Regional convergence clubs in Europe: A Markov Chain Monte Carlo approach

Jan Mutl* and Manfred Fischer†

June 29, 2010

Abstract

The objective of this paper is to present a spatial econometric framework and methodology for club convergence testing using a growth regression approach that accounts for both spatial autocorrelation as well as spatial heterogeneity. Unlike previous literature, we use in-sample information to sort the regions into clubs according to the best fitting maximum likelihood (ML) configuration using a Markov Chain Monte Carlo algorithm. The methodology is applied to a sample of 255 NUTS-2 regions across 25 European countries and provides evidence for the club-convergence hypothesis in Europe.

Keywords: Regional income growth, club convergence, Markov Chain Monte Carlo, spatial econometrics
JEL Classification: C23, O47, O52, R11

*Jan Mutl: Institute for Advanced Studies, Stumpergasse 56, 1060 Vienna, Austria. Email: mutl@ihs.ac.at

†Vienna University of Economics and Business

1 Introduction

Regional income convergence – or what the European Commission calls regional cohesion – is a primary policy objective in the European Union [EU], and seen as vital to the success of other policy objectives, such as the single market, monetary union, EU competitiveness and enlargement (see European Commission 2004). But increasing economic integration and enlargement themselves impose major shocks on the European regions, and potentially threaten the achievement of regional income convergence in Europe (see Martin 2001).

Stimulated by this policy interest, the convergence debate has become one of the foremost topics in regional economic research in recent years. Empirical research has proceeded in many directions, using different definitions and methodologies (see Magrini 2004 for a recent survey of literature). Most of the empirical literature has focused on the conditional convergence hypothesis. Convergence, however, may hold not for all regional economies, but within groups of economies sharing similar characteristics (Azariadis and Drazen 1990). This evidence is referred to as *club convergence hypothesis* which implies that economies in a set of regions may converge with each other, in the sense that they tend towards a common steady state level in the long run, but that there is no convergence across different sets of regions.

So far, relatively little explicit attention has been paid to the question of systematically identifying regional convergence clubs in Europe. This paper presents a spatial econometric framework for club convergence analysis and testing using a growth regression approach that – unlike previous studies (see Baumont et al. 2003; Fischer and Stirböck 2006 among others) – accounts for both spatial and temporal autocorrelation as well as spatial heterogeneity. We do not need *a priori* criteria to identify regional clubs, but use in-sample information to sort the regional economies into clubs according to the best maximum likelihood [ML] fitting configuration using a Markov Chain Monte Carlo (MCMC) algorithm.

The units of observation are NUTS-2 regions[1] within Europe which the European Commission has chosen as targets for the convergence process and defined as the geographical level at which the persistence or disappearance of income inequalities should be measured. These regions belong to EU-25 countries, excluding Cyprus and Malta. Also included are the NUTS-2 re-

We would like to thank ...for helpful comments and suggestions. Financial support from the Austrian Science Fund (FWF), grant no. P19025-EU is gratefully acknowledged.

[1] NUTS-2 regions are defined on the basis of normative rather than functional criteria. We are aware of the problems associated with the NUTS classification, but data on functionally defined economic regions are not publicly available.

gions in Switzerland and Norway, although both countries are only members of the European Economic Areas and not the EU.

We focus on an output based measure of growth and use gross value added, *gva*, as proxy for regional output. *gva* is the net result of output at basic prices less intermediate consumption valued at purchasers' prices, and measured in accordance with the European System of Accounts 1995. The data relate to the time period from 1995 to 2004 when economic recovery in Eastern Europe gathered pace. The time period is relatively short due to a lack of reliable figures for the regions in Eastern Europe[2] (Fischer and Stirböck 2006).

We find that *ad hoc* specifications of convergence clubs, such as clustering regions from eastern European countries, can lead to the rejection of the hypothesis of existence of no convergence clubs in Europe. Our MCMC results than convey that although inference on which regions belong the different convergence clubs is very imprecise, it is possible to obtain robust estimates of the speed of convergence as well as effects of other explanatory variables on the per capita income growth. In particular, we find evidence of relatively fast income convergence in Europe with half lifes of about 9 to 12 years.

The layout of the rest of the paper is as follows. First, Section 2 provides some more details on the econometric framework and testing methodology. Second, Section 3 describes the data set and outlines the results of the spatial econometric analysis. Finally, Section 4 concludes.

2 Econometric framework

2.1 Model Specification

The empirical methodology we advocate and implement for club convergence testing is based on a spatial version of the familiar test equation for conditional convergence that incorporates both substantive and nuisance sources

[2] The political changes since 1989 have resulted in the emergence of new or re-established states (the Baltic states, the Czech Republic, Slovakia and Slovenia) with only a very short history as sovereign national entities. In most of these states historical regional data series do not exist. Even for states such as Hungary and Poland that existed for much longer time periods in their present boundaries, the quality of data referring to the period of central planning imposes serious limitations on analysing regional growth over larger time periods. This is closely related to the change in accounting conventions, from the material product balance system to the European System of Accounts 1995.

of spatial autocorrelation.

$$\begin{aligned} g_i &= \rho \sum_{j=1}^{n} w_{ij} g_j + \mathbf{Z}_i \boldsymbol{\delta} + u_i, \\ u_i &= \lambda \sum_{i=1}^{n} m_{ij} u_j + \varepsilon_i \end{aligned} \tag{2.1}$$

for $i = 1, \ldots, n$ regions. g_i is the average per capita income growth rate of region, $\mathbf{Z}_i$ is a $1 \times (k+2)$ vector of exogenous variables including the constant, the initial income level and a set of k conditioning variables, and $\boldsymbol{\delta}_i$ is the associated $(k+2) \times 1$ vector of regression parameters.

Equation (2.1) accounts for spatial externalities in the process of growth. The degree of interregional interdependence is described by the spatial autocorrelation parameter ρ, with $1 \leq \rho < 1$. The parameter is assumed to be identical for each region but the net effect on the growth rate of region i depends on the relative connectivity between this region and its neighbors j. We represent the connectivity between a region i and all the regions belonging to its neighborhood by the exogenous friction terms w_{ij}, for $j = 1, .., n$ and $j \neq i$. These terms are assumed to be non-negative, non-stochastic and finite, with the properties $0 \leq w_{ij} \leq 1$ if $i \neq j$, $w_{ii} = 0$, and $\sum_{j=1}^{n} w_{ij} = 1$ for $i = 1, ..., n$.

The disturbances u_i in equation (2.1) are no longer assumed to be *iid.* Rather they are assigned a spatial autoregressive structure instead of being iid. [see the second equation in (2.1)]. λ is the parameter that captures the influence of space. This parameter is, however, a nuisance parameter in the sense that it is not assigned any meaningful economic interpretation. The terms m_{ij}, for $i, j = 1, .., n$, represent exogenously determined non-negative, non-stochastic and finite terms where $0 \leq m_{ij} \leq 1$ if $i \neq j$, $m_{ii} = 0$, and $\sum_{j=1}^{n} m_{ij} = 1$ for $i = 1, ..., n$.

Stacking the variables over the different cross-sectional units, we obtain the following matrix representation of our model:

$$\begin{aligned} \mathbf{g} &= \rho \mathbf{W} \mathbf{g} + \mathbf{Z} \boldsymbol{\delta} + \mathbf{u}, \\ \mathbf{u} &= \lambda \mathbf{M} \mathbf{u} + \boldsymbol{\varepsilon}, \end{aligned} \tag{2.2}$$

where $\mathbf{W}$ and $\mathbf{M}$ are now $n \times n$ matrices containing the spatial weights, $\mathbf{g}$ is an $n \times 1$ vector of the dependent variable, $\mathbf{Z}$ is $n \times (k+2)$ vector of explanatory variables and $\mathbf{u}$ and $\boldsymbol{\varepsilon}$ are $n \times 1$ vectors of spatially correlated disturbances and innovations respectively.

