = {} & A(AB^{-1})^{m-1}\Delta Y_{t-m} + \Delta v_t + AB^{-1}\Delta v_{t-1} + \ldots + (AB^{-1})^{m-1}\Delta v_{t-(m-1)} + \\ & \beta_{\Delta X_t} + AB^{-1}\beta_{\Delta X_{t-1}} + \ldots + (AB^{-1})^{m-1}\beta_{\Delta X_{t-(m-1)}}. \end{aligned} \tag{6}$$

Using $\Delta v_{t-j} = v_{t-j} - v_{t-j-1}$ (j=0,…,m-1), the m disturbance terms in this equation may be rewritten as

$$v^* = v_t + (\Pi - I_N)v_{t-1} + (\Pi - I_N)\Pi v_{t-2} + ... + (\Pi - I_N)\Pi^{m-2}v_{t-(m-1)} - \Pi^{m-1}v_{t-m}. \qquad (7)$$

where $\Pi = AB^{-1} = (\tau I_N + \eta W)(I_N - \delta W)^{-1}$.

Assume for the moment that the model does not contain exogenous explanatory variables (β=0), i.e, the second line of equation (6) is discarded. Since the successive values of v_t are uncorrelated, we then have

$$E(B\Delta Y_t) = A\Pi^{m-1}\Delta Y_{t-m} \text{ and } \quad Var(B\Delta Y_t) = \sigma^2 V_b, \qquad (8)$$

where the N×N matrix V_b is defined as

$$V_b = I_N + (\Pi - I_N)(I_N - \Pi\Pi')^{-1}(\Pi - I_N)' - (\Pi - I_N)\Pi^{m-1}(I_N - \Pi\Pi')^{-1}\Pi'^{m-1}(\Pi - I_N)' - (\Pi\Pi')^{m-1} \qquad (9)$$

Note that if the matrix W is symmetric, V_b reduces to the simpler form

$$V_b = 2*(I_N + \Pi)^{-1}(I_N + \Pi^{2m-1}). \qquad (10)$$

Depending on wether the process has reached stationarity or not, we may subsequently assume either that (cf. Hsiao, Pesaran, and Thamiscioglu 2001):

[I] The process has started long ago (m approaches infinity) and $|\Pi| < 1$. Under this assumption, $E(\Delta BY_1) = 0$, while V_b reduces to $V_b = 2*(I_N + \Pi)^{-1}$.

[II] The process started in the past, but not too far back from the 0th period, and $E(B\Delta Y_1) = \pi_0 e_N$, where π_0 is a fixed but unknown parameter to be estimated.[4] In words, the expected *changes* in the initial endowments of the spatial units follow a first-order spatial autoregressive lag model, just as the original model, which is the same across all spatial units. Note that this assumption, although restrictive, does not impose the even stronger restriction that all spatial units should start from the same initial endowments.

It can be seen that the first assumption reduces to the second one when $\pi_0 = 0$, $|\Pi| < 1$, and m is sufficiently large so that the term Π^m becomes negligible. Therefore, we consider the unconditional log-likelihood function of the complete sample under the more general assumption [II].

Writing the residuals of the model as $\Delta v_1 = B\Delta Y_1 - \pi_0 I_N$ for t=1 and $\Delta v_t = B\Delta Y_t - A\Delta Y_{t-1}$ for t=2,…,T, we have $Var(\Delta v_1) = \sigma^2 B^{-1} V_b B'^{-1}$,

[4] This parameter may also be considered the first time dummy that disappeared by taking first differences.

$\mathrm{Var}(\Delta v_t) = 2\sigma^2 B^{-1}B'^{-1}$ (t=2,…,T), $\mathrm{Covar}(\Delta v_t, \Delta v_{t-1}) = -\sigma^2 B^{-1}B'^{-1}$ (t=2,…T), and zero otherwise. This implies that the covariance matrix of Δv can be written as $\mathrm{Var}(\Delta v) = \sigma^2[(I_T \otimes B^{-1})H_{V_b}(I_T \otimes B'^{-1})]$, by which the NT×NT matrix $H_V|_{V=V_b}$ is defined as

$$H_V \equiv \begin{bmatrix} V & -I_N & 0 & . & 0 & 0 \\ -I_N & 2\times I_N & -I_N & . & 0 & 0 \\ 0 & -I_N & 2\times I_N & . & 0 & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & 2\times I_N & -I_N \\ 0 & 0 & 0 & . & -I_N & 2\times I_N \end{bmatrix}, \tag{11}$$

with its submatrix in the first block-row and first block-column set to the N×N matrix V, in this case $V=V_b$.

We now add back explanatory variables to the model. If X_t is strictly exogenous and to be generated by a stationary process in time, we have $E\Delta X_t = 0$ and thus again $E(B\Delta Y_1) = A\Pi^{m-1}\Delta Y_{t-m}$. This expectation is determined under assumption [II]. By contrast, $\mathrm{Var}(B\Delta Y_1)$ is undetermined, because lagged values of ΔX_{t-j} if t=1 are not observed. This implies that the probability function of ΔY_1 is also undetermined. The panel data literature has suggested different approximations for these lagged values leading to different optimal estimation procedures. The leading approximation in the early dynamic panel literature is of Bhargava and Sargan (1983), which is also applied in Hsiao, Pesaran and Thamiscioglu (2002). They suggest predicting the unobserved values of X by all the observed exogenous variables in the model subdivided by time over the observation period. Recently, Nerlove and Balestra (1996) and Nerlove (1999 or 2000) introduced another approximation. Elhorst (2005) has compared both approximations and has concluded that the Nerlove and Balestra approximation outperforms the Barghava and Sargan approximation.[5]

Starting with a regression equation formulated in levels (instead of first differences), Nerlove and Balestra (1996) and Nerlove (1999 or 2000) suggest replacing the variance of unobservable X variables by Σ_X, where Σ_X denotes the covariance matrix of the observable explanatory variables X. This covariance matrix may be determined from the sample data in advance and then used to calculate the unknown variance of $\sum_{j=0}^{m-1}\Pi^j\beta_{\Delta X_{t-j}}$ (see the second line

[5] He estimated a dynamic demand model for cigarettes based on a long (30 years) and a short panel (5 years) of 46 US states and compared the short-term and long-term elasticities. The elasticities obtained from short panel estimations appeared to be closer to those obtained from long panel estimations when using the Nerlove and Balestra approximation. The root mean squared error of predictions at a forecast horizon of one year, five years and ten years also appeared to be smaller when using the Nerlove and Balestra approximation.

of eq. 6). Suppose that each explanatory variable X_{tk} (k=1,…,K) follows a well-specified common stationary time series model

$$X_{tk} = \tau_{X_k} X_{t-1k} + \gamma_t, \quad \text{where} \quad \gamma_t \sim N(0, \sigma^2_{\gamma X_k} I_N). \tag{12}$$

Then the sum term over the unobservable X_t has a well-defined variance $\Sigma_{\Sigma X}$, which is a function of β and $\tau_{X_k}, \sigma^2_{\gamma X_k}$ (k=1,…,K). Although it would be possible to determine the resulting log-likelihood function based on $\Sigma_{\Sigma X}$, this covariance matrix depends on so many parameters that its practical value in empirical applications is almost nil (unless K is very small). Nerlove and Balestra (1996) and Nerlove (1999 or 2000) have pointed out that it is not necessary to go that far. Since we are not really interested in the parameters τ_{X_k} and $\sigma^2_{\gamma X_k}$ (k=1,…,K), we can suppress these parameters and restrict the log-likelihood to the remaining parameters. While omitting estimation of τ_{X_k} and $\sigma^2_{\gamma X_k}$ (k=1,…,K) leads to a loss of efficiency, the ML estimates obtained in this way remain consistent as long as the random variables have well-defined variances and covariances, which they will if the explanatory variables are generated by a stationary process.

