The Regional Economics Applications Laboratory (REAL) is a cooperative venture between the University of Illinois and the Federal Reserve Bank of Chicago focusing on the development and use of analytical models for urban and regional economic development. The purpose of the **Discussion Papers** is to circulate intermediate and final results of this research among readers within and outside REAL. The opinions and conclusions expressed in the papers are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Chicago, Federal Reserve Board of Governors or the University of Illinois. All requests and comments should be directed to Geoffrey J. D. Hewings, Director, Regional Economics Applications Laboratory, 607 South Matthews, Urbana, IL, 61801-3671, Phone (217) 333-4740, Fax (217) 244-9339.
Web page: www.uiuc.edu/unit/real

LOCAL VERSUS GLOBAL CONVERGENCE IN EUROPE: A BAYESIAN SPATIAL ECONOMETRIC APPROACH

by

Cem Ertur, Julie Le Gallo and James P. LeSage

REAL 03-T-28 July, 2003

Local versus Global Convergence in Europe: A Bayesian Spatial Econometric Approach

Cem Ertur
LATEC UMR CNRS 5118 Université de Bourgogne
Pôle d'Economie et Gestion, B.P. 26611,
21066 Dijon Cedex, France
cem.ertur@u-bourgogne.fr

Julie Le Gallo
Regional Economics Applications Laboratory
University of Illinois at Urbana-Champaign
220 Davenport Hall
607 South Mathews Avenue
Urbana, IL 61801-3671
jlegallo@uiuc.edu

James P. LeSage
University of Toledo
Department of Economics
Toledo, OH 43606
jlesage@spatial-econometrics.com

July 31, 2003

Abstract

Numerous studies have pointed to the econometric problems introduced by heterogeneity in cross-sectional data samples used to explore convergence suggested by neo-classical growth models. We introduce a local concept of convergence along with a Bayesian locally linear spatial estimation method to address these problems. The method allows global and local β-convergence to be viewed in a continuous fashion. Inference regarding global convergence can be treated as a mixture distribution arising from local β-convergence estimates from each region in the sample. Taking this approach eliminates the need to specify sub-samples and regimes as well as parameter variation schemes that have been used to model heterogeneity. We illustrate the method using a sample of 138 European regions.

KEYWORDS: locally linear estimation, robust, outliers, heteroscedastic

JEL: C11,R11,R12

1 Introduction

Since the pioneering contribution of Baumol (1986) and the more formal contributions of Barro and Sala-I-Martin (1991, 1992, 1995) and Mankiw et al. (1992), numerous studies have examined the β-convergence hypothesis based on the neoclassical growth model (Solow, 1956). Numerous empirical studies involving cross-sectional samples of countries and regions appear in the macroeconomic and regional science literature. The prediction of the neoclassical growth model (Solow, 1956) is that the growth rate of an economy will be positively related to the distance that separates it from its own steady state.

Making the simplistic assumption that economies are structurally similar, characterized by the same steady state, and differing only in their initial conditions, we should see *unconditional convergence* to the same steady state. In this case, we would see low income economies grow faster than those with high incomes and eventually catch up in the long run. Under the more realistic scenario, where economies have different steady states that are conditional on identifiable structural differences, it should be possible to draw econometric inferences regarding *conditional convergence*. This requires that we appropriately condition on structural differences that give rise to differences in steady states. In empirical practice, it is difficult to measure and model structural differences, and in theory heterogenous structures suggest heterogeneity in steady states as well as the structural factors on which we need to condition our econometric models (Durlauf, 2000, 2001; Brock and Durlauf, 2001).

Econometric models rely on a cross-section of countries or regions, using the average growth rate of per capita GDP (y) over a given time period as the dependent variable. These models rely on an explanatory variables matrix, $X = [\iota \;\; y_0]$ consisting of a constant (ι) as well as the initial level of log per capita GDP (y_0), and the associated parameter vector $\gamma = [\alpha \;\; \beta]'$ as shown in (1).

$$\begin{aligned} y &= X\gamma + \varepsilon \\ \varepsilon &\sim N(0, \sigma_\varepsilon^2 I_n) \end{aligned} \tag{1}$$

Most often, least-squares estimation is used to determine the sign and significance of the parameter β, for the case of unconditional β-convergence. For conditional β-convergence, a matrix of explanatory variables that purport to measure and control for structural differences is introduced in (1).

Typical variables suggested by Mankiw et al. (1992) in an augmented Solow growth model were: human and physical capital, saving rates and population growth rates. Additional variables to control for structural differences might include: the ratio of public consumption to GDP, the ratio of domestic investment to GDP, terms of trade, the fertility rate, the degree of political instability etc. (see Barro and Sala-I-Martin, 1995). In fact, more than 90 such variables have been included in cross-country regressions using international data sets in the empirical growth literature as surveyed by Durlauf and Quah (1999).

Heterogeneity in the structure of economies suggest that conditioning attempts that rely on smoothly varying variables to describe economic structure might fail to achieve the appropriate conditioning needed to produce valid inferences regarding conditional β-convergence. A theoretical motivation for heterogeneity can be found in endogenous growth theory (Azariadis and Drazen, 1990) as well as the neoclassical model with heterogenous structure (Galor, 1996).

Econometric methods that attempt to directly accommodate heterogeneity offer an alternative approach to the problem of estimation and inference. Partitioning the cross-sectional sample into regimes based on income levels or other structural characteristics is one approach to modeling heterogeneity (Desdoigts, 1999; Durlauf and Johnson, 1995). Allowing for explicit parameter variation over the sample represents another (Durlauf, Kourtellos and Minkin, 2001). In both cases, model specification issues beyond those involving which explanatory variables to include in the model arise. For the case of a multiple regime model, decisions must be made regarding how to partition the cross-sectional sample, and for varying parameter models, a specification for this variation must be set forth.

We propose:

1. use of a spatial autoregressive lag structure in place of a matrix of explanatory variables,

2. a locally linear Bayesian spatial model to accommodate heterogeneity.

With regard to 1), the spatial lag structure plays the role of a lagged dependent variable in time-series models, accounting for variation in the dependent variable arising from latent or unobservable variables. In the case of our spatial lag, these latent factors are correlated among cross-sectional observations located nearby in geographic space. A prior belief that latent spatial/geographical factors can explain a large part of observed structural variation in economies eliminates many decisions that arise when devising

econometric strategies for modeling heterogenous phenomena. As support for this prior, we note that empirical studies have found evidence of spatial autocorrelation in the residuals of traditional models (Conley and Ligon, 2002; Fingleton, 1999; Le Gallo, Ertur and Baumont, 2003; Moreno and Trehan, 1997). In addition, theoretical models from economic geography point to factors such as technology diffusion, factor mobility and trade which all have a strong geographic dimension which might interact with growth processes as in (Kubo, 1995 or Martin and Ottaviano, 1999, 2001). Another motivation for inclusion of the spatial lag of the dependent variable is that this variable highlights a spatial spillover effect, where the growth rate in each region is affected by those of neighboring regions after conditioning on initial per capita GDP levels.

Regarding 2), our locally linear spatial model partitions the cross-sectional sample observations by treating each location along with surrounding locations as a sub-sample. This reduces the need to make arbitrary decisions regarding how to partition the sample observations, but allows for variation in the parameter estimates across all observations. Our proposed method presumes that similarities in legal and social institutions as well as culture and language might act to create spatially local uniformity in economic structures, leading to similar spatial locality in rates of convergence. We think it useful to define the concept of *local convergence*, which we use to refer to a situation where rates of convergence in economic growth rates are similar for observations located at nearby points in space. In other words, there exists spatial clustering in the magnitudes of the β-convergence parameter estimates. It should be noted that our locally linear spatial estimation method does not impose apriori similar rates of convergence for spatially neighboring observations. Rather, we estimate β-convergence parameters for each region/observation in the sample and then examine these estimates in an effort to assess whether there is empirical support for our concept of *local convergence*. This represents an important difference between our approach and a spatially varying parameter estimation scheme that imposes spatial similarity on the estimates. For an example of the latter approach see LeSage (2003).

Another benefit of the locally linear model we propose is that expanding the sub-sample size around each locality results in a limiting model where the sub-sample size expands to include all observations in the cross-sectional sample. This produces locally linear econometric estimates that vary systematically as the sub-sample size increases towards the global estimates one would achieve using the entire sample. It allows a systematic assessment of the mapping between the locally linear estimates that accommodate hetero-

geneity and estimates based on the global sample reflecting homogeneity. This allows us to assess empirical evidence in support of local convergence in light of the more traditional global convergence approach.

Section 2 describes global versus local spatial autoregressive estimation, and compares our spatial autoregressive locally linear estimation approach to more traditional locally linear methods based on non-spatial least-squares. Bayesian Markov Chain Monte Carlo (MCMC) estimation of the model is taken up in section 3 and the model is applied to a sample of 138 European regions in section 4.

2 Estimation and inference regarding convergence

A spatial autoregressive model that can be used to produce regression estimates in the presence of spatial dependence in the cross-section of observations representing regions or countries is described in section 2.1. Section 2.2 extends this model to allow for a sequence of locally linear parameter estimates associated with each observation (country or region) in the data sample. The relationship between our proposed approach and other locally linear modeling methods is taken up in section 2.3.

2.1 Global spatial autoregressive estimates

To accommodate spatial dependence in the growth rates of regions or countries reflected in the cross-sectional dependent variable y, we might produce estimates using the spatial autoregressive model (SAR) shown in (2). This model includes what is known as a *spatial lag* of the dependent variable, (see Anselin, 1988).

$$y = \rho W y + X\gamma + \varepsilon \tag{2}$$

This model conventionally assumes that $\varepsilon \sim N(0, \sigma^2 I_n)$, but we will have more to say about this later. The vector y, matrix X and parameter vector γ are as described for the model in (1).

