Testing for Convergence Clubs in Income per-capita: A Predictive Density Approach.

Fabio Canova *

Universitat Pompeu Fabra; University of Southampton; and CEPR

First draft, July 1997
This revision, June 1999

Abstract

The paper proposes a technique to jointly tests for groupings of unknown size in the cross sectional dimension of a panel and estimates the parameters of each group, and applies it to identifying convergence clubs in income per-capita. The approach uses the predictive density of the data, conditional on the parameters of the model. The steady state distribution of European regional data clusters around four poles of attraction with different economic features. The distribution of income per-capita of OECD countries has two poles of attraction and each group has clearly identifiable economic characteristics.

JEL Classification No.: C11, D90, O47

Key words: Heterogeneities, Panel Data, Predictive Density, Income Inequality.

---

*I would like to thank three anonymous referees, the editor of this journal, Bruce Hansen, Hashem Pesaran, Russell Cooper, Christopher Croux, Albert Marcet and the participants at seminars at Universitat Pompeu Fabra, the University of Southampton, Universite de Paris I-MAD, IGIER, University of Jerusalem, the 1997 NBER-NSF Summer workshop on "Empirical Methods in Macroeconomics" and the 1998 CEPR workshop "New Approaches to the Study of Business Cycles" for comments and suggestions. A previous version of the paper has circulated with the title " Testing for heterogeneity in the cross sectional dimension of a panel: a predictive density approach".

We share the uncommonness of being different. J.P. Roche

1 Introduction

Recent theories of growth and development have suggested that the distribution of income per-capita of countries and/or regions may display convergence clubs, i.e. a tendency for the steady states distribution to cluster around a small number of poles of attraction (see e.g. Ben David (1994), Quah (1996a), or Galor (1996)). This tendency may be induced by several factors: the existence of some threshold level in the endowment of strategic factors of production; non-convexities or increasing returns; similarities in preferences, technologies; government policies, which become more similar over time within certain groups (e.g. EEC or East Asian countries). While there is anecdotal evidence supporting the view that clustering is an important feature of world income data, to the best of my knowledge, only Durlauf and Johnson (1995) using regression tree analysis, Quah (1996a) and Desdoigts (1998) using nonparametric methods, have attempted to formally document whether this tendency exits in the actual data.

This paper proposes a technique to examine whether the distribution of income per-capita displays convergence clubs. The approach is general, determines the number of groups and the location of the break points when the appropriate ordering of the units in the cross section is unknown and, at the same time, allows to estimate the parameters characterizing the distribution of each group in a unified manner. The approach is based on the predictive density (marginal likelihood) of the data, conditional on the parameters, and has appealing features for both Bayesian and classical analysts.

The suggested technique can be viewed as a natural extension of the standard testing approach used to determine the number of heterogeneous groups in a cross section (see e.g. Goldfeld and Quandt test) when the number of groups, the location of the breaks and the ordering of units are unknown. However, instead of assuming that the regression coefficients are the same for all units belonging to one group, as it is the case with switching regressions or regression tree analyses, I allow for a further layer of heterogeneity within groups. This second layer of heterogeneity takes the form of an exchangeable prior which restricts the coefficients of the units in a group to have the same distribution. Hence, while standard exchangeable approaches assume that the coefficients of the statistical model of *all* cross-sectional units have the same distribution, I restrict the behavior of coefficients within a group, but I allow the distribution of the coefficients of units in different groups to differ. Because exchangeability over the entire cross-section implies that the steady state distribution of income per-capita is unimodal, while exchangeability within groups implies that the steady state distribution may display multiple basins of attraction, testing for the presence of convergence clubs can be fruitfully examined by checking which of these two assumptions is more appropriate.

Once the optimal number of groups and the location of the break points in the cross section have been established, I provide a simple way to estimate the parameters of each group and to conduct inference. The approach I employ lies within the Empirical Bayes tradition: I use predictive densities to estimate the parameters and posterior analysis to draw conclusions about functions of the coefficients of the model. Posterior inference is appealing because it gives us a compact way to summarize both subjective and objective uncertainty about economically interesting functions of the coefficients of the model (convergence rates, long run multipliers, steady states distribution, etc.).

The methodological contribution of this paper is linked to a number of articles, both in the classical and the Bayesian tradition, testing for the existence of a unknown break points in time series, see e.g. Ploberger, et. al. (1989), Bai (1997) and Polasek and Rei (1997) and to the Empirical Bayes tradition of constructing posterior estimates of the coefficients of a model by plugging-in ML type-II estimates of the parameters of the prior (see Morris (1983), Berger (1985) or Efron (1996)). The approach is also related to those of Forni and Reichlin (1997), who attempt to estimate a reduced number of common latent factors from large dynamic cross sectional data, and of Hansen (1997a,b), who examines estimation and testing problems in threshold models for cross sections, time series or static panels. The most significant difference between the approach of the paper and the one of the latter author, apart from the classical vs. Bayesian perspective, is that in Hansen's work the threshold between groups is observable and exogenous - so that the problem is to obtain useful estimates of the threshold parameter - while here the threshold index is either unknown or unobservable and could even be endogenous. Finally, the testing procedure shares similarities with classification/cluster analyses (see e.g. Mardia, Kent and Bibby (1980)). Three features distinguish the proposed approach from existing ones: I use regression models with serially correlated data; I allow the number of break points to be unknown; and I assign units to groups so as to maximize the predictive ability of the model.

I employ European regional income per-capita data from the NUTS2 data set of Eurostat and OECD national income per-capita from the Summer and Heston data set to determine whether the income distribution shows any tendency toward club convergence. Recent theories of economic growth have suggested that the initial conditions of income per-capita and of the average human capital; the dispersion of the distribution of income and education within units; and the geographical location may determine the position of a unit in the steady state distribution and the club it will join. Unfortunately, most of this information is not available at regional level. Therefore a search for clubs is conducted ordering units in the cross section according to five different criteria: (i) the ranking of income per-capita relative to European average prevailing in a pre-sample period, with poor units coming first; (ii) the magnitude of the average per-capita income relative to the

European average in the sample, with poor units coming first; (iii) the magnitude of the average growth rate of income in the sample, with poor units coming first; (iv) the ranking of income per-capita in the pre-sample period, scaling per-capita income by the national average; (v) the ranking of income per-capita in the pre-sample period, scaling per-capita income of "southern" regions (Mediterranean regions and Ireland) and of "northern" regions (the others) by their own respective average. At country level, the initial conditions of income per-capita and human capital and their within-country dispersions are available so that it is possible to examine the likelihood of convergence clubs using these indicators to order the data. In addition, I search for groupings along size and geographical dimensions: putting G-3 countries first and then the rest, ordered according to the size of the economies; putting European countries before the rest, with Mediterranean countries and Ireland preceding other European countries in the order.

I find that the ordering based on the ranking of the initial conditions of income per-capita scaled by the European average in the pre-sample period is the one which maximize the predictive power of the model for both data sets. With that ordering, there is a natural clustering of units in four groups of regional income per-capita and two groups of national income per-capita. In both cases clubs are characterized by different parameters controlling the speed of adjustment to the steady state and the mean level of per-capita steady state income relative to the average. More precisely, poor units converge faster to their steady state than rich ones and they tend to cluster around a pole of attraction which is substantially below the average (see also Quah (1996b)). The dispersion of steady states around each basin of attraction is significant suggesting that clustering is more prevalent than convergence even within groups. I show that even though groups have different long run mobility indices, there is substantial immobility in the ranking of units within groups, confirming the strong persistence in inequality found by Canova and Marcet (1995). As a consequence of the persistence of the initial income characteristics and of the immobility in ranking, the steady state distribution of income per-capita will become polarized. Since poor units are also the ones with low initial average human capital; with distributions of income and education which are more polarized; and are geographically or economically located in the "South" of the industrialized world, the results provide a bleak picture over the possibility of equalizing income per-capita both in Europe and in OECD countries over the near future.

The rest of the paper is organized as follows. The next section describes the details of the testing approach to find groups in the cross sectional dimension of a panel when the number of groups, the location of the break points and the ordering of units is unknown. Section 3 provides a technique to estimate the parameters and to conduct posterior inference on functions of the coefficients of the dynamic model. Section 4 provides the link between growth theory and the proposed econometric procedure, emphasizing measurable factors which may determine club convergence. Section 5 examines the existence of convergence clubs in European regional and OECD national income per-capita data. Section 6 concludes.

2 The Testing Procedure

The starting point of the analysis is the a-priori belief that there may be significant heterogeneities in the cross section of a panel and a natural clustering of units around certain poles of attraction, in the sense that the coefficients of the statistical model are more similar within each group than across groups. For example, if units i and j belong to a group, the vector of coefficients of the model for the two units may have the same mean and the same dispersion. However, if units i and j do not belong to the same group, the vector of coefficients of the two units may have different means and different dispersions.

For the sake of generality, I assume that the ordering of cross sectional units which naturally gives rise to clustering is unknown. In practice, clustering in income per-capita may be linked to geographical, economic or sociopolitical factors and modern growth theory provides a restricted set of ordering to be examined. Let N be the size of the cross section, T the size of the time series, and $m = 1, 2, \ldots N!$ the particular ordering of the units of the cross section. It is assumed that there may be $q = 1, 2, \ldots Q$ break points in the cross section, Q given. Each of the resulting $q + 1$ groups is characterized by a statistical model of the form:

$$Y_{it} = \alpha_i + \rho_i(\ell)Y_{it-1} + \theta_i(\ell)W_{t-1} + u_{it} \quad (1)$$

$$\beta_i = \beta^p + \epsilon_i^p \quad (2)$$

where $i = 1, \ldots, n^p(m)$; $p = 1, \ldots, q+1$, $u_{it} \sim (0, \sigma^2_{u_i})$, $\epsilon_i^p \sim (0, \Sigma_p)$, $\rho_i(\ell)$ and $\theta_i(\ell)$, are polynomials in the lag operator of order r and d, $\beta_i = [\alpha_i, \rho_{i1}, \ldots \rho_{ir}, \theta_{i1}, \ldots, \theta_{id}]$ is the vector of coefficients of unit i, $n^p(m)$ is the number of units in group p, given the m-th ordering of the cross section, $\sum_p n^p(m) = N$. I assume that Y_{it} is a vector of dimension s for each unit i, while W_{t-1} is a vector of exogenous variables of dimension v affecting all units of the cross section with a period lag. In (2), I assume that the vector of coefficients for each i is random and that the coefficients of the $n^p(m)$ units belonging to group p have the same mean and same covariance matrix. This situation will be termed exchangeable structure within group. Furthermore, I assume that the exchangeable structure may differ across groups: the coefficients of units belonging to different groups may be drawn from distributions with different parameters. Equations (1)-(2) therefore captures in a simple way the idea that there may be clustering of units within groups but that groups may drift apart over time, implying heterogeneous dynamics in the cross section. For the rest of the paper I refer to β^p and Σ_p as the hyperparameters of the model.