Following Nerlove and Balestra, but then for a regression equation formulated in first differences and including time period fixed effects, $\mathrm{Var}(B\Delta Y_1)$ might be approached by (see also eq. 8)

$$\mathrm{Var}(B\Delta Y_1) = \sigma^2 V_b + \mathrm{Var}(\sum_{j=0}^{m-1} \Pi^j \beta_{\Delta X_{1-j}}) = \sigma^2 (V_b + \frac{1}{\sigma^2} \Sigma_{\Sigma\Delta X}), \tag{13a}$$

$$\text{where } \Sigma_{\Sigma\Lambda X} = (I_N - \Pi)^{-1}(I_N - \Pi^m)\theta' \Sigma_{\Delta X} \theta (I_N - \Pi'^m)(I_N - \Pi')^{-1}, \tag{13b}$$

$$\theta = [\lambda_2, ..., \lambda_T, \beta_1', \beta_2', \beta_3', \beta_4']', \tag{13c}$$

$$\Delta X = \begin{bmatrix} e_N & ... & 0 & \Delta X_2 & \Delta X_1 & W\Delta X_2 & W\Delta X_1 \\ . & . & . & . & . & . & . \\ 0 & ... & e_N & \Delta X_T & \Delta X_{T-1} & W\Delta X_T & W\Delta X_{T-1} \end{bmatrix} \tag{13d}$$

Let $V_{NB} = V_b + 1/\sigma^2 \Sigma_{\Sigma\Delta X}$ with V_b specified as in (9) or (10), then the covariance matrix of Δv can be written as $\mathrm{Var}(\Delta v) = \sigma^2[(I_T \otimes B^{-1}) H_{V_{NB}} (I_T \otimes B'^{-1})]$, by which the matrix $H_V |_{V=V_{NB}}$ is defined by (11). This implies that the joint probability function of the complete sample of size T is

$$f(\Delta Y_1,...,\Delta Y_T) = (2\pi\sigma^2)^{-T/2} \,|(I_T \otimes B^{-1})H_{V_{NB}}(I_T \otimes B'^{-1})|^{-1/2} \times \exp(-\frac{1}{2\sigma^2}\Delta v'(I_T \otimes B^{-1})H_{V_{NB}}(I_T \otimes B'^{-1})\Delta v). \quad (14)$$

To obtain the log-likelihood, we utilize three sets of properties. First, we have $|(I_T \otimes B^{-1})H_{V_{NB}}(I_T \otimes B'^{-1})| = |H_{V_{NB}} \otimes B^{-1}B'^{-1}| = |H_{V_{NB}}|^N |B^{-1}B'^{-1}|^T$ (Magnus and Neudecker, 1988, p.29), so that

$$\log|H_{V_{NB}} \otimes B^{-1}B'^{-1}|^{-1/2} = -\frac{1}{2}[N\log|H_{V_{NB}}| - 2T\log|B|] = -\frac{N}{2}\log|H_{V_{NB}}| + T\log|B|. \quad (15)$$

Properties of the matrix H_V are (Elhorst, 2005): (i) The determinant is $|H_V| = |I_N - T\times I_N + T\times V|$; (ii) The inverse is $H_V^{-1} = (1-T)(H_0^{-1} \otimes D^{-1}) + ((H_1^{-1} - (1-T)H_0^{-1}) \otimes (D^{-1}V)$, where $D = I_N - T\times I_N + T\times V$ and the inverse matrices $H_0^{-1} = H_V^{-1}|_{V=0}$ and $H_1^{-1} = H_V^{-1}|_{V=I}$ can easily be calculated and are characterized by a specific structure; and (iii) H_V^{-1} can be partitioned in T block-rows and T block columns, by which the submatrix $H_V^{-1}(t_1,t_2)$ (t_1,t_2=1,…,T) equals $H_V^{-1}(t_1,t_2) = (1-T)H_0^{-1}(t_1,t_2)\times D^{-1} + (H_1^{-1}(t_1,t_2) - (1-T)H_0^{-1}(t_1,t_2)\times(D^{-1}V)$. The last equation is used to obtain the matrix H_V^{-1} computationally.

Properties of the matrix W, assuming that its characteristic roots denoted by ω_i (i=1,…,N) are known, are (Griffith 1988, p. 44, table 3.1): (i) if W is multiplied by some scalar constant, then its characteristic roots are also multiplied by this constant; (ii) if δI is added to W, where δ is a real scalar, then δ is added to each of the characteristic roots of W; (iii) the characteristic roots of W and its transpose are the same; (iv) the characteristic roots of W and its inverse are inverses of each other; and (v) if W is powered by some real number, each of its characteristic roots is powered by this same real number.

Using these properties, the log-likelihood function is obtained as

$$\mathrm{LogL} = -\frac{NT}{2}\log(2\pi\sigma^2) + T\sum_{i=1}^{N}\log(1-\delta\omega_i) - \frac{1}{2\sigma^2}\Delta e' H_{V_{NB}}^{-1}\Delta e$$
$$-\frac{1}{2}\sum_{i=1}^{N}\log\left[1 - T + \frac{2T(1-\delta\omega_i)}{(1-(\delta-\eta)\omega_i+\tau)}(1+(\frac{\tau+\eta\omega_i}{1-\delta\omega_i})^{2m-1}) + T\frac{\theta'\Sigma\theta}{\sigma^2}\frac{(1-\delta\omega_i)^2}{(1-(\delta+\eta)\omega_i-\tau)^2}(1-(\frac{\tau+\eta\omega_i}{1-\delta\omega_i})^m)^2\right], \quad (16)$$

where $\Delta e = \begin{bmatrix} B\Delta Y_1 - \pi_0 1_N \\ B\Delta Y_2 - A\Delta Y_1 - \beta_{\Delta X_2} \\ . \\ B\Delta Y_T - A\Delta Y_{T-1} - \beta_{\Delta X_T} \end{bmatrix}$, $E(\Delta e \Delta e') = \sigma^2 H_{V_{NB}}$.

This log-likelihood function is well-defined, satisfies the usual regularity conditions and contains 4K+T+4 unknown parameters to be estimated: $\beta_1, \beta_2, \beta_3, \beta_4, \pi_0, \lambda_2, ..., \lambda_T, \tau, \delta, \eta$ and σ^2. An appropriate value of m should be chosen in advance. None of the parameters can be solved analytically from the first-order maximizing conditions. This implies that a numerical iterative procedure must be used to find the maximum for all the parameters simultaneously.

