The Poor Stay Poor: Non-Convergence across Countries and Regions

Fabio Canova *
Universitat Pompeu Fabra, Universita' di Modena and CEPR
and
Albert Marcet
Universitat Pompeu Fabra, CREI

First Version, October 1995
Current Version: June 1999

April 19, 2000

*We would like to thank two anonymous referees, Manolo Arellano, Andrew Barnard, Steve Durlau Paul Evans, Thryphon Kollintzas, Ana Lamo, Robert Litterman, Teresa Garcia-Mila, Frank Portiei Danny Quah and Xavier Sala-i-Martin the participants of seminars at Ohio State University, Bosto College, European University Institute, Universitat Pompeu Fabra, Universita' di Catania, Universit di Modena, IGIER, University of Aarhus, LSE, University of Southampton, University of Manchestei University of Bern, CEMFI, the CEPR Conference 'Empirical Macroeconomics: What Methods an for What Problems', Bruxelles, 1995; the Conference 'Growth and Business Cycles in Europe, Hydr Greece, 1996; the 1995 SEDC Meetings, Barcelona; the 1996 Winter Meetings of the North America Econometric Society, San Francisco; the 1996 European Meetings of the Econometric Society, Istambu the conference "Macroeconometric Modelling", Paris 1997, for very helpful comments and suggestion and Rosa M. Fernandez for help in organizing the data. An early version of this paper was prepared fc the FEDER project 'The Effects of the Unified Market on Regional Development' administered by th Institut d'Anàlisi Econòmica, Universitat Autònoma de Barcelona. Financial support from DGICYT CIRIT and CNR grants, as well as from CREI is gratefully acknowledged. Part of Marcet's research wa undertaken while he was a visitor at the Federal Reserve Bank of Minneapolis.

Abstract

We study the issue of income convergence with a Bayesian estimator. Our approach permits the estimation of different steady states and speed of adjustment for each cross sectional unit, while pooling information from other units. When this feature is allowed, we find that the adjustment of each unit to (its own) steady state is much faster than previously estimated but that cross sectional income inequalities will only be reduced by a small amount by the passage of time. We argue that the slow convergence rate to a common level of per-capita income found in the convergence literature is due to a 'fixed effect bias'. The cross country distribution of the steady states is largely explained by the cross country distribution of initial conditions and not by standard conditioning variables.

Key Words: Convergence, Income Inequality, Persistence, Panel Data, Prior Distribution.

Jel classification numbers: C11, C23, D90, 047.

Era bella, di una dura bellezza bruna che nessuno notava, perche' troppo frequente in quell'ambiente e in quella epoca.
Marguerite Yourcenar

1 Introduction

The issue of convergence of per-capita incomes across economic areas is an old one. Are income inequalities disappearing as time goes by? Do poor regions stay poor?. The issue, it seems, is whether the initial inequality of income is expected to persist or not.

This issue has been placed at the forefront of economic research in the last few years, for example, by Barro and Sala-i Martin (BS) (1991) and (1992). They analyze the available data with cross-section regressions and conclude that convergence obtains when using a sample of countries, US states or even European regions roughly at the very slow rate of 2% a year. In the case of US states or European regions, convergence is to a common level of per-capita income, but countries appear to converge to a common steady state only after conditioning with proxies for human capital and government policy. A large literature has ensued, exploring these issues in different data sets and with different statistical methods, but the main conclusions of BS have been, by and large, confirmed[1]. Roughly speaking, these results support the view that, as long as countries follow "adequate" policies on human capital accumulation, size of government sector, etc., differences in per-capita income between economic areas will slowly disappear as time goes by. Also, it supports the view that conditioning variables do not affect income of regions in the long run.

Typically, this literature attempts to explain income growth by aggregating growth rates over the sample period, and then performing a cross-section regression with *one* observation per unit (either country or region). Such an approach is problematic for three reasons: first, it wastes information, since unit-specific time variations in growth rates are ignored in the estimation process; second, it prevents the estimation of a steady state for each unit separately, which causes a number of conceptual and econometric distortions; third, it forces the use of a definition of convergence that is not related directly to the idea of persistence of inequality that we describe in the first paragraph.

In this paper we provide a formal definition of convergence that allows us to analyze whether income inequalities are persistent. We then propose a Bayesian procedure to estimate the speeds of adjustment and the steady states that uses the information available for all periods and all cross sectional units. Our prior is based on the belief that the parameters of the statistical model in different units have "similar" but not necessarily

[1] See, for example, Mankiw, Romer and Weil (1992), Barro and Lee (1994), Sala-i-Martin (1995) or Durlauf and Johnson (1995). The main conclusions and the estimates are even consistent when the analysis is performed for regions which are under very different political and economic system than Europe or the US (see, e.g. the case of China in Rivera-Batiz (1993) or the one of Japan in Shioji (1993)).

identical values. This allows an efficient use of all information without imposing unrealistic assumptions that may cause various types of biases. Once steady state estimates are obtained for each unit we can explore, in a second step, whether they are the same and, if not, what variables determine the cross sectional distribution of steady states. In particular, we can examine whether, over the cross-section, initial conditions matter for estimated steady states income and, in this manner, perform a meaningful test of persistence in inequality. This estimation/testing strategy is applied to two data sets: yearly per capita income of European regions and of European OECD countries.

Three major findings arise from our analysis:

- The average estimate of the speed of adjustment is much higher than that found in the literature - on average about 11% for countries and 23% for regions - with each unit converging to its own steady state. These estimates imply a much more reasonable capital share in a neoclassical production function (of the order of 0.10-0.30).
- The hypothesis that the steady state is the same for all cross sectional units (unconditional convergence) is rejected for both regions and countries.
- The initial position in the income distribution is, by far, the most important determinant of the position in the distribution of estimated steady state income. Poorer regions and countries stay, on average, poor; over time, income differences are reduced only by a small amount.

Our study also shows that, when the prior forces all parameters to be exactly equal in all units - a case rejected in formal testing, but the one that most closely resembles the cross-sectional approach of BS - a systematic distortion emerges that causes the average estimated speed of adjustment to be biased downward and, surprisingly, of the order of 1 or 2%. Thus, we explain the previous estimates as arising from a fixed effects bias, well known in the panel data literature (see e.g. Hsiao (1986) and, more recently, with a different flavor, Pesaran and Smith (1995)); such a value is mechanically obtained when observations from heterogeneous units are pooled as if their data generating process were the same.

Our work is linked to a number of papers in the literature. Quah (1993)-(1994) has used a non-parametric procedure to examine the evolution of income distributions across time; he provides descriptive statistics, but no formal testing of the importance of initial conditions. We share with Quah the preoccupation for exploiting the information in all periods, as well as the use of per-capita income scaled by the average (over the cross section) per-capita income. On the other hand, we share with BS the use of a tightly parameterized model which allows for testing of hypotheses. Our definition of convergence is related to that of Bernard and Durlauf (1996) in that it focuses on the limiting behavior of per capita income. Parente and Prescott (1993) analyze informally the data and also argue that the evidence is consistent with persistence of inequality. The point that the average speed of adjustment may have been underestimated by BS is also made by Evans (1997) and Caselli, Esquinel and Lefort (1996) using standard panel data estimators.

The rest of the paper is organized as follows. Section 2 provides a definition of convergence and links it to those existing in the literature. Section 3 discusses the statistical model, the estimation and testing strategy. Section 4 describes the data. Section 5 comments on estimates of the speed of adjustment to the steady state, tests the equality of estimated steady states and studies whether income inequalities are persistent. Section 6 examines whether sources of misspecification and econometric biases may affect the results. Section 7 concludes.

2 A Definition of Persistence in Inequality

We are interested in studying the following issue: *is there a tendency for the income of initially poor units to become similar to the income of initially rich units as time passes? or, should we expect that the poor stay (relatively) poor?*. In the former case we would say there is convergence, in the latter that there is persistence of inequality. More formally, we would say that there is persistence of inequality if there is little mobility across income levels and the differences between, say, the first and last decile do not disappear as time goes by.

To properly state the issue at stake, we first provide a definition that most closely formalizes the above ideas. We assume that observations are collected across units and time. The evolution of per-capita income for all units is determined by a doubly indexed stochastic process $\{Y_t^i\}$, where $i \in I$ indexes units, and $t = 0, 1, ...$ indexes time. The set I can be the first n integers, the unit interval, etc. The initial values $\{Y_0^i\}_{i\in I}$ are realization of a random variable with a, possibly non-degenerate, distribution. It is convenient to study (the log of) each unit's per-capita income relative to the aggregate, i.e. $y_t^i = \log\left(\frac{Y_t^i}{Y_t}\right)$, where Y_t represents the aggregate per-capita income over all units at each t[2]; in section 3 we will argue that modelling this variable has some advantages for both theoretical and econometric reasons.

Let $w^i \equiv \lim_{t\to\infty} E_0 y_t^i$, where the limit is assumed to exist and E_0 is the expectation operator conditional on $\{y_0^i, X^i\}_{i\in I}$. In words, w^i represents the best guess about the long run mean of a given region, given the initial condition and some additional characteristics of the unit X^i. Somewhat loosely, we will refer to w^i as 'the steady state'.

Definition 1 $\{Y_t^i\}$ *displays unconditional persistence of inequality if there is function f such that*

$$E\left(w^i|y_0^i\right) \equiv f(y_0^i) \tag{1}$$

and f is increasing.