Model (1)-(2) is sufficiently general to include several models studied in the panel data literature as special cases. For example, a standard switching regression model is obtained by setting $\epsilon_i^p = 0, \ \forall i$. A fixed effect model is obtained by restricting $\rho_i = \rho$ and $\epsilon_i^p = 0, \ \forall i$ while a random effect model is obtained by setting $\rho_i = \rho$ and $\epsilon_i^p = [\epsilon_{i1}^p, 0, \ldots, 0]$. Also, for future reference, I take the alternative to (1)-(2) to be a model with homogeneous dynamics in the cross section. In this case $Q = 1$ and I replace equation (2) with

$$\beta_i = \beta + \epsilon_i \qquad i = 1, \ldots N \tag{3}$$

where $\epsilon_i \sim N(0, \Sigma)$. In other words, in the alternative β and Σ are the same for all i, so that there is an exchangeable structure across all units of the cross section. The limiting case of this alternative is a pooled model which can be obtained by setting $\epsilon_i^p = 0, \ \forall i$.

The task of the paper is two fold. First, I am interested in providing a framework for verifying the hypothesis that there are heterogeneities in the cross section in a situation where the number of groups, the location of the breaks (and consequently the number $n^p(m)$ units in each group) and the permutation m, which naturally give rise to the clustering, are unknown. Once I have established the number of groups, the location of the breaks and the ordering of the cross section optimally, I will be concerned, at a second stage, with the problem of estimating the hyperparameters (β^p, Σ_p) for each p and $\sigma_{u_i}^2$ for each i. These parameters are assumed to be unknown to the investigator and are needed to construct posterior estimates of the β_i which can then be used for inference and forecasting.

Let Y be a $(N * T * s) \times 1$ vector of the LHS variables in (1) ordered to have the N cross sections for each $t = 1, \ldots T, \ s$ times, X be a $(N * T * s) \times (N * k)$ matrix of the regressors, $k = s * r + v * d + 1$, β be a $(N * k) \times 1$ vector of coefficients of the model, U a $(N * T * s) \times 1$ vector of disturbances, β_0 a $(q + 1) * k \times 1$ vector of means of β, A be a $(N * k) \times (q + 1) * k$ matrix, $A = diag\{A_p\}$, where A_p has the form $\iota \bigotimes I_k$ where I_k is a $k \times k$ identity matrix and ι is a $n^p(m) \times 1$ vector of ones. For given m, we rewrite $(1) - (2)$ as:

$$\begin{aligned} Y &= X\beta + U \qquad\qquad U \sim (0, \Sigma_u) && (4) \\ \beta &= A\beta_0 + E \qquad\qquad E \sim (0, \Sigma_E) && (5) \end{aligned}$$

where the dimension of Σ_u is $(N*T*s) \times (N*T*s)$ and $\Sigma_E = diag\{\Sigma_p\}$ is a $(N*k) \times (N*k)$ matrix. Using (5) into (4) we arrive at

$$Y = \tilde{X}\beta_0 + W \qquad\qquad W \sim (0, \Sigma_W) \tag{6}$$

where $\tilde{X} = X * A$ and $\Sigma_W = X\Sigma_E X' + \Sigma_u$. In (6) I have eliminated (integrated out) the vector of coefficients β and expressed the dependent variable Y as a linear combination of the X's and of the hyperparameters β_0 with errors which have an heteroschedastic structure. Since $q + 1 << N$, this operation has effectively reduced the dimensionality of the model. For the moment assume that β_0, Σ_E, Σ_u are known. Our approach to group units proceeds in several steps.

First, given an ordering of the units of the cross section, I examine how many groups there are using the sequential testing approach described below. Second, given an ordering of units and the optimal number of groups I attempt to find the location of the break points by maximizing the predictive density (marginal likelihood) of the model with respect to the location of the breaks. Third, I iterate on the first two steps, varying the ordering of units in the cross section. I choose as optimal the ordering that maximizes the predictive density.

To be precise, let $L(Y|H_0)$ be the predictive density of the data under the assumption that the hyperparameters are the same in each subgroup, i.e. $\beta_0 = \iota_1 \bigotimes \gamma_0$ where ι_1 is a $(q+1) \times 1$ vector of ones and γ_0 a $k \times 1$ vector, and $\Sigma_p = \Sigma$, $\forall p$. Let $L(Y^p|H_q, n^p(m), m)$ be the predictive density for group p, under the assumption that there are q break points, with $n^p(m)$ observations in each group, for ordering m and let $L(Y|H_q, n^p(m), m) = \prod_{p=1}^{q+1} L(Y^p|H_q, n^p(m), m)$ be the total predictive density for the sample under the assumption that there are q break points. Define the following quantities:

- $L^+(Y|H_q, n^p(m), m) = \sup_{i \in [h_1(p), h_2(p)]} L(Y|H_q, n^p(m), m)$,
- $L^\dagger(Y|H_q, n^p, m) = \sup_{m \in [1, N!]} L^+(Y|H_q, n^p(m), m)$,
- $L^{Aq}(Y|H_q, n^p, m) = \sum_{i=h_1(p)}^{h_2(p)} \pi_i^p L(Y|H_q, n^p(m), m)$,

where π_i^p is the prior probability that, for group p, there is a break at $i = h_1(p), \ldots, h_2(p)$ where $h_1(p) \geq 1$; $h_2(p) \leq n^p(m)$. The first expression gives the maximized value of the predictive density with respect to the location of break points for each q; the second, the maximized value of the predictive density, once the location of the break point and the ordering of the data are chosen optimally. The last expression gives the average predictive density under the assumption that there are q breaks. Here the average is calculated over all possible locations of the break points, using the prior probability that there is a break point in each location as weight. In general, unless there are compelling reasons not to do so, ignorance about the location of the break points leads us to assume that π_i^p is uniform over each p.

To examine the hypothesis that the dynamics of the cross section are heterogenous one can use either a posterior odds (PO) ratio, a Wilks likelihood ratio (WL) criteria (see e.g. Efron (1996)) or the modified likelihood ratio (ML) of Hansen (1997a). I consider first the null hypothesis that there are no break points against the alternative that there are at most Q breaks and then, if the alternative is more likely, sequentially test a series of hypotheses where the null is that there are $q-1$ break points and the alternative that there are q break points, $q = 1, \ldots Q$. Given an ordering m, the three statistics for the first hypothesis are:

$$PO(m) = \frac{\pi_0 L(Y|H_0)}{\sum_{q=1}^{Q} \pi_q L^{Aq}(Y|H_q, n^p, m)} \tag{7}$$

$$WL(m) = -2 * log(\frac{L^+(Y|H_Q, n^p, m)}{L(Y|H_0)}) \tag{8}$$

$$ML(m) = |\Sigma_w|^{-1} * (L(Y|H_0) - L^+(Y|H_Q, n^p, m)) \tag{9}$$

where π_0 is the prior probability that there are no breaks and π_q is the prior probability that there are q breaks. H_0 is preferred to H_1 when $PO(m) > 1$; rejected when $WL(m)$ exceeds an asymptotic confidence level obtained from a $\chi^2(Qk)$ random variable or ML(m) exceeds the asymptotic confidence level for the distribution tabulated by Hansen. The statistics for the hypotheses that there are $q-1$ vs. q breaks in the cross section are:

$$PO(m, q-1) = \frac{\pi_{q-1} L^{A(q-1)}(Y|H_{q-1}, n^p(m), m)}{\pi_q L^{A(q)}(Y|H_q, n^p(m), m)} \tag{10}$$

$$WL(m, q-1) = -2 * log(\frac{L^+(Y|H_q, n^p(m), m)}{L^+(Y|H_{q-1}, n^p(m), m)} \tag{11}$$

$$ML(m, q-1) = |\Sigma_w|^{-1} * (L^+(Y|H_{q-1}, n^p, m) - L^+(Y|H_q, n^p, m)) \tag{12}$$

Similarly, q breaks are preferred when $PO(m, q-1) < 1$; the hypothesis of $q-1$ breaks rejected when $WL(m, q-1)$ exceeds an asymptotic confidence level obtained from a $\chi^2(k)$ random variable or when ML(m,q-1) exceeds the tabulated values. We can also test the null hypothesis that there are q break points at particular locations against the alternative that there is a further break point at a particular location i using a posterior odds ratio of the form:

$$PO(m, q-1*) = \frac{\pi_{q-1}\ L^+(Y|H_{q-1}, n^p(m), m)}{\pi_q\ \pi_i^p\ L^+(Y|H_q, n^p(m), m)} \tag{13}$$

Note that when $\pi_q = \pi_{q+1} = 0.5$ (13) corresponds to the PIC criteria of Phillips and Ploberger (1994).

To put the testing problem in an alternative perspective, one can ask what is the prior probability on each of the null hypotheses one must entertain so that his/her beliefs will not be overturned by the data. For example, it may be of interest to know how much confidence one should have on the hypothesis that the sample is homogeneous so that a overall exchangeable prior is sufficient to characterize the data. This prior probability, which I call $\hat{\pi}$ can be found for any of the hypotheses considered by setting PO in (7)-(10)-(13) equal to 1 and solving the three equations for $\hat{\pi}_0$, $\hat{\pi}_q$, $\hat{\pi}_q$, respectively.

The testing procedure I have described leaves the value of Q unspecified. Following Hartigan (1975), I suggest to select Q using the rule of thumb $Q << \sqrt{(N/2)}$.

To find the location of the break point, given that there are q breaks, I assign units to groups so as to provide the highest total predictive density, i.e. I compute $L^{+}(Y|H_q, n^p, m)$. Since there are m possible permutations of the cross section over which to search for clustering I take the optimal permutation rule of units in the cross section to be the one which achieves $L^{\dagger}(Y|H_q, n^p, m)$. Hence, given Q and q, the criteria to optimally classify units in groups is:

$$\sup_m \sup_i L(Y|H_q, n^p(m), m) \tag{14}$$

Bai (1997) shows that proceeding sequentially in testing for breaks, i.e. test first for one break against no breaks; then conditional on the results of the first test, test for the existence of one break in each of the two subsamples and so on, produces consistent estimates of the number and the location of the breaks. However, when there are multiple groups and one tests for the presence of two groups only, the estimated break point is consistent for *any* of the existing break points and its location depends on which of the breaks is "stronger". If this is the case Bai suggests to refine the estimate of the break points. That is, if two breaks are identified at i_1 and i_2, it is convenient to reestimated i_1 over $[1, i_2]$ and i_2 or $[i_1, N]$. Each refined estimator of the location of the break has then the same properties as the estimator obtained in the case the sample has a single point.