The nxn matrix W is a row-standardized spatial weight matrix. While a number of ways exist to specify W, a common specification sets $W_{ij} > 0$ for observations $j = 1 \ldots n$ sufficiently close (as measured by some metric) to observation i. For example, we might rely on observations that are spatially contiguous to observation i, those that have borders in common, or we might use the five nearest neighbors measured by distance from the centroids of each location. By construction, the main diagonal of W is set to zero to

preclude an observation from directly predicting itself. Row-standardization of the matrix W scales each element in the matrix so that the rows sum to unity, producing an explanatory variable Wy that reflects the average of growth rates from neighboring observations. The scalar parameter ρ measures the influence of the variable, Wy on y.

As in the case of time-series, use of this *spatial lag* should account for latent or unobservable influences that take on a spatial character. This approach ameliorates the need to specify structural economic explanatory variables, where data availability is constrained. One could of course include such variables in the matrix X in (2) along with an associated parameter vector if this information were available.

This model can be estimated using maximum likelihood methods (see Anselin, 1988) assuming that there is a homogeneous relationship between y and X across the spatial sample of observations. The estimated scalar parameter $\hat{\rho}$ could be used to test for the presence of significant spatial dependence in the sample of cross-sectional growth rates. If this parameter is not significantly different from zero, the model in (2) collapses to the simple least-squares model in (1). The scalar parameter estimate $\hat{\beta}$ contained in the parameter vector γ could be used to produce an inference regarding convergence that we label *global convergence*. Inferences based on this parameter represent a conclusion regarding convergence or non-convergence that averages over sample data evidence from the entire sample of countries or regions.

As an illustration, we provide estimates for the model in (1) based on least-squares alongside maximum likelihood estimates for the SAR model in (2) in Table 1. These estimates were based on a sample of growth rates in real GDP for 138 European regions over the period from 1980 to 1995 (see section 4.1 and the data appendix for a detailed description of the sample data). The SAR model used a spatial weight matrix W based on using the ten nearest neighbors to each region in the sample. Results based on a spatial weight matrices formed using eight to twelve nearest neighbors to each region were similar to those reported in the table.

From the homoscedastic model estimates reported in the table, we see strong evidence of spatial dependence as indicated by the estimate of $\hat{\rho} = 0.75$, that is significant at the 99 percent confidence level. The table also illustrates a difference between the magnitude of the least-squares β and that from the SAR model, pointing to differing rates of convergence. The least-squares estimate suggests more rapid convergence than that from the spatial model.

Another issue that plagues growth regressions is non-constant variance

across the sample of countries or regions. Table 1 also presents estimates for the least-squares and SAR model based on a Bayesian heteroscedastic linear model proposed by Geweke (1993) and a spatial autoregressive variant of this model suggested by LeSage (1997). These models allow the disturbances to take the form $\varepsilon \sim N(0, \sigma^2 V)$, where V is a diagonal matrix containing variance scalars $v_1, v_2, \ldots, v_n$, estimated using Markov Chain Monte Carlo (MCMC) methods. Prior information regarding the variance scalars v_i takes the form of a set of n independent, identically distributed, $\chi^2(r)/r$ distributions, where r represents the single parameter of the χ^2 distribution. This allows us to estimate the additional n non-zero variance scaling parameters v_i by adding only a single parameter r, to the model.

The specifics regarding the prior assigned to the v_i terms can be motivated by considering that the mean equals unity and the variance of the prior is $2/r$. This implies that as r becomes very large, the terms v_i will all approach unity, resulting in the non-zero variance scalars taking the form $V = I_N$, the traditional assumption of constant variance across space. On the other hand, small values of r lead to a skewed distribution permitting large values of v_i that deviate greatly from the prior mean of unity. The role of these large v_i values is to accommodate outliers or observations containing large variances by down-weighting these observations. In the context of spatial modeling, outliers or aberrant observations arise due to "enclave effects", where a particular region exhibits divergent behavior from nearby areas. Geweke (1993) shows that this approach to modeling the disturbances is equivalent to a model that assumes a Student$-t$ distribution for the errors. We note that this type of distribution has frequently been used to deal with sample data containing outliers, (e.g., Lange, Little and Taylor (1989)). In practice, one can either assign an informative prior for the parameter r based on the exponential distribution centered on a small value, or treat this as a hyperparameter in the model, set to a small value, say 4 to 7. The estimates in table 1 are based on $r = 4$.

These robust estimates suggest lower values for the convergence parameter β in both the least-squares and SAR models. It should be noted that the probabilities associated with these estimates represent Bayesian $p-$levels suggested by Gelman, Carlin, Stern and Rubin (1995) as an analogue to conventional marginal probabilities used for $t-$statistics for models estimated using MCMC methods. These point to a β estimate for the heteroscedastic SAR model that is not significantly different from zero at the 0.01 level of significance, pointing to a lack of convergence.

To summarize this discussion, inferences regarding convergence based on what we choose to label global estimates that presume homogeneity in

Table 1: Least-squares versus spatial model estimates

Homoscedastic model estimates				
Variable	OLS	OLS t-marginal probability	SAR	SAR t-marginal probability
Constant	0.1294	0.0000	0.0580	0.0009
1980 log-level	-0.0078	0.0000	-0.0048	0.0033
Spatial lag, Wy			0.7502	0.0000
Bayesian Heteroscedastic model estimates				
Variable	OLS	OLS p-level	SAR	SAR p-level
Constant	0.0994	0.0000	0.0416	0.0016
1980 log-level	-0.0045	0.0058	-0.0031	0.0164
Spatial lag, Wy			0.7681	0.0000

the relationship across the sample of regions or countries are likely to be sensitive to outliers and to influences such as spatial dependence that have the potential to bias least-squares estimates. For this reason, we propose a locally linear spatial autoregressive model described in the next section. This model is capable of producing inferences regarding our concept of *local convergence*.

2.2 Locally linear spatial autoregressive estimates

To accommodate both spatial dependence and heterogeneity, we produce estimates using $n-$ models, where n represents the number of cross-sectional sample observations, using the locally linear spatial autoregressive model in (3). This homoscedastic model was proposed by Pace and LeSage (2002) and labelled spatial autoregressive local estimation (SALE).

$$U(i)y = \rho_i U(i)Wy + U(i)X\gamma_i + U(i)\varepsilon \tag{3}$$

Where $U(i)$ represent an nxn diagonal matrix containing distance-based weights for observation i that assign weights of one to the m nearest neighbors to observation i and weights of zero to all other observations. This results in the product $U(i)y$ representing an mx1 sub-sample of observed GDP growth rates associated with the m observations nearest in location (using Euclidean distance) to observation i. Similarly, the product $U(i)X$ extracts a sub-sample of explanatory variable information based on m near-

est neighbors. The SALE model assumes $\varepsilon_i \sim N(0, \sigma_i^2 U(i) I_n)$, but we will have more to say about this later.

The scalar parameter ρ_i measures the influence of the variable, $U(i)Wy$ on $U(i)y$. We note that as $m \to n$, $U(i) \to I_n$ and these estimates approach the global estimates based on all n observations that would arise from the SAR model in (2).

We extend the SALE model to accommodate non-constant variances, by introducing $\varepsilon_i \sim N(0, \sigma_i^2 U(i)V), V = \text{diag}(v_1, v_2, \ldots, v_n)$. We label this model BSALE, Bayesian spatial autoregressive local estimation. The specifics of this extension are described in section 3.

2.3 Comparison with other spatially local estimation methods

McMillen (1996) and McMillen and McDonald (1997) introduced a form of spatial non-parametric locally linear weighted regression (LWR) which Brunsdon, Fotheringham and Charlton (1996) term geographically weighted regressions (GWR). This approach to modelling spatial dependence relies on separate models estimated using a sub-sample of the data based on observations nearby each observation. The motivation for this approach is that if spatial dependence arises due to inadequately modeled spatial heterogeneity, LWR can potentially eliminate this problem. These models often rely on the estimated parameters to detect systematic variation in the relationship being examined over space.

This type of model is shown in (4), where $M(i)$ represent an nxn diagonal matrix containing distance-based weights for observation i that reflect the distance between observation i and all other observations.

$$M(i)^{1/2} y = M(i)^{1/2} X\gamma_i + M(i)^{1/2}\varepsilon_i \tag{4}$$

The subscript i on γ_i indicates that this kx1 parameter vector is associated with region i. The LWR model produces n such vectors of parameter estimates, one for each region/observation. These estimates are calculated using:

$$\hat{\gamma}_i = (X'M(i)X)^{-1}(X'M(i)y) \tag{5}$$

A number of alternative approaches have been proposed to construct the distance-based weights for each observation i contained in the vector on the diagonal of $M(i)$. As an example, McMillen suggests a tri-cube weighting function:

$$\text{diag}(M(i)) = (1 - (d_i^j/d_i^m)^3)^3 \quad \text{I}(d_i^j < d_i^m) \tag{6}$$

Where d_i^j represents the distance between observation j and observation i, d_i^m represents the distance between the mth nearest neighbor and observation i, and I() is an indicator function that equals one when the condition is true and zero otherwise. In practice, the number of nearest neighbors used (often referred to as the 'bandwidth') is determined with a cross-validation procedure, typically a prediction criterion based on excluding a single observation.