[2] Note that Y_t is total income of all regions divided by total population over all regions.

Definition 2 $\{Y_t^i\}$ *displays persistence of inequality, conditional on variables* X_i *if there is a function* f *such that*

$$E\left(w^i|y_0^i, X^i\right) = f(y_0^i, X^i) \tag{2}$$

and $f(., X^i)$ *is increasing for all possible values of* X^i.

In both definitions, the expectation is taken with respect to the distribution across units. The first definition implies that initially rich units are expected to stay relatively rich, while the second definition looks at persistence of inequality among regions with the same characteristics X^i.[3]

Because we measure w^i in expectations (as t gets large), we concentrate on differences in units' income that persist through time on average, i.e. income differences due to temporary shocks are disregarded. Also, since E in (1) and (2) is the average over all units, it can happen that there is persistence of inequality even though some initially poor units become very rich (economic miracles, according to the terminology of Parente and Prescott (1993); what we require is that, on average over units, the long run mean is lower for initially poor units.

A corresponding definition of unconditional (conditional) *convergence* is the following:

Definition 3 $\{Y_t^i\}$ *converges unconditionally (conditionally) if the function* f *in (1) (the function* $f(., X)$ *in (2)) is equal to zero (the derivative if* $f(., X)$ *is equal to zero).*

It is possible to find stochastic processes that display no convergence and no persistence of inequality; this will happen whenever the limit w^i is not well defined, or when f is decreasing.

2.1 An example to compare definitions.

There are many convergence definitions available in the literature each of them meaning different things and focusing on different aspects of the distribution of income per-capita. Here we show, by means of an example, how our definition relates to other convergence definitions. A more verbal discussion appears in the summary below. We will see that other definitions do not always allow for the kind of distinction we want to study. Consider the following model:

$$y_t^i = \nu y_0^i + \rho\, y_{t-1}^i + \epsilon_t^i \tag{3}$$

for $i = 1, ...n$, where ν, ρ are given constants, $\{\epsilon_t^i\} \sim i.i.d(0, \sigma_\epsilon^2)$ across time and units, and the distribution of initial conditions $\{y_0^i\}_{i=1}^n$ is independent of the ϵ's.

[3] There may be some confusion because we use 'unconditional' in the first definition referring to the conditional expectations E_0 that define w^i. The term 'unconditional' is in the sense of BS, i.e., in the sense of not conditioning on the characteristics X^i.

This model allows for initial conditions to influence the future in a permanent w through the parameter ν; the parameter ρ captures the dependence of y_t^i on past shoc and it is another source of dependence on initial condition. This example is chosen for tractability, because it allows to make the comparison in a simple way and it is not t model we will later estimate.

Clearly, if $|\rho| < 1$,

$$w^i \equiv \lim_{t \to \infty} E_0\left(y_t^i\right) = \lim_{t \to \infty} \frac{(\nu - \rho^t(\nu + \rho - 1))y_0^i}{(1 - \rho)} = \frac{\nu y_0^i}{1 - \rho} \equiv f(y_0^i). \qquad ($$

where $(1 - \rho)$ is the speed of adjustment of $E_0\left(y_t^i\right)$ to the limit w^i.

In the rest of this section, we determine for what values of ρ and ν we have convergen in the process (3) under alternative definitions. In order to obtain an unambiguous answ about whether there is persistence of inequality or convergence according to our definitic we assume that $\nu \geq 0$ and $-1 < \rho < 1$ for the rest of this section[4].

2.1.1 Our Definition

It is clear from equation (4) that, if $\nu > 0$, there is persistence of inequality, but if $\nu =$ there is convergence in the sense of Definition 3 above. The fraction $\frac{\nu}{1-\rho}$ is the proporti of initial income of unit i that has a permanent effect on this unit's income; inequality reduced as time goes by if this fraction is less than one.[5]

To summarize, convergence obtains when we have $\nu = 0$; in all other cases there persistence of inequality. This delivers the first row of table 1.

2.1.2 σ-convergence

Let $\Sigma_t \equiv \frac{1}{n} \sum_{i=1}^{n} \left(y_t^i\right)^2$ be the dispersion of y_t; σ-convergence obtains if $\Sigma_t \leq \Sigma_0$ for t lar enough (see, e.g. Quah (1996) and Barro and Sala-i-Martin (1992)). In the example (: we have:

$$y_t^i = \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j}^i + y_0^i \left(\nu \sum_{j=0}^{t-1} \rho^j + \rho^t\right)$$

[4]As we explained before, there are processes for which there is neither persistence of inequality 1 convergence. This happens in example (3) if *i)* $|\rho| < 1$ and $\nu < 0$, or if *ii)* $\rho \leq -1$. In case *i)* the funct: f as given by (4) is strictly decreasing; the model will satisfy the biblical prophecy 'the last will be fir: In case *ii)*, if $\rho = -1$, there is a cycle in the sense that $E_0(y_t^i) = E_0(y_{t-1}^i)$; and if $\rho < -1$ and $y_0^i >$ then, as $\lim_{t \to \infty} y_{2t}^i = \infty$ (i.e., for t even) and $\lim_{t \to \infty} y_{2t+1}^i = -\infty$ (i.e., for t odd); therefore, the lir w^i is not well defined for case *ii)*.

[5]Throughout the paper we use the term 'reduction (or increase) in inequality' to mean that inequality between the initially poor and the initially rich regions is expected to go down (up). Forma we require the derivative of f with respect to y_0^i to be smaller (larger) than one.

so that

$$\frac{1}{n}\sum_{i=1}^{n}(y_t^i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=0}^{t-1}\rho^j\varepsilon_{t-j}^i\right)^2 + \left(\nu\sum_{j=0}^{t-1}\rho^j+\rho^t\right)^2\frac{1}{n}\sum_{i=1}^{n}(y_0^i)^2+$$

$$+2\nu\left(\sum_{j=0}^{t-1}\rho^j+\rho^t\right)\sum_{j=0}^{t-1}\rho^j\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{t-j}^i y_0^i\,.$$

For $n \to \infty$, using the fact that the ε's are serially uncorrelated across i's and t's, an that they are independent of y_0^i, we have

$$\Sigma_t = \left(\sum_{i=0}^{t-1}\rho^{2j}\right)\sigma_\epsilon^2 + \left(\nu\sum_{j=0}^{t-1}\rho^j+\rho^t\right)^2\Sigma_0\,. \quad (5$$

If $|\rho| < 1$, we have

$$\frac{\sigma_\epsilon^2}{1-\rho^2}+\Sigma_t \to \frac{\nu^2\Sigma_0}{(1-\rho)^2} \quad \text{as } t\to\infty. \quad (6$$

Hence, σ-convergence obtains if and only if

$$\Sigma_0 \geq \frac{\nu^2\Sigma_0}{(1-\rho)^2}+\frac{\sigma_\epsilon^2}{1-\rho^2} \quad (7$$

Persistence of inequality is only partially related to the restriction (7). We can have σ convergence with persistence of inequality (when $\nu > 0$, $|\frac{\nu}{1-\rho}| < 1$, and Σ_0 is sufficientl large), and we can have that σ-convergence fails but there is convergence in our sens (when $\nu = 0$ and $\sigma_\varepsilon^2/(1-\rho^2)$ is sufficiently large)[6]. If $|\frac{\nu}{1-\rho}| > 1$ the above inequality fail so that convergence fails under both definitions.

This situation is summarized in the second row of table 1.

2.1.3 β-convergence

The concept of β-convergence, favored by BS, requires that, on average, units startin out poorer display faster growth. In our example, given T,

$$y_T^i = \beta y_0^i + \eta^i \quad (8$$

where $\eta^i \equiv \sum_{j=0}^{T-1}\rho^j\epsilon_{t-j}^i$ and $\beta \equiv \rho^T + \nu\sum_{j=0}^{T-1}\rho^j$. Clearly, β-convergence obtains in (8) i $\beta < 1$: in this case, units that start out below average (i.e., $y_0^i < 0$) have $E(y_T^i) > y_0^i$, an poorer units grow faster than units which start above average ($y_0^i > 0$). β-convergenc fails if $\beta \geq 1$.

[6]This is a concrete example of the argument made by Sala-i-Martin (1995) that studying the evolutio of the dispersion is not the same as studying the position of each unit within a distribution. He argue this point with an example taken from sports classifications.

We have that $\beta < 1$ for T large if and only if $\frac{\nu}{1-\rho} < 1$. Therefore, we can have β-convergence coexisting with persistence of inequality when $\nu > 0$. Hence, failure of β-convergence is sufficient but not necessary for persistence of inequality. The third row of table 1 describes these cases.

2.1.4 Conditional β-convergence

The idea of conditional β-convergence is to examine if poorer units grow faster after conditioning on certain observed characteristics X^i. This hypothesis is often tested, with a sample of T years, by running a cross-section regression of the form

$$y_T^i = \phi X^i + \beta \, y_0^i + \, \eta^i \tag{9}$$

for $i = 1, \ldots, n$. Conditional convergence implies $\phi \neq 0$ and $\beta < 1$. Notice that here y_0^i can not be included in X^i, as this would cause perfect multicollinearity. If the X^i are good indicators of the initial condition or of the income levels in periods $t = 1, ..., T-1$ (as they often are, since the characteristics X^i are often measured as averages of X_t^i between $t = 0$ and $t = T$), it is likely that the hypothesis $\phi = 0$ will be rejected even when it holds true because X_i are correlated with the residuals. Therefore, conditional convergence can coexist with persistence in inequality. The fourth row of table 1 presents these cases.