The major stumbling block to the application of the procedure I have described is the dimensionality of the maximization problem. When no information is available on the ordering of the units in the cross section, it becomes imperative to calculate the predictive density for N! ordering. Clearly, when N is moderately large, this is an impossible task given existing computer technology. However, this is not a binding constraint for many applications since economic theory typically suggests to researchers which orderings should be tried and this considerably reduce the computational complexity of the problem. Note also that, even in the case economic theory is silent and one engages in an unstructured search, the maximization of (14) requires a considerably smaller number of evaluation than N!, since many ordering are equivalent from the point of view of the predictive density. That is, once a particular grouping is found, searching for groups can be shwredly conducted by reassigning units across groups around this local maxima.

An example may clarify the issue. Suppose N=4 so that we have a total of 24 possible ordering to examine. Suppose we have started with the ordering 1234 and found two groups: 1 and 234. Then all permutations of 234 with unit 1 coming ahead, i.e. 1243, 1342, etc., need not to considered as they give the same predictive density (see the appendix for a confirmation of this result in a Monte Carlo context). Similarly permutations which leave unit 1 last need not to be tried, i.e. 2341, 2431, etc. This first pass reduces the number of ordering to be examined to 13. But this is not the end. By trying another ordering, say 4213, and finding, for example, two groups: 42 and 13, we can further eliminate all the ordering which simply consist of permutations of the elements of each group, i.e. 4132, 2341, etc.. It is easy to verify that once four carefully selected ordering have been tried and, say, two groups found in each trial, we have exhausted all possible combinations, as far as the predictive density is concerned. The example is rigged so that at each stage we find two groups. When this is not the case, the number of ordering to be examined is larger, but it does not exceed hN where h is the maximum number of breaks found with any of the permutation.

3 Inference

Once the "best" ordering of the units in the cross section, the number of break points and their locations have been determined, I will be interested in estimating the unknown matrix Σ_u and the hyperparameters contained in the vector β_0 and in the matrix Σ_E. Let $\omega = [\beta_0', vec(\Sigma_E)', vec(\Sigma_u)']$ be the vector of parameters of the model. The predictive density $L^\dagger$ can now be used as a function of ω, for fixed Y, and maximized to construct the best possible fit of the model to the data. Maximizing $L^\dagger(\omega|Y, H_q, n^p(m), m)$ with respect to ω yields the so-called ML-type II estimator for the ω vector (see Berger (1985)) which is the starting point for obtaining posterior estimates of the coefficients of the dynamic model for each individual unit.

There are several ways to obtain estimates of ω under the assumption that the errors in $(1)-(2)$ are normally distributed, or under the more general assumption that the errors are drawn from a distribution in the exponential family (see e.g. Efron (1996)). For, example, if the u's are normally distributed, the vector ω can be estimated as (see Maddala (1991)):

$$\hat{\beta}^p = \frac{1}{n^p(m)} \sum_{j=1}^{n^p(m)} \beta^j_{ols} \tag{15}$$

$$\hat{\Sigma}_p = \frac{1}{n^p(m)-1} \sum_{j=1}^{n^p(m)} (\beta^j_{ols} - \hat{\beta}^p)(\beta^j_{ols} - \hat{\beta}^p)' - \frac{1}{n^p(m)} \sum_{i=j}^{n^p(m)} (x_j x_j')^{-1} \hat{\sigma}^2_j \tag{16}$$

$$\hat{\sigma}^2_j = \frac{1}{T-k}(y_j' y_j - y_j' x_j \beta^j_{ols}) \tag{17}$$

where $p = 1, \ldots, q+1$; $i = 1, \ldots, N$; x_j is the matrix of regressors and y_j the vector of dependent variables for unit j in the panel and β^j_{ols} is the OLS estimator of β^j obtained using only the information for unit j.

Given these estimates for the hyperparameters, one can construct Empirical Bayes (EB) posterior point estimates for the β vector by plugging-in estimated values in standard formulas, i.e.:

$$\hat{\beta} = (X'\hat{\Sigma}_u^{-1} X + \hat{\Sigma}_E^{-1})^{-1}(X'\hat{\Sigma}_u^{-1} Y + \hat{\Sigma}_E^{-1} A \hat{\beta}_0) \tag{18}$$

Alternatively, Smith (1973) shows that, under normality of the u's and the ϵ's and after imposing a diffuse prior on the ω, it is possible to jointly estimate ω and the posterior mean of β as follows:

$$\hat{\beta}^p = \frac{1}{n^p(m)} \sum_{j=1}^{n^p(m)} \beta_j^* \tag{19}$$

$$\hat{\Sigma}_p = \frac{1}{n^p(m) - k - 1}[R + \sum_{j=1}^{n^p(m)} (\beta_j^* - \hat{\beta}^p)(\beta_j^* - \hat{\beta}^p)'] \tag{20}$$

$$\hat{\sigma}_j^2 = \frac{1}{T+2}(y_j - x_j\beta_j^*)'(y_j - x_j\beta_j^*) \tag{21}$$

$$\beta_j^* = (\frac{1}{\hat{\sigma}_j^2}x_j'x_j + \hat{\Sigma}_p^{-1})^{-1}(\frac{1}{\hat{\sigma}_j^2}x_j'x_j\beta_{ols}^j + \hat{\Sigma}_p^{-1}A_0\hat{\beta}^p) \tag{22}$$

where $p = 1, \ldots, q+1$; $i = 1, \ldots, N; j = 1, \ldots, n^p(m)$; R is a diagonal matrix with small positive entries used here as in ridge-like estimators to insure that estimates of the dispersion matrix for each group are positive definite.

Note that while the first approach only requires OLS estimates for each unit, so that posterior estimates can be computed in two steps, in the second approach estimates of the prior parameters and of the posterior mean of β are obtained jointly using an iterative approach.

It is typically the case that the normal posterior distribution whose mean is given in (18) or in (22) has a covariance matrix which underestimates the covariance matrix obtained from a fully hierarchical Bayes approach. This is because no allowance is made for the fact that the hyperparameters have been estimated and that the number of units in each group may be small. In this situation it is typical to correct the posterior distribution to eliminate the bias in the confidence intervals for β by either explicitly taking the uncertainty in the estimates of Σ_E and β_0 into account or by bootstrapping confidence intervals directly and taking the conditional mean of the empirical distribution as the relevant confidence interval (see e.g. Morris (1983) or Carlin and Gelfand (1990)). In many applications, among which the one presented here, researchers are not necessarily interested in the spread of the posterior distribution of β's but rather they may want to study functions of the posterior mean of β. In this case, no correction is necessary and EB estimates in (18) or (22) provide reliable point estimates (see Berger (1985)).

3.1 A comparison with the existing literature

As mentioned in the introduction, our testing-classification-estimation approach share features with existing procedures and improves over them in some dimensions.

For example, one advantage of the procedure over the regression-tree analysis of Durlauf and Johnson (1995) is that heterogeneity within groups is allowed while their procedure makes the extreme assumption that all the heterogeneities disappear once one sorts units into groups. In other words, their estimates represent within group averages of the underlying individual coefficients. On the other hand, their approach allows to look for breaks in more than one dimension at the time, while this is somewhat cumbersome in the approach presented here. Nevertheless, as we will see in section 5, it is possible to combine information about breaks obtained ordering units in different ways at the inferential stage.

Relative to the graphical techniques adopted by Quah (1996a) and (1996b), the approach allows formal verification of the existence of groups in the cross section. Quah's approach, on the other hand, requires less stringent assumptions than I am making here (e.g. the coefficients of the dynamic model could be time varying in his setup).

The procedure is also related to the one of Paap and Van Dijk (1994) who use a mixture of normal densities to characterize the multimodal distribution present in their data and assign units to groups using a decision-making Bayesian rule. The analysis conducted here under normality can be interpreted as attempting to fit a mixture of Q normal distributions to the data, where both Q and the number of units in each group is unknown. Desdoigts (1998) uses a non-parametric (projection pursuit) method to find a set of economic characteristics which allow him to sort units into groups. However, he groups units in the cross section using differences in the regressors of model (1), while here differences across groups have to do with the parameters of the distribution of the coefficients of the dynamic model, not with its regressors.

Finally, the approach is also related to standard clustering and classification techniques (see e.g. Mardia, Kent and Bibby (1980)). Contrary to these techniques I use a regression framework with serially correlated data; I allow groups to have different covariance matrices; I do not restrict a-priori the number of groups (only the maximum number of grouping is chosen a-priori) and I use the predictive density, as opposed to the within group variance, as classification device.

I have run a Monte Carlo exercise to examine the ability of the procedure to detect breaks in the cross sectional dimension of a panel and of unbiasedly estimating the hyperparameters with simple DGPs. The results are presented in some details in the Appendix. It turns out that, if the ordering is correctly specified, the predictive density approach I have suggested is able to correctly detect the number and the location of breaks when there are simple or multiple breaks in the data. However, the posterior odds ratio appears to be slightly biased downward when no heterogeneities are present. This suggests that a conservative strategy to avoid the proliferation of groups is to choose a prior odds to slightly favor the null of no breaks even in situations where no prior information is available. When the ordering is unknown, the maximization of the predictive density over permutations recovers the best ordering of units in the cross section, and once the ordering is found, the number and the location of the breaks is correctly identified. Estimates of the hyperparameters are biased when the size of the time series is small: mean parameters are downward biased and variance parameters upward biased. When $T \geq 30$ most of these biases disappear.

4 Linking the Econometric Approach and Growth Theory

While I have argued that without any prior information about the ordering of the units, a brute force approach to the maximization of the predictive density is feasible, even though computationally demanding when N is large, it is also the case that economic theory provides information that restricts the number of interesting permutations one should try. It does so by providing indicators which may determine which unit will belong to which group.

For the case of convergence clubs, existing growth theory has suggested many mechanisms that may lead to such an outcome. Galor (1996) provides a thoughtful and compact summary of the major implications of various theoretical models, stressing the economic indicators which may produce club convergence. To provide the necessary link between the theory and the implementation of the proposed approach and some guidelines to interpret the results, I next briefly summarize the causes of clubs convergence and the indicators which can be useful to order units in the cross sectional dimension of the panel.