Unfortunately, this procedure also includes the coefficients of the time period fixed effects. Due to the inclusion of the density function of the cross-section of first observations, the model is no longer linear in its parameters. Consequently, it is not allowed to determine the means of the Y and X variables separately of each cross-section of observations at a particular point in time, transform these variables by subtracting out the appropriate cross-section means, and then to maximize the above log-likelihood function for 4K+4 parameters based on the transformed data.

2.5 INVERTIBILITY AND STATIONARITY

To derive the log-likelihood function, we have made the assumption that the matrix $B = (I_N - \delta W)$ is invertible and thus that the determinant of B, $|B| = |I_N - \delta W| = \prod_{i=1}^{N}(1 - \delta\omega_i)$, the second term in the log-likelihood function, is not zero. Let ω_i be the N ordered characteristic roots of W such that $\sum_{i=1}^{N}\omega_i = 0$. A non-zero determinant does not constrain δ, except that $\delta \neq 1/\omega_i$ (i=1,…,N). However, invertibility requires that

$$(I_N - \delta W)^{-1} = I_N + \delta W + \delta^2 W^2 + ... \tag{17}$$

is a convergent series expansion and this immediately implies that $1/\omega_{min} < \delta < 1/\omega_{max}$ (Ord, 1981). For ease of interpretation, W is often standardized such that the elements of a row sum to one. The elements of a row-standardized weight matrix thus equal $w_{ij}^S = w_{ij} / \Sigma_j w_{ij}$. This ensures that all weights are between 0 and 1 and facilitates the interpretation of the operations with the weight matrix as an averaging of neighboring values. For a row-standardized weight matrix, the largest eigenvalue is always +1. A side effect of row standardization is that the resulting matrix is likely to become asymmetric, even though the original matrix is symmetric. As an alternative, the elements of W may be divided by its largest eigenvalue of ω_{max}, $W^S = 1/\omega_{max} W$. This has the effect that the characteristic roots of W are also divided by ω_{max}, as a result of which $\omega^S_{max} = 1$, just like the largest characteristic root of a row-

standardized matrix. The advantage of this alternative standardization is that the spatial weight matrix is kept symmetric. We have seen that this considerably simplifies the matrix V_b.

Since a space-time data set has two dimensions, it is possible to consider asymptotic behavior as $N\rightarrow\infty$, $T\rightarrow\infty$, or both. Generally speaking, it is easier to increase the cross-section dimension of a space-time data set. If as a result $N\rightarrow\infty$ is believed to be the most relevant asymptotics, it is not necessary to assume $|\Pi|<1$ as long as T is fixed. By contrast, Nerlove (1999) has pointed out that the cross-section of first observations conveys a great deal of information about the process generating the data since these observations reflect how that process has operated in the past. Thus, conditioning on the cross-section of first observations is an undesirable feature, especially when the time dimension of the space-time data set is short.

When it is assumed that $|\Pi|=|AB^{-1}|=|(\tau I_N+\eta W)(I_N-\delta W)^{-1}|<1$, the process generating the data is stationary in time. A more detailed description of this condition is given in figure 1 and in the first row of table 1. If W is standardized, $\omega_{max}=1$ and $-1<\omega_{min}<0$. It can be seen that these conditions are also captured by the log-likelihood function (16) in that it is not defined for parameter values that do not satisfy these conditions. Figure 1 illustrates that the relationship between the spatial and temporal parameters introduces constraints that go beyond the standard condition $|\tau|<1$ in time-series models and the standard condition $1/\omega_{min}<\delta<1/\omega_{max}$ in spatial cross-section models.

The critical condition for stationarity in space is that the row and column sums should not diverge to infinity at a rate equal to or faster than the rate of the sample size N in the cross-section domain (Lee, 2002).[6] This condition should hold before the spatial weight matrix W is standardized.

2.6 SIMPLER MODELS

If restrictions are imposed on the parameters, the general model reduces to a simpler model. Figure 2 summarizes nine econometric models subsumed by (1), some of which are frequently used in applied research, notably model 4, 6, 8, 9 and 10. This are models that either exploit the time-series dimension or the spatial dimension of the data. The number of studies considering both serial and spatial effects, such as model 1, 2, 3, 5 and 7, is still rather small. If restrictions are imposed on the response parameters of the general model, the stationarity conditions also simplify. Table 1 reports the stationarity conditions for the restricted models.

In addition to the restrictions given in figure 1, we may also exclude spatial and/or time period fixed effects. This implies that forty model are considered: each model in figure 1 with

[6] To limit the correlation of sample observations across different spatial units to a manageable degree Kelejian and Prucha (1999) assumed that the row and column sums of W are uniformly bounded. Griffith and Lagona (1998) established that the correlation between two spatial units should converge to zero as the distance separating them increases to infinity.

Figure 1 *The area on which the log-likelihood function is defined for τ, δ and η*

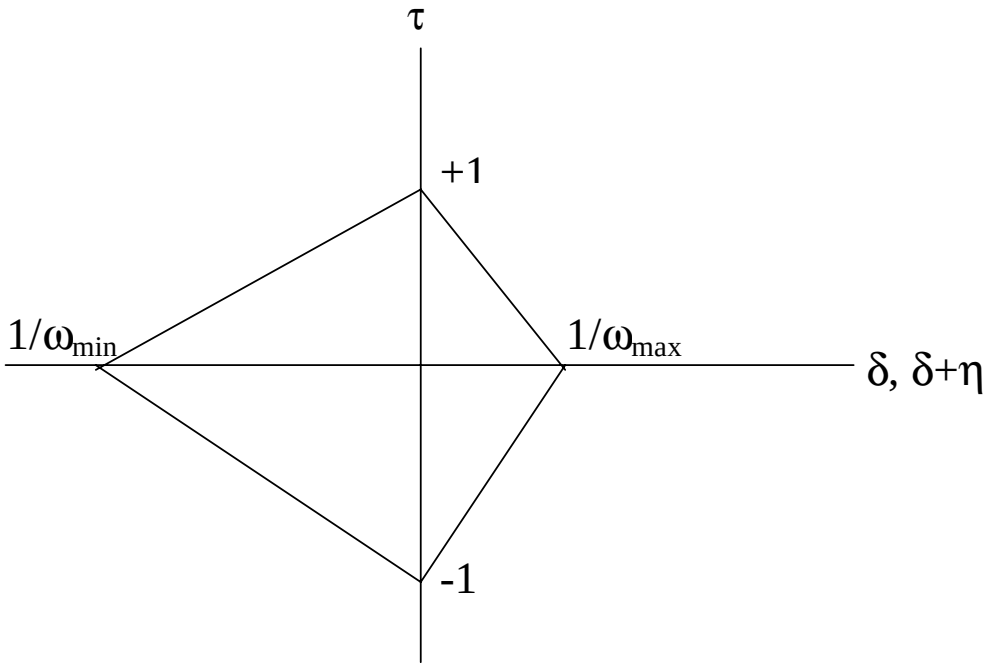
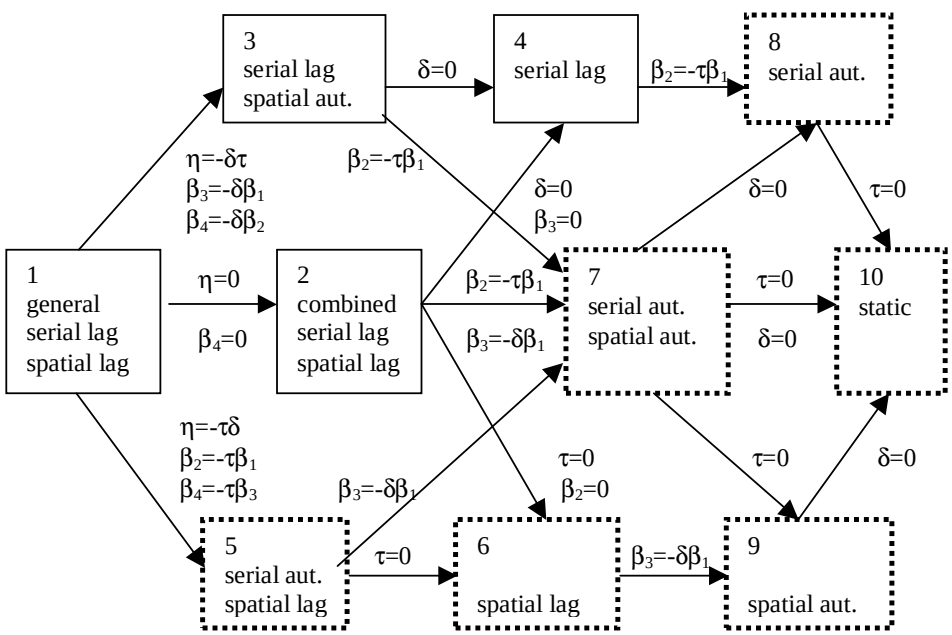