2.1.5 Unit root convergence

Bernard and Durlauf (1996) define absence of convergence as a situation where the differences $y_t^i - y_t^j$ contain unit roots. Clearly, this is only a sufficient condition for persistence of inequality; for example, when ν >0 and 0< ρ <1 there is persistence of inequality and no unit roots. Furthermore, their definition only allows for pairwise comparisons. The fifth row of table 1 describes these cases.

2.1.6 Summary

For all parameter values considered (i.e., $\nu \geq 0$ and $-1 < \rho < 1$), Table 1 indicates which definitions of convergence obtain. The sign $\times$ indicates that convergence obtains according to the definition in the corresponding row; an empty box indicates no convergence; in the cases where additional conditions on parameters other than ν and ρ are needed, this is indicated.

It is clear that various definitions are not strongly related to ours. The concept of $\beta-$(conditional β-)convergence is the one with the largest number of matches with the first row, but it misses in the second column, which is precisely the case that our statistical analysis found relevant. β-convergence, by its own nature, can not distinguish between reduction in inequality or convergence (the first and second columns in this table), while our definition does. The concept of σ-convergence can only distinguish between the two first columns under some additional conditions. The conclusion is that, even though existing

Table 1: **Relationship between Definitions of Convergence**

	$\lvert\rho\rvert < 1$		
	$\nu = 0$	$0 < \frac{\nu}{1-\rho} < 1$	$\frac{\nu}{1-\rho} > 1$
our definition	$\times$		
σ-conv.	$\times$ (if $\Sigma_0 < \frac{\sigma_\epsilon^2}{1-\rho^2}$)	$\times$ (if (7))	
β-conv.	$\times$	$\times$ (T large)	
Conditional β-conv.	$\times$ ($\phi \neq 0$)	$\times$ ($\phi \neq 0$, T large)	
Unit root conv.	$\times$	$\times$	$\times$

definitions focus on important aspects of the data, they are not necessarily appropriate for studying the issue of persistence in inequality. What is needed is a definition of convergence and a statistical procedure that it is able to distinguish between reduction of inequality and convergence.

3 Model Specification

We now specify a flexible statistical model which allows us to formally test for persistence of inequality. We assume that y_t^i follows:

$$y_t^i = a^i + \rho^i \ y_{t-1}^i + \ \epsilon_t^i \tag{10}$$

where ϵ_t^i has mean zero and it is independent across i's and t's. Using the *relative* of per-capita income y_t^i instead of plain per-capita income Y_t^i as our basic variable, alleviates problems of serial and residual cross-unit correlation: since recessions and expansions affect the whole aggregate of regions (or countries) ϵ_t^i would have been serially and cross-sectionally correlated had we used $\log(Y_t^i)$ instead of y_t^i in (10). Also, we do not need to specify a process for growth (whether it is trend or unit-root with drift); all that is needed is that no region grows forever at a faster rate than the average. Appendix 1 provides a setup that formalizes this idea and shows that model (10) is consistent with unit-specific and aggregate transitory shocks and a trend that is common to all units; the persistence of transitory variations within each unit is governed by ρ^i and an i.i.d. shock. Appendix 2 shows that (10) is also consistent with the standard neoclassical growth model and the empirical specification used by BS and other researchers.

Our setup has two important advantages over alternative specifications: first, it allows for a more efficient use of the information contained in the time dimension of the panel since per-capita income for all t's is used to estimate the parameters of the model. Second, and perhaps more importantly, we do not force either the parameters or the steady states to be the same for each unit (or to be the same function of observed conditioning variables) as it is done in cross-section regressions. For example, BS (1992) showed that the theoretical speed of adjustment to the steady state depends on the parameters of preferences and technologies which may differ across units. However their empirical analysis constrains the speed of adjustment to the steady state to be the same for each unit and, depending on the specification, either assume that a^i are constant across i (so that the steady states are the same) or that they are a constant function of the observed characteristics of the unit. Because our approach allows the estimation of the steady states directly, we can separately examine three issues: (i) what is the speed of adjustment of each unit to its own steady state; (ii) whether steady states are all the same, a question having to do with the validity of a version of the neoclassical growth model, and if not (iii) whether there is persistence in inequalities.

It is straightforward to check from (10) that, as long as $|\rho| < 1$, the speed of adjustment of each unit to its own steady state is given by $1 - \rho^i$, and that the expected steady state value of y_t^i is $w^i = \frac{a^i}{(1-\rho^i)}$.

The main problem with our model specification is that, typically, there are too many parameters for each cross sectional unit relative to the number of time series observations. When t is small and (a^i, ρ^i) are estimated using only observations on unit i, estimates will have large standard errors and their small sample distribution may strongly deviate from the asymptotic one [7].

Our approach is to impose a Bayesian prior on the parameters and to combine it with the sample information to construct posterior estimates. This procedure does not require the stringent assumption that the coefficients of the statistical model are the same for each unit to undertake meaningful estimation. The prior assumes that the speed of adjustment and the intercept of the model do not differ too much across units; more precisely, our prior distribution satisfies

$$(\rho^j - \rho^i) \sim N(0, \sigma_\eta^2) \qquad \forall i \neq j \qquad (11)$$

$$(a^j - a^i) \sim N(0, \sigma_\nu^2) \qquad \forall i \neq j \qquad (12)$$

(11)-(12) do not require any 'a priori' belief about the *level* of each set of coefficients. To see this, notice that (11)-(12) imply $F(\beta^j|\beta^i) \sim N(\beta^i, \Sigma_\beta)\ \forall j$, where $F(\cdot|\cdot)$ is the conditional prior distribution and that the marginal prior distribution for $\beta^i = \{a^i, \rho^i\}$ is left unspecified.

Setting σ's equal to zero is equivalent to imposing equality of coefficients across units therefore pooling estimates of the parameters towards their cross sectional mean. Hence,

[7] Microeconometricians encounter similar problems when dealing with panels of data and offered some solutions. For example, Arellano and Bond (1991) assume that the constant term a^i (the unit specific fixed effect) differs across i's, while the coefficients on other regressors are assumed to be the same for all i. In Chamberlain (1984) the intercept is not allowed to vary across units either but the variability of the error term is allowed to be unit specific. Under these assumptions an equation like (10) is written in quasi first-difference form and estimated wiht IV or GMM procedures.

setting all σ's to zero in (11)-(12) would roughly replicate the cross sectional analysis performed by BS or Mankiw, Romer and Weil (1992) (MRW). If the σ's tend to infinity, the β^i are believed a-priori to bear no information for β^j so that estimated parameters of different regions are very similar to those obtained applying OLS to (10) for each unit separately. Finally, if σ's are positive finite numbers, estimates of β in one unit will influence, but be different from, estimates of β in other units. Hence, for finite σ's, estimates of the parameters are constructed using information coming from both the cross-section and the time series dimensions of the panel.

The idea of constructing posterior estimates trading-off the information contained in the cross-section and the time series dimensions is tightly related to the literature on "exchangeability priors" discussed, e.g., in Lindley and Smith (1972). A similar prior was used by several other authors (see e.g. Garcia-Ferrer et. al. (1987), Zellner and Hong (1989), Marcet (1991)). The above studies find that the imposition of this type of prior on the coefficients of a cross-section time-series model improves its out-of sample forecasting ability.

Were $\sigma \equiv [\sigma_\eta, \sigma_\nu]$ given, posterior estimates of the coefficients could be easily obtained with an augmented least square procedure. Let N be the size of the cross section. Define

$$\eta^i = \rho^i - \rho^{i+1} \tag{13}$$
$$\nu^i = a^i - a^{i+1} \tag{14}$$

and let $u^i = [\eta^i, \nu^i]'$ for $i = 1, ...N$, $u = [u^1, u^2, \dots u^{N-1}]'$, $\beta = [\beta^1, \dots, \beta^N]$, R be a $2(N-1) \times 2N$ matrix with ones on the main diagonal and minus one on the second upper diagonal. Rewrite (13)-(14) as $R\beta = u$ where $u \sim N(0, \Omega)$ and where Ω has the following structure

$$\begin{aligned}
cov(\eta^i, \eta^j) &= \sigma_\eta^2 && \text{if } j = i \\
&= -0.5\ \sigma_\eta^2 && \text{if } j = i \pm 1 \\
&= 0 && \text{otherwise} \\
cov(\nu^i, \nu^j) &= \sigma_\nu^2 && \text{if } j = i \\
&= -0.5\ \sigma_\nu^2 && \text{if } j = i \pm 1 \\
&= 0 && \text{otherwise} \\
cov(\nu^i, \eta^j) &= 0 && \text{for all } i,\ j
\end{aligned} \tag{15}$$

Notice that in (13)-(14) we have written the restrictions for adjacent units only (i.e. there are only $N-1$ restrictions instead of the $N(N-1)/2$ restrictions in (11)-(12)) without loss of generality. It turns out that, if (15) is imposed, the prior is symmetric, in the sense that it is invariant to the ordering of the units. Such a prior implies that $var(\rho^j - \rho^i) = \sigma_\eta^2$; $var(\alpha^j - \alpha^i) = \sigma_\nu^2$ and that $cov(\rho^i - \rho^j, \rho^k - \rho^l) = cov(\alpha^i - \alpha^j, \alpha^k - \alpha^l) = cov(\alpha^i - \alpha^j, \rho^k - \rho^l) = 0$ for any four-tuple of distinct (i, j, k, l).