Basic neoclassical growth models, with production functions exhibiting decreasing returns to scale to the capital-labor ratio, exogenous population growth and fixed saving rate may generate convergence clubs in, at least, two circumstances: when saving rates out of wage and interest income differ with the first being larger; when the economy features heterogenous agents. The first assumption may be a consequence of heterogenous factor endowment across individuals and life-cycle considerations, while the second one, for example, is a standard feature of OG models. In both cases, multiplicity of stationary equilibria occurs and the distribution of initial income per-capita determines the asymptotic club to which a particular unit will belong.

The incorporation of empirically important elements such as human capital or fertility in the basic neoclassical growth model, along with some type of market imperfections (externalities, imperfectly competitive markets, non-convexities, and so on) produces additional channels which stregthen the possibility of club convergence. For example, social increasing returns with respect to human capital accumulation or capital market imperfections together with non-convexities in the production of human capital generate convergence clubs. In this case units which are similar in their characteristics and in their initial level of income may cluster around different steady state equilibria because they have different endowments of human capital (see e.g. Azariadis and Drazen (1990)). In some cases, it may be the within unit distribution of human capital which determines the steady state around which units may cluster (see Galor and Zeira (1993)). The within unit distribution of initial income may also be the reason for why units converge to different clubs: capital market imperfections together with some fixed cost in production may generate this outcome (see Quah (1996b)). A model with endogenous fertility, as in Barro and Becker (1989), can also produce convergence clubs. In this case the initial conditions with respect to number of children and the level of human capital dictate the steady state equilibria in which a unit will settle. In other versions of a such a model, it is the initial level of distribution of income which determines the distribution of the steady state level of output per-capita and fertility rates.

Finally, Quah (1996a) suggests that club convergence may due to informational externalities which may occur at either state or neighborhood level. That is, units which are either members of the same nation, share some borders, or belong to geographically homogenous areas, may tend to cluster together because information flows more easily across units with these characteristics. Hence the geographical location of a unit determines the convergence club it will join. This local externality hypothesis substantially differs from those which use increasing returns to scale in some factor of production and may generate converge clubs even under standard assumptions about preferences and technologies.

To summarize, the theoretical literature has provided at least four different indicators which may be used to order units along the cross section: the initial level of income, the initial level of human capital; the initial distribution of income per-capita and human capital within the unit. Furthermore, geographical indicators can be used to scale per-capita income data and/or to organize units in the cross section.

5 Are There Convergence Clubs?

In this section I study whether convergence clubs exist in income per-capita with two goals in mind. First, I would like to examine the compatibility of income per-capita data with modern growth theory with multiple steady states [1]. Second, I would like to better understand the statistical properties of income per-capita data. In particular, I am interested in examining what kind of heterogeneities the data displays, whether the average adjustment properties to the steady state and the average steady state are group dependent, and whether different groups display difference persistence of inequalities, in the sense that the relative ranking in the initial distribution is more important in

[1] What I examine here is a somewhat strong version of this hypothesis. A weaker version would predict the existence of convergence clubs in the distribution of *growth rates* of income per capita.

determining the relative ranking in the steady state distribution for some groups than others. I study these issues using two different data sets: European regional income per-capita from the Eurostat database and OECD national income per-capita from the Summer and Heston database.

5.1 European Regional Income per-capita

The data set used in this subsection covers 144 European NUTS2 units and refers the period 1980-1992 [2]. Canova and Marcet (1995) show that an AR(1) model with region specific parameters captures sufficiently well the dynamics and leaves the residuals close to a normal white noise when the data is scaled in each period by the European average.

They also show that for this data set income inequalities are persistent, in the sense that there is very little evidence that the income of poor and rich regions will become more similar as time progresses, and that the estimated distribution of steady states is far from collapsing to a single point. Hence, their setup offers the natural ground for examining whether there is any tendency toward clustering. For $t = 1, \ldots, T, \quad i = 1, \ldots, N$ I use a model of the type

$$y_{it} = \alpha_i + \rho_i y_{it-1} + u_{it} \qquad u_{it} \sim N(0, \sigma_i^2) \tag{23}$$

where y_{it} is the per-capita income of each region relative to the European average. I also assume that $\beta_i = [\alpha_i, \rho_i]'$ can be represented as:

$$\beta_i = \beta^p + e_i^p \qquad e_i^p \sim N(0, \Sigma_p) \tag{24}$$

Given that N=144 I allow, at most, 6 groups (i.e. Q=5). For regional data there are very few indicators which can be used to order units according to the suggestions of recent growth theories. For example, no indicators of the average regional human capital (or its distribution) at the beginning of the sample is available, nor do I have regional measures of dispersions of income per-capita. Given that the sample covers the 1980's and that I am examining European regions belonging to EEC countries, I conjecture that differences along these dimensions are unlikely to provide relevant information to group units into convergence clubs [3].

Given these limitations, I search for clubs ordering the cross section according to: (i) the magnitude of per-capita income relative to the European average in 1979, with poor regions coming first; (ii) the magnitude of per-capita income relative to the national

[2] Roughly speaking the NUTS2 classification corresponds to regions. NUTS1 refers to larger territorial units (the "North", the "Centre" and the "South") while NUTS3 provides data at provincial level.

[3] As an informal check of this conjecture, I separatedly examined the case of regions in Italy and Spain, for which educational data are available (see Boldrin and Canova (1998) for a description of this data). I find that regional differences in average human capital and in the distribution of human capital are small and typically unrelated to the time path of income per-capita in the sample.

average in 1979, with poor regions coming first; (iii) the magnitude of locally scaled income per-capita in 1979 (Mediterranean regions and Ireland are scaled by their average and northern regions are scaled by their average), with poor regions coming first; (iv) the magnitude of the average share of per-capita income relative to the European average in the sample, with poor regions coming first; (v) the magnitude of the average growth rate of per-capita income in the sample, with regions growing slower coming first.

The first ordering attempts to capture the effect that initial conditions may have on the steady state distribution of income per-capita; the next two orderings try to verify whether geographical externalities (either at country or at south-north level) may be important to determine the position of a unit in the steady state distribution or its basin of attraction; the last two classifications attempt to study the importance of threshold externalities, here proxied by the size of the share of income per-capita in Europe or its growth rate. Note that if geographical externalities are important, any tendency toward convergence clubs that may appear with ordering (i), should be weakened or disappear with ordering (ii) or (iii).

Among these five orderings, I find that the first one maximizes the predictive density of the data. Given this ordering I identify three break points, corresponding to units 15, 23 and 120, and, consequently, four groups in the data. Within the first 15 units there are ten regions of Greece, four of Portugal and one of Spain (Extremadura); in the second group there are four regions of Greece, three of Spain and one of Italy (Calabria); finally, the last group includes regions from nine different countries but the majority are German (9) and Northern Italian (5). Note that the fourth and fifth orderings produce 4 and 3 groups, whose composition is very similar to these ones. Hence, the splitting that the algorithm produces is highly suggestive of the fact that European regions cluster into homogeneous groups along the poor-rich, south-north dimensions.

In figure 1 I provide graphical evidence of the existence of groups with the first ordering by plotting the predictive density as a function of the location of the break point, together with the predictive density obtained when there are no breaks (the dotted line). The first panel refers to the full sample, the next two panels to the two subsamples obtained separating units according to the first optimal splitting. To interpret the graphs note that the entries on the horizontal axis gives the location of the break and those on the vertical axis the value of the predictive density. Therefore, entry 23 on the horizontal axis in the first panel indicates that assigning units 1-23 to the first group and units 24-144 in the second gives a value of L^+ of 4943 (as compared to the value of 4863 when no breaks are allowed) Similarly, the second panel indicates the need of further splitting group 1 in two subgroups (1-15 and 16-23) and this split gives a value of L^+ of 679 (as compared to the value of 670 when no breaks are allowed). Finally, splitting group 2 in two subgroups (24-120, 121-144) gives a value for L^+ of 4325 (as compared to a value of 4272 when no breaks are allowed).

Table 1 presents the results of testing various hypotheses using the posterior odds ratio. Also reported are the prior odds ratio for each of the hypotheses of interest and $\hat{\pi}$, the minimum value of the prior probability on the null needed so that the data will not reject it. Three features of the table should be noted. First, the overall fit of a model with three breaks is significantly better than the one without breaks: the predictive density is substantially higher and a PO ratio favors the hypothesis of heterogeneities. Second, in going from one to three breaks, the relative improvement is nonmonotonic and, for the third break, the posterior odds ratio does not provide enough support for the alternative hypothesis that there are three breaks. Third, and as a consequence of the above, we need to have progressively weaker a-priori expectations on the null as the number of break points we are testing for increases.

While the statistical evidence in favor of three breaks is not overwhelming, economic differences appear to be relevant. I present estimates of β^p for the whole sample and for each of the four selected groups in table 2. It is clear that the four groups can be identified by both the value of the intercept and of the slope of the model (23). For example, the first group is characterized by very slow average persistence in relative income per-capita (low ρ^p) and below average mean intercept (low and negative α^p). At the opposite, the last group is characterized by higher average persistence and above average mean intercept (high ρ^p and positive α^p). Interestingly the central group, which contains the largest number of units, has a mean value for the persistence parameter which is higher than that of the last group.

The within group dispersion of hyperparameter estimates, varies substantially across groups. For example, differences in the persistence parameter are small in the second group (0.04) but large in the last one (0.64). For three of the four groups the dispersion of the persistence parameter within subgroups is substantially smaller than the dispersion obtained by (weakly) pooling together all units with an exchangeable prior, suggesting an overall reduction of the residual heterogeneity once groups are identified. For the last group the dispersion parameter is large, probably because the sample is small and there are few outliers (Dutch oil producing regions). In general, it appears that the last group is heterogeneous and requires a further subdivision. However, the procedure was unable to locate any further break in this group. Finally, except for the second group, there is no evidence that the dispersion of the coefficients around the mean is negligible, stressing the need to control for residual heterogeneities once groups are identified.

To summarize the features of the posterior distribution of the β, I report three economically interesting functions of the coefficients of the dynamic model: a scatter plot of speeds of adjustment to the steady state ($1 - \hat{\rho}_i$) against the initial condition y_{i0} for each of the four groups; the mean and the dispersion of estimated steady states for each group; and a long run mobility index.