2.1.1 Spatial Durbin Model

We now show how an explicit treatment of the possibility of omitted variables can lead to the spatial Durbin model that we employ in our empirical analysis. We follow LeSage and Pace (2008) and demonstrate that under specific assumptions (elaborated below) the model in (2.1) can be shown to be equivalent to a spatial Durbin specification (see, e.g. also the derivations in section 2.1 in LeSage and Fischer, 2008). Suppose that not all the k variables in $\mathbf{Z}$ are included in an empirical estimation and, for concreteness only the first k_1 are considered. We denote $n \times k_1$ matrix of the included variable by $\mathbf{Z}_1$ and the $n \times k_2$ matrix of excluded variables as $\mathbf{Z}_2$. Of course we have $k_1 + k_2 = k$. We denote the corresponding coefficients $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$. We assume that the omitted variables are themselves spatially correlated and that there is some correlation among the included and omitted variables. In particular, we assume that $\mathbf{Z}_2$ is generated as

$$\mathbf{Z}_2 = \lambda_2 \mathbf{M}_2 \mathbf{Z}_2 + \mathbf{Z}_1 \boldsymbol{\Phi} + \boldsymbol{\nu}, \tag{2.3}$$

where $\mathbf{M}_2$ is an $n \times n$ matrix of spatial weights, $\boldsymbol{\Phi}$ is a $k_1 \times k_2$ matrix of (direct) correlation coefficients among $\mathbf{Z}_1$ and $\mathbf{Z}_2$, and $\boldsymbol{\nu}$ is an $n \times k_2$ matrix of innovations. Thus, the variables $\mathbf{Z}_2$ are generated by

$$\mathbf{Z}_2 = \left(\mathbf{I}_n - \lambda_2 \mathbf{M}_2\right)^{-1} \left(\mathbf{Z}_1 \boldsymbol{\Phi} + \boldsymbol{\nu}\right), \tag{2.4}$$

and after substituting into our model, we obtain

$$\begin{aligned} \left(\mathbf{I}_n - \rho \mathbf{W}\right) \mathbf{g} &= \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \qquad (2.5) \\ &= \mathbf{Z}_1 \boldsymbol{\delta}_1 + \mathbf{Z}_2 \delta_2 + \left(\mathbf{I}_n - \lambda \mathbf{M}\right)^{-1} \boldsymbol{\varepsilon} \\ &= \mathbf{Z}_1 \boldsymbol{\delta}_1 + \left(\mathbf{I}_n - \lambda_2 \mathbf{M}_2\right)^{-1} \left(\mathbf{Z}_1 \boldsymbol{\Phi} + \boldsymbol{\nu}\right) \delta_2 + \mathbf{u}, \end{aligned}$$

and hence

$$\left(\mathbf{I}_n - \lambda_2 \mathbf{M}_2\right)\left(\mathbf{I}_n - \rho \mathbf{W}\right) \mathbf{g} = \left(\mathbf{I}_n - \lambda_2 \mathbf{M}_2\right) \mathbf{Z}_1 \boldsymbol{\delta}_1 + \left(\mathbf{Z}_1 \boldsymbol{\Phi} + \boldsymbol{\nu}\right) \boldsymbol{\delta}_2 + \left(\mathbf{I}_n - \lambda_2 \mathbf{M}_2\right) \mathbf{u}, \tag{2.6}$$

or after collecting terms

$$\begin{aligned} \mathbf{g} &= \lambda_2 \mathbf{M}_2 \mathbf{g} + \left(-\lambda_2 \rho\right) \mathbf{M}_2 \mathbf{W} \mathbf{g} + \rho \mathbf{W} \mathbf{g} \qquad (2.7) \\ &\quad + \mathbf{Z}_1 \left(\boldsymbol{\delta}_1 + \boldsymbol{\Phi} \boldsymbol{\delta}_2\right) \\ &\quad + \left(-\lambda_2\right) \mathbf{M}_2 \mathbf{Z}_1 \boldsymbol{\delta}_1 \\ &\quad + \boldsymbol{\nu} \boldsymbol{\delta}_2 + \left(\mathbf{I}_n - \lambda_2 \mathbf{M}_2\right)\left(\mathbf{I}_n - \lambda \mathbf{M}\right)^{-1} \boldsymbol{\varepsilon}. \end{aligned}$$

Thus the model can be thought of as having potentially higher order spatial lags ($\mathbf{M}_2 \mathbf{W} \mathbf{g}$), as well as spatial lags of the (a combination of the) explanatory variables ($\mathbf{M}_2 \mathbf{Z}_1 \boldsymbol{\delta}_1$).

Note that there is a possibility of a convenient simplification of the model. If one assumes that spatial autocorrelation in the disturbances and the omitted variables is identical (i.e. $\lambda = \lambda_2$ and $\mathbf{M} = \mathbf{M}_2$), the overall disturbance term becomes uncorrelated and simplifies to $\boldsymbol{\nu}\boldsymbol{\delta}_2 + \boldsymbol{\varepsilon}$. Under an additional assumption that the spatial weights are same in all components, i.e. $\mathbf{W} = \mathbf{M} = \mathbf{M}_2$, we obtain a simplified model:

$$\mathbf{g} = \rho_1 \mathbf{W}\mathbf{g} + \rho_2 \mathbf{W}^2 \mathbf{g} + \mathbf{Z}_1 \boldsymbol{\beta}_1 + \mathbf{W}\mathbf{Z}_1 \boldsymbol{\beta}_2 + \boldsymbol{\eta}, \tag{2.8}$$

where $\boldsymbol{\eta} = \boldsymbol{\nu}\boldsymbol{\delta}_2 + \boldsymbol{\varepsilon}$ is assumed to be independent in the cross-sectional dimension. Furthermore, observe that the second order spatial lag is in our application with normalized spatial weights matrices numerically almost close to a zero matrix, and we hence consider a possibility that it does not enter in the estimations.[3]

2.1.2 Club Convergence

Our main working hypothesis is that the parameters of the empirical model will differ for regions with different (say institutional) characteristics. If we were able to observe these characteristics and there were no spatial dependencies in the data, we could sort the regions in our sample into different groups and estimate each group separately. This approach might not be feasible for two reasons. Firstly, it is not *a priory* clear how to distinguish regions with 'similar characteristics,' and, secondly, the fact that regions from different convergence clubs interact with each other, makes it impossible to estimate each group in a separate regression.

We thus estimate all of the regions from different convergence clubs in a single regression. We conjecture that regions from the different convergence clubs will differ in the income level to which they converge, as well as their speed of convergence. However, we maintain that the influence of other determinants is the same across the convergence clubs. In particular, we assume that the spatial dependence has the same strength for all regions in our sample.[4]

Defining a dummy variable, say d_i indicating a membership in a particular convergence club and assuming for simplicity that $\rho_2 \mathbf{W}^2 \approx \mathbf{0}$, our model thus

[3] Note that for row-normalized spatial weights, the parameters ρ and λ_2 must be less then one in absolute value. Hence the expected coefficient at the second-order spatial lag ($-\rho\lambda_2$) is likely to be small in absolute value.

[4] As a special case, our specification thus contains the specification arising due to approximation errors of a nonlinear dynamics where the poor regions are farther away on the nonlinear convergence trajectory and thus after linearization they appear to have a faster speed of convergence.

in the case of two convergence clubs becomes (the full model with a general number of convergence clubs is specified in the appendix):

$$g_i = \rho_1 \sum_{j=1}^{n} w_{ij} g_j + [(1-d_i)\,\alpha_1 + \alpha_2 d_i] + [(1-d_i)\,\beta_1 + \beta_2 d_i] \ln y_{0i} + \mathbf{X}_i \boldsymbol{\gamma} + \eta_i. \tag{2.9}$$

Note that in contrast to equation (2.1), we now explicitly introduce the constant term (α_1 or α_2), the log of the initial income level (lny_{0i} with associated parameter β_1 or β_2) as regressors, collecting the rest of the explanatory variables (that is $\mathbf{Z}_1$ and its spatial lags) in a $1 \times h$ vector $\mathbf{X}_i$ with an associated $h \times 1$ vector of parameters $\boldsymbol{\gamma}$. Note that the difference between a region belonging to the convergence club and a region outside of the convergence club is assumed to be only in the speed of convergence and the baseline steady-state growth rate. The influence of the other explanatory variables is assumed to be of the same magnitude for all regions. In terms of the parameters of the model, the values for a region that is not in the convergence club, the constant has a value of α_1, while for a region in the convergence club the value is α_2. Analogically, the effect of log of the initial income level is β_1 and β_2 respectively.