The posterior mean of β is

$$\beta = (X'\Sigma^{-1}X + R'\Omega^{-1}R)^{-1}(X'\Sigma^{-1}Y) \tag{16}$$

where $Y^i = \{y_1^i, \ldots, y_t^i\}$, $X^i = [1, Y_{-1}^i]$, $X = diag[X^1, \ldots, X^n]$, $Y = [Y^1, \ldots, Y^n]$, where 1 is a vector of dimension t, and $\Sigma = diag\{\sigma_\epsilon^2\}$.

The discussion so far assumed that Σ and Ω, which contain the parameters regulating the trade-off between the information contained in the time-series and the cross-section dimensions of the panel, were known. In a standard Bayesian approach one imposes a prior on Σ and Ω and uses the resulting hierarchical structure to construct posterior estimates along the lines of Lindlay and Smith (1972) and Smith (1973). Rather than taking the Bayesian approach literally, we start the empirical analysis in section 5 by computing the mode of the posterior of the data for different values of σ's, given the best possible scale invariant estimator for Σ of the form $S = s^2 * I$ where $s^2 = \frac{(Y-X\beta)'(Y-X\beta)}{N(T-2)}$, and explore how different choices of σ match with the data. This is in the spirit of the specification searches of Leamer (1979) and Sims and Uhlig's (1991) 'helicopter tour' and gives us information about the shape of the posterior density. In addition, we also approximate prior maximum likelihood estimates of σ obtained by numerically maximizing the likelihood of Y. Posterior inference can be conducted plugging in S and estimates of Ω in (16) in an Empirical Bayes fashion.

In discussing the issue of unconditional convergence we will be interested in examining whether estimates of the steady state of relative per-capita income differ across units. That is, we need to examine the hypothesis that $a^i/(1-\rho^i) = a^j/(1-\rho^j)$ $\forall i,\ j$ versus the alternative composite hypothesis that they are different. Given the one-to-one correspondence between the values of σ and the steady states the two hypotheses can also be formulated as $\sigma = 0$ (equal steady states) versus $\sigma \neq 0$ (different steady states). To verify this hypothesis in a manner which is consistent with our approach we employ the Posterior Odds ratio (PO) (see e.g. Leamer (1979) or Sims (1988)). The details of the testing approach are discussed in appendix 3. Using a standard zero-one loss function, the hypothesis that the steady states are the same is preferred if it has higher posterior probability than the alternative. To provide a different point of view we also compute the largest prior probability on the alternative so that the data would not reject the null, i.e how much confidence should we have in the null so that the data does not overturn our prior beliefs. We call this measure π^*. Small values of this statistic indicate, that unless the alternative is a-priori very improbable, the data would always reject the null. Finally for those who feel uncomfortable with Bayesian approaches, we also provide a likelihood ratio test for the null hypothesis of equality of steady states.

If the null of unconditional convergence is rejected, we would like to know what variables explain the cross sectional distribution of estimated steady states. For this purpose we model the dependence of steady states in the following way:

$$w^i = \delta + \gamma y_0^i + \omega X^i + u^i \qquad (17$$

where w^i is the steady state of unit i, y_0 is the initial (scaled) level of per-capita incom in region i; the vector X^i includes, as in BS or MRW, variables proxying for difference in technologies, government policies, and human capital. This model allows us to dis tinguish a number of hypothesis, having to do with persistence of inequality, conditional convergence and the presence of 'miracles'; this is interesting, since a number of models in the growth literature have implications for each of these hypothesis. Whenever $\gamma > 0$, the distribution of the initial level of income determines the cross sectional distribution of the steady states, i.e. income inequalities are persistent. The magnitude of $\hat{\gamma}$ provides an indication of how persistent inequalities are. A small positive $\hat{\gamma}$ suggests that the gap between the rich and the poor will eventually be reduced even though it will not disappear. At the opposite end, a $\hat{\gamma}$ which is positive and close to one implies that inequalities are extremely persistent. A negative $\hat{\gamma}$, on the other hand, indicates the realization of the 'biblical prophecy ', i.e. the steady state income of initially poor will be uniformly higher than that of the initially rich. Whenever $\omega \neq 0$, in addition to $\gamma > 0$, factors other than the initial conditions explain the distribution of the steady states.

The $\bar{R}^2$ of a regression when $X^i = 0$ can be interpreted as synthetic measure of long run mobility. A high $\bar{R}^2$ indicates that the ordering of units in the steady state distribution is the same as the ordering in the initial distribution. A low $\bar{R}^2$, on the other hand, suggests that individual units may move up or down in the ranking of income distribution. If this occurs when $\hat{\gamma}$ is positive, average income differences persist even if some individual units may experience miracles and busts.

4 The Data

In this study we employ two data sets. The first has not been used (to our knowledge) in the recent literature on convergence and it will be the center of our attention; the second is well known and is used as a benchmark for comparison with other studies.

The first data set consists of per-capita income for European regions of 14 member countries, calculated from the population and GDP data of the Regio data set of Eurostat. Using the Eurostat nomenclature, the regional disaggregation we use corresponds to Nuts-2 level for all countries except Ireland, Denmark and the United Kingdom where, because of lack of data, we revert to Nuts-1 level. Roughly speaking, level 2 includes two or three times as many regions as level 1, depending on the country. Very small regions, such as Açores (Portugal) or Martinique (France), were excluded. GDP is measured with the Purchasing Power Standard as provided by the Eurostat. Since we use the ratio of regional to aggregate per-capita income there is no need to convert nominal into real income.

Even though we have data from 1975 to 1992 many data points were missing for the first few years. To maximize the number of units for which Nuts-2 level data was available, we only used observations for the period 1980-92; about twenty data points for this time period were missing and were linearly interpolated. This leaves a total of 144 regions and 1728 data points.

Using data at Nuts-2 level is important because a higher level of aggregation is too coarse for a meaningful discussion of regional convergence. As an example, the regions of Aragón and Euskadi (Basc Country) are placed together in the 'Northwest' Spanish region at Nuts-1 level, even though the first is largely an agricultural region that has been loosing population through migration for most of this century, while the opposite is true of the second; Euskadi is traditionally wealthier (its per-capita income is about 23% larger than Aragón's in 1981) and deep cultural, historical, linguistic and political differences cause these regions to have different autonomous governments. The Nuts-2 level, however, properly distinguishes among these regions. For another example, all of continental Portugal constitutes one Nuts-1 region while there are clear economic differences between e.g., Algarve and Alentejo.

Since the statistical procedure we propose has not been used in the recent literature, we also apply it to the per-capita real GDP measured in international prices from Summers and Heston (SH) (1991) data set. This data set is well known to students of economic growth. We limit attention to 17 Western European countries (Austria, Belgium, Denmark, Finland, France, West Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom); this choice guarantees that all units are very close in terms of institutions and economic structure, and it makes it more likely that the hypothesis of convergence will be accepted [8].

In both data sets we verified that model (10) accurately describe the data. For the Regio data (SH data) set, the hypothesis that an AR(1) model should be preferred to an AR(2) is not rejected in 141 (16) cases using both the AIC and the SIC criteria. Moreover we found that the residuals of (10) are hardly correlated across regions (the three largest correlations are 0.23, 0.21, 0.21) and the largest correlation in the sample of countries is 0.17. Finally, we examined whether the business cycle component, captured by Y_t, is really common to all regions by running a regression of the form:

$$Y_{it} = a^i + \rho^i Y_{it-1} + \eta^i Y_t + e_t \qquad (18)$$

and verifying whether $\eta^i = 1, \forall i$. In both samples the null hypothesis is not supported in one case only.

5 The Empirical Results

5.1 Speed of adjustment to the Steady State

Our first set of results is contained in table 2 which reports, for selected settings of σ, the value of the posterior mode, the average estimate and the cross sectional dispersion of the AR parameter of model (10) and of the steady state. The first panel reports results obtained with the Regio data, the second those obtained with the Summers and Heston

[8] To make sure that the results did not depend on our choice of countries, we also considered the sample of 21 OECD countries standardly used in the literature. No substantial changes in the conclusions emerged.

data. Here, the comparison of different prior specifications is informal; we attempt to summarize the data in the light of our model; more formal tests are provided in subsection 5.3.

First of all, notice that our estimations differ from the standard cross-sectional regressions of say, BS, in three features: *i)* we use income of the units relative to aggregate income (i.e., we use y_t^i instead of Y_t^i), *ii)* we use all the data for all years, *iii)* when $\sigma > 0$ we allow for different parameter values for each unit.

The first row in table 2 (the case $\sigma_\eta = \sigma_\nu = 0.000001$), corresponds to only introducing features *i)* and *ii)* and forces the coefficients for each unit to be the same, as is implicitely done in cross-section regressions. Surprisingly, we approximately obtain the eerily ubiquitous average speed of adjustment of about 2% per year with both data sets. This is important because it implies that features *i), ii)* alone can not account for the different results we obtain, when we add *iii)*.