With the scatter plots I am interested in verifying whether the magnitude of the slope between $1 - \hat{\rho}_i$ and y_{i0} varies with the group and, in particular, whether units with below average initial conditions adjust faster or slower to the steady state than units with above average initial conditions. Recall that the standard neoclassical growth model has the property that the speed of adjustment does not depend on the initial conditions. The second statistic provides information on the core question of this paper, i.e. whether the identified groups do cluster around different steady states. The mobility index, on the other hand synthetically measures, given a particular position in the initial income distribution, the likelihood of switching income classes in the long run (i.e. it measures the likelihood of "miracles and busts"). Such an index also highlights whether inequalities are persistent, a result which is of interest to policymakers concerned with, e.g., the evaluation of transfer programs to underdeveloped regions. In the exercise I consider only two classes (above and below average income at the beginning of the sample and in the steady states) [4] and the mobility index for the two states (long-run) Markov chain is calculated as $M = 1 - p_{11} - p_{22}$ where p_{ii} is the estimated probability of staying in the class where a unit starts, $i = 1, 2$. Notice that $-1 \leq M \leq 1$, with values greater than zero indicating mobility across the two classes and values less than zero supporting the idea that there is persistence of inequalities.

Figure 2 indicates that indeed there are striking differences in the relationship between speeds of adjustment and initial conditions of the four groups. While for the first two groups the slope is strongly negative (estimates are 0.7-0.9), the slope for the third group is still negative but smaller in magnitude, while the slope for the fourth group is slightly positive even though not significantly so. Notice also that there is a number of regions in the last two groups which have speeds of adjustment which are either negative or greater than one, indicating possible non-stationary or oscillatory posterior dynamics.

Table 3 confirms that the identified groups do constitute different convergence clubs. The means of the steady states are different across groups (given equal prior odds, the posterior probability that they are equal is negligible for every pair except the first two) while the dispersion of steady states around these means varies with the group. The economic significance of these differences is substantial. For example, the mean steady state of the first group is around 45% of the European average and the mean of the fourth group is about 15% above the European average. Also, the steady state distribution is far from collapsing for all but group 2 and there is a substantial reduction of the steady state dispersion once units are appropriately grouped.

Given these results, one would like to know what are the characteristics of the units belonging to each group. For example, one may be interested in knowing if there will be any mobility in the steady states ranking (relative to the initial conditions) or if club convergence occurs in a situation of immobility in the ranking. The mobility index for the

[4] Changing the threshold from the mean to the median do not change the qualitative features of the results.

whole sample, equal -0.24 , suggests a very weak tendency to transit from the position where the units start: the tendency to transit is much stronger for units which starts above the mean, while poor regions tend to stay uniformly poor, i.e. busts are more probable than miracles. The four groups clearly display different mobility characteristics. In the first group there is a strong tendency to stay in the low income class and in the second group there is complete immobility. The third group mirrors, with minor numerical differences, the tendencies of the whole sample but 67% of the units starting above the average end up below it in the steady state. The fourth group also shows a tendency to slump and about 50% of the units which started above average are expected to be below average in the steady states (curiously, most are French and German regions!).

Few interesting general economic conclusions can be drawn from the analysis. First, the income dynamics of initially poor regions tend to be different from those of the initially rich. Second, there is very little tendency for the poor to move up in the income distribution ladder while the initially rich may fall back into mediocrity. Third, and as a consequence of the above, the steady state distribution of income per-capita may become more polarized with few very rich units and the rest clustered in few groups below the average. Fourth, the low mobility in the income distribution ladder of the majority of poor and very rich units, confirms the results of Canova and Marcet (1995) concerning the persistence in inequalities in regional per-capita data.

Quah (1996a) has argued that once geographical externalities are taken into account the tendency toward convergence clubs weakens. Does this occur in our sample of regional data? The answer is partially positive. In Figure 3 I plot the predictive density as a function of the location of the break when regional income per-capita is scaled by national income per-capita and ordered according to the magnitude of the scaled initial conditions. There is evidence of only one significant break (producing two groups with units 1-93 and 94-141), but now the hyperparameters of the two groups are more similar. For example, the AR parameters has a mean of 0.597 in the first group and 0.713 in the second. Moreover, differences in estimated steady states are much smaller than those obtained when per-capita income is scaled with the European average and the dispersion around the two steady states is substantially reduced. Hence, there is some evidence that geographical and/or informational externalities are present: once these effects are taken into account the number of clubs is smaller and the economic differences among them significantly reduced.

5.2 OECD National Income per-capita

Following Canova and Marcet (1995), the model used to capture the time series characteristics of this data set is (23)-(24), where now N=21, time runs from 1951 to 1985 and at most 3 groups are allowed (i.e. Q=2). Contrary to the case of regional data, useful information to order units in the cross section dimension of the panel is available at the country level. Hence, I search for clubs ordering units according to: (i) the magnitude of the per-capita GDP relative to the OECD average in 1950, with units having poor initial conditions coming first, (ii) the magnitude of the average human capital in 1950, measured as in Barro and Lee (1994), ordering units increasingly in their average endowment of human capital, (iii) the dispersion of income distribution in 1950 (Gini index from the Luxemburg Income Study), with units displaying high dispersions coming first; (iv) the dispersion of the distribution of human capital in 1950, (measured as the sum of the percentage of the population with primary and university educations using Barro and Lee data), with units displaying high dispersions coming first; (v) a center-periphery classification of the world economy (G-3 first and then the rest), (vi) a geographically criteria with European nations first and rest of the world afterward and Mediterranean countries preceding northern European countries in the order. Note that in these latter two classifications income per-capita is scaled by OECD average and organized the according to geographical and/or neighborhood dimensions [5].

When one break is allowed the maximized value of L^+ for the six classifications is 2436, 2423, 2433, 2423, 2420 and 2415 respectively, suggesting that the predictive power of the model is maximized when units are ordered according to the initial conditions of income per-capita. Consistent with the results of Durlauf and Johnson (1995) the procedure prefers initial output over literacy rates as the most useful variable to identify breaks in the data. However, differences in L^+ across alternative classifications are small since the ordering of units in the cross section is very similar in at least four cases. That is, countries which have low initial income conditions also have low average initial human capital, a distribution of income with high dispersion and are geographically located in the "South"' of the developed world.

Given this ordering of the data, the posterior odds ratio establishes the presence of only one break in the cross sectional dimension of this panel, with a value of 0.979, given equal prior odds on the null of one group and the alternative of two groups. In figure 4 I plot L^+ as a function of the location of the break point for the best ordering together with the predictive density obtained in the case of no breaks (dotted line). The first group contains the five poorest units (Turkey, Portugal, Greece, Spain and Ireland) and the second group the rest.

Estimates of the hyperparameters for the two groups are $\beta^1 = [-0.162, 0.824]$ and $\beta^2 = [0.0004, 0.958]$, suggesting a much faster a-priori average rate of convergence toward

[5] Substitution of the size of the population holding 50% of national wealth for Gini indices and the sum of the inverse of the percentage of the population with primary and the inverse of the percentage of population with secondary education for the percentage of the population holding primary and university degrees does not change the results. The ordering obtained with these new indices are practically identical to those I use.

a pole of attraction which is below the OECD average for the countries in the first group. Note that pooling the two groups together produces estimates of β of $[-0.035, 0.881]$. The dispersion of estimates is small but non-negligible (in particular, the dispersion of estimates of the AR parameter is 0.02 in the first group and 0.05 in the second group) suggesting the presence of measurable heterogeneities within groups. In other words, it appeasrs that clustering is more prevalent than convergence even after optimally splitting the sample.

The posterior characteristics of the two clubs differ. The average posterior estimate of the steady state for the countries in the first group is -0.7647 and for the countries in the second group is 0.0498. This difference is statistically and economically large: in particular, it implies that there will be a permanent discrepancy in the average per-capita income of units in the two groups of about 60%. The dispersion of estimated steady states around these poles of attraction is smaller than the one obtained when all units are (weakly) pooled together. However, differences of about 15-20% in steady state income per-capita in each group are still possible. Finally, the mobility characteristics of the two groups are similar: apart for few exceptions, the ranking of units in the income distribution changes very little over time and countries which were poor at the beginning of the sample are still the poorest in the steady state. What is interesting about this last observation is the fact that there is no evidence that the economic boom which took place in Ireland in the 1990's and allowed the country to move up in the OECD income distribution ladder was forthcoming.

In sum, in agreement with what Quah (1996b) and Durlauf and Johnson (1995) have detected for a larger sample of countries, I find that clustering along the poor-rich dimension is prevalent in this data set. Moreover, countries which were initially poor were also those having below average initial human capital, large income and educational inequalities and were located in the "South" of the developed world. These initial characteristics are very persistent and produce polarization in the steady state distribution of income. The policy implications of these outcomes are striking: unless some major changes occur the initially poor will remain poor forever and they will tend to cluster around a basin of attraction which is substantially below the OECD average.

6 Conclusions

This paper describes a procedure to examine the likelihood of convergence clubs in the distribution of income per-capita. It proposed a unified approach to testing, estimation and inference when the number of groups, the location of the breaks and the ordering of units in the cross sectional dimension of the panel is unknown. Such an approach has a number of applications, apart from the one considered in this paper. For example, it could be used to examine the differential response of firms to monetary policy shocks or the international transmission of shocks across fixed and flexible exchange rate regimes. In general, the simplicity of the procedure, its easiness of implementation and the good properties it demonstrates in a simple Monte Carlo exercise make it a candidate to deal with the issue of grouping in a number of microeconomic and macroeconomic fields.

The procedure I suggest has its cornerstone in the predictive density of the data, conditional on the hyperparameters of the model. The use of predictive densities has a long tradition in Bayesian econometrics and provides a simple and appealing framework where interesting hypotheses can be verified. What is appealing about predictive densities is that, once hypotheses concerning the number of groups present in the data are examined, the location of the breaks, the best permutation in the data and the hyperparameters of the model can be easily estimated by simply considering the predictive density as function of the quantities of interest. Once the hyperparameters are selected, inference can be conducted in a standard Empirical Bayes fashion and the properties of functions of the posterior estimates of the coefficients of the model can be examined once we plug-in hyperparameters estimates in the appropriate formulas.

I search for clubs usign income per-capita data from European regions and OECD countries. I find that there are heterogeneities in European regional per-capita income and a tendency of the steady state distribution to cluster around four poles of attractions characterized by different dynamics, different posterior mean steady states and different mobility features. Similarly, OECD national per-capita income data presents two convergence clubs with poor countries clustering below the mean of the income distribution.

One word of warning in interpreting the results in light of theories of economic growth is useful. The paper has demonstrated that the *scaled* distribution of regional and national income per-capita shows a tendency to cluster around few poles of attraction when ordered according to the initial conditions of income per-capita and that, even within the endogenously selected groups, level convergence is a rare phenomenon. Clearly these results do not imply that the *unscaled* level of per-capita income shows these features and neither they have anything to say about the existence of a steady state distribution of per-capita income in levels or in growth rates. Furthermore, they do not suggest that one type of economic theory (endogenous growth) is to be preferred to another one (exogenous type) or viceversa, since both theories can generate outcomes which are consistent with the findings of the paper.