In a stacked notation, the model can be written as

$$\mathbf{g} = \rho_1 \mathbf{W}\mathbf{g} + (\boldsymbol{\iota}_n - \mathbf{d})\,\alpha_1 + \alpha_2 \mathbf{d} + \beta_1 [(\boldsymbol{\iota}_n - \mathbf{d}) \odot \ln \mathbf{y}_0] + \beta_2 (\mathbf{d} \odot \ln \mathbf{y}_0) + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{2.10}$$

where $\boldsymbol{\iota}_n$ is a $n \times 1$ vector of ones and $\odot$ denotes element-by-element multiplication (Hadammar product). The variables y_i and $\mathbf{X}_i$ are stacked in the $n \times 1$ vector $\mathbf{y}$ and the $n \times h$ matrix $\mathbf{X}$.

2.2 Model Interpretation

That fact the model includes spatial lag of the dependent variable complicates somewhat interpretation and comparison to the case of no spatial autocorrelation. See e.g. LeSage (2009) for details. In essence, spatial autocorrelation introduces nonlinearity and hence in order to compare to the linear alternatives, we have to consider partial derivatives of the solution of the model instead of simply comparing estimated coefficients. We will now elaborate on this issue as it pertains to the calculation of speed of convergence in particular.

Observe that in equation (2.10), in absence of spatial interactions, we have for a given convergence club that

$$\mathbf{g} = \frac{\ln \mathbf{y}_T - \ln \mathbf{y}_0}{T} = \alpha_s + \beta_s \ln \mathbf{y}_0 + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{2.11}$$

where $s = 1, 2$ indicates the convergence club membreship, $\mathbf{y}_t$ is a vector of per capita income level of the regions and the time indexes 0 and T indicate the initial and terminal times in our sample. After rearranging we obtain

$$\ln \mathbf{y}_T = T\alpha_s + (1 + T\beta_s) \ln \mathbf{y}_0 + T\mathbf{X}\gamma + T\boldsymbol{\eta}, \tag{2.12}$$

and hence the coefficient β_s has the usual interpretation. In particular negative values imply that the influence of $\ln \mathbf{y}_0$ disappears over time and that the model exhibits (conditional) convergence. The speed of convergence is usually measured by its half-life calculated as $-\ln(2) / \ln(1 + \beta_s)$. However, this interpretation and calculation of half-life of convergence must be modified when spatial interactions are present. With spatial lags, we have for a given convergence club:

$$(\mathbf{I}_n - \rho_1 \mathbf{W}) \mathbf{g} = (\mathbf{I}_n - \rho_1 \mathbf{W}) \frac{\ln \mathbf{y}_T - \ln \mathbf{y}_0}{T} = \alpha_s + \beta_s \ln \mathbf{y}_0 + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{2.13}$$

and hence

$$(\mathbf{I}_n - \rho_1 \mathbf{W}) \ln \mathbf{y}_T = T\alpha_s + [(\mathbf{I}_n - \rho_1 \mathbf{W}) + T\beta_s \mathbf{I}_n] \ln \mathbf{y}_0 + T\mathbf{X}\gamma + T\boldsymbol{\eta}, \tag{2.14}$$

or after solving for $\ln \mathbf{y}_T$:

$$\ln \mathbf{y}_T = \left[\mathbf{I}_n + T\beta_s (\mathbf{I}_n - \rho_1 \mathbf{W})^{-1}\right] \ln \mathbf{y}_0 + (\mathbf{I}_n - \rho_1 \mathbf{W})^{-1} (T\alpha_s + T\mathbf{X}\gamma + T\boldsymbol{\eta}). \tag{2.15}$$

Thus the model will exhibit convergence when the largest absolute eigenvalue of the entire matrix $\left[\mathbf{I}_n + \beta (\mathbf{I}_n - \rho_1 \mathbf{W})^{-1}\right]$ is less then one (we will denote this by $\lambda_{\max}$).[5] Note that only when $\rho_1 = 0$, this simplifies to the previous condition that β_s should be negative. When spatial interactions are present ($\rho_1 \neq 0$), we must determine whether $\lambda_{\max} < 1$ and if this condition holds , we can then calculate the half-life of convergence as $-\ln(2) / \ln(\lambda_{\max})$. Observe that in general, $\lambda_{\max}$ depends on β_s, as well as the strength of spatial interactions (ρ_1) and the spatial configuration ($\mathbf{W}$ and its eigenvalues). When $\mathbf{W}$ is normalized so that its largest eigenvalue is equal to one (and the largest negative eigenvalue is less than one) and has zeros on the main diagonal, we can obtain an analytical expression for $\lambda_{\max}$ as in LeSage and Fischer (2008), equation (13):

$$\lambda_{\max} = \frac{1 + \beta_s - \rho_1}{1 - \rho_1}. \tag{2.16}$$

[5] Observe that if this condition holds for $\left[\mathbf{I}_n + \beta (\mathbf{I}_n - \rho_2 \mathbf{W})^{-1}\right]$, it will be automatically satisfied for $\left[\mathbf{I}_n + T\beta (\mathbf{I}_n - \rho_2 \mathbf{W})^{-1}\right]$.

In this case the convergence is again present if $\beta_s < 0$ (assuming that the model is not exploding in space, i.e. that $\rho_1 < 1$). However, the speed of convergence is still influenced by the strength of spatial interactions.

There are still some caveats that apply when using $-\ln(2)/\ln(\lambda_{\max})$ as the half life of convergence. This quantity reflects the time needed for the space-time system to reach its steady state and hence represents the half life of convergence of a typical region. However, there might be a considerable degree of heterogeneity hidden in the average. Depending on the exact form of the connectivity represented in the $\mathbf{W}$ matrix, there could be regions that take longer/shorter to converge to the steady state.

To explore this issue, we can think of equation (2.15) as a stochastic difference equation for $\mathbf{y}_t$ where the value of t increases in T increments:

$$
\begin{aligned}
\ln \mathbf{y}_{qT} &= \left[\mathbf{I}_n + T\beta_s\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\right] \ln \mathbf{y}_{(q-1)T} + \left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\left(T\alpha_s + T\mathbf{X}\gamma + T\boldsymbol{\eta}_{qT}\right) \\
&= \left[\mathbf{I}_n + T\beta_s\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\right]^q \ln \mathbf{y}_{(q-1)T} \qquad (2.17) \\
&\quad + \sum_{r=0}^{q-1}\left[\mathbf{I}_n + T\beta_s\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\right]^r \left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\left(T\alpha_s + T\mathbf{X}\gamma + T\boldsymbol{\eta}_{(q-r)T}\right)
\end{aligned}
$$

where $q = 1, 2, ...$ Assuming that system is stable, i.e. that $\lambda_{\max} < 1$, the matrix $\left[\mathbf{I}_n + T\beta\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\right]^q$ converges to a matrix of zeros as $q \to \infty$. The steady state is solution hence

$$
\begin{aligned}
\ln \mathbf{y}_{\infty} &= \sum_{r=0}^{\infty}\left[\mathbf{I}_n + T\beta_s\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\right]^r \left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\left(T\alpha_s + T\mathbf{X}\gamma + T\boldsymbol{\eta}_{(q-r)T}\right) \\
&= \left[\mathbf{I}_n - \mathbf{I}_n - T\beta_s\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\right]^{-1}\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\left(T\alpha_s + T\mathbf{X}\gamma\right) + \boldsymbol{\eta}_{\infty} \\
&= \frac{1}{T\beta}\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\left(T\alpha_s + T\mathbf{X}\gamma\right) \\
&= \frac{\alpha_s + \mathbf{X}\gamma}{-\beta_s} + \boldsymbol{\eta}_{\infty}, \qquad (2.18)
\end{aligned}
$$

where

$$
\boldsymbol{\eta}_{\infty} = \sum_{s=0}^{\infty}\left[\mathbf{I}_n + T\beta_s\left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1}\right]^s \left(\mathbf{I}_n - \rho_1 \mathbf{W}\right)^{-1} T\boldsymbol{\eta}_{(q-r)T}. \qquad (2.19)
$$

Note that the steady state values of each region depend only on the characteristics of the region and not on the spatial constellation of the system. However, the transition paths of the different regions to the steady state depend, among other things, on characteristics of their neighbors and the spatial weights matrix. This derivation of the steady state income level also allows us to give interpretation to the constant term α_s as allowing for differences in unconditional long run income levels between the two convergence clubs.