Second, when we allow for heterogeneity in parameters across units the average speed of adjustment increases up to about 23% a year with the Regio data set and to about 11% a year with the Summers and Heston data set. For a similar sample of OECD countries BS and MRW estimated the average speed of adjustment to be of the order of 1.4-1.8% when using cross sectional regressions, while Evans (1995), Caselli, Esquinel and Lefort (1996), Le Pen (1996) find values of the order of 4-9% a year using panel data techniques. Ourestimates imply that that the capital share is in the interval [0.1-0.3], a range which is more reasonable than the one obtained by BS or MRW [9].

It is instructive to provide an intuitive explanation for these results. Consider a situation where the "true" model has different intercepts, but similar ρ's, and the steady state is positively correlated with the initial condition across units. Figure 1 represents equation (10) for three units under these assumptions, and a likely cloud of points in a finite sample. If one traces *one* regression line through this cloud of points (or through the average value for each unit), as is done when $\sigma = 0$, the estimated ρ will be much higher than the average of the true ones. This phenomenon is well known among microeconometricians as the 'fixed effect bias' and occurs whenever heterogeneity is not appropriately accounted for. This bias does not disappear as more time series or cross sectional observations are collected. The pervasiveness of average speed of adjustment around 2% previously obtained may therefore be the result of a biased estimation procedure that ignores fixed effects present in the data.

Overall, the average speed of adjustment increases uniformly with σ and when $\sigma = 1$

[9] Barro and Sala-i-Martin (1995, p.52) show that the speed of adjustment is linked to the parameters of the neoclassical production function by the relationship

$$1 - \rho = (1 - \theta)(\delta + n + x) \tag{19}$$

where δ is the depreciation rate of capital, n the growth rate of population, x the labor augmenting technological progress and θ is the capital share in the production function. Assuming $\delta = [0.10, 0.12]$, $n = [0.01, 0.02]$, $x = [0.02, 0.05]$ we obtain the range [0.1-0.3] for θ.

results are close to those obtained with $\sigma = \infty$. Therefore, by varying the σ vector from zero to one we can explore the trade-off between the information contained in the cross-section and in the time-series dimension of the panel. To investigate such a trade-off, consider two intermediate cases where we are imposing either that ρ^i is the same for each unit and differences in steady states are solely due to unit specific fixed effects ($\sigma_\nu = 1.0, \sigma_\eta = 0.000001$) or that there is no unit specific fixed effect and that steady states differ because of different speeds of adjustment ($\sigma_\nu = 0.000001, \sigma_\eta = 1.0$).

For the Regio data set, forcing ρ^i to be the same for each i (seventh row) reduces the value of the mode of the posterior. However, the speed of adjustment is still on average about 13% and the dispersion of the posterior steady state estimates is still large. When we eliminate the individual effect but we allow the speeds of adjustment to differ (third row), the reduction is much larger, and the average speed of adjustment is only about 1.2% a year. For the 17 European countries the results are similar, although less spectacular quantitatively. Restricting ρ^i to be the same across i causes a drop in the value of the mode of the posterior and a drop in the average estimated speed of adjustment. However, leaving out individual effects while letting the ρ^i be country specific induce larger drops. Therefore, our results are consistent with figure 1: forcing the a's to be the same causes a larger distortion then setting the ρ's to be the same.

The last row of each panel reports approximate ML estimates of σ, which are not too different from $\sigma_v = \sigma_\eta = 1.0$ for both data sets, a situation where there is a minimum amount of pooling across units [10]. In both cases, however, we find that there are units for which $\hat{\rho}^i > 1$, implying divergence of per-capita income, a result that does not make sense give our model specification. In figures 2 and 3 we therefore present results using $\sigma_v = 1.0$ and $\sigma_\eta = 0.5$ for both data sets, the values of the hyperparameters which are closest to the 'optimal' ones and keep all estimates of ρ^i below 1.0. Panel A of both figures plots the posterior speeds of adjustment against the initial conditions for each unit in the two data sets. Speeds of adjustment vary from a low 1-2% (Nord-Pas-de-Calais (France), Luxemburg, Drenthe (Netherland) and Yorkshire (UK)) up to almost 80% (North Portugal, Voreio Aigaio and Kentriki Makedonia (Greece)) in the Regio data set and from 1% (Switzerland) to 33% (Turkey) in the Summers and Heston data set. Note also that there is a weak negative relationship between the initial conditions and the speeds of adjustment.

5.2 Unconditional Convergence

The distribution of estimated steady states clearly depends on the value of σ and tends to be more dispersed as σ increases. With our preferred choice ($\sigma_v = 1.0, \sigma_\nu = 0.5$), the

[10] We resorted to maximize $m(Y|\sigma)$ numerically with a set of three successive grids since this density of both data set is very flat for a large set of values for σ and analytical routines find different peak values depending on the initial conditions. The average value of the estimated ρ, however, is not affected by these changes.

dispersion of estimated steady states is substantial. Panel B of figures 2 and 3 provides a histogram of estimated steady states for each data set. The histogram is organized so that regions are grouped in eight classes of steady-state per-capita income (up to 40%, 41-55%, 56-70%, 71-85%, 86-100%, 101-115%, 126-130%, above 131%) and countries in five income classes (25-50%, 51-75%, 76-100%, 101-125%, above 126%) where 100 is the average income of each data set (the steady state level of y_t^i which would obtain if there was unconditional convergence). It is clear that the estimated steady state distribution for both data sets is far from collapsing toward its central (or any) value. Hence, it is very unlikely that unconditional convergence holds in our data sets. Table 3 summarizes some formal tests on the hypothesis of unconditional convergence. This table reports the values of the PO ratio, of π^* and the p-value of the likelihood ratio test for the hypothesis that the steady states are the same. Regardless of the statistic used, the alternative seems more likely than the null. Particularly informative is the reported value of π^* (i.e. the maximum value of the prior probability on the alternative needed to accept the null hypothesis that units have the same steady state). For both data sets, unless we assume a-priori that the alternative is impossible, the null hypothesis will always be overturned by the data.

5.3 Sub-sample Instabilities

Next, because of possible structural breaks in the Summer and Heston data set, we explore the issue of subsample instabilities. Consistent with the literature, we split the sample in two with 1965 as a breaking date. Also, previous studies have suggested that in the 1980's convergence is weak. To examine this possibility, we also consider the sample 1950-1979 and compare the results with those obtained for the 1950-1985 sample.

The qualitative features of previous results are confirmed for different subsamples (see table 5 and figures 4-6): forcing the steady-states to coincide drives the average speed of adjustment down and for the best specification the average estimated speed of adjustment is substantially larger then the one found in the literature. Also, the estimated distribution of steady states is far from collapsing to a signle point. However, while in the 1966-1985 subsample the quantitative results are in agreement with those for the 1950-1985 sample, the other two subsamples (1950-1965 and 1950-1979) display interesting differences. First, the "best" specification is one where the speed of adjustment is a-priori pooled toward a common value (pooling being stronger in the 1950-1966 sample) while it is optimal to leave some heterogeneity in the intercept. Second, while Bayesian statistics favor the alternative hypothesis that the estimated steady states differ, the likelihood ratio test is unable to reject the null in both samples. Finally, by comparing the results of the 1950-1979 sample with those of the 1950-1985 sample we can conclude that the 1980's were indeed a period where heterogeneities become more marked, in agreement with Blanchard and Katz (1992). More importantly, these heterogeneities turned out to emerge more strongly in the speed of adjustment which, consistent with the results of our

Regio data set, became more dissimilar across countries.

5.4 Persistence in Inequality

Since our results indicate that the estimated distribution of steady states is non-degenerate regardless of the sample used, we examine which variables account for the cross sectional dispersion in estimated steady states using equation (17). The analysis of this issue in this section will be, again, informal, since in order to provide appropriate Bayesian statistics, the parameters in (17) would have to be estimated jointly with the σ, ρ, a's

Equation (17) allows for differences in the steady states through a set of variables X^i, capturing differences in technologies or policies. Significant effects have been found in the literature, especially for samples of countries, from the introduction of proxies for human capital and government expenditure in the regressions. Most of the literature concludes that these effects are rather small for OECD countries and absent for regions (see e.g. BS (1992) and Sala-i-Martin (1995)). The results of section 5.1, however, indicate that *some* variable must be having an effect on the level of steady states.

One candidate to explain the limiting distribution of steady states can be found by inspecting panel C of figures 2 and 3, which plots estimated steady states against initial income levels for the two data sets. It is clear that estimated steady states appear to have a strong *positive* connection with the initial conditions and the ranking in the initial distribution is largely maintained.

Table 4 studies the strength of this association. Because of the lack of disaggregated data on X^i for European regions, we only examine unconditional persistence of inequality with the Regio data set. For the 17 countries of the Summers and Heston data set, we first examine how important are initial conditions to explain the cross sectional distribution of steady states and second, whether the inclusion of variables proxying for human capital (the secondary education variable used by Barro (1991)), for differences in saving behavior (the investment/output ratio used by MRW (1992)) and for government policies (share of government expenditure in GNP used by Barro (1991)), change the essence of our results.