Figure 2 *Restricted econometric models subsumed by the general model*



both spatial and time period fixed effects, with only spatial fixed effects, with only time period fixed effects, or without fixed effects.

Figure 2 suggests that the researcher begins with a theory-based model specification that is viewed as being correct in terms of explanatory variables, goodness of fit, correctness of signs, and significant t-statistics. Theory may provide insight into the specification for functional form, but it provides almost

Table 1 *Conditions on restricted models to ensure stationarity and invertibilty*

1. Spatial and serial lag general	$\lvert\tau\rvert<1-(\delta+\eta)\omega_{max}$	if $\delta+\eta\geq 0$,	$1/\omega_{min}<\delta<1/\omega_{max}$
	$\lvert\tau\rvert<1-(\delta+\eta)\omega_{min}$	if $\delta+\eta<0$,	
2. Spatial and serial lag combined	$\lvert\tau\rvert<1-\delta\omega_{max}$	if $\delta\geq 0$,	$1/\omega_{min}<\delta<1/\omega_{max}$
	$\lvert\tau\rvert<1-\delta\omega_{min}$	if $\delta<0$	
3. Serial lag and spatial autocorrelation	$\lvert\tau\rvert<1$		$1/\omega_{min}<\delta<1/\omega_{max}$
4. Serial lag	$\lvert\tau\rvert<1$		-
5. Spatial lag and serial autocorrelation	$\lvert\tau\rvert<1$		$1/\omega_{min}<\delta<1/\omega_{max}$
6. Spatial lag	-		$1/\omega_{min}<\delta<1/\omega_{max}$
7. Spatial and serial autocorrelation	$\lvert\tau\rvert<1-\delta\omega_{max}$	if $\delta\geq 0$,	$1/\omega_{min}<\delta<1/\omega_{max}$
	$\lvert\tau\rvert<1-\delta\omega_{min}$	if $\delta<0$	
8. Serial autocorrelation	$\lvert\tau\rvert<1$		-
9. Spatial autocorrelation	-		$1/\omega_{min}<\delta<1/\omega_{max}$
10. Static	-		-

none about dynamic structure (the parameters values on lagged variables in space and/or time) and about the impact of time-invariant and/or spatial-invariant variables. Once the reseracher is satisfied with the specification of the theory-based model, a so-called simplification research is carried out, in which simpler special cases are tested against the general econometric specification at the left-hand side of Figure 2. Since special cases are obtained by placing restrictions on the parameters of the model, statistical tests for such restrictions may be employed for this purpose. This testing-down or general-to-specific procedure has some dangers: the initial model may not be general enough; multicollinearity and other data problems may limit the generality of the initial model; and there is no standard or best sequence of testing and the final model may well depend on the order in which tests are carried out. On the other hand, since this procedure captures both serial and spatial effects, it may throw more light on which of these two effects is more important, an issue that has hardly been investigated up to now.

Not all models have to be estimated in first differences. If the model does not contain spatial fixed effects (μ_i), then it is not necessary to take first differences. Consequently, it would be more efficient to determine the ML estimator of the model formulated in levels and taking into account the density function of the first cross-section of observations also in levels, as first-differencing diminishes the number of observations available for estimation by one for every spatial unit. If we repeat the whole procedure set out in section 2.3 but then for the general formulated in levels and leaving spatial effects aside, the log-likelihood function turns out to be[7]

[7] See Elhorst (2001), but only for models without time period fixed effects and without an approximation for lagged values of the exogenous explanatory variables X.

$$\text{LogL} = -\frac{N(T+1)}{2}\log(2\pi\sigma^2) + (T+1)\sum_{i=1}^{N}\log(1-\delta\omega_i) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} e_t' e_t - \frac{1}{2\sigma^2} e_0'(\tilde{V} + \frac{1}{\sigma^2}\Sigma_{\Sigma X})e_0$$
$$+\frac{1}{2}\sum_{i=1}^{N}\log\left[1 - \frac{(\tau+\eta\omega_i)}{(1-\delta\omega_i)^2} + \frac{\theta'\Sigma\theta}{\sigma^2}\frac{(1-\delta\omega_i)^2}{(1-(\delta+\eta)\omega_i-\tau)^2}(1-(\frac{\tau+\eta\omega_i}{1-\delta\omega_i})^m)^2\right], \tag{18a}$$

$$e_0 = Y_0 - \delta W Y_0 - \pi_0,$$
$$e_t = Y_t - \delta W Y_t - \lambda_t e_N - X_t\beta_1 - X_t\beta_2 - WX_t\beta_3 - WX_{t-1}\beta_4, \quad t > 0, \tag{18b}$$

$$\tilde{V} = (I_N - \Pi^m)(I_N - \Pi\Pi')^{-1}, \tag{18c}$$

$$\theta = [\lambda_1, ..., \lambda_T, \beta_1', \beta_2', \beta_3', \beta_4']', \tag{18d}$$

$$X = \begin{bmatrix} e_N & ... & 0 & X_1 & X_0 & WX_1 & WX_0 \\ . & . & . & . & . & . & . \\ 0 & ... & e_N & X_T & X_{T-1} & WX_T & WX_{T-1} \end{bmatrix}. \tag{18e}$$

Since we have one observation more on every spatial unit, we have T+1 instead of T. Recall that it is assumed that Y_0 and X_0 are observable. $\Sigma_{\Sigma X}$ may be computed just as in equation (13b) but then for X formulated in levels (18e) instead of first differences (13d).

If the model does not contain a serially lagged dependent variable (Y_{t-1}), then it is also not necessary to take first differences. This is the case with models 5-10 in figure 2 (the boxes with dashed lines). The most general model of this group of models is a spatial autoregressive distributed lag model with serial autocorrelation and spatial fixed effects (model 5). The log-likelihood function has been derived in Elhorst (2001).