Pace and LeSage (2002) point out that LWR methods exhibit a trade-off between increasing the sample size to produce less volatile estimates that contain increasing spatial dependence. Selecting a smaller sample size reduces the spatial dependence, but at the cost of increased parameter variability that impedes detection of systematic patterns of parameter variation over space. They argue that the SALE method eliminates this problem by extending the LWR approach to include a spatial lag of the dependent variable, which accommodates spatial autocorrelation likely to arise as the sub-sample size is increased. They argue that inclusion of the spatial autoregressive term in the model results in improves prediction and stability of the parameter estimates, decreasing the sensitivity of performance to the bandwidth that is typically observed.

There is a cost associated with introducing the spatial lag since the SALE model requires maximum likelihood methods, whereas the LWR model relies on least-squares. However, Pace and LeSage (2002) present an efficient recursive approach for maximum likelihood estimation of the n spatial autoregressive models for problems involving large numbers of observations and illustrate the method for a sample of 3,107 US counties. Most cross-sectional samples of countries or regions used in the empirical convergence literature involve considerably smaller samples.

These smaller samples give rise to another problem with local estimation methods, pointed out by LeSage (2003). Aberrant observations or outliers arising from spatial enclave effects or shifts in regime can exert a large impact on the locally linear estimates. Since these sub-sample estimates may be based on a small number of observations, and the sample data observations are re-used when estimates are produced for each point in space, a single outlier can contaminate estimates covering large areas or sub-regions of the spatial sample. This may create an artifact that resembles a regime shift or spatial clustering pattern in the estimates for β (or in β as well as the parameters on control variables in the conditional $\beta-$convergence model).

Intuitively, a single outlier will re-appear in sub-samples constructed using neighboring locations needed to produce estimates for each point in the spatial sample. This allows a single outlier to produce a contagion effect that can impact estimates for an entire region of the sample.

In the next section, we set forth the BSALE model that can accommodate outliers by down-weighting these observations.

3 Bayesian spatial autoregressive local estimation

For each spatial autoregressive model based on a sub-sample of size m, we specify our model as shown in (7), where the nxn diagonal matrix $U(i)$ assigns a weight of unity to the m nearest neighbors to observation i, and zero weight to all other observations.

$$\begin{aligned} U(i)y &= \rho U(i)Wy + U(i)X\gamma + U(i)\varepsilon \\ U(i)\varepsilon &\sim N(0, \sigma^2 U(i)V), \qquad V = \text{diag}(v_1, v_2, \ldots, v_n) \end{aligned} \tag{7}$$

The mxm matrix W represents a spatial weight matrix with row-sums normalized to unity. The weight matrix used in our empirical application was constructed using the six nearest neighbors to each region in the sample. Estimation results based on a first-order contiguity weighting matrix were also examined. The number of neighbors ranged from a low of just 3 first-order contiguous neighbors up to 10 contiguous neighbors with an average around 6 neighbors. Estimates from the model based on a first-order contiguity weighting matrix were nearly identical to those reported here based on the six nearest neighbors. We note that using 10 nearest neighbors in the formulation of W places a constraint on the smallest local sample size that can plausibly be used during estimation. It seems advisable to assign non-zero weights using the matrix $U(i)$ for at least 20 or 30 observations to provide an adequate amount of sample data on which to base estimates of ρ, β, V and σ. This in part motivated our choice of 6 nearest neighbors, and the restriction to 20 observations as the smallest sample size we consider.

This locally linear Bayesian variant of the basic spatial autoregressive model introduces a set of variance scalars $(v_1, v_2, \ldots, v_n)$, that represent unknown parameters that need to be estimated. This allows us to assume $\varepsilon \sim N(0, \sigma^2 U(i)V)$, where $V = \text{diag}(v_1, v_2, \ldots, v_n)$, but we note that only m of the variance scalars v_i take on non-zero values. As noted, this approach to robust modeling in the face of non-constant variance or outliers was in-

troduced by Geweke (1993) for a least-squares model, and LeSage (1997) for the spatial autoregressive model.

Equation (8) provides a formal statement of the prior distributions used in the model.

$$\begin{aligned} \pi(\gamma) &\sim N(c, T) \\ \pi(r/v_i) &\sim \text{IID}\chi^2(r) \\ \pi(1/\sigma^2) &\sim \Gamma(d, \nu) \\ \pi(\rho) &\sim U[-1, 1] \end{aligned} \tag{8}$$

Prior distributions assigned to the parameters α, β, σ and ρ in the model take the form of diffuse prior distributions. Given the small sub-sample sizes that we will be working with, informative prior information is likely to exert a substantial impact on the posterior distributions for these parameters. Given our interest in drawing inferences regarding β based on the sample data, diffuse prior assignments seem reasonable for these parameters in our application. This can be accomplished for the parameters γ by setting the vector of the prior means c to zero, and the prior variance-covariance $T = I_k \cdot (1e + 10)$. A diffuse prior on σ is associated with settings of: $d = \nu = 0$. We rely on a uniform prior for ρ ranging from -1 to 1.

Prior information regarding the variance scalars v_i takes the form of a set of n independent, identically distributed, $\chi^2(r)/r$ distributions, where r represents the single parameter of the χ^2 distribution. This allows us to estimate the additional m non-zero variance scaling parameters v_i by adding only a single parameter r, to the model. We use the same value for the hyperparameter r for all sub-samples during estimation.

The other aspect of our Bayesian SALE model is selection or setting of the sub-sample size m. As already noted, variation in this will create a host of parameter outcomes that are: highly volatile over the spatial sample for small values of m; and nearly constant taking on values near the global estimates as $m \to n$. This issue typically arises with locally linear non-parametric estimation methods, and cross-validation methods are often used to select an optimal sub-sample size. A plausible range for sub-sample size consideration might be $(1/4)n < m < (3/4)n$, so that sub-sample sizes are at least 1/4 the number of observations but less than 3/4 of the entire sample. Of course, these ranges could be changed depending on the size of the sample data. A related problem is that inference regarding the parameters is conditional on the sub-sample size selected.

One advantage of the SALE method is that a mapping of the parameter estimates is provided that allows an examination of the sensitivity of inferences with regard to choice of sub-sample size. We can examine the sequence of estimates for sub-sample sizes ranging from $m = (1/4)n$ to $m = (3/4)m$ in an effort to see whether inferences would differ as the sub-sample size varies. This is the approach we take here.

A cross-validation approach in this setting might involve use of the estimates for observation i based on a sub-sample size m to predict "fringe observations", those that border the sub-sample of m observations. This would represent a spatial analogue to one-step-ahead predictions in time-series. A Bayesian solution to the problem of sub-sample size selection would be to mix over estimates based on alternative sub-sample sizes to produce posterior estimates that reflect uncertainty with regard to the choice of sub-sample size. Unfortunately, this requires determination of weights that would be used in mixing over the estimates from alternative sub-sample sizes. These weights should be based on posterior probabilities associated with models arising from the various sub-sample sizes, but this would require integration over sub-sample sizes, which would be treated as a parameter in the model. This would lead to computationally expensive calculations.

We demonstrate that inference regarding convergence versus non-convergence is not sensitive to sub-sample sizes ranging from 40 to 100 observations, which roughly corresponds to $(1/4)n$ and $(3/4)n$.

3.1 Estimation of the model

The parameters γ, V, σ and the spatial lag parameter ρ in the heteroscedastic SAR model can be estimated by drawing sequentially from the conditional distributions of these parameters, a process known as "alternating conditional sampling", or Markov Chain Monte Carlo sampling.

To illustrate how this works, let $\theta = (\theta_1, \theta_2)$, represent a parameter vector and $p(\theta)$ denote the prior, with $L(\theta|y, X, W)$ denoting the likelihood. This results in a posterior distribution $p(\theta|D) = c \cdot p(\theta)L(\theta|y, X, W)$, with c a normalizing constant. Consider the case where $p(\theta|D)$ is difficult to work with, but a partition of the parameters into two sets θ_1, θ_2 is easier to handle. Given an initial estimate for θ_1, which we label $\hat{\theta}_1$, suppose we could easily estimate θ_2 conditional on θ_1 using $p(\theta_2|D, \hat{\theta}_1)$. Denote the estimate, $\hat{\theta}_2$ derived by using the posterior mean or mode of $p(\theta_2|D, \hat{\theta}_1)$. Assume further that we are now able to easily construct a new estimate of θ_1 based on the conditional distribution $p(\theta_1|D, \hat{\theta}_2)$. This new estimate for θ_1 can be used to construct another value for θ_2, and so on. On each pass through the sequence

of sampling from the two conditional distributions for θ_1, θ_2, we collect the parameter draws which are used to construct a joint posterior distribution for the parameters in our model. Gelfand and Smith (1990) demonstrate that sampling from the sequence of complete conditional distributions for all parameters in the model produces a set of estimates that converge in the limit to the true (joint) posterior distribution of the parameters. That is, despite the use of conditional distributions in our sampling scheme, a large sample of the draws can be used to produce valid posterior inferences regarding the joint posterior mean and moments of the parameters.

To implement this estimation method, we need to determine the conditional distributions for each parameter in our BSALE model. The conditional distribution for γ follows from the insight that given $\rho, \sigma, U(i), V$ and m, we can rely on results from LeSage (1997), and define $A = (I_n - \rho W)$ to arrive at:

$$\begin{aligned} p(\gamma|\rho,\sigma,V,m) &\sim N(\bar{g}, \sigma^2 G) \\ \bar{g} &= (X'U(i)V^{-1}X + \sigma^2 T^{-1})^{-1}(X'U(i)V^{-1}Ay + \sigma^2 T^{-1}c) \\ G &= \sigma^2(X'U(i)V^{-1}X + \sigma^2 T^{-1})^{-1} \end{aligned} \tag{9}$$

We see that the conditional for γ is a multivariate normal distribution from which it is easy to sample a vector γ.