The evidence contained in the table is overwhelming: the main determinant of the position of a unit in the steady state distribution is its position in the initial income distribution. For regional data, the estimate of γ is close to 0.8, suggesting that, on average, the income gap between the rich and the poor will be reduced in the limit by 20%, and 33% of the variations of the cross section distribution of steady states is explained solely by initial conditions. For European countries the estimate is about 0.5 and the initial conditions alone explain 47% of the cross sectional distribution of steady states for the sample 1950-1985 and other conditioning variables add no significant explanatory power to the regressions. More importantly, none of the conditioning variables appears to be correlated with the steady states once the effect of initial conditions is accounted for - eliminating the initial conditions from the regression does not change the estimates of the coefficients for the other conditioning variables - therefore denying that the existence of a correlation between the initial conditions and the X's. Overall these results are consistent with Easterly et. al. (1993)'s conclusion that policy variables play a small role in explaining the pattern of growth rates and indicate that there is persistence of inequality. In the language of Levine and Renelt (1992), only the relative position in the initial distribution is a robust determinant of the relative position of a unit in the distribution of steady states.

The same results hold for the 1966-1985 subsample of the 17 countries of the Summers and Heston data set. It is remarkable that in this subsample the initial conditions alone explain about 85% of the cross sectional dispersion of steady states and that the estimate of γ is close to 0.8, indicating only a minor reduction in income inequality in the limit. For the 1950-1965 subsample the initial conditions are important but now government share in GNP appears to be an important determinant of the steady state distribution. Finally, for the 1950-1979 sample, no variable appears to explain the small cross sectional differences in estimated steady states.

5.5 Summary

Our results can be summarized as follows: (i) different countries and regions converge to different steady states; this happens in both data sets and in most sub-samples. (ii) The estimated average speed of adjustment varies with σ's in a way that is consistent with the fixed-effect bias described in figure 1. For our preferred choice of σ's, the estimated average speed of adjustment is significantly higher than previously estimated and there is considerable dispersion of estimates across units, dispersion which increases in the 1980's. (iii) In four of the five samples, the initial conditions are the most important determinant of the estimated distribution of the steady state of per-capita income; according to our definitions, there is persistence of inequality. Income inequality was reduced to some extent from the 1950 to the 1970, it persisted intact for most of the 1970's and it increased over the 1980's. This is true even when we account for differences in government variables, human capital etc.. A country (region) which is initially below the average per-capita income will eventually expect the gap with respect to richer countries (regions) to narrow somewhat but not to improve its relative standing in the cross sectional distribution. Hence, with some exceptions, the poor stay almost as poor as they were at the beginning.

6 What could be wrong? Misspecification and Biases

Our results are substantially at odds with those commonly found in the literature but we have provided an explanation (the fixed-effects bias depicted in Figure 1) that makes consistent ours and previous results. It is important to challenge our conclusions, however, to see whether misspecifications or econometric biases intrinsic in our estimation/testing procedure can account for the results.

In section 3 we have justified the use of per-capita income relative to the average per-capita income of the cross section by the simple aggregation model of appendix 1 and by the fact that, scaled in this way, the stochastic process for income per-capita of different units is well represented by an AR(1) process. It may be worth examining whether results would change when an alternative normalization which preserves the AR(1) properties for the scaled variable is used. We therefore repeat the estimation process for the sample of European regions scaling each unit at each point in time by its country mean [11]. In terms of the model of Appendix 1 this implies that there is one trend common for the regions of each country. With this normalization we can roughly examine whether there is any tendency for regional steady states to cluster around their own country mean, a result consistent with some of the findings of BS (1991).

The results of this experiment are presented in table 6 and in figure 7 and confirm previous conclusions. Few additional features are worth noting. First, the σ vector which maximizes the mode of the posterior is very large and the posterior is flat up to $\sigma = \infty$ so that estimation by OLS equation by equation provides the best possible fit to the data. Hence, knowledge of the α and ρ for one region does not provide information for α and ρ in another region. Put it in another way, with this scaling, income per-capita at regional level behaves as if there were no regional interdependencies. Second, for the best choice of σ's, the average speed of adjustment increases to about 36%. Finally, the hypothesis that the estimated steady states for all regions of one country are equal has practically no support in the data. From figure 7 we see that the cross sectional distribution of steady states is far from collapsing to a single point and that the relationship between the initial and steady state distribution of per-capita income is strong with little tendency toward reducing inequality (the slope is 0.8).

The presence of measurement error may constitute a serious problem for our time series approach to estimate steady states and for our cross sectional tests of persistence of inequality. It is well know that if y_t^i is measured with error, estimates of ρ are downward biased (i.e. the estimated speed of adjustment is higher than the true one) with the magnitude of the bias depending on the serial correlation properties of the measurement error and on the variability of its innovations relative to the variability of innovations in y_t^i. To quantify the extent of the problem for our two data sets we ask the following question. Suppose that the true speed of adjustment to the steady state is 2% per year. What properties should the measurement error have to obtain estimates of 23% (Regio data) or 12% (Summer and Heston data)? Table 7 presents the results allowing for serially correlated measurement error. In the most favorable outcome (strongly serially correlated measurement error), the variability of innovations in the measurement error should be 1/6 (1/3) of the variability in innovations in y_t^i, which is large by any standards, given that we are considering European GNP's. When the measurement error is i.i.d, the variability of its innovations should be about 40% (70%) of the variability of innovations in y_t^i.

[11] Since data for Denmark, Luxemburg and Ireland is available only at country level, we exclude them from the sample for this experiment.

Measurement error could also bias the cross sectional regressions estimates but *agains* our finding that the distribution of initial conditions is an important determinant of the distribution of steady states. If measurement errors are present, e.g., in the initial conditions, the conditioning variables in X_i are likely be correlated with the error term, making it difficult to accept the hypothesis that the initial conditions are the most important variable in those regressions.

To summarize, measurement error is unlikely to be the reason for both the high average estimates of the speed of adjustment and the strong persistence of inequality we found.

One additional potential problem with our cross sectional regressions is that steady states are estimated on a short sample. That is, the left hand side variable of (17) is measured with an error which, in a short sample, maybe correlated with initial conditions. This makes the disturbance correlated with the regressors and therefore biases estimates of γ and ω we obtain. In principle, it is hard to measure how important is this problem in our setup. One way to assess the magnitude of the bias is the following: let the steady states be estimated using data from $\bar{t}$ to T where $\bar{t} > 1$. Then, if the average ρ is about 0.8, taking $\bar{t} \geq 5$ should almost entirely eliminate the correlation between the initial conditions and the error in the estimated steady states (even though the precision of the estimates may decline). We have reestimated our cross sectional regressions using steady states estimates obtained using $\bar{t} = 2, 3, 4, 5$. None of the qualitative conclusions we obtained are affected. In particular, it is still true that there is a very strong tendency for inequality to persist [12].

Finally, even when measurement error is absent, OLS estimates of ρ are downward biased and estimates of a are upward biased in small samples. That is, although OLS equation by equation produces consistent estimates, their small sample distribution may strongly deviate from the asymptotic one; our Bayesian estimates may inherit some of this bias. We have conducted some work on the small sample properties of our estimators. Monte-Carlo results indicate that our basic results are not due to small sample biases. Overall, it appears that the use of cross-sectional information substantially alleviates the downward bias typical of time-series estimation of ρ. For example, when the cross section is large (say, $N \geq 100$), the OLS bias is cut by more then 50-60% even with samples with 12 observations. Moreover, there is no evidence that the spread of the estimated steady state distribution is biased in one direction or another, suggesting that our conclusion that inequality persists is not due to econometric biases.

7 Conclusion

The modern literature on convergence concludes that there is convergence for regions and countries at a very slow speed. Limiting steady states may differ because of differences

[12] One alternative approach would be to specify model (10) as in, e.g. Schotman and Van Djik (1991) and have the constant term directly depend on the initial conditions.

in technologies or government policies but the effects of variables proxying for these differences are weak for countries and absent for regions. The policy conclusion seems to be that, either because of current redistribution policies or because of neoclassical-growth-model convergence, poor regions should be "patient" enough and wait for inequalities to slowly disappear. Also, they should set certain policy variables (the conditioning variables) close to those of richer countries.

The conclusions this paper offers are very different: we find fast speeds of adjustment to a non-degenerate distribution of steady state levels of per-capita income where inequalities largely persist. A poor region can expect the gap between its initial level of income and the average to be reduced by only 20% in the limit. The conclusion seems to be that current redistribution and development policies, such as the PAC, the ESDF and the Cohesion programs of the EEC, are not working to the full extent; rich regions can be taxed more heavily for solidarity reasons but not in the hope that these transfers will foster development of the poor regions. Poor regions cannot expect to become as well off as rich ones unless structural changes occur; controlling the conditioning variables is not sufficient for inequality to disappear. The fact that some regions grew very fast is consistent with our finding, and it can be due to pure luck or to unmeasured variables.

We argued that the restrictions that cross section regression approaches impose on the data are strongly rejected in formal testing and this accounts for the differences in the conclusions; previous results can also be accounted for by the fact that the definition of convergence that had been used is not necessarily related to the issue of persistence of inequality. The opposite story (given the estimated rates of adjustment, moviments are so slow that can't be seen in such a short sample) can not be sustained since, given the initial distribution, a 2% rate of adjustment and the sample sizes we are considering, γ in equation (17) would be only of the order of 0.2-0.3. We have also shown that, by exploring the data with a flexible Bayesian procedure, we can find the best model specification, examine the features of the data in a systematic way and formally verify various propositions.

We would like to comment briefly on the issue of which theoretical growth model is consistent with our results. Our setup does not distinguish between exogenous vs. endogenous type of models. All we can say is that to match the stylized facts we uncovered, the model must have a reduced form like (10) with two parameters (a^i, ρ^i) which vary across units. Moreover, it should produce a non-degenerate steady state income distribution whose ranking mirrors the ranking of the initial income distribution even though income inequality may be partially reduced as time goes by.