3. NUMERICAL ILLUSTRATION: REGIONAL UNEMPLOYMENT

As an application, we estimate the unemployment rate as a function of the labor force participation rate and the employment growth rate at regional level. This relationship is the cornerstone of a complete model of the regional unemployment rate (Elhorst, 2003b).

On investigating the order of integration, most studies have found that the unemployment rate and the participation rate at regional level are integrated of order 0 (Blanchard and Katz, 1992; Martin, 1997; Baddeley et al., 1998; Pehkonen and Tervo, 1998). By contrast, the level of employment is often found to be integrated of order 1, and only its growth rate is integrated of order 0 (see Blanchard and Katz (1992) for regional US data and Decressin and Fatás (1995) for regional EU data).

The idea to analyze the unemployment rate, the participation rate and the employment growth rate within one framework stems from a seminal paper of Blanchard and Katz (1992). They found that booms and slumps are best described as transitory accelerations or slowdowns of employment growth. Growth eventually returns to normal, but the level of employment is permanently affected. These transitory changes in growth lead to transitory fluctuations in unemployment, participation and, to the extent that regional wages are flexible on a regional level, a change in the real wage. Suppose a region experiences a positive

demand shock. Initially, one would expect the unemployment rate to fall and the wage rate to rise. The lower unemployment rate and the higher wage rate trigger two adjustment mechanisms. First, labour supply increases through more labor force participation, net inward migration and net inward commuting within a region. Although from a theoretical viewpoint the unemployment rate can also have a positive effect on labour force participation, known as the additional worker effect (instead of the discouraged worker effect), the latter effect dominates empirical research on labour force participation (see Elhorst, 1996). Second, labor demand decreases, because a higher wage makes a region less attractive to firms. The effect of the unemployment rate is uncertain. On the one hand, a lower unemployment rate implies a smaller pool of workers from which to choose and this can prove inattractive to firms. On the other hand, an increase in labour demand in backward regions may stop the selective out-migration of high-skilled workers. This may lessen the adverse effects related to geographical concentrations of high unemployment and may counteract the downward spiral effect of economically depressed regions, experiencing increasing difficulty keeping pace with economically thriving regions. Decressin and Fatás (1995) have found that people moving in and out of the labor force is the most important of these adjustment mechanisms to transitory changes in employment growth in the EU. For this reason, we limit the supply side of our model to the participation rate and the demand side of our model to the employment growth rate.

In contrast to Blanchard and Katz (1992) and Decressin and Fatás (1995), we do not estimate a trivariate system of equations where, in each equation, one of the variables depends on the other variables. By contrast, while Blanchard and Katz (1992) and Decressin and Fatás (1995) regress the unemployment rate on its past value and on current and past values of the labor force participation rate and the employment growth rate in the own region only, we extend the model with current and past neigboring values of the unemployment rate, the labor force participation rate and the employment growth rate.

Our data set consists of 904 observations of 113 regions across 9 EU member countries over the period 1989-1996. When taking first differences, the number of observations reduces to 791, while T becomes 7. The data are harmonized series produced by Eurostat intended to be comparable among EU member states as well as to give a consistent picture of unemployment, participation and employment over time. We have used Eurostat's regional division on the NUTS2 level. The countries with their number of regions between brackets are: Denmark (1), (West-)Germany (28), the Netherlands (12), Belgium (9), Luxembourg (1), France (21), Spain (16), Portugal (5) and Italy (20). These countries and regions form a contiguous area.[8] The spatial weight matrix W used in the estimations is a symmetric normalized binary contiguity matrix.

[8] Isolated regions have been left aside, such as Berlin in the former East-Germany, and the islands of Canarias (Spain), Acores and Madeira (Portugal).

The estimation results obtained for the general model, with m set to 16, are recorded at left side of table 2. The results are consistent with three stylized facts of regional labor market data in general and the regional unemployment rate in particular.

Stylized fact 1-In the EU, regional unemployment rates are strongly correlated over time. The correlation coefficients of the unemployment rate observed in single regions over time are large and diminish slightly over time. In our data set, even the mutual correlation coefficient of observations seven years apart is greater than 0.9. This explains the significant value of 0.489 (T-value 24.62) of the lagged unemployment rate in the regression equation.

Stylized fact 2-Regional unemployment rates parallel the national unemployment rate. Figure 3 charts the regional unemployment rates in deviation of their national counterparts at the beginning and at the end of the observation period for the 113 regions in the sample. The result, as well as the R-squared of regressing the y-values on the x-values, show that there is a tendency for regional unemployment rates to increase and to decrease together in different regions along the national evolution of this variable over time.

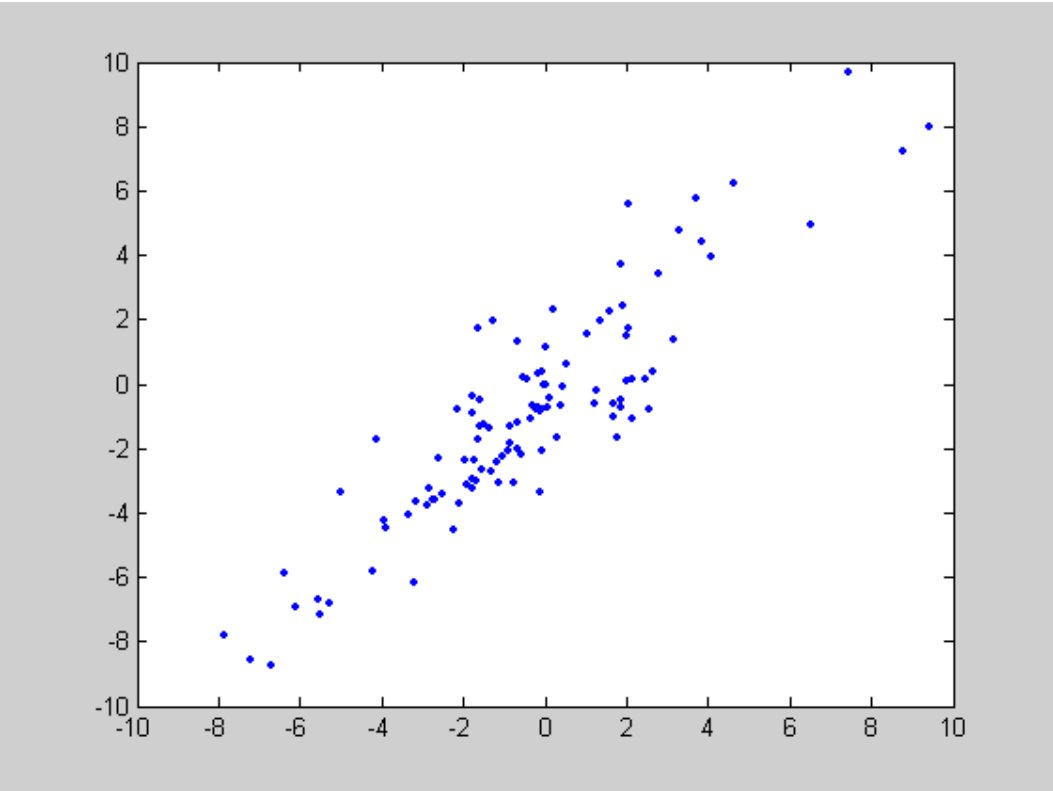


Figure 3 *Regional unemployment in deviation of the national unemployment rate in 1989 on the x-axis versus its counterpart in 1996 on the y-axis (y-value=1,075*x-value, R^2=0.85)*