The conditional distribution for σ given the other parameters, takes the form (see Gelman, Carlin, Stern and Rubin, 1995):

$$\begin{aligned} p(\sigma^2|\gamma,\rho,V,m) &\propto (\sigma^2)^{-(\frac{m}{2}+d+1)}\exp\left[-\, e'U(i)V^{-1}e + \frac{2\nu}{2\sigma^2}\right] \\ e &= (I_n - \rho W)y - X\gamma \end{aligned} \tag{10}$$

which is proportional to an inverse gamma distribution with parameters $(m/2) + d$ and $e'U(i)V^{-1}e + 2\nu$. Again, this would be an easy distribution from which to sample a scalar value for σ.

Geweke (1993) shows that the conditional distribution of V given the other parameters is proportional to a chi-square density with $r + 1$ degrees of freedom. Specifically, we can express the conditional posterior of each v_i as:

$$p(\frac{e_i^2 + r}{v_i}|\beta,\rho,\sigma^2,v_{-i},m) \sim \chi^2(r+1) \tag{11}$$

where $e_i = U(i)e$, with e as defined in (10), and $v_{-i} = (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$ for each i. Again, this represents a known distribution from which it is easy to construct a scalar draw.

Finally, the conditional posterior distribution of ρ takes the form shown in (12).

$$\begin{aligned} p(\rho|\beta, \sigma, V, m) &\propto |A|(s^2)^{-(m-k)/2} \\ s^2 &= (AU(i)y - Xb)'U(i)V^{-1}(AU(i)y - Xb)/(m-k) \end{aligned} \tag{12}$$

A problem arises here in that this distribution is not one for which established algorithms exist to produce random draws. We can however rely on univariate numerical integration of the conditional posterior of ρ. Applying a log transformation to the conditional posterior in (12), and the Barry and Pace (1999) Monte Carlo estimator for the log determinant in (12) over a grid of $j = 1, \ldots, q$ values for the parameter ρ ranging from -1 to 1, we can express $s(\rho)^2$ as a vector over the grid of ρ values. This takes the form:

$$\begin{pmatrix} s(\rho_1)^2 \\ s(\rho_2)^2 \\ \vdots \\ s(\rho_q)^2 \end{pmatrix} \propto \begin{pmatrix} \text{Ln}|I_n - \rho_1 W| \\ \text{Ln}|I_n - \rho_2 W| \\ \vdots \\ \text{Ln}|I_n - \rho_q W| \end{pmatrix} - (m-k)/2 \begin{pmatrix} \text{Ln}(\phi(\rho_1)) \\ \text{Ln}(\phi(\rho_2)) \\ \vdots \\ \text{Ln}(\phi(\rho_q)) \end{pmatrix} \tag{13}$$

where $\phi(\rho_j) = e_o'U(i)e_o - 2\rho_j e_d'U(i)e_o + \rho_j^2 e_d'U(i)e_d, j = 1, \ldots, q$ and we use $\tilde{y}, \tilde{X}$ to denote the products $\sqrt{V^{-1}}y, \sqrt{V^{-1}}X$, resulting in:

$$\begin{aligned} e_o &= \tilde{y} - \tilde{X}\gamma_o \\ e_d &= W\tilde{y} - \tilde{X}\gamma_d \\ \gamma_o &= (X'V^{-1}X)^{-1}X'V^{-1}y \\ \gamma_d &= (X'V^{-1}X)^{-1}X'V^{-1}Wy \end{aligned} \tag{14}$$

This produces a simple numerical integration problem that can be solved rapidly using Simpson's rule. We arrive at the entire conditional distribution using this numerical integration approach, and then produce a draw from this distribution using "inversion". Keep in mind that on the next pass through the MCMC sampler, we need to integrate the conditional posterior again. This is because the distribution is conditional on the changing values for the other parameters v_i, β, σ in the model, which obviously produce an altered expression for s^2 in the conditional distribution for ρ.

3.2 The MCMC sampler

By way of summary, an MCMC estimation scheme involves starting with arbitrary initial values for the parameters which we denote $\gamma^0, \sigma^0, V^0, \rho^0$, for a fixed value of the sub-sample size m and observation i. We then sample sequentially from the following set of conditional distributions for the parameters in our model for sub-sample size m and observation i.

1. $p(\gamma|\sigma^0, V^0, \rho^0, m)$, which is a multinormal distribution with mean and variance defined in (9). This updated value for the parameter vector γ we label γ^1.

2. $p(\sigma|\gamma^1, V^0, \rho^0, m)$, which is chi-squared distributed with $m+2d$ degrees of freedom as shown in (10). Note that we rely on the updated value of the parameter vector $\gamma = \gamma^1$ when evaluating this conditional density. We label the updated parameter $\sigma = \sigma^1$ and note that we will continue to employ the updated values of previously sampled parameters when evaluating the next conditional densities in the sequence.

3. $p(v_i|\gamma^1, \sigma^1, v_{-i}, \rho^0, m)$ which can be obtained from the chi-squared distribution shown in (11). Note that this draw can be accomplished as a vector, providing greater speed.

4. $p(\rho|\gamma^1, \sigma^1, V^1)$, which we sample using the numerical integration approach set forth above.

We now return to step 1) employing the updated parameter values in place of the initial values $\gamma^0, \sigma^0, V^0, \rho^0$. On each pass through the sequence we collect the parameter draws which are used to construct a joint posterior distribution for the parameters in the model associated with sub-sample size m and observation i. The average of these draws for the parameters γ, σ, ρ, V represent the mean of the posterior distributions for the model associated with observation i and sub-sample size m.

We repeat this process for models based on varying the sub-sample size m to produce estimates that vary as a function of the sub-sample size.

4 Convergence of European regions

We illustrate the BSALE method using a sample of 138 European regions and data covering the period 1980 to 1995. These local estimation results and inferences regarding convergence are compared to the global estimates and inferences presented in section 2.1.

4.1 The sample data

Data limitations remain a serious problem in the European regional context. Harmonized and reliable data allowing consistent regional comparisons are scarce, in particular for the beginning of the time period under study. There is clearly a lack of appropriate or easily accessible data that could be used to measure and control for structural differences considered by conditional $\beta-$convergence models. This represents a departure from the cross-country studies of Barro and Sala-I-Martin (1995) or Mankiw, Romer and Weil (1992) which rely on an extensive international data set.

We use the log of European regional per capita GDP over the period 1980-1995 expressed in ECUs, the former European Currency Unit, replaced by the Euro in 1999. The data are extracted from the EUROSTAT-REGIO database, which is widely used in empirical studies of European regions, see for example López-Bazo et al. (1999), Neven and Gouyette (1995), Quah (1996) among others. Our sample includes 138 regions in 11 European countries over the 1980-1995 period: Belgium (BE:11), Denmark (DK:1), France (FR:21), Germany (DE:30), Greece (GR:13), Luxembourg (LU:1), Italy (IT:20), the Netherlands (NL:9), Portugal (PT:5) and Spain (ES:16) in NUTS2 level and the United Kingdom (UK:11) in NUTS1 level (see the data appendix for more details). NUTS is the French acronym for Nomenclature of Territorial Units for Statitistics used by Eurostat. In this nomenclature NUTS1 refers to European Community Regions and NUTS2 to Basic Administrative Units.

It is worth mentioning that our sample is far more consistent and encompasses more regions than the one initially used by Barro and Sala-I-Martin (1991, 73 regions; 1995, 90 regions) and Sala-I-Martin (1996a, 73 regions; 1996b, 90 regions) where different sources and different regional breakdowns where mixed. For example, for the sample of 90 regions used by Barro and Sala-I-Martin (1995) mixed: i) GDP data collected by Molle (1980) for the pre-1970 period, ii) Eurostat data for the recent period and iii) personal income data from Banco de Bilbao for Spanish regions. Button and Pentecost (1995) also report these problems. Moreover the smaller 73 region data set is largely confined to prosperous European regions belonging to Western Germany, France, United-Kingdom, Belgium, Denmark, Netherlands and Italy, excluding Spanish, Portuguese and Greek regions, which are less prosperous. This may result in a selection bias problem raised by DeLong (1988). Armstrong (1995) attempted to overcome these problems by expanding the original Barro and Sala-I-Martin (1991) 73 region data set to less prosperous southern regions using a more consistent sample of 85 regions.

Despite improvements in our sample, we are aware of shortcomings in the database we use, especially concerning the adequacy of the regional breakdown adopted. The regional breakdown might give rise to a form of the ecological fallacy problem (King, 1997; Anselin and Cho, 2000) or "modifiable areal unit problem" well known to geographers (Openshaw and Taylor, 1979, Arbia, 1989). The choice of the NUTS2 level as our spatial scale of analysis may appear arbitrary and may have some impact on our inferences. Use of NUTS2 regions may reflect a spatial scale that is not fine enough to adequately capture unobserved heterogeneity, leading to the ecological fallacy problems noted above. On the other hand, the finer spatial scale of NUTS2 versus NUTS1 could lead to spurious spatial autocorrelation that could arise as an artifact of slicing homogenous zones. Despite these theoretical considerations regarding the appropriate spatial scale, we are constrained in our empirical illustration by data availability. In addition to the availability issue, we note that the NUTS2 level has been used in European regional development policy considerations. Since reform in 1989, NUTS2 is the level at which eligibility for Objective 1 Structural Funds is determined (see: The European regions: Sixth periodic report on the socio-economic situation in the regions of the European Union, European Commission, 1999). By way of conclusion, we simply note that our empirical results regarding convergence among regions should be interpreted with caution due to these issues arising from spatial scale considerations.