Even though our empirical results and our predictions about regional inequality are rather striking, we should offer several words of caution before taking the conclusions literally. First, the Regio data set has not been examined sufficiently by academics to guarantee its reliability; some questions have been raised on the way it is constructed. The fact that we obtain similar results with both data sets comforts us about possible incongruities present in this data. Second, the time span of the Regio data set is short so the prior may have a substantial influence on posterior estimates. "Misspecification" of th prior may therefore cause distortions. Third, by its own nature, the issue of convergenc is an exercise in asymptotic extrapolation, which is well known to be unreliable. Becaus this sin is committed by all studies, we let the reader decide what to do with the entir empirical literature on convergence.

Appendix 1:A Model of Aggregation over units.

In this appendix we show that the model of equation (10) is internally consistent. Let $\overline{y}_t^i = Y_t^i/Y_t$ (or, equivalently, $y_t^i = \log \overline{y}_t^i$) satisfy

$$\overline{y}_t^i = a^i + \rho^i \ \overline{y}_{t-1}^i + \epsilon_t^i \tag{20}$$

The ϵ's are i.i.d. across time and units, have mean zero, and their support is such that $\overline{y}_t^i$ is always positive [13].

Assume there is a continuum of regions $i \in [0,1]$. Clearly, $\{Y_t^i\}$ is consistently determined from (20) and *any* process for average per-capita income $\{Y_t\}$ simply by setting $Y_t^i = y_t^i \ Y_t$. Since aggregate income is defined as $\int_0^1 Y_t^i \ di = Y_t$, in order to insure that the model is well defined, we have to show that $\int_0^1 y_t^i \ di = 1$ for all t. As long as the process (20) generates ratios that are consistent, we can specify a process for aggregate output independently. The process for $\{Y_t\}$ is left unspecified in the applied part of the paper; therefore, our empirical results are consistent with a $\{Y_t\}$ displaying any pattern for aggregate growth or aggregate business cycles shocks.

Now to check consistency, we need to make

Assumption 1 *The random variables ρ^i, ϵ_t^i and $\frac{a^i}{1-\rho^i}\overline{y}_0^i$ are all mutually independent across i's, and satisfy $\int_0^1 \overline{y}_0^i di = \int_0^1 \frac{a^i}{1-\rho^i} \ di = 1$.*

To show that $\int_0^1 y_t^i \ di = 1, \forall t$ notice that (20) can be rewritten as

$$\overline{y}_t^i - \frac{a^i}{(1-\rho^i)} = \rho^i \left(\overline{y}_{t-1}^i - \frac{a^i}{(1-\rho^i)}\right) + \epsilon_t^i =$$

$$\sum_{j=0}^{t-1} \left(\rho^i\right)^j \ \epsilon_{t-j}^i + \left(\rho^i\right)^t \left(\overline{y}_0^i - \frac{a^i}{(1-\rho^i)}\right) \tag{21}$$

Taking integrals in (21) and using Assumption 1, we have

$$\int_0^1 \left(\overline{y}_t^i - \frac{a^i}{(1-\rho^i)}\right) \ di = \sum_{j=0}^{t-1} \int_0^1 \left(\rho^i\right)^j \ di \ \int_0^1 \epsilon_{t-j}^i \ di + \int_0^1 \left(\rho^i\right)^t \ di \ \int_0^1 \left(\overline{y}_0^i - \frac{a^i}{(1-\rho^i)}\right) \ di =$$

$$\sum_{j=0}^{t-1} \int_0^1 \left(\rho^i\right)^j \ di \ 0 + \int_0^1 \left(\rho^i\right)^t \ di \ \int_0^1 (1 \ - 1) = 0 \tag{22}$$

Using Assumption 1 and the previous equation, we have for all t

$$\int_0^1 \overline{y}_t^i \ di = \int_0^1 \frac{a^i}{1-\rho^i} \ di = 1 \tag{23}$$

[13] A slight difference with the equation estimated in the paper is that, here we do not use the logs of the ratios. Using the logs is done to insure non-negativity of the process under normality.

for all t, as desired.

This model, then, allows for a fairly rich pattern of regional fluctuations, with a (possibly correlated) aggregate shock shared by all units, different cycles in different units, and growth in all units. Also, it allows for one region to be systematically poorer or richer than the average through differences in a^i's. The only restrictive assumption is that persistence of idiosyncratic shocks enters only through an AR(1) process and that the ratio $\overline{y}^i_t$ is not affected by the aggregate shock, so that no units are allowed to respond more strongly than others to aggregate shocks.

Let's suppose that aggregate output follows a random walk with drift:

$$Y_t = Y_{t-1}\ \delta\ \eta_t \tag{24}$$

where η_t is a stationary process that may be serially correlated, its log has mean zero, and is independent of all other random variables. In this particular case, multiplying both sides of (20), by Y_t we have

$$Y_t^i = a^i Y_t + \rho^i\ Y_{t-1}^i + \left(\rho^i\ Y_{t-1}^i(1 - \delta\eta_t) + \epsilon_t^i Y_t\right) \tag{25}$$

The presence of the aggregate shock causes the residual in this equation (the term in parenthesis) to be highly correlated across regions and across time and correlated with the regressors. It is because of this undesirable property of the residuals of the equation for Y_t^i that some authors have avoided estimation of convergence regressions with levels of per-capita income and panel data (see e.g. Pesaran and Smith (1995)).

Appendix 2: Linking the Statistical and the Solow Model

In this appendix we show that model (10) is consistent with both the standard neocl sical growth model and the estimable specification employed by Barro and Sala-i-Mar (1992).

The neoclassical growth model where the production function displays constant retu to scale implies the following equation describing out of steady state dynamics (see (1992)):

$$\log[\hat{y}_i(t)] = \log[\hat{y}_i(0)]e^{-\beta_i t} + \log[\hat{y}_i^*](1 - e^{-\beta_i t}) \quad (\text{:}$$

where β_i is the parameter controlling the speed of adjustment to the steady state (wh depends on the parameter preferences, technologies and population), $\hat{y}_i(t)$ is output unit of effective labor at time t and $\hat{y}_i^*$ is output per unit of effective labor in the stea state.

The estimable specification BS employ is:

$$\frac{1}{T}\log[\frac{y_i(t_0+T)}{y_i(t_0)}] = B_i - \log[y_i(t_0)]\frac{(1-e^{-\beta_i T})}{T} + u_{i,t_0,t_0+T} \quad (\text{:}$$

where u_{i,t_0,t_0+T} is a distributed lag of u_i for t between 0 and T, where $B_i = z_i$ $\frac{(1-e^{-\beta_i T})}{T}[\log(y_i^*) - z_i t_0]$ and z_i is the rate of exogenous labor augmenting technologi progress. Using discrete time notation the above equation can be written as:

$$\log(y_T^i) = \alpha + \rho^T \log(y_0^i) + \gamma\ X^i + \epsilon^i \quad ($$

where $t = 0, 1, ..., T$ and the variables X^i are introduced to allow for shifts in the limit steady state means of y_t^i across i. Roughly speaking, this model implies that, if $0 < \rho <$ the mean of $\log(y_T^i)$ converges monotonically to $(\alpha + \gamma\ X^i)/(1-\rho)$ as T becomes la If each period represents a year, the speed of adjustment to the steady state is $(1-\rho^1$ a year.

If we now let $a_i = \alpha + \gamma X^i$ our model specification is consistent with BS model. N however that, to estimate the speed of adjustment, BS restrict $B_i = B$ and $\beta_i = \beta\ \forall i$

Appendix 3: Testing equality of the steady states

The procedure we have used to examine whether or not steady states are equal consists of the following steps:

- linearize the steady state for each unit i around the average steady state of the cross section.
- Set up the null hypothesis of equality of linearized steady states, with the alternative being that they are different. This amount to a linear restriction on the vector of coefficients of the model.
- Use the fact that the posterior distribution of the $2N \times 1$ vector of parameters β is normal (conditional on σ with mean given by equation (16) in the paper and variance $(X'\Sigma^{-1}X + R'\Omega^{-1}R)^{-1}$ to derive the posterior distribution of the relevant linear combination of parameters under the null and the alternative.

Specifically, let $w^i = \frac{\alpha_i}{1-\rho_i}$ be the steady state for unit i. Taking a first order Taylor expansion around the average steady state we have:

$$\frac{\alpha_i}{1-\rho_i} = [\frac{\alpha}{1-\rho}]^A + \frac{1}{1-\rho^A}(\alpha - \alpha^A) - \frac{\alpha^A \rho^A}{1-\rho^A}(\rho - \rho^A) + Z$$

where A stands for average and Z is the remainder. Disregarding higher order terms, the null hypothesis is therefore:

$$\frac{1}{1-\rho^A}(\alpha^j - \alpha^i) - \frac{\alpha^A \rho^A}{1-\rho^A}(\rho^j - \rho^i) = 0 \quad \forall i, j$$

or $R\beta = 0$ where $\beta = [\beta_1, \ldots \beta_N]'$, $\beta_i = [\alpha_i, \rho_i]'$ and R is a matrix with $\frac{1}{1-\rho^A}$ and $\frac{\alpha^A\rho^A}{1-\rho^A}$ in the appropriate positions. Given that the posterior of $\beta|\sigma$ is normal, the posterior of $R\beta|\sigma$ is also normal. Let σ take on a finite number of values and let the probability $P(\sigma_i) = \theta_i$. Recall that $\sigma = 0$ corresponds to the hypothesis that the posterior steady states are the same. To compute the posterior odds ratio we then have to compute the density of $R\beta|\sigma = 0$ and the weighted sum of densities of $R\beta|\sigma \neq 0$.