To evaluate region specific half life of convergence, we calculate a series of vectors $\ln \mathbf{y}_{qT}$ under the assumption that $\boldsymbol{\eta}_t = \mathbf{0}$ for $t = T, 2T, 3T, ...$ We then search at which q reaches a particular element of this vector half of the value of the corresponding element of the vector $\frac{\alpha + \mathbf{X}\gamma}{-\beta}$. The result is a vector of region specific convergence half lifes.

2.3 Model Estimation

If we could observe club membership, i.e. if we knew the values of the variable d_i, we could proceed to estimate the model in first differences by a (quasi) maximum likelihood method (QML). See the appendix for the likelihood function. A set of initial estimates is easily obtained (as in Mutl and Pfaffermayr, 2009) using instrumental variables (IV) estimation based on a set moment conditions that combine the suggestions for the dynamic panel data models (cf. Arellano and Bond, 1991) together with moments used for spatial models (e.g. Kelejian and Prucha, 1998). Thus a set of valid moment conditions include those based on the observation that the first difference of the innovations is uncorrelated with the space-time lags of the dependent variable, where the time lag operates at least twice. Additional moment conditions follow from the assumption that the space-time lags of the explanatory variable are also uncorrelated with the innovations. See the appendix for the specification of the available moment conditions and exact definition of the initial estimators.

However, we also want to account for the fact that we are uncertain about the club membership of the different regions. One possibility is to treat the entries in $\mathbf{d}$ as additional (discrete) parameters and make inference on these using the available data. A maximum likelihood based estimation approach would require us to estimate the model for all possible assignments of the regions among the convergence clubs and then calculate a weighted average of the estimates. Observe that this would require estimating 2^n different models and hence is not practically feasible.

In this paper we thus propose to use an algorithm that strategically moves through the large model space and only estimates models that have a high probability mass.[6] Note that our problem is somewhat similar to the choice of which variables to include when there is a potential large set of relevant variables. For these class of problems a Markov Chain Monte Carlo approach

[6] We define a 'model' as given by a specific set of values of the dummy variable $\mathbf{d}$.

is often used (XX cite references). Our approach here is essentially an adaptation of the Metropolis-Hastings algorithm to our model. The similarities of the variable choice problem and convergence club membership choice are apparent. In both cases a particular model (a state of the Markov chain) is characterized by a vector of binary variables. Each model has also well defined (classical) maximum likelihood associated with it and there is an estimation method yielding a parameter estimates as well as an estimate of how likely this particular model is the true model[7].

Less apparent are the differences. One advantage in the case of regions is that we have more *a priori* knowledge about how these regions are related to each other than is typically the case for variables. Regions do not live in a vacuum but instead are geographically related and we can use this to our advantage. If we, for example, conjecture that regions geographically close to each other are more likely to be in the same convergence club, we might be able to move through the model space more efficiently.

There are also some drawbacks for the case of regions. Dropping or adding one region to a particular convergence club is not likely to produce any noticeable changes to the estimates, nor fit of the model. Furthermore, it is not *a priori* clear that using the full sample as a single club leads to a statistically significant worse fit.[8] As a result, the traditional Metropolis-Hastings algorithm might fail to work very well.

In particular, we need to modify the steps of the algorithm that propose an alternative candidate model and choose a new model based from the set of candidate models. The alternation needs to make sure that the acceptance rule for candidates models is stricter than in the usual case and at the same time the set of alternative candidate models in one step of the algorithm is much larger. These changes are necessary in order to achieve reasonable convergence properties of the chain.

We now describe our proposed algorithm in more detail. We will denote the state of the Markov Chain after s iterations by values in a vector $\mathbf{d}^s$; in contrast the true values are simply $\mathbf{d}$.

(i) Start with an initial model, given by values of the $n \times 1$ vector $\mathbf{d}^0$, and set the iteration counter to $s = 0$.

(ii) Estimate the parameters of the model $\mathbf{d}^s$ and calculate its likelihood.[9]

[7] Assuming that the 'true' model is among the alternatives considered.

[8] The traditional $\beta-$convergence literature maintains exactly that all regions are from a single convergence club and hence finding a test that rejects the hypothesis of a single club would cast doubts about the validity of their results.

[9] In our case the maximized value of the likelihood function but this can be replaced to a posterior likelihood if explicitly Bayesian approach is desired.

(iii) For each region calculate the probability of being in the convergence club 1 in a new candidate model as $p_i = \sum_{i=1}^{n} r_{ij} d_j$, where r_{ij} is a function of the distance of the regions i and j (this may or may not be the same as the spatial weights w_{ij} or m_{ij}). Scale the values p_i so that the expected number of ones is equal to a metaparameter q.

(iv) Construct an alternative candidate model $\mathbf{d}^{alt}$ by setting $\mathbf{d}^{alt} = \mathbf{d}^{s}$ and then choosing randomly (with equal probabilities) one of the three following possibilities to modify it:

  (a) (birth step) Select a single region for which $d_i^s = 0$ and set $d_i^{alt} = 1$. The probability of selecting a particular region is proportional to p_i.

  (b) (death step) Select a single region for which $d_i^s = 1$ and set $d_i^{alt} = 0$. The probability of selecting a particular region is proportional to p_i.

  (c) (move step) Select a single region for which $d_i^s = 0$ and set $d_i^{alt} = 1$. Select a single region for which $d_j^s = 1$ and set $d_j^{alt} = 0$. The probability of selecting a particular region is proportional to p_i.

(iv) Estimate the remaining parameters (ρ_1, α_1, α_2, β_1, β_2, $\boldsymbol{\gamma}$) of the candidate model $\mathbf{d}^{alt}$ and calculate its likelihood value.

(v) Accept the new model $\mathbf{d}^{alt}$ with probability $\ln L_{alt} - \ln L_s$, where $\ln L_{alt}$ and $\ln L_s$ are the maximized values of the log-likelihood functions for the $\mathbf{d}^{alt}$ and $\mathbf{d}^{s}$ models respectively. In case of acceptance of the $\mathbf{d}^{alt}$ model, increase iteration count s by one, set $\mathbf{d}^{s} = \mathbf{d}^{alt}$, and move to step (ii), or terminate the algorithm if desired number of replications has been achieved. If the model $\mathbf{d}^{alt}$ was not accepted, move to step (iv).

Our MCMC algorithm has the following control parameters:

- q – the expected number of regions in each convergence club (step 3)
- r_{ij} – proportional to the probability that regions i and j are in the same convergence club (step 3)
- the functional form that translates a likelihood of each model to the probability of it being selected as the state of the Markov chain in the next iteration (step 5)

We run the MCMC algorithm for a pre-specified number (10,000 in our case) of iterations as given in the stopping rule in step (v). Note that the MCMC algorithm that moves through the model space is in the end used to approximate the likelihood function that is concentrated with respect to the convergence club membership. However, we are also able to use the saved estimated parameters along with a likelihood of the associated models visited by the Markov chain to construct weighted averages of all parameters of the model, including the dummy variable $\mathbf{d}$. Hence we are able to provide a likely distribution of each element d_i, i.e. the inferred probability of a particular region being in a particular convergence club.

In our case we use the maximized value of the likelihood function as the weight and hence our weighted estimates have the interpretation of classical model averages. This is an alternative to perhaps more commonly used Bayesian model averages that could be in our case easily obtained by using the posterior likelihood of each model as weights. Observe that the classical model averages differ from the unconditional (classical) maximum likelihood estimates of the parameters in $\mathbf{d}$ that are given by the value of the vector $\mathbf{d}$ for a most likely model (in our case of those visited by the MCMC algorithm). These unconditional ML estimates are also reported among our results. However, providing confidence intervals for these would be problematic and hence we prefer to rely on the averaged estimates in our conclusions.