4.2 Estimation results

The first point we illustrate using our estimation results regards the statistical significance of the spatial dependence parameter ρ. Locally linear non-parametric models attempt to eliminate this dependence by relying on small sample sizes, where spatial dependence would be small or non-existent. We present kernel density estimates of the distribution of 138 estimates for ρ from the SALE model based on sub-sample sizes of $m = 20, 30, 40$ in Figure 1. Even in the case of the small sub-sample size of $m = 20$ shown in Figure 1 we see a multi-modal distribution of the 138 estimates for ρ, suggesting a great deal of variation in spatial dependence across the sample of European regions. The mean of these estimates is -0.07, near zero, lending support to the notion that locally-linear methods based on small sub-samples can overcome spatial dependence. However, there are a number of regions where the spatial dependence estimate appears to take on large (positive or negative) values, indicating the presence of spatial dependence between elements of the y vector. This would have an adverse impact

on the estimates of β for a number of regions in the 138 region sample. The impact of non-zero ρ values in the spatial autoregressive model is similar to that arising from simultaneity, resulting in biased and inconsistent estimates of β (see Anselin, 1988).

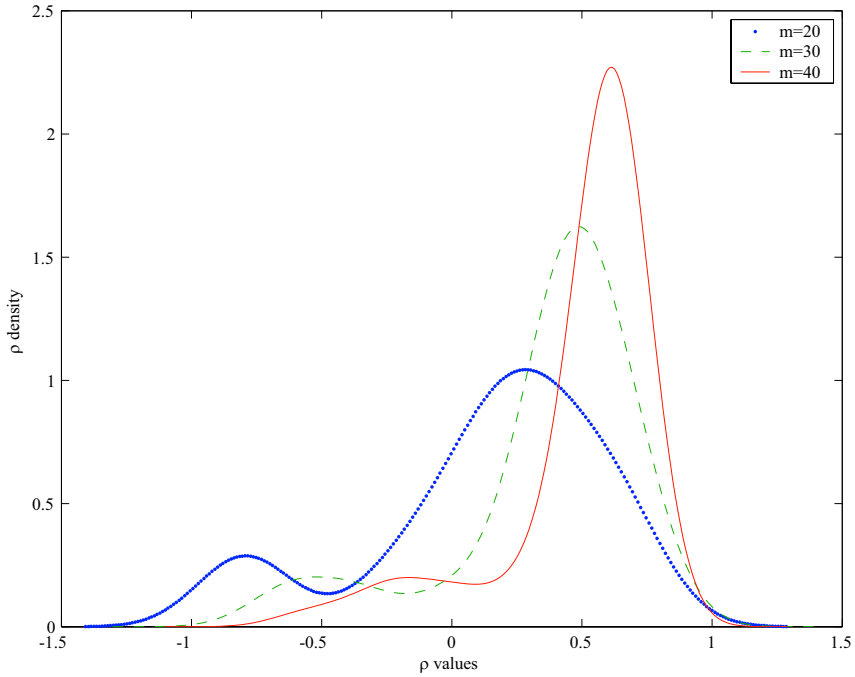

Figure 1: Distribution of ρ estimates for $m = 20, 30, 40$

The distribution of ρ estimates for sub-sample sizes of $m = 30$ and $m = 40$ shown in Figure 1 more clearly point to larger positive modal values, suggesting that spatial dependence increases as the sub-sample size increases, as we would expect. In these cases, the majority of estimates for β would be subject to the biasing impact of spatial dependence among the y values. The mean of 138 estimates for ρ based on sub-sample sizes 30 to 80 ranged from 0.35 for the small sub-sample size of 30, up to 0.71 for the large sub-sample size of 80. In this context, it should be noted that the "fringe observation" cross-validation criterion mentioned earlier pointed to an optimal sub-sample size around 80 observations.

There is also variation in the amount of spatial dependence as we move across countries, shown in Figure 2, where individual estimates for ρ are

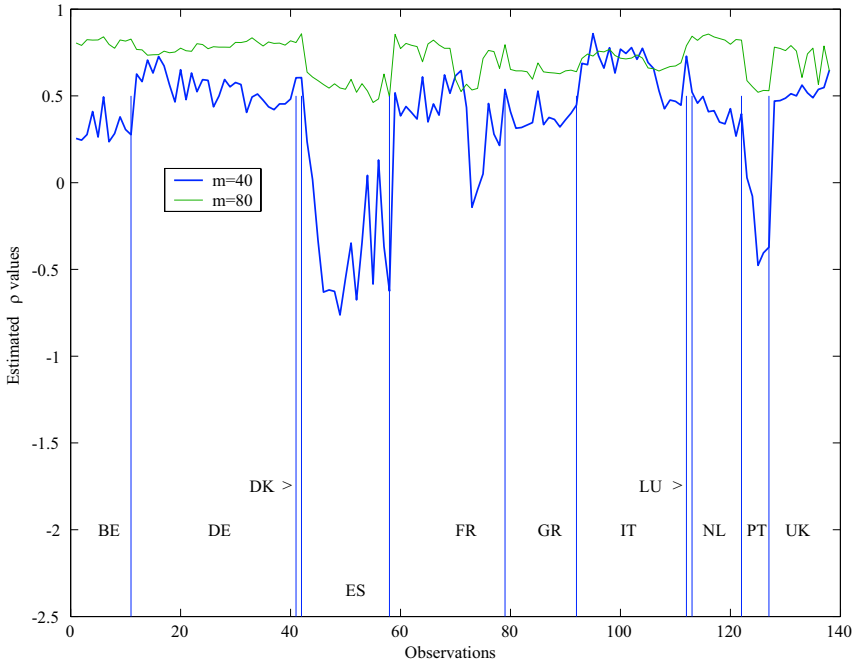

Figure 2: Estimates for ρ based on $m = 40, 80$

shown. Observations associated with countries are delimited by vertical lines in the figure, and estimates based on a sub-sample size of 40 and 80 are shown in the figure. It should be clear that spatial dependence of a sufficiently large magnitude to create bias in least-squares estimates arises even for the relatively small sub-sample size of 40.

These results suggest inclusion of the spatial lag of the dependent variable serves two useful purposes. First, it acts as a parsimonious proxy for unobserved latent spatial influences that are typically modelled by adding numerous explanatory variables to the model. Second, it allows increasing the sub-sample size used to produce locally linear estimates, which can stabilize the estimates and allow identification of spatial patterns or regimes. This can be done without introducing bias in the estimates for β that typically arises when larger sub-samples are used in local spatial estimation methods such as GWR.

Estimates for the convergence parameter β are shown in Figure 3, where again observations associated with countries are delimited by vertical lines in the figure. A set of three estimates based on sub-sample sizes of 60, 70 and

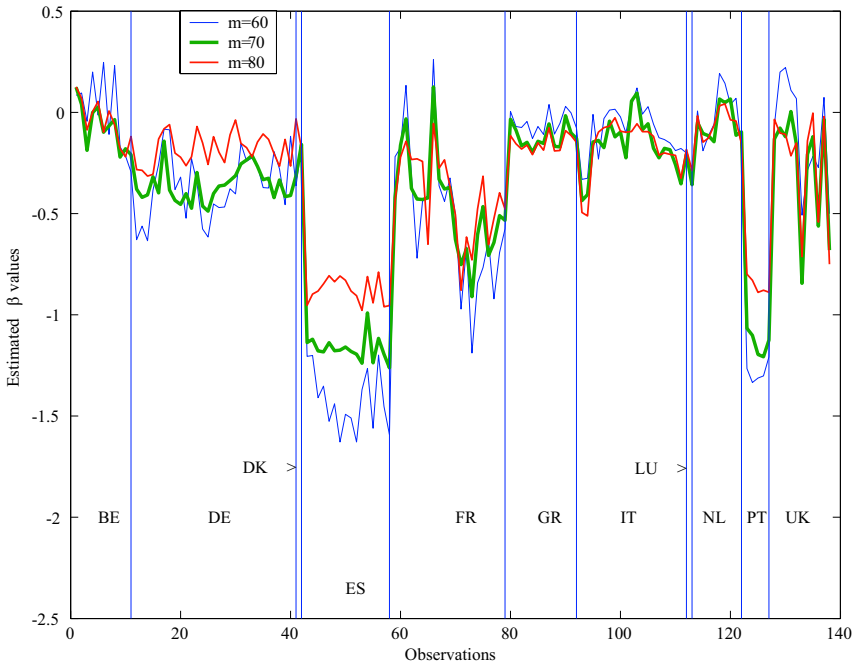

Figure 3: Estimates for β based on $m = 60, 70, 80$

80 are presented in the figure. Country-level differences are apparent in the figure, where we see estimates change abruptly as we move from one country to another. In addition to distinct variation in the convergence parameter between countries, there is also substantial variation between regions within a country in some cases.