To investigate the extent to which labor market shocks are shared by all regions and how unemployment, participation and employment growth adjust to labor demand shocks which are region-specific, Blanchard and Katz (1992) and Decressin and Fatás (1995) apply a cyclical sensitivity model. This model explains the regional unemployment rate by the

national unemployment rate, $u_{reg}= a_0+a_1u_{nat}$. A similar approach is also used for the participation rate and the employment growth rate. Thirlwall (1966) originally introduced the cyclical sensitivity model formulated in first differences, and Brechling (1967) in both levels and logarithms. The central point of this type of model is the parameter a_1 that measures cyclical sensitivity, the extent to which a region's unemployment rate changes when the national rate changes. Naturally, this type of model only makes sense if a regression equation is estimated separately for each region; otherwise, a_1 will be equal to unity.[9] Blanchard and Katz (1992) and Decressin and Fatás (1995) use the cyclical sensitivity model to divide the regional unemployment rate into one short-run and one long-run (also called non-cyclical, equilibrium or frictional) component. The long-run component, $u_{reg,long}=u_{reg}-a_1u_{nat}$, and similar components for the participation rate and the employment growth rate are eventually taken up in their trivariate model.

Objections to this approach are the instability of the cyclical component a_1 to the chosen estimation period (Dunn, 1982; Owen and Gillespie, 1982; Byers, 1990; Chapman, 1991) and the absence of any explanation why a_1 should vary across regions (Chapman, 1991; Martin, 1997). Another objection is the separation of the two model stages. In the first stage, a Brechling-Thirlwall type model is estimated for each single region to construct the long-run component, and in the second stage, an explicative model of the long-run component is estimated for all regions taken together. It is more likely that these stages are interdependent, since short-term shocks may have major long-run structural impacts through hysteresis effects, and, conversely, structural shocks may change the cyclical dynamics of certain regions (see Baddeley et al., 1998). To address this problem, one may better pool time series data of different regions and of different variables into one model and include time dummies to correct for common trends along the observations over time, either linear of cyclical.

The results obtained for the time period fixed effects according to this method point to two significant ($\alpha = 0.05$) shocks shared by all regions over the observation period; a positive shock in 1993 (T-value 9.04) and a negative shock in 1995 (T-value –2.77). Given that for most regions a cyclical component of unity could not be rejected by the data, Blanchard and Katz (1992) eventually decided to take the unemployment rate, the participation rate and the employment growth rate in deviation of their national national counterparts each year, i.e., a_1=1, which is computationally equivalent with our approach of adding a set of time period fixed effects to the regression equation (except for Y_0 and X_0). However, Decressin and Fatás (1995) did not.

Stylized fact 3-Regional unemployment rates are correlated across space. In section 2.2 we discussed a number of reasons why spatial lags appear in econometric equations. The value of Moran's I for spatial autocorrelation among regional unemployment rates amounts to 0.59 (T-value 9.86), among participation rates to 0.62 (T-value 10.41), and among employment growth rates to 0.16 (T-value 2.87). The problem of this test is that it does not

[9] Provided that all regions within the country are being investigated and weighted by the size of the labour force. If this is not the case, a_1 will be different from unity.

correct for serial dependence among the observations, while Elhorst (2001) has found that serial dynamic effects tend to be more important than spatial dynamic effects in regional labor market data. The proposed econometric model addresses this problem. The estimation results show that spatial dependence is an issue that, even when serial dynamic effects are accounted for, should not be ignored. Although smaller than the coefficient of the past value of the serially lagged dependent variable (urate[-1]), the coefficient of the current value of spatially lagged dependent variable (W*urate) is substantial, 0.392, and significant (T-value 12.74). By contrast, the coefficient of the past value of the spatially lagged dependent variable (W*urate[-1]) turns out to be negligible and insignificant. Some analysts have been troubled with the idea that the spatial autoregressive interaction between Y and WY is instantaneous (see Upton and Fingleton, 1985: 369). Instead, they suggest a model in which the autoregressive response is allotted a period in which to take effect, $Y_t=\eta WY_{t-1}$. The advantage of this specification is that the Jacobian term, $|I-\delta W|$, which is the result of transforming the estimation model from the error term into the dependent variable, disappears. This should considerably simplify estimation of the model by maximum likelihood. As we can see, the last statement is true for the conditional log-likelihood function but not for the unconditional log-likelihood function. If $\delta=0$, we indeed get rid of the second right-hand side term of eq.(16), but not of the the fourth right-hand term. Other analysts do not seem to have problems with the idea that Y_t in one spatial unit is regressed on Y_t in other spatial units depending on a spatial weight matrix W, $Y_t=\delta WY_t$. By starting with $Y_t=\delta WY_t+\eta WY_{t-1,}$ the data can help to find the most appropriate model. In this particular case, we find that WY_t is a better choice.

Another interesting result emerges when we test for the joint significance of the time dummies. A Wald test whether eliminating the time dummies is acceptable on the data must be rejected. If we nonetheless eliminate them from the regression equation, the coefficient of WY_t increases at the expense of the coefficient of Y_t. The explanation for this striking result is that the spatially lagged dependent variable partially replaces the time dummies. From a mathematical viewpoint, time dummies can be written as an alternative spatially lagged dependent variable WY_t, where all the elements of the spatial weight matrix equal 1/N. Note that this also concerns the diagonal elements, which in the traditional spatial context are assumed to be zero, since no region can be viewed as its own neighbor. To correct for cross-section dependence, Pesaran (2002) suggests to run a regression augmented with the cross-section averages of the regressand and the regressors. The estimator of this model appears to be the same as the standard time period fixed effects estimator. Pesaran's finding can be seen as another, more formal, proof that the inclusion of time period fixed effects and the use of spatially lagged dependent variable are two different approaches that both deal with the same problem of region-invariant unobserved differences among time periods. Our finding that both the time dummies and the coefficient of the spatially lagged dependent variable (WY_t) are significant points out that they do not overlap and that both have a different interpretation. These interpretations are as follows. Regional labor market variables tend to increase and

decrease together in different regions along the national evolution of these variables over the business cycle. These business cycle effects are captured by the time dummies. In the long term, after the effects of labor supply and demand shocks have settled, regional labor market variables return to their equilibrium values. In equilibrium, regional labor market variables might still be spatially correlated in that neighboring values are more similar than those further apart. This type of spatial correlation is the type of spatial dependence spatial econometricians are actually looking for.