3 Data and Results

In principle, we could treat the club and variable selection in a common procedure. However, this might be an ambitious task and instead we turn to previous studies to guide us in the choice of the explanatory variables that we include in the model. In particular, based on the results of model averaging in LeSage and Fischer (2008), we include initial income, human capital, population density, area, sectoral employment, as well as spatial lags of these variables.[10] The sources of data for the variables are summarized in the following table:

[10] These variables had inclusion probabilities higher than 5% in Lesage and Fischer (2008).

variable		source	unit
regional income	GVA	EuroStat	Euro
employment	EMPLOY	Cambridge Econometrics	1,000
area	AREA	Eurostat and INSEE (for France)	km^2
population	POP	Eurostat	1,000
human capital	HEDU	ISCED 1997	%
employment in:			
agriculture	AGRITE	Cambridge Econometrics	%
manufacturing	MANUTE	Cambridge Econometrics	%
construction	CONSTR	Cambridge Econometrics	%
market services	MASETE	Cambridge Econometrics	%
non-market services	NMSETE	Cambridge Econometrics	%

variable	description
GVA	total value of gross value added
EMPLOY	
AREA	area (including inland waters) in 2003
POP	
HEDU	Share of population (from 15 years of age on) with higher education (classes 5-6)
AGRITE	Proportion of employment in agriculture to total employment
MANUTE	Proportion of employment in manufacturing to total employment
CONSTR	Proportion of employment in constructing to total employment
MASETE	Proportion of employment in market services to total employment
NMSETE	Proportion of employment in non market services to total employment

Our sample covers 255 NUTS 2 regions in Austria, Belgium, Switzerland, Czech Republic, Germany, Denmark, Estonia, Spain, Finland, France, Greece, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Netherlands, Norway, Poland, Portugal, Sweden, Slovenia, Slovakia and the UK. Our data has annual frequency and covers the years 1995 to 2003.

As an initial starting point, we work with the hypothesis that the regions in the eastern European countries (Czech Republic, Estonia, Hungary, Lithuania, Latvia, Poland, Slovenia and Slovakia) form one convergence club. We estimate the full data sample without any convergence club as well as the model where the regions in the convergence club are from the eastern European countries. We report our initial estimated coefficients in Table 1.

Table 1: Initial Estimates

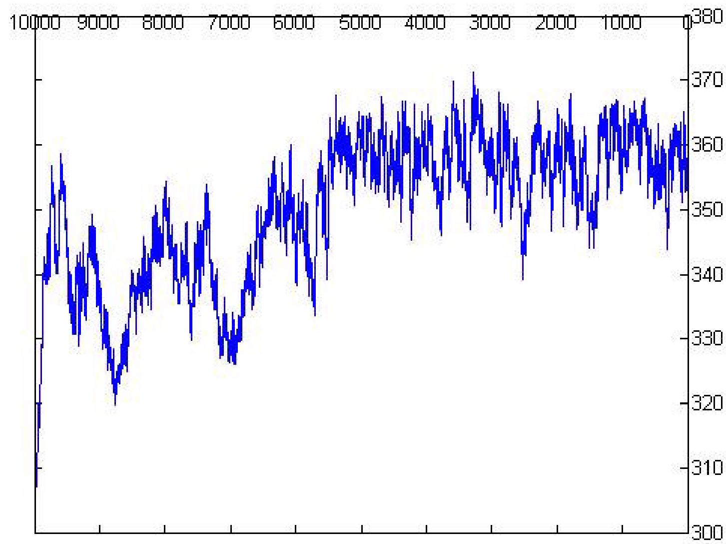
Variable	CEE ML Estimates	No Club ML Estimates
Wy	$0,90$	$0,95$
α_1	$3,08$	$1,96$
$\alpha_2 - \alpha_1$	$-1,14$	
β_1	$-0,23$	$-0,12$
$\beta_2 - \beta_1$	$0,11$	
EMPLOY	$0,34$	$0,17$
AREA	$-0,23$	$-0,90$
POP/AREA	$-0,24$	$-0,82$
HEDU	$0,07$	$0,07$
GFIXCF	$-0,11$	$-0,09$
SPECIND	$-0,22$	$-0,68$
DIVIND	$0,02$	$0,14$
W*(initial GVA)	$-0,01$	$-0,01$
W*(HEDU)	$0,08$	$0,16$

The estimates imply that the estimated coefficient of initial GVA level for Eastern Europe is by 0,1120 higher than for the rest of the regions. This implies lower speed of convergence for Eastern Europe. The implied half lifes are 4.9 years compared to 2.4 years in the rest of the sample while the half life of convergence for the model without any clubs is 2.2 years. Note that the extreme values for the half life of convergence are driven by high estimates of the spatial autocorrelation that are in both case close to 0.9.

Although the speed of convergence implied by the initial estimates seems to be higher than what is usually found in the literature, we note that the estimated effect of the rest of the explanatory variables has plausible signs for both models. However, they differ notably between the two models. We calculate the likelihood ratio statistics to distinguish among the two models and arrive at a p-value of 99% (using chi-square distribution with two degrees of freedoms). This can be interpreted as a strong evidence in favor of existence of a convergence club that could coincide with eastern European regions.

Turning to our MCMC results, we will now investigate to what extend are these findings robust to perturbations of the definition of the convergence clubs. Figure 1 shows the evolution of the measure of fit (logarithm of the maximized value of the likelihood function) over the accepted replications in the chain.

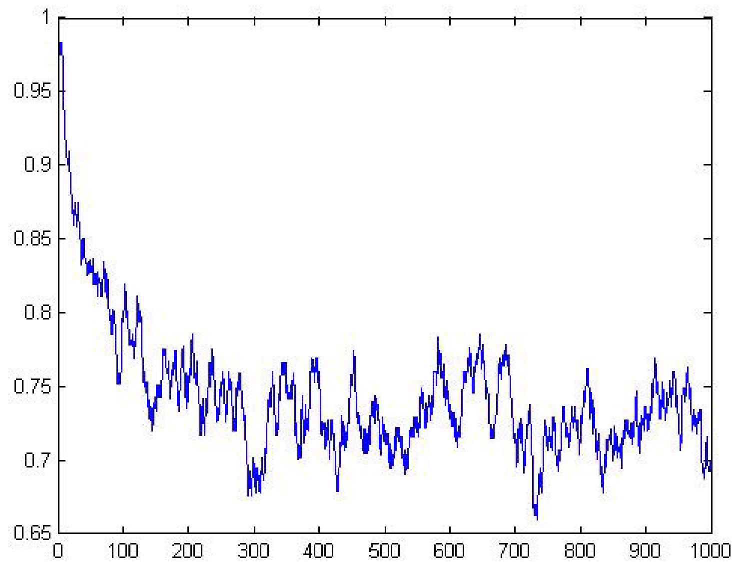
Figure 1: Evolution of the ML value



a

We observe a substantial improvement in the fit relative to the initial model. This indicates that eastern Europe is not unique convergence club that would be accepted relative to a baseline model without convergence clubs. Next, we note that our chain seems to converge after about 5,000 replications. To judge whether our chain has converged, we calculate the correlation of the last accepted model with the previous 1,000 accepted replications of the chain and plot these in Figure 2.

Figure 2: Correlation of the last 1,000 replications



We observe values that are almost all higher than 0.7 implying that in the last 1,000 replications, about two thirds of the regions are not changing their membership in the convergence clubs and the chain only oscillates around the boundary of the two clubs. We confirm this directly by counting regions that remain unchanged in 90% of the last 1,000 values of the vector **d** in the chain and find that this characterizes 189 of the 255 regions. Of these 189 regions, 119 regions never change their club membership in the last 1,000 replications.