Codes for implementing the procedure are written in RATS4.2 and are available from the author on request.

Appendix

In this appendix I present the results of a Monte Carlo exercise designed to examine the properties of the testing procedure to uncover breaks and estimation approach for the hyperparameters with data displaying the same statistical properties and sample sizes similar to those considered in section 5. For this reason I generate times series for N=144 units, each of length T=13, and assume that the data generating process is:

$$\begin{aligned} y_{it} &= \alpha_i + \rho_i y_{t-1} + e_{it} && (25)\\ \beta_i &= \beta^1 + u_i^1 \quad u_i^1 \sim N(0, \Sigma_1) \text{ if } \quad \text{i} \leq 50 && (26)\\ \beta_i &= \beta^2 + u_i^2 \quad u_i^2 \sim N(0, \Sigma_2) \text{ if } \quad 51 \leq \text{i} \leq 144 && (27) \end{aligned}$$

where $\beta_i = [\alpha_i, \rho_i]$, $\beta_1 = [0.3, 0.8]$, $\beta_2 = [-0.3, 0.4]$, $\Sigma_1 = \text{diag}(0.052, 0.255)$, $\Sigma_2 = \text{diag}(0.102, 0.155)$, $var(e_{it}) = 0.1$ if $i \leq 51$ and $var(e_{it}) = 0.15$ otherwise, and I assume that the initial conditions satisfy: $y_{i,0} \sim U[-0.10, 0.10]$.

On the panel of simulated data I estimate both AR(1) and AR(2) models for each unit $i = 1, \ldots, 144$ and apply the testing procedure to examine whether there is a break in the cross section using data in the order I have generated it. The posterior odds ratio, giving equal chance to the possibility that there is one break and there are no breaks, is 0.9771 for the AR(1) and 0.9846 for the AR(2). The hypothesis that there is a further group produced a posterior odds ratio of 1.0124 and 1.0165 for the two models, confirming the presence of one break only. Figure A.1 plots L^+ as a function of the location of the break $\bar{i}$ for the AR(1) model (the first panel) where $h_1^1 = 10$, $h_2^1 = 135$. The peak is achieved at $\bar{i} = 50$, implying the presence of two groups comprising units 1-50 and 51-144, and there are no other peaks within the range I explore. Repeating the experiment 100 times, I find that in 100% of the cases the posterior odds ratio reveals the presence of two groups in the cross section (with an average value of 0.9797) and the predictive density is maximized in 75% of the time at $\bar{i} = 50$ (96% of the times for $\bar{i} \in [49, 51]$). The average posterior odds ratio for the hypothesis that there are three groups in the generated data is 1.0114.

Next, I conducted three experiments: first, I randomized the order of the units within the two groups before estimation is undertaken. This did not change any of the results confirming that, absent any information on the appropriate ordering of the data, the number of actual permutations to be tried is substantially less than N!. Second, I reshuffled the entire cross section, taking the first 20 units of the time series and putting them last. In this case the ordered data displays three groups with breaks at i=30 and i=124. Estimating an AR(1) model on the data, the posterior odds ratio finds 2 breaks, and the predictive density is maximized at $\bar{i}_1 = 30$ (see plot as function of $\bar{i}$ in panel 2 of figure A.1). The pattern displayed by the predictive density in this case is very well known from the break point literature (see Bai (1997) or Hansen (1997b)) and conveys information suggesting that there are three groups in the sample. In fact, conditional on having a break at $\bar{i}_1 = 30$, the posterior odds ratio for a second break is 0.9933 and the location of the break is $\bar{i}_2 = 124$. Repeating this experiment 100 times I find that the average posterior odds ratio for the hypothesis of no breaks against 2 breaks is 0.991, that the neighborhood $\bar{i}_1 \in [29,\ 31]$ is identified as the first break point 80% of the times and that the neighborhood $\bar{i}_2 \in [123,\ 125]$ is identified as first break point in 12% of the cases (average PO ratio for the hypothesis of one break is 0.992). Conditional on having a break at $\bar{i}_1 = 30$, the latter neighborhood is identified as the second break in 80% of the cases (average PO ratio for the hypothesis of two breaks is 0.993).

The design of this second experiment also allows to examine the power of the test when the cross-sectional data is not properly ordered. That is, suppose that the DGP is such that there are only two groups in the data, but an econometrician has available data ordered in a way which may be different from the correct one. Would the procedure be able to recognize the optimal permutation of the units in the cross section, select the correct ordering with only two groups and find the location of the break point? To provide an idea of the properties of the approach in this case I assume that there is a break at unit 56 and reshuffle blocks of 28 units, so that I allow 5! combinations (120 trials) over which to search for the optimal ordering. Figure A.2 plots L^+ and the selected location of break point as a function of the permutation $m = 1, \ldots 120$. There is a plateau in L^+, corresponding to the 12 permutations which correctly put the first two blocks first and the next three last and for the remaining cases L^+ declines slowly. Notice also that for all permutations L^+ is substantially higher than the likelihood under the null (the dotted line in the graph). Also, the procedure correctly identifies the location of the break in those 12 cases when L^+ is maximized.

Finally, I study the properties of the testing procedure when the cross section is homogeneous (the parameters for the two groups are those for β_1, Σ_1 and e_1). The posterior odds ratio for the hypothesis of 0 versus 1 breaks gives a value of 1.0001 and the predictive density as a function of $\bar{i}$ produces a plateau with very little difference between the minimum and the maximum values (see the third panel of figure A.1). Replicating the experiment 100 times I find that the average posterior odds ratio giving equal prior probabilities to the null and the alternative of two groups, is 0.9997 with several cases giving a value greater than 1. The distribution of the break point is practically uniform in the interval $[10, 135]$, confirming the results obtained with one experiment only.

Estimates of the hyperparameters of the model are, in general, biased. In particular, the average values across 100 experiments, in the baseline case are $\beta^1 = [0.3614, 0.6931]$, $\beta^2 = [-0.3660, 0.2682]$, indicating that estimates of ρ are downward biased and, as a consequence, estimates of α are upward biased. This appears to be due to the small time series size of each cross section: if I increase the sample size to T=36 (the size of the time series with OECD country data), most of these biases disappear. The variances of all the estimated coefficients are also upward biased by 25-50%. Again, the bias drops to 10-15% when T=36. When there are three groups in the cross section results are similar even though average estimates of the hyperparameters of the third group are more biased, probably because of the small number of units in this group. Finally, when the cross section is homogeneous, average estimates (across replications) are still biased but by a smaller amount (average β = [0.3816, 0.7372]) while variances of the estimated coefficients are practically identical to those obtained in the baseline case.

Overall, the results indicate that the testing procedure has reasonable size and power properties against the particular alternative I consider. It also appears to be able to identify multiple groups and the location of the breaks with sufficient precision, even when the data is not correctly ordered. However, since the posterior odds ratio appears to be slightly biased downward when there are no heterogeneities, a conservative strategy would be to choose a critical value for the posterior odds ratio which is slightly less than one. Alternatively, one could choose to give to the null hypothesis a slightly higher probability to start with, say 50.5%. Also, one should be aware that when there are multiple groups, the posterior odds ratio may be very close to one, given equal prior odds, especially if the peaks in the predictive density are not very sharp.

When the time series size of each cross section is small, estimates of the autoregressive parameters are downward biased and averaging over the cross section does not help since estimates of all the units are downward biased (see also Pesaran and Smith (1995)). When the size of each cross section is greater than 30, estimates of the hyperparameters obtained by maximizing the predictive density of the data are sufficiently precise while estimates of the dispersion of the prior distribution are still significantly biased.

References

[1] Azariadis, C and Drazen, A. (1990), "Threshold externalities in economic development", *Quarterly Journal of Economics*, 105, 501-526.

[2] Bai, J. (1997), "Estimation of multiple breaks one at a time", *Econometric Theory*, 13, 315-352

[3] Barro, R and Becker, G. (1989) "Fertility Choice in a Model of Economic Growth", *Econometrica*, 57, 481-501

[4] Barro, R and Lee, (1994) Data set for a panel of 138 Countries, manuscript.

[5] Boldrin, M. and Canova, F. (1998), "Economic Policies to Foster Growth", manuscript.

[6] Ben David, D., (1994), "Convergence Clubs and Diverging Economies", CEPR working paper 922.

[7] Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer and Verlag.

[8] Canova, F. and Marcet, A. (1995), "The Poor Stay Poor: Non-convergence Across Countries and Regions", CEPR working paper 1215.

[9] Carlin, B. and Gelfand, A. (1990), "Approaches for Empirical Bayes Confidence Intervals", *Journal of the American Statistical Association*, 85, 105-114.

[10] Desdoigts, A. (1998), "Pattern of Economic Development and the Formation of Clubs", Universite' d'Evry-Val D'Essonne, EPEE, working paper.

[11] Durlauf, S. and Johnson, P. (1995), "Multiple Regimes and Cross Country Growth Behavior", *Journal of Applied Econometrics*, 10, 365-384.

[12] Efron, B. (1996), "Empirical Bayes Methods for Combining Likelihoods", *Journal of the American Statistical Associations*, 91, 538-565 (with discussion).

[13] Forni, M. and Reichlin, L. (1997), "Let's get real: Dynamic Factor Analytical Approach to disaggregated Business Cycles", *Review of Economics Studies*, 65, 453-474.

[14] Galor, O. (1996), "Convergence? Inference from Theoretical Models", *Economic Journal*, 106, 1056-1069.

[15] Galor, O. and Zeira, J (1993), "Income Distribution and Macroeconomics, *Review of Economic Studies*, 60, 35-52.

[16] Hansen, B.E. (1997a), "Sample Splitting and Threshold Estimation", Boston College, manuscript.

[17] Hansen, B.E. (1997b), "Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference", Boston College, manuscript.

[18] Hartigan, J. (1975) *Clustering Algorithms*, Wiley, New York, N.Y.

[19] Maddala, G.S. (1991), "To Pool or not to Pool: That is the Question", *Journal of Quantitative Economics*, 7(2), 255-264.

[20] Maddala, G.S. and Wu, S. (1996), "Cross Country Growth Regression: Problems of Heterogeneity, Stability and Interpretation", Ohio State University, manuscript.

[21] Mardia, K., J. Kent and J. Bibby (1980) *Multivariate Analysis*, Academic Press

[22] Morris, C.N. (1983), "Parametric Empirical Bayes Inference: Theory and Applications", *Journal of American Statistical Association*, 78, 47-59.