Samples of draws generated during MCMC sampling can be used to produce estimates for the standard deviations of the parameter β, and associated confidence intervals. It should be noted that the estimates suffer from sample re-use as in the case of other locally linear non-parametric estimation methods. Sample observations from neighbors are re-used to produce estimates for each location, and in the case of neighboring observations the amount of sample overlap would be substantial. This inhibits our ability to interpret these measures of dispersion in estimate outcomes in a strict statistical sense. Nonetheless, we provide a graphical depiction of the β estimates based on a sub-sample size of 80 observations along with two standard deviation intervals in Figure 4. We simply note that convergence indicated by negative and significant values of β is likely for the EU regions in Spain and

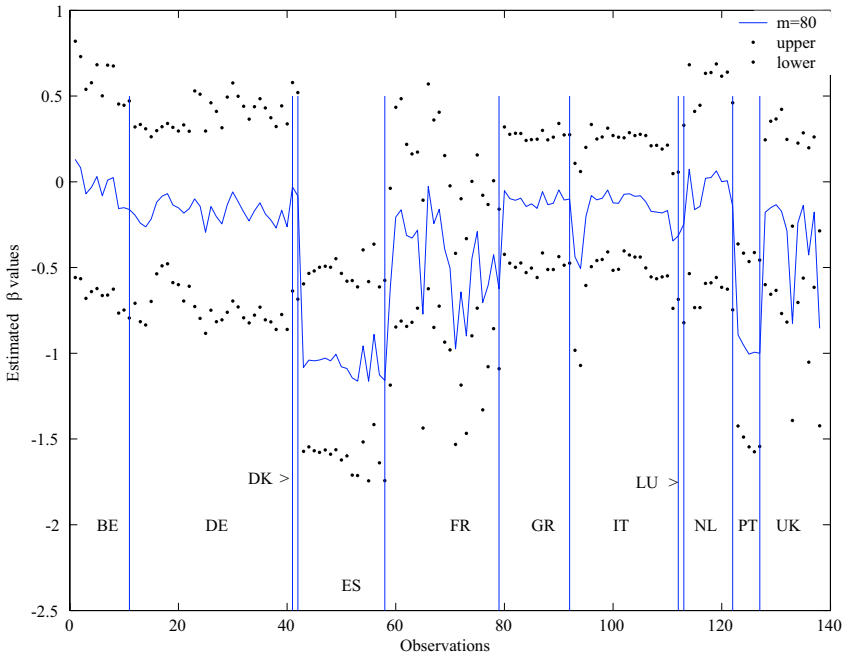

Figure 4: Upper and lower confidence intervals for β based on $m = 80$

Portugal as well as some regions in France. For observations associated with these regions, the estimates for the convergence parameter β are negative, and the upper confidence interval lies below zero, suggestive of significant negative values for this parameter.

Individual β and ρ estimates for the 138 regions based on a sample size of 80 are presented in Table 2 along with standard deviations constructed using the MCMC draws. Regions where the estimate for β is negative and more than two standard deviations away from zero are flagged in the table with the symbol $\star$. These parameter estimates would be consistent with convergence. For the case of the parameter ρ, all values were more than two standard deviations away from zero, so no symbols were added to the table. There are no cases where the positive coefficient values for β are more than two standard deviations away from zero, indicating divergence of the region from surrounding regions.

It is interesting to note that in Table 2, only 31 of the 138 locally linear spatial autoregressive estimates for β are negative and significant (more than

two standard deviations from zero), consistent with an inference of convergence. These regions tend to be spatially clustered in Spain, Portugal and southern France as shown in Figure 5. Use of global least-squares and SAR model estimates such as those presented in Table 1 of section 2.1, do not allow for this type of distinction. The β parameter estimates based on the four global models would lead to an inference of global convergence in three of the four cases presented in Table 1, the exception being the heteroscedastic SAR model. The concept of local convergence in conjunction with the BSALE model proposed here provide a great deal of additional information regarding the nature of convergence in growth rates across a spatial sample of observations. The BSALE estimates suggest that convergence is taking place for some regions in our sample, but not others.

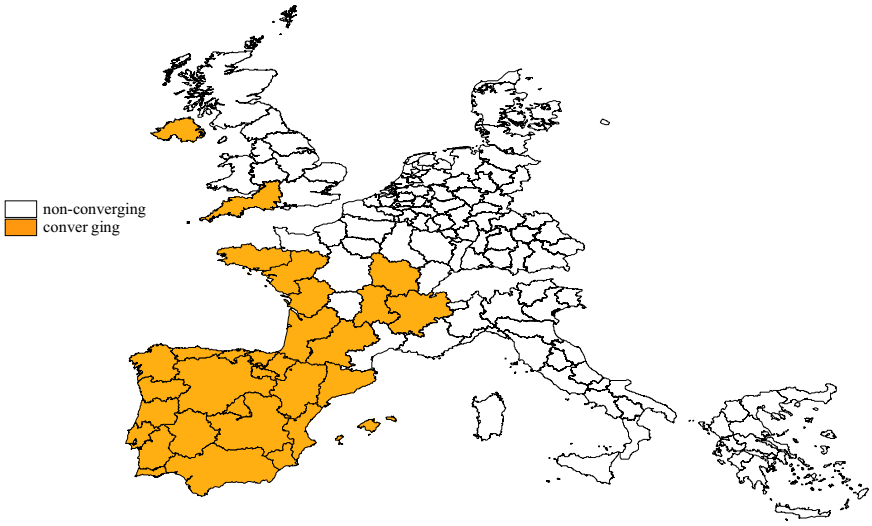

Figure 5: Converging and non-converging regions

5 Conclusions

We argue that problems created for conventional convergence regressions by shifts in regime as one moves across the spatial regions can be accommodated by a Bayesian spatial autoregressive locally linear estimation approach. Additional problems that arise due to non-constant variance and outliers can also be ameliorated using this approach. We define a local convergence concept and provide an estimation method that we label BSALE to draw

inferences regarding this notion of convergence. We demonstrate that inferences regarding convergence differ when using the BSALE methodology and more traditional SAR models based on the entire sample.

One aspect of this methodology is reliance on a spatial autoregressive model to account for latent unobservable factors that influence economic growth, but are not typically accounted for in $\beta-$convergence models. We argue that as in the case of lagged dependent variables in time-series modelling, spatial lags can filter adverse impacts arising from excluded variables. Another key facet of our BSALE approach is the use of a robust Bayesian variant of the spatial autoregressive local estimation (SALE) model set forth in Pace and LeSage (2002). This type of locally linear sub-sample estimation produces estimates that converge to robust Bayesian spatial autoregressive estimates based on the entire sample as the size of the sub-sample increases towards use of all observations. This allows practitioners to avoid use of a single bandwidth or sub-sample size on which they will ultimately proceed to draw inferences. The continuous nature of the mapping between locally linear and global estimates allows one to consider the role of sub-sample size on the resulting conclusions regarding convergence. For our sample of 138 European regions we find that conclusions regarding convergence are similar for sub-sample sizes varying from roughly one-fourth to three-fourths of the sample size.

There are several areas where the approach set forth here could be extended or enhanced. These methods could be extended to the case of a spatial Durbin model, where spatial lags of the initial levels are included as an explanatory variable in the model. Spatial error models where the disturbances are modelled as following a spatial autoregressive process would be another extension of the approach. A place for enhancement would be a formal method for identifying the optimal sub-sample size to use in the Bayesian SALE estimation method.

6 Data Appendix

The data are extracted from the EUROSTAT-REGIO database. This database is widely used in empirical papers dealing with European regions (e.g. Neven and Gouyette 1995; Quah 1996b; López-Bazo et al. 1999, Beine and Jean-Pierre, 2000). Eurostat is the Statistical Office of the European Communities. Its task is to provide the European Union with statistics at a regional level that enable comparisons between countries and regions. These statistics are used by the European Commission and other European Institutions to design, implement and analyze Community policies. The REGIO database is the official source of harmonized annual data at the regional level throughout the 1980-1995 period for the European Union.

In this paper, we use Eurostat 1995 nomenclature of statistical territorial units, which is referred to as NUTS (a French acronym used by Eurostat for Nomenclature of Territorial Units for Statistics). The aim is to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. In this nomenclature, NUTS1 refers to European Community Regions while NUTS2 defines Basic Administrative Units. For practical reasons having to do with data availability and implementation of regional policies, this nomenclature is based primarily on the institutional divisions currently in force in the Member States following "normative criteria". Eurostat defines these criteria as following: "normative regions are the expression of political will; their limits are fixed according to the tasks allocated to the territorial communities. according to the size of population necessary to carry out these tasks efficiently and economically, and according to historical and cultural factors" (Eurostat 1999, p. 7). It excludes territorial units specific to certain fields of activity or functional units (Cheshire and Carbonaro 1995) in favor of regional units of a general nature. The regional breakdown adopted by Eurostat appears therefore as one of the major shortcomings of the Regio database.

We use the series E2GDP based on ESA79 and expressed in ECUs per inhabitant over the 1980-1995 period for 138 regions in 11 European countries as follows: United Kingdom (UK:11) at NUTS1 level and Belgium (BE:11), Denmark (DK:1), France (FR:21), Germany (DE:30), Greece (GR:13), Luxembourg (LU:1), Italy (IT:20), Netherlands (NL:9), Portugal (PT:5) and Spain (ES:16) at NUTS2 level. For the United Kingdom NUTS1 is used because there is no official counterpart to NUTS2 units which are drawn up only for the European Commission use as groups of counties. This explains data non-availability at NUTS2 level throughout the period for this country. Luxembourg and Denmark may be considered as NUTS2 regions according

to Eurostat. Our preference for NUTS2 level rather than NUTS1 or NUTS3 levels, when data is available, is based on European regional development policy considerations: this is the level at which eligibility under Objective 1 of Structural Funds is determined (European Commission 1999). Our empirical results are certainly conditioned by this choice and could be affected by missing regions and different levels of aggregation. They must therefore be interpreted with caution.