We now turn to the question of how robust are the estimated effects of the explanatory variables and in particular the speed of convergence with respect to the specification of the convergence club. Table 2 reports the averaged estimates from 10,000 replications of the MCMC sampler. The estimates were weighted by a measure of fit (maximized value of the likelihood). The second column reports the estimated coefficients of the most likely model. The rest of the columns of Table 2 report quantiles of the estimates of the 22,487 models that were required to obtain 10,000 accepted steps in our chain. We label the spatial lag of the dependent variable by W*y and the element by element products of the vector **d** with the constant and the initial GVA as D*constant and D*(initial GVA).

Table 2: MCMC Estimates

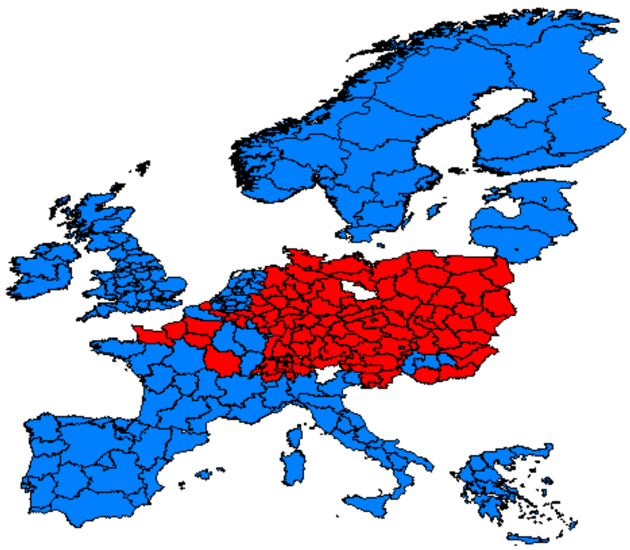
Variable	Average	max ML	25%	75%
$\mathbf{Wy}$	$0,64$	$0,58$	$0,67$	$0,88$
α_1	$3,15$	$3,16$	$2,63$	$3,30$
$\alpha_{2-}\alpha_1$	$-0,67$	$-0,65$	$-0,89$	$-0,44$
β_1	$-0,25$	$-0,25$	$-0,27$	$-0,20$
$\beta_2-\beta_1$	$0,05$	$0,05$	$0,03$	$0,08$
EMPLOY	$0,30$	$0,29$	$0,25$	$0,34$
AREA	$-0,27$	$-0,26$	$-0,30$	$-0,19$
POP/AREA	$-0,27$	$-0,25$	$-0,30$	$-0,19$
HEDU	$0,07$	$0,07$	$0,05$	$0,08$
GFIXCF	$-0,03$	$-0,03$	$-0,07$	$-0,02$
SPECIND	$-0,21$	$-0,14$	$-0,69$	$0,27$
DIVIND	$0,12$	$0,12$	$0,10$	$0,16$
$\mathbf{W}$*(initial GVA)	$-0,05$	$-0,06$	$-0,05$	$-0,01$
$\mathbf{W}$*(HEDU)	$-0,19$	$-0,23$	$-0,19$	$0,11$

The results show that the estimated coefficients are quite robust to the specification of the convergence club (the spatial lag of the human capital variable being the exception). The implied speed of convergence now also becomes closer to what other studies report. The most likely model implies half lifes of 9.9 and 12.4 years for the two clubs. We also calculate the half lifes of convergence for all the 10,000 accepted models and then average these out using the model fits as weights. This yields similar results and implies averaged half lifes 8,6 and 10,8 years.

We now turn to identification of the regions that belong to the two convergence clubs. In order to normalize our results, we label the regions that belong to the same club as the region of Brussels as the baseline (these receive the value of zero in the vector $\mathbf{d}$). We also transform the maximized values of the likelihood functions to weights in an analogy to calculation of the contribution of each model to the posterior in a Bayesian analysis. We find that of the 10,000 models, only 458 receive a weight higher than in a uniform weighting. Of these, 93 have a weight higher than 0.1 percent and 13 higher than 1 percent. However, the best fitting model receives about 14 percent of the total weight and the three best fitting models receive together 37 percent of the weight while the top 13 models account for 63 percent of the weights. Thus there is apparently some degree of uncertainty of which convergence club model fits the data best. However, when we calculate inclusion probabilities of each region as a weighted average of values of the variable $\mathbf{d}$ over all accepted replications (see Table 3 and Figure 3), we find that these do not differ substantially to the best fitting model. The correlation coefficient between the vectors $\mathbf{d}$ of these two models is 0.98.

Table 3: Club Membership

Figure 3: Club Membership



To find out to which extent are the different models identifying a convergence club, we compute the correlation coefficients of the top 50 models (accounting for 80 percent of the weights) and plot these in Figure 3. We find that all the values are above 0.8, indicating again a substantial agreement on club membership of most regions. As above, we confirm this by counting regions that do not change club membership in all these 50 models. This characterizes 188 of the 255 regions. The top 13 models agree on 226 regions.

We therefore conclude that our methodology has managed to identify two groups of core regions that should not be included in one sample when investigating growth convergence in Europe. Next, we want to investigate whether our results are statistically significant. Remember that when examining the initial model (where the club was set to be the eastern European regions), we were able to reject the null of no club convergence using likelihood ratio test. If we recalculated the same test statistic for any of the better fitting models we found when running our MCMC simulations, we would also reject the null of no club convergence. However, this testing procedure is incorrect for two reasons. First, it does not account for the fact that the alternative (unrestricted models) were obtained as to maximize fit. Second, likelihood ratio test can only be applied to testing problems where the parameter spaces are open sets and this assumption cannot be satisfied when we treat the zero/one entries in the vector **d** as parameters. We therefore resort to calculation information criteria. When we penalize for the additional parameters α_2 and β_2 and treat the n bits of information in **d** as one additional parameter, we obtain a value BIC in the best fitting model to be -2.57 compared to a value of -2.05 for the initial model and -2.02 for the model without convergence clubs, suggesting that the results of our MCMC estimation are the most prefferable. We obtain the same conclusions using AIC or HQC information criteria.

Table 4: Model Selection Criteria

	max ML	Initial CEE	No Club
AIC	-2,79	-2,25	-2,20
BIC	-2,59	-2,05	-2,01
HQ	-2,71	-2,17	-2,12
LR Test	137,95	18,65	
p-value	0,00	0,00	

Our result hence indicate that the club convergence model with the club identified by our MCMC methodology is the most parsimonius characterization of the data. Our motivation in doing the empirical study has been to find out the properties of the growth convergence process in Europe. The different models suggest different convergence half lifes as discussed in the text above and summarized in Table 5 below.

Table 5: Convergence Half Lifes

	Average	max ML	Initial CEE	No Club
Club 1	8,6	9,9	2,4	2,2
Club 2	10,8	12,4	4,9	2,2

As discussed in Section 2.2, these convergence half lifes can potentially hide a large degree of heterogeneity in the growth convergence process. We therefore calculate the region specific convergence half lifes using our methodology from Section 2.2 and report the results in Figures 4 and 5. We find that this is indeed the case and that traditional measure of convergence half life of 10 years implies region specific half lifes of 1 to 50 years.