[23] Pesaran, H. and Smith, R. (1995) "Estimating Long Run Relationships for Dynamic Heterogeneous Panels", *Journal of Econometrics*, 68, 79-113.

[24] Paap, R. and van Dijk, H. (1994), "Distribution and Mobility of Wealth of Nations", Erasmus University, manuscript.

[25] Phillips, P.C.B. and Ploberger, W (1994), "Posterior Odds Testing for Unit Root with data-based Model Selection", *Econometric Theory*, 10, 774- 808.

[26] Polasek, W. and Ren, L. (1997), " Structural Breaks and AR modelling with marginal likelihoods", University of Basel, manuscript.

[27] Ploberger, W., Kramer, W. and Kontrus, K. (1989), "A new test for the structural stability in the linear regression model", *Journal of Econometrics*, 40, 307-18.

[28] Quah, D. (1996a), "Convergence Empirics across Economies with some Capital Mobility ", LSE, Center for Economic Performance, Discussion paper 257.

[29] Quah, D. (1996b), "Regional Convergence Clusters across Europe", CEPR working paper 1286.

Hypotheses	Location Break(s)	Prior odds ratio	Posterior odds ratio	$\hat{\pi}$
$H_0 : Q = 0$		1.013	0.546	0.646
$H_1 : 3 \geq Q > 0$				
$H_0 : Q = 0$		1.013	0.991	0.522
$H_1 : Q = 1$	23			
$H_0 : Q = 1$	23	1.013	0.994	0.521
$H_1 : Q = 2$	23, 120			
$H_0 : Q = 2$	23, 120	1.013	1.005	0.518
$H_1 : Q = 3$	23, 120, 15			

Notes: The column labelled $\hat{\pi}$ reports the minimum prior probability on the null needed so that the data will not overturn it. In each case the null has prior probability equal to 0.505.

Table 1: Hypotheses testing for the presence of groups

	α^p	σ^2_α	ρ^p	σ^2_ρ	$\sigma_{\alpha,\rho}$
Overall	-0.086	0.060	0.725	0.298	-0.071
Group 1 (units 1-15)	-0.598	0.102	0.251	0.155	-0.031
Group 2 (units 16-23)	-0.368	0.019	0.534	0.048	-0.042
Group 3 (units 24-120)	-0.032	0.0004	0.686	0.193	-0.008
Group 4 (units 121-144)	0.116	0.052	0.629	0.641	0.023

Notes: The table reports ML- type II estimates of the hyperparameters obtained maximizing the predictive density of the data, viewed as function of the hyperparameters.

Table 2: Estimated values of the hyperparameters of the prior

Table 3: Posterior Estimated Steady States

Sample	Mean	Dispersion
Overall	-0.2712	0.6234
Group 1	-1.3171	0.3883
Group 2	-0.6369	0.0751
Group 3	-0.1390	0.1468
Group 4	0.2922	0.2308

Notes: The steady state for each region is computed as $\lim_{T\to\infty} \frac{\alpha_i*(1-\rho_i)^{T+1}}{1-\rho_i} + \alpha_i^T y_{i0}$ where α_i and ρ_i are posterior estimates. The column named "Dispersion" reports the standard deviation of steady states around the mean value.

Table 4: Mobility Index

	Overall	Group1	Group 2	Group 3	Group 4
M	-0.24	-0.41	-0.50	-0.18	0.00
P_{11}	0.83	0.91	1.00	0.85	0.00
P_{12}	0.17	0.09	0.00	0.15	0.00
P_{21}	0.59	0.00	0.00	0.67	0.50
P_{22}	0.41	0.00	0.00	0.33	0.50

Notes: The M statistics is given by $M = 1 - P_{11} - P_{22}$. P_{11} is the probability that the unit starts below average and ends up below average in the steady state, P_{22} is the probability that the unit starts above and ends up above average in the steady state, P_{12} and P_{12} are the probabilities that the unit transits from a state to the other. In the case the group is unbalanced, so that all units in the group are initially in one income class, the statistics M is computed as $M = 0.5 - P_{ii}$ where P_{ii} is the diagonal value different from zero.