We exclude Groningen in the Netherlands from the sample because of anomalies related to North Sea Oil revenues, which substantially increase its per capita GDP (as in Neven and Gouyette 1995). We also exclude the Canary Islands and Ceuta y Mellila, which are geographically isolated. Corse, Austria, Finland, Ireland and Sweden are excluded because data is not available for the whole 1980-1995 period in the EUROSTAT-REGIO database. Berlin and East Germany are also excluded for well-known historical and political reasons.

Some authors use per capita GDP expressed in purchasing power standards (PPS), i.e. adjusted for purchasing power parity (e.g. Armstrong 1995; López-Bazo et al. 1999), whereas we restrict our analysis to per capita GDP expressed in ECUs. The choice between per capita GDP expressed in ECUs or in PPS is rather complicated. Indeed, for international and interregional comparisons, per capita GDP values expressed in national currencies should be converted to a common currency (ECU for European countries). This conversion is carried out by means of official exchange rates but, for different reasons, these exchange rates don't mirror the real purchasing power of a given currency in the economic area of a country and their use does not provide a reliable indication of the volume of goods and services produced and consumed in different countries. To overcome this drawback, an exchange rate based on purchasing power parity is often used. Values obtained this way are called purchasing power standards (PPS). However, it is worth stressing that the construction of regional accounts in purchasing power parity that are comparable across space and time is very complicated and can raise serious problems. First, this conversion should be based on regional purchasing power parity but, due to data non-availability, this adjustment is calculated on the basis of national price levels and so does not take into account regional differences in prices, which can be significant, particularly when there are wide variations in income between regions. Second, per capita GDP expressed in PPS can change in one economy relative to another not only because of a difference in the rate of GDP growth in real terms but also because of a change in relative price levels. This complicates the analysis of changes over time insofar as a relative increase in per capita

GDP, which arises from a reduction in the relative price level or from a re-estimation of the PPS adjustment might have slightly different implications than one which results from a relative growth in real GDP. Third, as pointed out by Vanhoudt et al. (2000) data obtained using this method are primarily designed to compare countries within the same year. Using these data in time-series comparisons could be problematic yielding "implausible results for annual average growth rates. This is often overlooked when analyzing determinants of economic growth in empirical exercises" (p.83). This problem does not seem to be taken into account, and the fact that the definition of PPP has changed with every enlargement of the EU is also ignored. Results obtained for growth rates using PPS should therefore be interpreted with caution.

Another problem concerning the use of per capita GDP for assessing regional disparities is that commuter flows affect comparisons between regions. In the case of city regions, surplus commuters ensure that production activity in these regions is higher than it would be with resident workers only. As a result, per capita GDP in these regions is generally overestimated, and that of the regions in which the commuters live is generally underestimated. However, we note that: this effect is significant only in a few cases (Brussels, Luxembourg, Hamburg, Darmstadt, Bremen, Ile-de-France), poor regions are generally not affected by this problem, and these commuter flows are globally negligible at the NUTS2 level of regional breakdown.

7 References

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

Anselin L., Cho W.K.T. (2000), "Spatial Effects and Ecological Inference", *Political Analysis*, 10:276-297.

Arbia G. (1989), *Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems*, Kluwer Academic Publishers, Boston.

Armstrong H. (1995), "Convergence among the Regions of the European Union". *Papers in Regional Science*, 74:143-152.

Azariadis C., Drazen A. (1990), "Threshold Externalities in Economic Development". *Quarterly Journal of Economics* 105:501-526.

Barry R., Pace R.K. (1999), "A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices", *Linear Algebra and its Applications*, 289:41-54.

Barro R.J., Sala-I-Martin X. (1991) "Convergence across States and Regions". *Brookings Papers on Economic Activity,* 1:107-182.

Barro R.J., Sala-I-Martin X. (1992), "Convergence". *Journal of Political Economy* 100:223-251.

Barro R.J., Sala-I-Martin X.(1995) *Economic Growth Theory*, McGraw-Hill, New York.

Baumol W.J. (1986), "Productivity Growth, Convergence and Welfare: What the Long Run Data Show". *American Economic Review* 76:1072-1085.

Brock W.A., Durlauf S.N. (2001), "Growth Empirics and Reality". *The World Bank Economic Review*, 15:229-272.

Brundson C., Fotheringham A. S., Charlton M.(1996), "Geographically weighted regression: a method for exploring spatial nonstationarity," *Geographical Analysis*, 28:281-298.

Button K.J., Pentecost E.J. (1995), "Testing for Convergence of the EU Regional Economies". Economic Inquiry, 33:664-671.

Conley T.G., Ligon E. (2002), "Economic Distance, Spillovers, and Cross Country Comparisons" Journal of Economic Growth, 7:157-187.

DeLong J.B. (1988), "Productivity Growth, Convergence and Welfare: Comment". *American Economic Review*, 78:1138-1154.

Desdoigts A. (1999), "Patterns of Economic Development and the Formation of Clubs". *Journal of Economic Growth* 4:305-330.

Durlauf S.N. (2000), "Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective". *Working Paper*. Department of Economics. University of Wisconsin.

Durlauf S.N. (2001), "Manifesto for a Growth Econometrics". *Journal of Econometrics* 100:65-69.

Durlauf S.N., Kourtellos A. and Minkin A. (2001) "The local Solow growth model", *European Economic Review*, 45:928-940.

Durlauf S.N., Johnson P.A. (1995), "Multiple Regimes and Cross-Country Growth Behaviour". *Journal of Applied Econometrics* 10:365-384.

Durlauf S.N., Quah D. (1999), "The New Empirics of Economic Growth" in *Handbook of Macroeconomics,* Taylor J. and Woodford M. (eds). Elsevier Science, North-Holland.

European Commission (1999) *The European Rregions: Sixth Periodic Report on the Socio-Economic Situation in the Regions of the European Union*, Luxembourg, Official Publication Office.

Fingleton B. (1999), "Estimates of Time to Convergence: An Analysis of Regions of European Union". *International Regional Science Review*, 22:5-34.

Gelman A., Carlin J.B., Stern H.S., Rubin D.B. (1995), *Bayesian Data Analysis*, London, Chapman & Hall

Galor O. (1996), "Convergence? Inferences from Theoretical Models". *Economic Journal* 106:1056-69

Geweke J. (1993), "Bayesian Treatment of the Independent Student t Linear Model". *Journal of Applied Econometrics*, 8:19-40.

King G. (1997), *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*, Princeton University Press, Princeton.

Kubo, Y., (1995), "Scale economies, regional externalities and the possibility of uneven regional development," *Journal of Regional Science*, 35:318-328.

Lange, K.L., R.J.A. Little, Taylor J.M.G., (1989) "Robust Statistical Modeling Using the t Distribution," *Journal of the American Statistical Association*, 84:881-896.

Le Gallo, J., C. Ertur C., Baumont C. (2003) "A Spatial Econometric Analysis of Convergence across European Regions, 1980-1995", *European Regional Growth*, B. Fingleton, (ed.), Springer-Verlag, Advances in Spatial Science, 99-130.

Le Gallo J., and C. Ertur (2003) "Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995", *Papers in Regional Science*, 82(2):175-201.

LeSage, J.P. (1997) "Bayesian Estimation of Spatial Autoregressive Models", *International Regional Science Review*, 20(1&2):113-129.

LeSage, J.P. (2003) "A Family of Geographically Weighted Regression Models", forthcoming in *Advances in Spatial Econometrics*, Luc Anselin, J.G.M. Florax and S.J. Rey (eds.) Springer-Verlag

López-Bazo E., Vayá E., Mora A., Suriñach J. (1999), "Regional Economic Dynamics and Convergence in the European Union". *Annals of Regional Science*, 33:343-370.

Mankiw N.G., Romer D., Weil D.N. (1992) A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107:407-437.

Martin P., Ottaviano G.I.P., (1999) "Growing Locations: Industry Location in a Model of Endogenous Growth," *European Economic Review*, 43:281-302.

McMillen, D.P. (1996) "One Hundred Fifty Years of Land Values in Chicago: A Nonparameteric Approach," *Journal of Urban Economics*, 40:100-124.

McMillen, D.P., McDonald. J.F. (1997) "A Nonparametric Analysis of Employment Density in a Polycentric City," *Journal of Regional Science*, 37:591-612.

Molle W. (1980), *Regional Disparity and Economic Development in the European Community*, Saxon House, Farnborough, Hantes.

Moreno, R. and Trehan, B. (1997) "Location and the growth of nations" *Journal of Economic Growth*, 2:399-418

Neven D., Gouyette C. (1995), "Regional Convergence in the European Community", *Journal of Common Market Studies*, 33:47-65.

Openshaw S., Taylor P. (1979), *A Million or so Correlation Coefficient: Three Experiments on the Modifiable Areal Unit Problem*, in Wrigley N. (Ed.), Statistical Applications in the Spatial Science, Pion, London.

Pace, R.K, LeSage. J.P. (2002) "Spatial Autoregressive Local Estimation," forthcoming in *Recent Advances in Spatial Econometrics*, Jesus Mur, Henri Zoller and Arthur Getis (eds.), Palgrave Publishers.

Quah D. (1996), "Regional Convergence Clusters across Europe", *European Economic Review*, 40:951-958.

Sala-I-Martin X. (1996a), "Regional Cohesion: Evidence and Theories of Regional Growth and Convergence". *European Economic Review*, 40:1325-1352.

Sala-I-Martin X. (1996b), "The Classical Approach to Convergence Analysis". *Economic Journal*, 106:1019-1036.

Solow R.M. (1956), "A Contribution to the Theory of Economic Growth". *Quarterly Journal of Economics* 70: 65-94.

Vanhoudt P, Matha T, Smith B (2000), "How productive are capital investments in Europe". *EIB Papers* 5:81-106.