Figure 4: Cummulative Plot of Region Specific Half Lifes

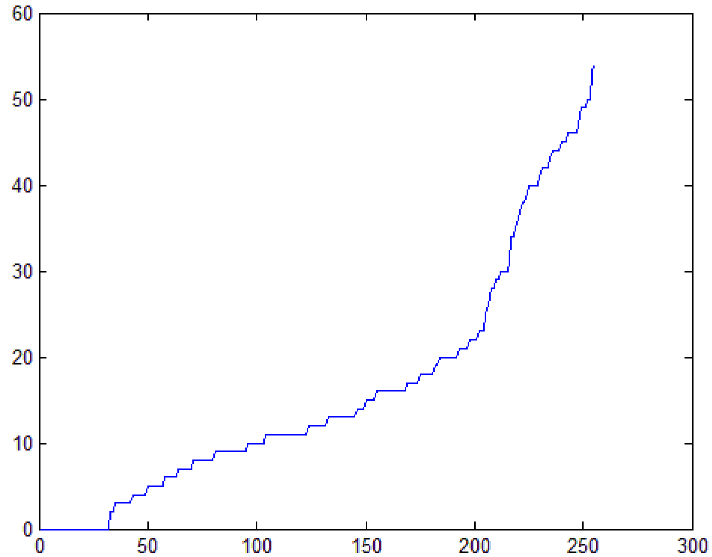
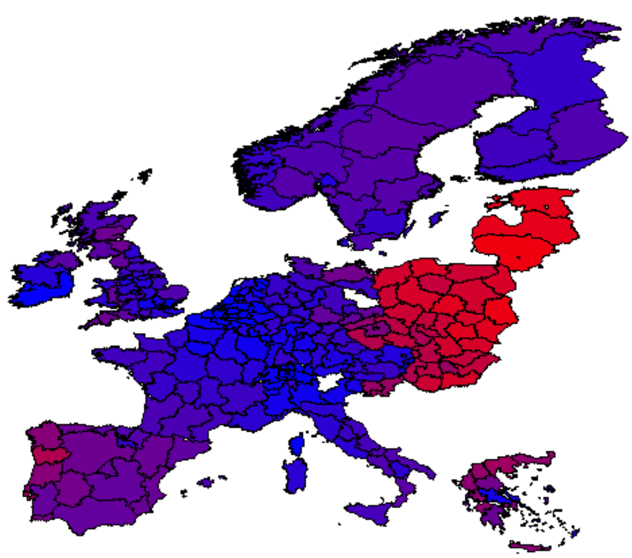


Figure 5: Region Specific Convergence Half Lifes



4 Conclusion

In this paper we introduce an empirical methodology for dealing with club convergence that uses in-sample information to sort the regions into the different convergence clubs. Our empirical model accounts for both spatial autocorrelation as well as spatial heterogeneity.

We apply our methodology to the sample of 255 NUTS-2 regions across 25 European countries. We initialize our procedure by assuming that the convergence club consists of regions from Central and Eastern Europe. We find that the likelihood ratio test statistics and information criteria favors this initial model over the null of no club convergence. However, the estimates of the model with the CEE as a club as well as a model without club convergence imply low convergence half lifes. We next run our MCMC simulations allowing for two convergence clubs and conlude that this produces the most parsimonious model according an array of information criteria. The estimates imply average convergence half life of about 10 years. We provide an alternative definition of region specific convergence half life and show that some regions will have convergence half lifes as long as 50 years. Our results indicate that allowing for the possibility of convergence clubs substantially changes the implied properties of the convergence process in Europe.

A Appendix

Convergence club model with general number of clubs

With R convergence clubs, we obtain the following model

$$g_i = \rho_1 \sum_{j=1}^{n} w_{ij} g_j + \sum_{r=1}^{R} \alpha_r d_i^r + \sum_{r=1}^{R} \beta_r d_i^r \ln y_{0i} + \mathbf{X}_i \boldsymbol{\gamma} + \eta_i, \tag{A.1}$$

where d_i^r is a dummy variable indicating whether the $i-th$ region belongs to the $r-th$ convergence club. The coefficients for the different convergent clubs are indexed by a corresponding subscript. The stacked model is then

$$\mathbf{g} = \rho_1 \mathbf{W}\mathbf{g} + \sum_{r=1}^{R} \alpha_r \mathbf{d}^r + \sum_{r=1}^{R} \beta_r \left(\mathbf{d}^r \odot \ln \mathbf{y}_0\right) + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{A.2}$$

or after collecting the dummy variables in an $n \times R$ matrix $\mathbf{D} = \left(\mathbf{d}^{1\prime}, .., \mathbf{d}^{R\prime}\right)'$ and the coefficients in $R \times 1$ vectors $\boldsymbol{\alpha} = (\alpha_1, .., \alpha_R)'$ and $\boldsymbol{\beta} = (\beta_1, .., \beta_R)'$:

$$\mathbf{g} = \rho_1 \mathbf{W}\mathbf{g} + \boldsymbol{\alpha}\mathbf{D} + \boldsymbol{\beta}\mathbf{D} \ln \mathbf{y}_0 + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\eta}. \tag{A.3}$$

Likelihood function of a club convergence model

The log-likelihood function of the model summarized in the above equation, assuming that $\boldsymbol{\eta} \sim N\left(\mathbf{0}_{n\times 1}, \boldsymbol{\Sigma}\right)$ is given by

$$\ln L_N(\boldsymbol{\theta}) = -\tfrac{n}{2} \ln\left(2\pi + \sigma^2\right) + \ln\left|\mathbf{I}_n - \rho \mathbf{W}\right| \tag{A.4}$$

$$+\tfrac{1}{2\sigma^2} \sum_{t=1}^{T} \left(\mathbf{g} - \rho_1 \mathbf{W}\mathbf{g} - \mathbf{Z}\delta\right)' \left(\mathbf{g} - \rho_1 \mathbf{W}\mathbf{g} - \mathbf{Z}\delta\right)$$

where the relevant parameters are collected in a vector $\boldsymbol{\theta} = (\sigma^2, \boldsymbol{\rho}_1, \boldsymbol{\delta}')'$. To speed up calculations in our implementation, we first concentrate all the parameters of the likelihood function except for ρ_1. The concentrated likelihood function is

$$\begin{aligned} \ln L_N(\rho) \quad = \quad & -\tfrac{n}{2} \ln(2\pi) + \ln\left|\mathbf{I} - \rho_1 \mathbf{W}\right| - \frac{1}{2n} + \frac{n}{2} \ln(n) \\ & -\tfrac{n}{2} \ln\left(\mathbf{g}'\mathbf{P}\mathbf{g} - 2\rho \mathbf{g}'\mathbf{P}\mathbf{W}\mathbf{g} + \rho^2 \mathbf{g}'\mathbf{W}'\mathbf{P}\mathbf{W}\mathbf{g}\right), \end{aligned} \tag{A.5}$$

where $\mathbf{P} = \left[\mathbf{I}_n - \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\right]$. Given a ML estimate of ρ_1, say $\widehat{\rho}$, from the concentrated likelihood function, we obtained the ML estimates of the remaining coefficients as

$$\begin{aligned} \widehat{\boldsymbol{\delta}}\left(\widehat{\rho}\right) &= \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\left(\mathbf{g} - \widehat{\rho}\mathbf{W}\mathbf{g}\right), \\ \widehat{\sigma}^2\left(\widehat{\rho}\right) &= \frac{1}{n}\left(\mathbf{g}'\mathbf{P}\mathbf{g} - 2\widehat{\rho}\mathbf{g}'\mathbf{P}\mathbf{W}\mathbf{g} + \left(\widehat{\rho}\right)^2\mathbf{g}'\mathbf{W}'\mathbf{P}\mathbf{W}\mathbf{g}\right). \end{aligned} \tag{A.6}$$

Moment conditions and IV estimation

Our model is essentially a dynamic panel data model. To see this we take the starting period to be $t-1$ and take $T = 1$. We re-write the model in terms of levels rather than growth rates (as in equation XX):

$$\ln\left(\mathbf{y}_t\right) = \rho_1\mathbf{W}\ln\mathbf{y}_t + \alpha + (1+\beta)\ln\mathbf{y}_{t-1} + (-\rho_1)\mathbf{W}\ln\mathbf{y}_{t-1} + \mathbf{X}_t\gamma + \boldsymbol{\eta},$$

where we now explicitly index the explanatory variables by a time subscript to exploit the fact that (some of) our explanatory variables are time-varying.

Thus the model can be thought of as explaining $\ln\mathbf{y}_t$ by its space ($\mathbf{W}\ln\mathbf{y}_t$), time ($\ln\mathbf{y}_{t-1}$) and space-time ($\mathbf{W}\ln\mathbf{y}_{t-1}$) lags. If one is willing to assume that the coefficient are constant over a particular window of time (9 years in our case), we can use the time series dimension of the panel to obtain initial set of IV estimates.