Figure A.1: Simulated Data

Predictive Densities

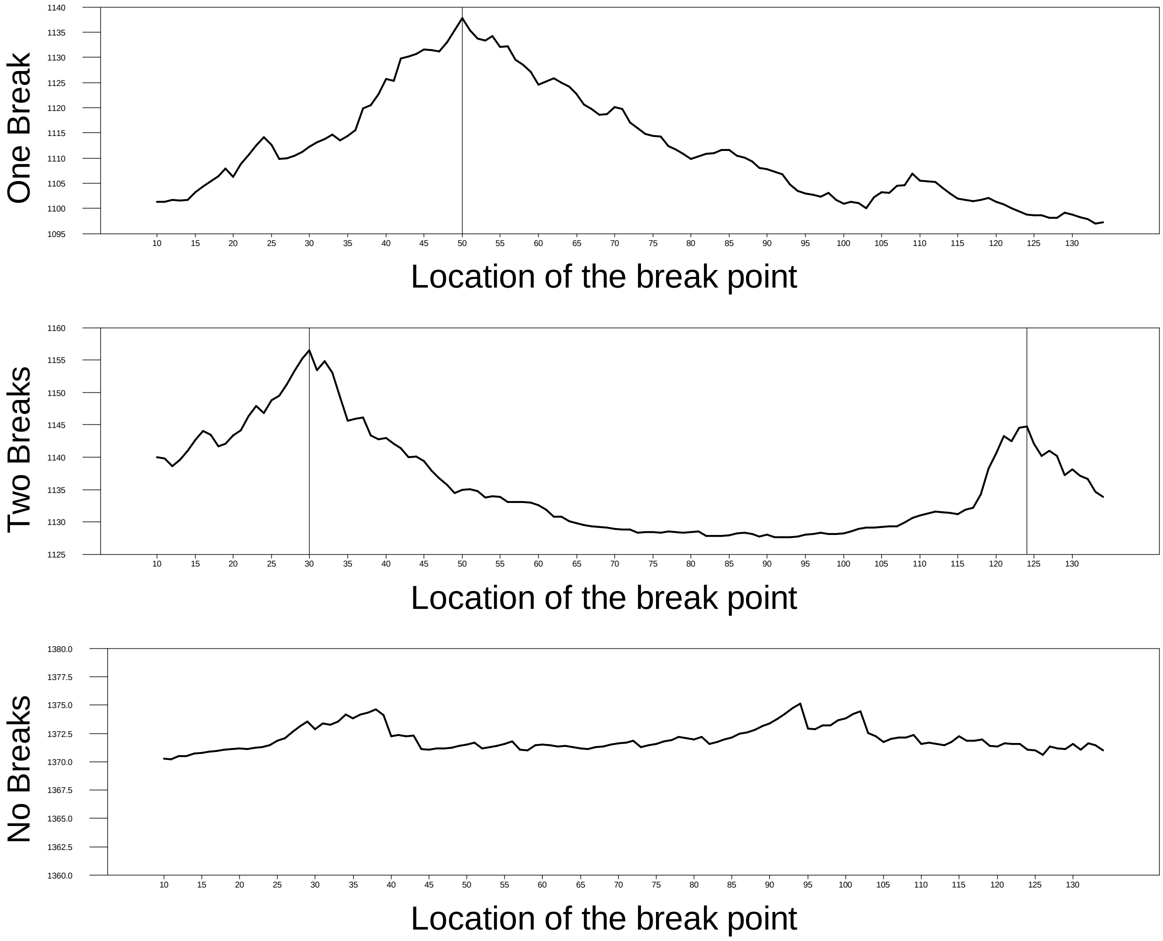


Figure A.2: DGP with one break

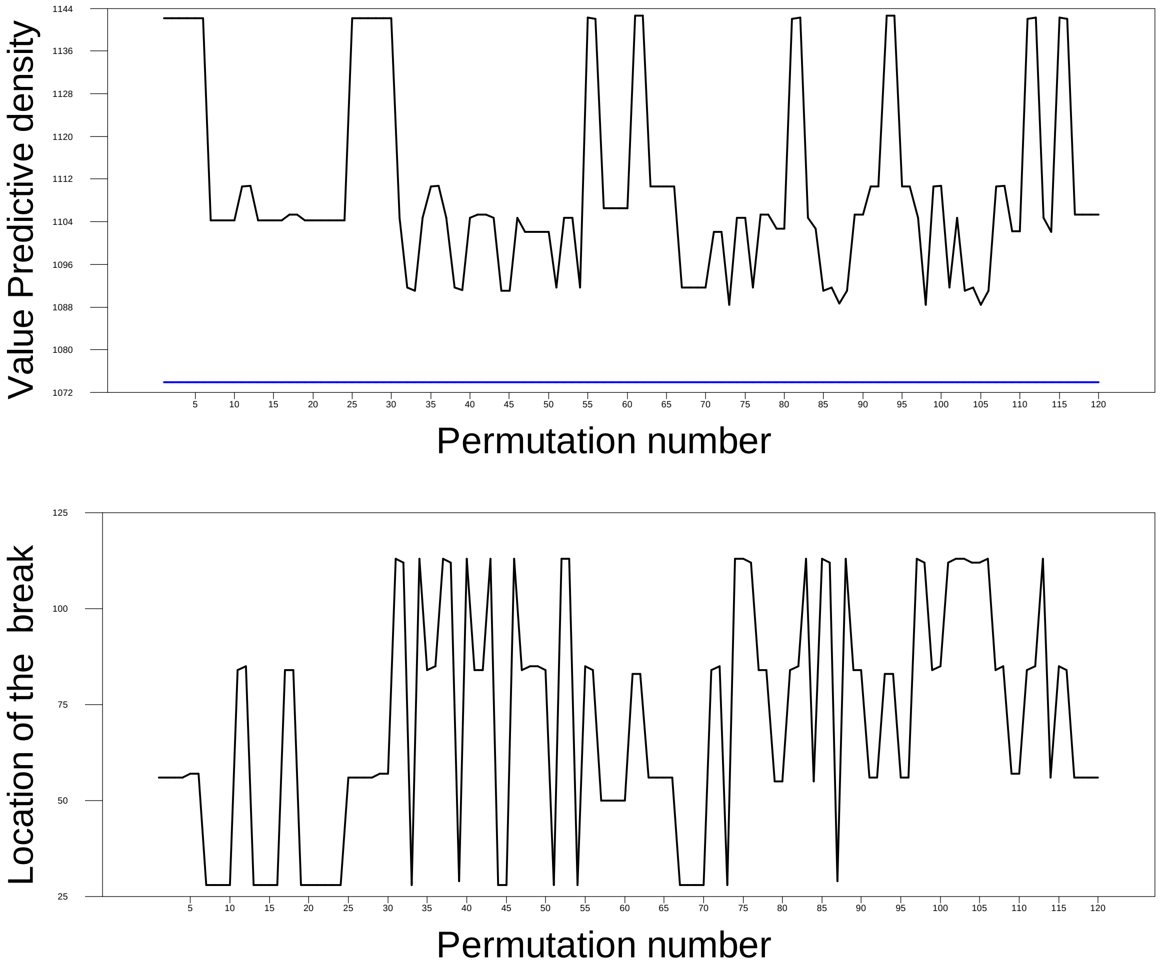


Figure 1:European Regional Data, European Scaling

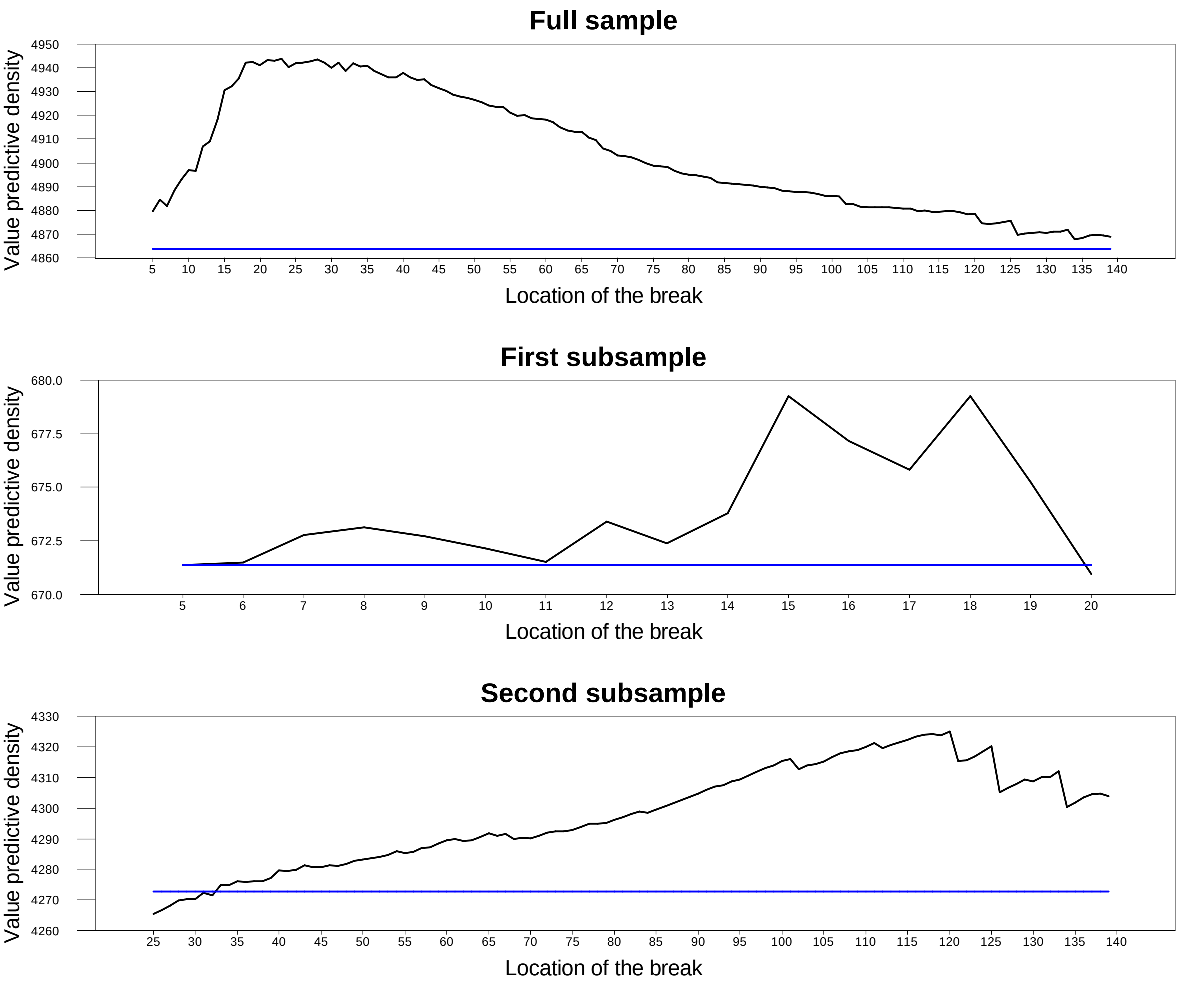


Figure 2: European Regional Data

Relationship Initial Conditions/Adjustement Rates

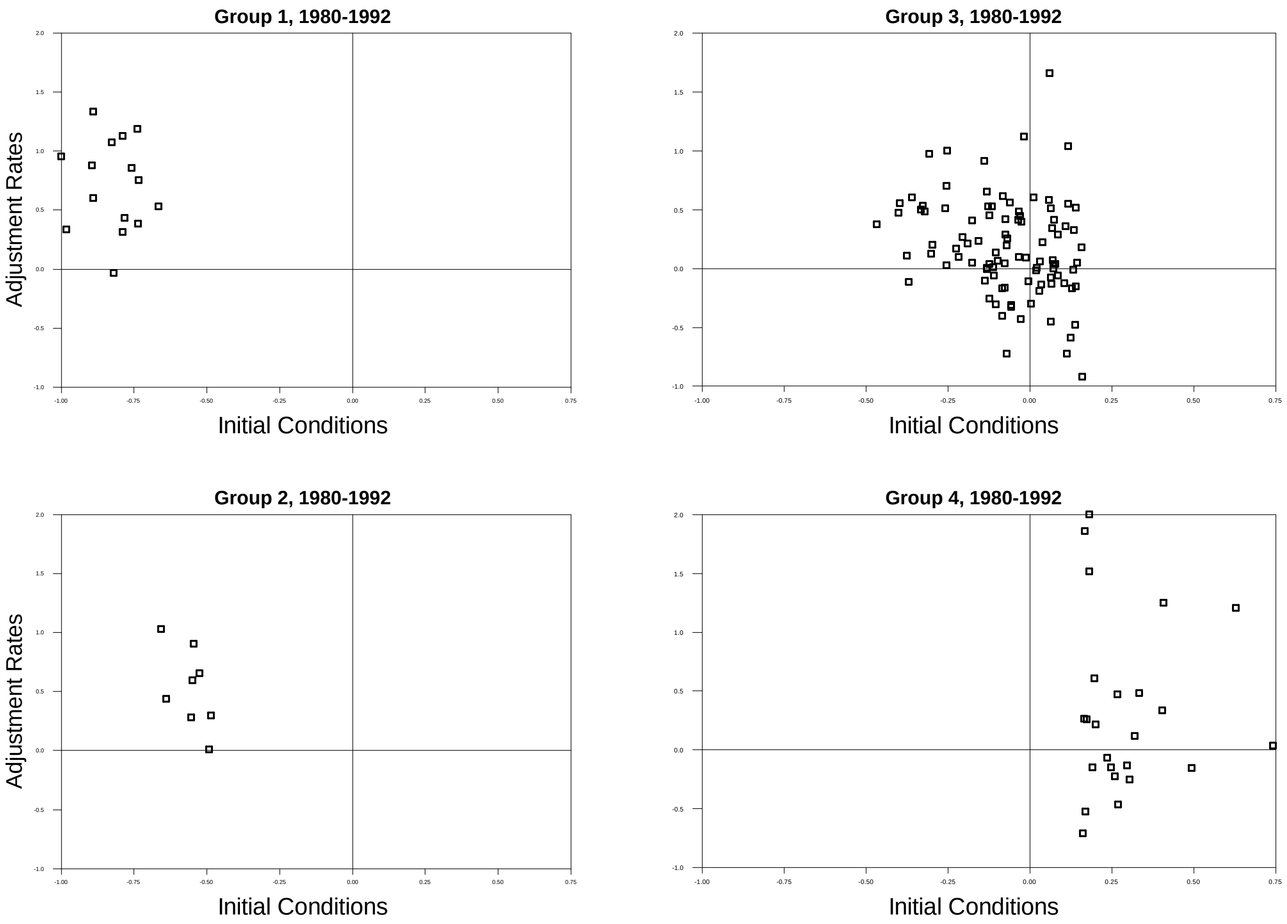
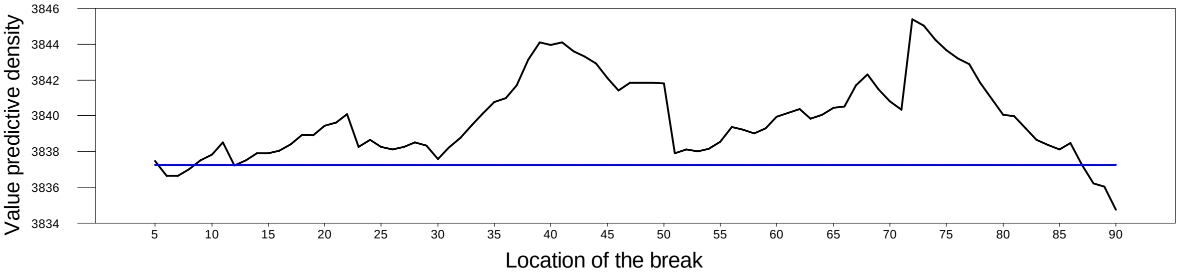


Figure 3:European Regional Data, National Scaling

Full sample

First subsample



Second subsample

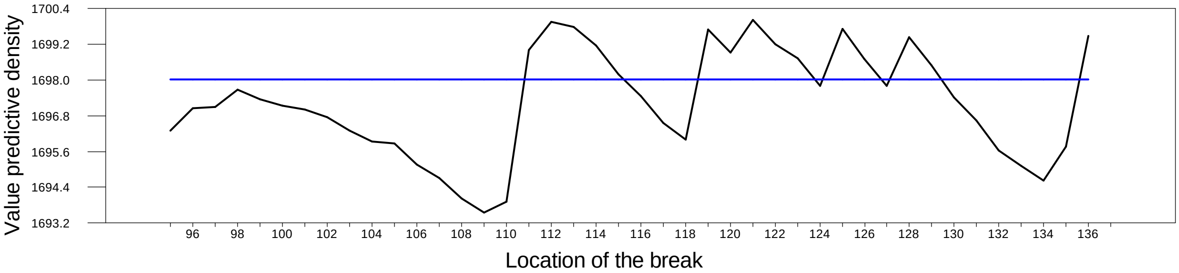


Figure 4: OECD National Data