When we look at the estimation results obtained for the participation rate and the emploment growth rate, we see a plausible model structure. The immediate or short-run effects of the current value of the participation in the own region (lfprate) and in neigboring regions (W*lfprate) are both positive, 0.185 and 0.058 respectively. The first of these two effects is also significant. By contrast, the effects of past values are both negative and significant, -0.157 and -0.254 respectively, as a result of which the long-run effects of the participation rate in the own and neighboring regions are negative. This might explain the controversy the literature has produced over the effect of the participation rate on the unemployment rate, both theoretically and empirically. According to Fleisher and Rhodes (1976), the effect is negative, since factors determining low participation rates in a particular region also reflect relatively low investments in human capital and low commitment to working life, resulting in a higher risk that people with these characteristics become unemployed. They are more likely to be laid off when employers reduce workforces and they are more likely to experience some unemployment when re-entering the labour force after temporary absence.

According to the accounting identity[10] and in contrast to Fleisher and Rhodes (1976), the effect of the participation rate on the unemployment rate should be positive; if the participation rate increases, the number of unemployed must rise, *ceteris paribus*. However, it is questionable whether this effect is also positive *mutatis mutandis*. Firstly, increased participation encourages the growth of more local jobs. Several studies predict that the growth of jobs almost fully compensates for the growth of the labour force, better known by the phrase '*people cause jobs*' (Layard, 1997), as a result of which the unemployment rate would hardly increase. Secondly, more jobs encourage more people to enter the labour market. In a detailed review of the empirical literature, Elhorst (2003b) records twelve studies that found a negative and significant effect of the participation rate (of males, females or the total population of working age) on the unemployment rate, one that found a negative but insignificant effect, four that found a positive but insignificant effect, and finally two that found a positive and also significant effect. In sum, our conclusion that the long-run effects are negative is in line with the majority of these studies.

[10] The accounting identity is one of the oldest models of regional unemployment determination. It holds true for every geographical labor market (local, urban, regional), and reads as $U_L = PW*L + NC - E$, with $\Delta PW = G + NM$, where U_L is the level of unemployment, PW is the working age population, L is the labour force participation rate, NC is net inward commuting, E is the level of employment, G is the balance between new entrants into, and departures from, the working age population, and NM is net inward migration.

Nevertheless, that the effect of participation on regional unemployment is an issue fraught with controversy is understandable, since the short-run effects appear to be positive.

The effect of employment growth appears to be negative. According to the accounting identity (footnote 10), the impact should be negative almost by definition. The first-largest and most significant effect is caused by current employment growth in the own region, -0.110 (T-value 10.13), while the second-largest and most significant effect is caused by current employment growth in neighboring regions, -0.051 (T-value 2.28). Naturally, one might think, but of the 16 studies reviewed in Elhorst (2003b) that have taken up employment growth as an explanatory variable of regional unemployment, only Burridge and Gordon (1981) and Molho (1995) also considered employment growth in neighboring regions. The third in row appears to be past employment growth in the own region, -0.029 (T-value 3.45), while past employment growth in neigboring regions turns out to be negligible.

The question that remains is whether the model can be simplified. We already saw that time dummies may not be eliminated. A bit difficult is a test for the joint significance of the spatial fixed effects, since the spatial fixed effects are not determined due to the transformation of the general model into first differences. Moreover, they cannot be estimated consistently, since their number increases as N increases. Nevertheless, if the spatial fixed effects of the general model are approached by $\mu_i = (\overline{Y}_{i.} - \overline{Y}_{..}) - \tilde{\beta}(\overline{X}_{i.} - \overline{X}_{..})$), with $\tilde{\beta} = (\tau, \delta, \eta, \beta_1, \beta_2, \beta_3, \beta_4)$ and X_t=(Y_{t-1},WY_t,WY_{t-1},X_t,X_{t-1},WX_t,WX_{t-1}), then we can test the joint significance of these fixed effects by performing an F-test on the residual sum of squares of this model and the model formulated in levels without spatial fixed effects according to eq. (18). The results of this latter model are recorded at the right side of table 2. The outcome of this F-test with (112,663) degrees of freedom amounts to 14.93, which indicates that the spatial fixed effects are jointly significant. In addition to this, we also tested whether or not the coefficient estimates of the urate, lfprate and egrowth variables in the general model (11 coefficients) reformulated in first differences and in the model formulated in levels without spatial fixed effects do differ systematically. This test is based on the idea that under the hypothesis that the spatial fixed effects have no explanatory power, the estimators of both models are consistent, while under the alternative, the estimator of the model reformulated in first differences is consistent, but the estimator of the model in levels is not. Since this Hausman type of test amounts to 31.48, the null hypothesis must be rejected. This again indicates that spatial fixed effects are jointly significant.

Finally, Wald tests are used to test for retrictions that would reduce the general model to a combined serial and spatial autoregressive distributed lag model (model 1 in figure 2), to a serial autoregressive distributed lag model with spatial autocorrelation (model 3) or to a spatial autoregressive distributed lag model with serial autocorrelation (model 5). Neither of these restrictions appear to be acceptable on the data.

4. CONCLUSIONS

The key conclusion for econometric modeling of space-time data that can be drawn from the analysis in this paper is the following. Econometric relationships estimated using space-time data better contain serially and spatially lagged dependent variables, serially and spatially lagged independent variables, and spatial and time period fixed effects from the outset. Tests carried out to investigate whether such a general model could be simplified to a model that is more frequently used in applied research or whether these fixed effects may be eliminated from the regression equation produced negative results without exception.

To overcome the inconsistencies associated with the traditional least squares dummy variables estimator of a model containing a serially lagged dependent variable as well as spatial fixed effects, the general model has been transformed into first differences to eliminate the spatial fixed effects and then the unconditional likelihood function has been derived taking into account the density function of the first-differenced observations on each spatial unit. This procedure yields a consistent estimator of the response parameters when the cross-sectional dimension N tends to infinity, regardless of the size of T, and provided that the row and column sums of the spatial weight matrix W do not diverge to infinity at a rate equal to or faster than the rate of the sample size N in the cross-section domain. Only the coefficients of the spatial fixed effects cannot be consistently estimated, since the number of these coefficients increases as N increases. To model the pre-sample values of the exogenous variables for the first-differenced observations on each spatial unit, we used the Nerlove and Balestra approximation.

Our case study on regional unemployment showed that the parameter estimates of the proposed model exhibit the following, often identified, characteristics of regional unemployment data: (1) Regional unemployment rates are strongly correlated over time; (2) Regional unemployment rates parallel their national counterpart; (3) Regional unemployment rates are correlated across space; (4) The long-run effect of the labor force participation rate on the unemployment rate is negative, not only in the own region but also in neigboring regions. By contrast, the short-run effect is positive. This finding explains the controversy in the literature over the effect of the participation rate on the unemployment rate, both theoreticaly and empirically; (5) The effect of employment growth is negative not only within the own region itself, but again also in neighboring regions.

REFERENCES

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

Baddeley, M., R. Martin and P. Tyler (1998) Transitory Shock or Structural Shift? The Impact of the Early 1980s Recession on British Regional Unemployment. *Applied Economics* 30, 19-30.

Baltagi, B.H. (2001) *Econometric Analysis of Panel Data*, 2d edition. Chichester: Wiley.

Beck, N. (2001) Time-Series-Cross-Section Data: What Have We Learned in the Past Few Years?. *Annual Review of Political Science* 4, 271-293.

Brechling, F. (1967) Trends and Cycles in British Regional Unemployment. *Oxford Economic Papers* 19, 1-21.