Table 2: Estimates for β and ρ

OBS	NUTS	β	σ_β	ρ	σ_ρ
1	be1	0.1265	0.3442	0.7673	0.0844
2	be21	0.0703	0.3283	0.7586	0.0818
3	be22	-0.0872	0.3123	0.7738	0.0789
4	be23	0.0039	0.3123	0.7603	0.0790
5	be24	0.0535	0.3251	0.7639	0.0811
6	be25	-0.0982	0.3067	0.7834	0.0762
7	be31	0.0070	0.3463	0.7605	0.0825
8	be32	-0.0606	0.3128	0.7704	0.0820
9	be33	-0.1795	0.3113	0.7614	0.0771
10	be34	-0.2197	0.3159	0.7881	0.0807
11	be35	-0.1197	0.3138	0.7745	0.0808
12	de11	-0.2825	0.2580	0.7235	0.0868
13	de12	-0.2864	0.2880	0.7233	0.0876
14	de13	-0.3147	0.2854	0.7211	0.0888
15	de14	-0.3066	0.2359	0.7171	0.0884
16	de21	-0.1317	0.2162	0.6610	0.1019
17	de22	-0.0809	0.2101	0.6621	0.0983
18	de23	-0.0605	0.2115	0.6452	0.0965
19	de24	-0.2011	0.2269	0.6834	0.0916
20	de25	-0.2221	0.2264	0.7110	0.0909
21	de26	-0.2637	0.2644	0.7006	0.0944
22	de27	-0.2195	0.2256	0.6976	0.0919
23	de5	-0.0696	0.3149	0.7506	0.0849
24	de6	-0.1511	0.3350	0.7412	0.0840
25	de71	-0.2574	0.3051	0.7147	0.0922
26	de72	-0.1204	0.3085	0.7033	0.0884
27	de73	-0.1966	0.3019	0.7137	0.0932
28	de91	-0.2483	0.2851	0.7458	0.0904
29	de92	-0.1110	0.3104	0.7221	0.0896
30	de93	-0.0381	0.3118	0.7522	0.0875
31	de94	-0.1534	0.3047	0.7928	0.0789
32	dea1	-0.1742	0.3297	0.7470	0.0806
33	dea2	-0.2194	0.3139	0.7528	0.0837
34	dea3	-0.1463	0.3112	0.7405	0.0795
35	dea4	-0.1064	0.3103	0.7356	0.0895
36	dea5	-0.1339	0.3171	0.7326	0.0850

OBS	NUTS	β	σ_β	ρ	σ_ρ
37	deb1	-0.2030	0.3111	0.7494	0.0842
38	deb2	-0.2660	0.3123	0.7439	0.0847
39	deb3	-0.1337	0.3013	0.7211	0.0844
40	dec	-0.2652	0.3078	0.7345	0.0850
41	def	-0.0324	0.3050	0.7382	0.0837
42	dk	-0.1784	0.3062	0.7973	0.0738
43	es11$\star$	-0.9529	0.2284	0.6055	0.0934
44	es12$\star$	-0.8981	0.2397	0.5882	0.0956
45	es13$\star$	-0.8828	0.2433	0.5757	0.0972
46	es21$\star$	-0.8472	0.2327	0.5866	0.0955
47	es22$\star$	-0.8067	0.2416	0.5926	0.1046
48	es23$\star$	-0.8367	0.2404	0.5903	0.1020
49	es24$\star$	-0.8080	0.2440	0.5880	0.1064
50	es3$\star$	-0.8296	0.2419	0.5913	0.1001
51	es41$\star$	-0.8826	0.2362	0.5826	0.0971
52	es42$\star$	-0.9053	0.2492	0.5776	0.1048
53	es43$\star$	-0.9778	0.2402	0.6198	0.1005
54	es51$\star$	-0.8109	0.2332	0.5509	0.1085
55	es52$\star$	-0.9414	0.2539	0.5366	0.1139
56	es53$\star$	-0.7892	0.2383	0.5178	0.1102
57	es61$\star$	-0.9604	0.2413	0.5983	0.0963
58	es62$\star$	-0.9545	0.2606	0.5344	0.1144
59	fr1	-0.4370	0.2779	0.7797	0.0784
60	fr21	-0.2209	0.3377	0.7213	0.0928
61	fr22	-0.1429	0.3156	0.7448	0.0878
62	fr23	-0.2334	0.2778	0.7045	0.0907
63	fr24	-0.2298	0.2587	0.6597	0.0919
64	fr25	-0.2424	0.2165	0.6067	0.0929
65	fr26$\star$	-0.6516	0.3209	0.7180	0.0993
66	fr3	-0.0549	0.3161	0.7810	0.0842
67	fr41	-0.2754	0.3073	0.7473	0.0839
68	fr42	-0.2344	0.3008	0.7195	0.0869
69	fr43	-0.3660	0.2929	0.7037	0.0900
70	fr51$\star$	-0.5063	0.2321	0.5635	0.1035
71	fr52$\star$	-0.8784	0.2505	0.5367	0.1021
72	fr53$\star$	-0.6159	0.2475	0.5418	0.1074

OBS	NUTS	β	σ_β	ρ	σ_ρ
73	fr61⋆	-0.7290	0.2477	0.5638	0.1026
74	fr62⋆	-0.4777	0.2150	0.5003	0.1194
75	fr63	-0.3155	0.2185	0.6185	0.0978
76	fr71⋆	-0.6536	0.2949	0.6966	0.0999
77	fr72⋆	-0.5235	0.2458	0.6629	0.1052
78	fr81	-0.3972	0.2135	0.5754	0.1105
79	fr82	-0.4783	0.2419	0.7016	0.0884
80	gr11	-0.1155	0.1817	0.6062	0.1079
81	gr12	-0.1547	0.1858	0.5879	0.1064
82	gr13	-0.1812	0.1857	0.6188	0.1065
83	gr14	-0.1617	0.1866	0.6220	0.1083
84	gr21	-0.2089	0.1844	0.5850	0.1114
85	gr22	-0.1501	0.1749	0.6405	0.1024
86	gr23	-0.1873	0.1858	0.6210	0.1102
87	gr24	-0.0759	0.1725	0.6008	0.1103
88	gr25	-0.1908	0.1757	0.6208	0.1074
89	gr3	-0.1886	0.1964	0.5836	0.1154
90	gr41	-0.0904	0.1919	0.5845	0.1097
91	gr42	-0.1135	0.1829	0.5821	0.1079
92	gr43	-0.1481	0.1723	0.6428	0.1060
93	it11	-0.4943	0.2629	0.6825	0.0961
94	it12	-0.5118	0.2725	0.7027	0.0926
95	it13	-0.1665	0.2006	0.6727	0.0970
96	it2	-0.0976	0.2019	0.7187	0.0873
97	it31	-0.0756	0.1800	0.6922	0.0909
98	it32	-0.0678	0.1812	0.7005	0.0925
99	it33	-0.0264	0.1749	0.6883	0.0919
100	it4	-0.0933	0.1903	0.6614	0.1012
101	it51	-0.0981	0.1916	0.6748	0.0990
102	it52	-0.0947	0.1589	0.6820	0.0978
103	it53	-0.0565	0.1714	0.6754	0.1023
104	it6	-0.0952	0.1680	0.6704	0.0973
105	it71	-0.0951	0.1769	0.6139	0.1048
106	it72	-0.1186	0.1912	0.5818	0.1137
107	it8	-0.2119	0.1814	0.6346	0.1078
108	it91	-0.2007	0.1810	0.6226	0.1045

OBS	NUTS	β	σ_β	ρ	σ_ρ
109	it92	-0.2078	0.1823	0.6307	0.1079
110	it93	-0.2123	0.1838	0.6351	0.1073
111	ita	-0.3251	0.1786	0.6641	0.0969
112	itb	-0.1837	0.1904	0.6836	0.0939
113	lu	-0.2880	0.2983	0.8040	0.0752
114	nl12	-0.0172	0.3131	0.7954	0.0742
115	nl13	-0.1470	0.2980	0.7908	0.0706
116	nl2	-0.1272	0.3108	0.7866	0.0755
117	nl31	-0.0648	0.3039	0.7947	0.0720
118	nl32	0.0327	0.3051	0.7826	0.0732
119	nl33	0.0414	0.3229	0.7659	0.0795
120	nl34	-0.0369	0.3167	0.7799	0.0827
121	nl41	-0.0431	0.3182	0.7821	0.0772
122	nl42	-0.1558	0.3109	0.7819	0.0816
123	pt11$\star$	-0.7989	0.2427	0.5621	0.0998
124	pt12$\star$	-0.8297	0.2405	0.5388	0.1013
125	pt13$\star$	-0.8889	0.2471	0.5222	0.1051
126	pt14$\star$	-0.8789	0.2725	0.5446	0.1086
127	pt15$\star$	-0.8882	0.2457	0.5369	0.1047
128	uk1	-0.0351	0.2141	0.6803	0.0871
129	uk2	-0.1025	0.2502	0.7078	0.0856
130	uk3	-0.1085	0.2554	0.6858	0.0928
131	uk4	-0.2146	0.3115	0.7332	0.0855
132	uk5	-0.1517	0.2779	0.6553	0.0957
133	uk6$\star$	-0.7126	0.2870	0.5241	0.1071
134	uk7	-0.1408	0.2375	0.6389	0.0942
135	uk8	-0.0041	0.2123	0.6752	0.0877
136	uk9	-0.5423	0.3013	0.5389	0.1107
137	uka	-0.0211	0.2172	0.6869	0.0847
138	ukb$\star$	-0.7504	0.2520	0.6019	0.0990