Observe that the contemporaneous spatial lag is an endogenous variable and should hence be instrumented. Substituting from the solution of the model:

$$\begin{aligned} \ln\mathbf{y}_t &= \left(\mathbf{I}_n - \rho_1\mathbf{W}\right)^{-1}\left[\alpha + (1+\beta)\ln\mathbf{y}_{t-1} + (-\rho_1)\mathbf{W}\ln\mathbf{y}_{t-1} + \mathbf{X}_t\gamma + \boldsymbol{\eta}\right] \\ &= \sum_{s=0}^{\infty}\left(\rho_1\mathbf{W}\right)^s\left[\alpha + (1+\beta)\ln\mathbf{y}_{t-1} + (-\rho_1)\mathbf{W}\ln\mathbf{y}_{t-1} + \mathbf{X}_t\gamma + \boldsymbol{\eta}\right], \end{aligned} \tag{A.7}$$

we obtain that valid set of instruments for $\mathbf{W}\ln\mathbf{y}_t$ includes (an independent subset of) $\mathbf{W}^{s+1}\ln\mathbf{y}_{t-1}$ and $\mathbf{W}^{s+1}\mathbf{X}_t$, for $s = 0, 1, ...$ Substitution for $\ln\mathbf{y}_{t-1}$ on the RHS in the time dimensions, shows that potential instruments also include $\mathbf{W}^{s+1}\ln\mathbf{y}_{t-1-r}$ and $\mathbf{W}^{s+1}\mathbf{X}_{t-r}\gamma$ for $s = 0, 1, ..$ and $r = 1, 2, ...$ However, these additional instruments do not provide more information that is already contained in $\mathbf{W}^{s+1}\ln\mathbf{y}_{t-1}$ and $\mathbf{W}^{s+1}\mathbf{X}_t\gamma$. Hence we do not consider these in our initial estimation.

The utilized instruments (for $s = 0, 1$) are associated with the moment conditions of the from

$$\begin{aligned} E\left[\eta_{it}\left(\ln\overline{y}_{j,t-1}\right)\right] &= 0, \\ E\left[\eta_{it}\left(\ln\overline{\overline{y}}_{j,t-1}\right)\right] &= 0, \end{aligned} \tag{A.8}$$

and

$$
\begin{aligned}
E\left[\eta_{it}\left(\ln \overline{\mathbf{X}}_{jt}\right)\right] &= \mathbf{0}, \\
E\left[\eta_{it}\left(\ln \overline{\overline{\mathbf{X}}}_{jt}\right)\right] &= \mathbf{0},
\end{aligned} \tag{A.9}
$$

where we denote by $\overline{y}_{j,t-1}$ the $j-th$ element of $\mathbf{W}\ln \mathbf{y}_{t-1}$, by $\overline{\overline{y}}_{j,t-1}$ the $j-th$ element of $\mathbf{W}^2 \ln \mathbf{y}_{t-1}$, and analogically for the spatial lags of $\mathbf{X}_t$. Since the model also includes $\ln \mathbf{y}_{t-1}$ and $\mathbf{X}_t$, as regressors, we additionally implicitly utilize moment conditions

$$
\begin{aligned}
E\left[\eta_{it}\left(\ln y_{j,t-1}\right)\right] &= \mathbf{0}, \\
E\left[\eta_{it}\left(\ln \mathbf{X}_{j,t-1}\right)\right] &= \mathbf{0}.
\end{aligned} \tag{A.10}
$$

References

[1] Anselin, L., 1988, *Spatial Econometrics: Methods and Models* (Kluwer Academic Publishers, Boston).

[2] Arellano, M. and S. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 1991, 277-297.

[3] Azariadis C, Drazen A (1990) Threshold externalities in economic development. Quarterly Journal of Economics 105(2): 501-526.

[4] Baltagi, B.H., 2008, *Econometric Analysis of Panel Data*, Wiley and Sons, Chichester, 4th edition.

[5] Baltagi , B.H. and Q. Li, 1992, A Note on the Estimation of Simultaneous Equations with Error Components, *Econometric Theory* 8, 113,119.

[6] Baumont C, Ertur C, LeGallo J (2003) Spatial convergence clubs and the European regional growth process, 1980-1995. In Fingleton B (ed) European regional growth. Springer, Berlin, Heidelberg, New York, pp. 131-158.

[7] Case, A. C., 1991, Spatial Patterns in Household Demand, *Econometrica* 59, 953-965.

[8] Chamberlain, G., (1982), Multivariate Regression Models for Panel, *Journal of Econometrics* 18, 5-46.

[9] Cornwell, Ch., Schmidt, P. and D. Wyhowski, 1992, Simultaneous Equations and Panel Data, *Journal of Econometrics* 51, 151-181.

[10] Egger, P., 2000, A Note on the Proper Econometric Specification of the Gravity Equation , *Economics Letters* 66, 25-31.

[11] European Commission (2004) A new partnership for cohesion. Convergence, competitiveness, cooperation. Third Report on Economic and Social Cohesion. Office for Official Publications of the European Communities, Luxembourg.

[12] Fischer MM, Stirböck C (2006) Pan-European regional income growth and club-convergence. Annals of Regional Science 40(4): 693-721.

[13] Greene, W., 2003, *Econometric Analysis*, 5th ed., Prentice Hall, New Jersey.

[14] Hausman, J.A., 1978, Specification Tests in Econometrics, *Econometrica* 46, 1251-1271.

[15] Horn, R.A. and Ch. R. Johnson, 1985, *Matrix Analysis*, Cambridge University Press, Cambridge UK.

[16] Kapoor, M., H.H. Kelejian and I.R. Prucha, 2007, Panel Data Models with Spatially Correlated Error Components, *Journal of Econometrics* 140, 97-130.

[17] Kelejian, H.H. and I.R. Prucha, 1998, A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, *Journal of Real Estate Finance and Economics* 17, 99-121.

[18] 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.

[19] Kelejian, H.H. and I.R. Prucha, 2007, Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances, forthcoming in the *Journal of Econometrics*.

[20] Kelejian H. H., I.R. Prucha and Y. Yuzefovich, 2004, Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results, in J. LeSage and R.K. Pace, *Advances in Econometrics*, New York: Elsevier, 163-198.

[21] Korniotis, G. M., 2008, Estimating Panel Models with Internal and External Habit Formation, forthcoming in the *Journal of Business and Economic Statistics*.

[22] Lee, L., 2003, Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances, *Econometric Reviews* 22, 307-335.

[23] Lee L. and Yu J.R., 2010, Estimation of spatial autoregressive panel data models with fixed effects, *Journal of Econometrics* 154, 165-185.

[24] LeSage J.P. and M.M. Fischer, 2008, Spatial Growth Regressions: Model Specification, Estimation and Interpretation, *Spatial Economic Analysis* 3:3, 275-304.

[25] LeSage J.P. and R:K. Pace (2009), *Introduction to Spatial Econometrics*, CRC Press/Taylor & Francis Group.

[26] Magrini S (2004) Regional (di)vergence. In Henderson J, Thisse J-F (eds) handbook of regional and urban economics. Elsevier, Amsterdam, pp. 2741-2796.

[27] Martin R (2001) EMU versus the regions? Regional convergence and divergence in Euroland. Journal of Economic Geography 1(1): 51-80.

[28] Mundlak, Y. (1978), On Pooling Time Series and Cross Section Data, *Econometrica* 46, 69-85.

[29] Mutl, J. (2006), *Dynamic Panel Data Models with Spatially Correlated Disturbances*, Dissertation, University of Maryland.

[30] Mutl, J. and M. Pfaffermayr (2010), Spatial Hausman Test, *Econometrics Journal*, forthcoming.

[31] Schmidt, P., 1976, *Econometrics*, Marcel Dekker, New York.

[32] Wooldridge, J., 2002, Econometric Analysis if Cross Section And Panel Data, MIT-Press, Cambridge MA.

[33] Yu, J. R. de Jong and Lee, L., 2007, Quasi-Maximum Likelihood Estimators For Spatial Dynamic Panel Data With Fixed Effects When Both n and T Are Large: A Nonstationary Case, Working paper, Ohio State University.

[34] Yu, J. R. de Jong and Lee, L., 2008, Quasi-Maximum Likelihood Estimators For Spatial Dynamic Panel Data With Fixed Effects When Both n and T Are Large, *Journal of Econometrics* 146, 118-134.