Bhargava, A. and J.D. Sargan (1983) Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods. *Econometrica* 51, 1635-1659.

Brueckner, J.K. (2003) Strategic Interaction among Governments: An Overview of Empirical Studies. *International Regional Science Review* 26, 175-188.

Burridge, P. and I. Gordon (1981) Unemployment in the British Metropolitan Labour Areas. *Oxford Economic Papers* 33, 274-297.

Byers, J.D. (1990) The Cyclical Sensitivity of Regional Unemployment: An Assessment. *Regional Studies* 24, 447-453.

Chapman, P.G. (1991) The Dynamics of Regional Unemployment in the UK, 1974-1989. *Applied Economics* 23, 1059-1064.

Cressie, N.A.C. (1991) *Statistics for Spatial Data*. New York, John Wiley.

Blanchard, O.J. and L.F. Katz (1992) Regional Evolutions. *Brookings Papers on Economic Activity*, no. 1, 1-75.

Das, D., H.H. Kelejian and I.R. Prucha (2003) Finite Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances. *Papers in Regional Science* 82, 1-26.

Decression, J. and A. Fatás (1995) Regional Labor Marker Dynamics in Europe. *European Economic Review* 39, 1627-1655.

Dunn, R. (1982) Parameter Instability in Models of Local Unemployment Responses. *Environment and Planning A* 14, 75-94.

Elhorst, J.P. (1996) Regional Labour Market Research on Participation Rates. *Tijdschrift voor Economische en Sociale Geografie* (Journal of Economic and Social Geography) 87, 209-221.

Elhorst, J.P. (2001) Dynamic Models in Space and Time. *Geographical Analysis* 33, 119-140.

Elhorst, J.P. (2003a) Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review* 26, 244-268.

Elhorst, J.P. (2003b) The Mystery of Regional Unemployment Differentials: Theoretical and Empirical Explanations. *Journal of Economic Surveys* 17, 709-748.

Elhorst, J.P. (2005) Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels. *Geographical Analysis* 37, 62-83.

Fleisher, B.M. and G. Rhodes (1976) Unemployment and the Labor Force Participation of Married Man and Women: A Simultaneous Model. *The Review of Economics and Statistics* 58, 398-406.

Griffith, D.A. (1988) *Advanced Spatial Statistics*. Dordrecht, Kluwer.

Griffith, D.A. and F. Lagona (1998) On the Quality of Likelihood-Based Estimators in Spatial Autoregressive Models when the Data Dependence Structure is Misspecified." *Journal of Statistical Planning and Inference* 69, 153-174.

Hamilton, J.D. (1994). *Time Series Analysis*. New Jersey, Princeton University Press.

Hendry, D.F. (1995) *Dynamic Econometrics*. Oxford, Oxford University Press.

Hsiao, C. (1986) *Analysis of Panel Data*. Cambridge, Cambridge University Press.

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu (2002) Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods. *Journal of Econometrics* 109, 107-150.

Johnston, J. and J. DiNardo (1997) *Econometric Methods*, 4th edition. New York, McGraw-Hill.

Kelejian H.H. and I.R. Prucha (1999) A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40, 509-533.

Lahiri, S.N. (2003) Central Limit Theorems for Weighted Sums of a Spatial Process under a Class of Stochastic and Fixed Designs. *Sankhya* 65, 356-388.

Layard, R. (1997) Preventing Long-term Unemployment: An Economic Analysis (with Discussion). In D.J. Snower and G. de la Dehesa (eds) *Unemployment Policy*, pp. 333-356. Cambridge University Press, Cambridge.

Lee, L.-F. (2002) Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models. *Econometric Theory* 18, 252-277.

Magnus, J.R. and H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester, John Wiley.

Martin, R. (1997) Regional Unemployment Disparities and their Dynamics. *Regional Studies* 31, 237-252.

Molho, I. (1995) Spatial Autocorrelation in British Unemployment. *Journal of Regional Science* 35, 641-658.

Nickell, S. (1981) Biases in Dynamic Models with Fixed Effects. *Econometrica* 49, 1417-1426.

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., K. Lahiri, L.-F. Lee and M.H. Pesaran (eds) *Analysis of Panels and Limited Dependent Variable Models*. Cambridge, Cambridge University Press.

Nerlove, M. (2000). Growth Rate Convergence, Fact of Artifact? An Essay on Panel Data Econometrics. In: Krishnakumar, J. and E. Ronchetti (eds) *Panel Data Econometrics: Future Directions*. Amsterdam, Elsevier.

Nerlove, M. and P. Balestra (1996). Formulation and Estimation of Econometric Models for Panel Data. In: Mátyás, L. and P. Sevestre (eds) *The Econometrics of Panel Data*, 2d revised edition. Dordrecht, Kluwer.

Ord, J.K. (1981) Towards a Theory of Spatial Statistics: A Comment. *Geographical Analysis* 13, 91-93.

Pehkonen, J. and H. Tervo (1998) Persistence and Turnover in Regional Unemployment Disparities. *Regional Studies* 32, 445-458.

Pesaran, M.G. (2002) Estimation and Inference in Large Heterogenous Panels with Cross Section Dependence. Cambridge, www.econ.cam.ac.uk/faculty/pesaran/public.htm.

Thirlwall, A.P. (1966) Regional Unemployment as a Cyclical Phenomenon. *Scottish Journal of Political Economy* 13, 205-219.

Upton G.J.G. and B. Fingleton (1985) Spatial Data Analysis by Example, volume 1. Wiley, Chichester.

Table 2 *Estimation results of the general model reformulated in first differences to eliminate spatial fixed effects and formulated in levels without spatial fixed effects. Dependent variable: urate*

Explanatory variables	First differences		Levels	
	Coefficient	T-value	Coefficient	T-value
urate[-1]	0.489	24.62	0.570	55.41
W*urate	0.392	12.74	0.309	18.13
W*urate[-1]	-0.012	-0.42	-0.090	-6.66
lfprate	0.185	6.31	0.144	3.17
W*lfprate	0.058	1.02	0.077	0.51
lfprate[-1]	-0.157	-5.53	-0.249	-3.27
W*lfprate[-1]	-0.254	-4.67	-0.098	-0.64
egrowth	-0.110	-10.13	-0.114	-4.71
W*egrowth	-0.051	-2.28	-0.038	-0.58
egrowth[-1]	-0.029	-3.45	-0.024	-0.96
W*egrowth[-1]	0.010	0.58	-0.005	-0.10
pienull	-0.501	-2.41	6.741	5.07
1990	-	-	10.071	1.43
1991	0.000	0.01	9.953	1.15
1992	0.191	1.92	9.979	1.27
1993	0.907	9.04	10.672	1.72
1994	0.047	0.54	10.725	1.81
1995	-0.252	-2.77	10.527	1.22
1996	0.171	1.93	10.716	2.05
σ^2	0.512	18.34	7.214	28.82
Wald - no time dummies	142.95	6 restrictions,rej.	14.87	6 restrictions for one intercept,rej.
Wald - model 2	24.30	2 restrictions,rej.	44.53	2 restrictions,rej.
Wald – model 3	54.06	5 restrictions,rej.	302.45	5 restrictions,rej.
Wald – model 5	131.91	5 restrictions,rej.	232.76	5 restrictions,rej.