The posterior odds (PO) ratio can be written as:

$$PO = 2 \cdot \log\left(\frac{1-\alpha}{\alpha}\frac{\phi(\chi)}{\Phi(\chi)}|V|^{-\frac{1}{2}}\right) \tag{29}$$

where α is the prior probability on the alternative hypothesis and $(1-\alpha)$ is the prior probability on the null hypothesis, $\phi(\chi)$ is the standard normal density of $\chi = R\beta|\sigma = 0$ $\Phi(\chi)$ a weighted average of standard normals at all points but χ, and V is the covariance matrix of the estimates the vector of steady states.

Two points need to be made: first, by selecting $\alpha < 1$, we are implicitly placing higher prior probability on the null hypothesis since α is spread over many possible alternative values. Second, since in this study we are dealing with small samples, we explicitly include

V in the criterion function. Asymptotically, $\log |V|^{-\frac{1}{2}}$ behaves like $\frac{1}{n\sqrt{T}}$ and is therefore negligible. By including it directly in PO, we take a stand on the fact that in our samples the estimates of the steady states may differ substantially from those obtained in large samples.

An alternative way of examining the equality of estimated steady states is to ask what is the largest (ex-post) prior probability that could be imposed on the alternative for the test to accept the null, given the data. Such a prior probability, which we denote by π^*, can be computed from (29) as:

$$\pi^* = \frac{1}{1+\exp(w)} \tag{30}$$

$$w = \log |V| + 2 * \log(\Phi(\chi)) + (n-1) * \log(2\pi) + \chi^2 \tag{31}$$

Finally, let $L(y|\sigma_\nu, \sigma_\eta)$ the best possible outcome under the alternative and $L(y|\sigma_\nu = \sigma_\eta = 0)$ the likelihood under the null. The (prior) likelihood test we perform is:

$$LIK = 2(\log(L(y|\sigma_\nu, \sigma_\eta)) - \log(L(y|\sigma_\nu = \sigma_\eta = 0)) \rightarrow \chi^2(2) \tag{32}$$

Rejection of the null in favor of the alternative indicates two important facts. First, that the best value of σ_ν and σ_η under the alternative are significantly more probable than those under the null, given available data. Therefore, estimates a and ρ implied by the alternative 'fit' the data better. Second, that the nonlinear combination of a and ρ determining the steady state obtained under the alternative is more likely than the nonlinear combination of a and ρ implied by the null from the point of view of the data.

References

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

[2] Barro, R.J. (1991), "Economic Growth in a Cross-Section of Countries", *Quarterly Journal of Economics*, CVI, 407-455.

[3] Barro, R.J. and Lee (1994), "Sources of Economic Growth", *Carnegie Rochester Conference Series on Public Policy*, 40, 1-46.

[4] Barro, R.J. and X. Sala-i-Martin (1991), "Convergence Across States and Regions", *Brookings Papers on Economic Activity*, II, 107-158.

[5] Barro, R.J. and X. Sala-i-Martin (1992), "Convergence", *Journal of Political Economy*, 100, 223-251.

[6] Berger, (1985) *Statistical decision Theory and Bayesian Analysis*, Springer Verlag.

[7] Bernard, A. and Durlauf, S. (1996), "Interpreting Tests of the Convergence Hypothesis", *Journal of Econometrics*, 71, 161-173.

[8] Blanchard, O. and Katz, L. (1992) "Regional Evolutions", *Brookings Paper on Economic Activity*, 1, 1-75.

[9] Caselli, F., Esquinel, G. and Lefort, F. (1996) "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics", *Journal of Economic Growth*, 1, 363-389.

[10] Chamberlin, G. (1984), "Panel Data", in Chapter 22 of the *Handbook of Econometrics*, Volume II, Z. Griliches and M. Intrilligator, eds., Amsterdam: North Holland Publishing Company.

[11] Durlauf, S. (1993), Nonergodic Economic Growth, *Review of Economic Studies*, 60, 349-366.

[12] Durlauf, S. and Johnson, P., (1995), "Multiple Regimes and Cross-Country Growth Behavior", *Journal of Applied Econometrics*, 10, 97-108.

[13] Evans, P. (1997), "How Fast Do Economies Converge", *Review of Economic and Statistics*, LXXIX, 219-226.

[14] Easterly, W., Kremer, M. Pritchett, L. and Summers, L. (1993), "Good Policy or Good Luck?", *Journal of Monetary Economics*, 32, 459-483.

[15] Garcia Ferrer, A., Highfield, R, Palm, F. and Zellner, A. (1987), "Macroeconomic Forecasting Using Pooled International Data", *Journal of Business and Economic Statistics*, 5, 53-67.

[16] Hsiao, C. (1986), *Analysis of Panel Data*, Econometric Society Monographs, Cambridge University Press.

[17] Judge, G, Griffith, W, Hill, R., Lutkepohl, and Lee, T, (1985), *The Theory and Practice of Econometrics*, New York, Wiley & Sons.

[18] Leamer, E. (1979), *Specification Searches*, New York, John Wiley & Sons.

[19] Le Pen, Y. (1996), "Estimating a Conditional Convergence Equation from Dynamic Heterogeneous Panel Model", University of Nantes, manuscript.

[20] Levine, R. and Renelt, D. (1992), "A Sensitivity Analysis of Cross Country Growth Regressions", *American Economic Review*, 82, 942-963.

[21] Lindley, D. and Smith, A. (1972), "Bayes Estimates for the Linear Model", *Journal of The Royal Statistical Society*, B 34, 1-41.

[22] Mankiw, N, Romer, D. and Weil, D., (1992), "A contribution to the Empirics of Economic Growth", *Quarterly Journal of Economics*, CVII, 407-437.

[23] Marcet, A. (1991); "Un Model de Sèries Temporals per a la Previsió de Recaptació Tributària per Comunitats Autònomes" ("A Time Series Model to Forecast Tax Collections in Autonomous Comunities"), in *El Finançament de les Comunitats Autònomes: Avaluació del Sistema Actual i Criteris per a la Seva Reforma*, ed. X. Calsamíglia, J.M. Esteban i C. Ponsati, Generalitat de Catalunya.

[24] Parente, S. and Prescott, E. (1993), "Changes in the Wealth of Nations", *Federal Reserve of Minneapolis, Quarterly Review*, Springs, 3-16.

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

[26] Quah, D. (1993), "Galton's Fallacy and Tests of the Convergence Hypothesis", *The Scandinavian Journal of Economics"*, 95, 427-443.

[27] Quah, D. (1994), "Empirical Cross-Section Dynamics in Economic Growth", *European Economic Review*, 37, 427-443.

[28] Quah, D. (1996), "Empirics for Economic Growth and Convergence", *European Economic Review*, 40, 1353-1376.

[29] Rivera-Batiz, L. (1993), "Convergence of Regional Income in China", UPF, working paper.

[30] Sala-i-Martin, X. (1995), "Regional Cohesion: Evidence and Theories of Regional Growth and Convergence", *European Economic Review*, 40, 1325-1352.

[31] Schotman, and Van Djik, H (1991), "A Bayesian Analysis of the Unit Root in Real Exchange Rates", *Journal of Econometrics*, 49, 195-238.

[32] Shioji, E. (1993), "Regional Growth in Japan", Yale University mimeo.

[33] Sims, C.A. (1988), "Bayesian Skepticism on unit Root Econometrics", *Journal of Economic Dynamic and Control*, 12, 463-474.

[34] Sims, C.A. and H. Uhlig (1991), "Understanding Unit Rooters: A Helicopter Tour", *Econometrica*, 1591-1600.

[35] Smith, A. (1973), "A General Bayesian Linear Model", *Journal of the Royal Statsitical Society*, ser B, 35, 67-75.

[36] Summers R. and A. Heston (1991),"The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988", *Quarterly Journal of Economics*, 106, 327-368.

[37] Zellner , A. and Hong, C. (1989), "Forecasting International Growth Rates Using Bayesian Shrinkage and Other Procedures", *Journal of Econometrics*, 40, 183-202.

Table 2: Average Estimated Parameters

Prior Parameters		European Regions			European Countries		
σ_ν	σ_η	Mode	ρ	ω	Mode	ρ	ω
.000001	.000001	3337.7	0.9848	-0.52	1410	0.9910	0.086
.000001	.100000	3364.0	0.9914	-0.94	1433	0.9832	0.418
			(0.018)	(0.541)		(0.012)	(0.455)
.000001	1.00000	3390.8	0.9840	-0.52	1439	0.9738	0.418
			(0.103)	(0.730)		(0.024)	(0.455)
.100000	.000001	3402.1	0.9401	-0.22	1447	0.9606	-0.056
				(0.390)			(0.415)
.100000	.100000	3404.1	0.9445	-0.26	1450	0.9551	-0.098
			(0.010)	(0.450)		(0.007)	(0.400)
.100000	1.00000	3438.0	0.9599	-0.47	1460	0.9398	-0.098
			(0.114)	(0.680)		(0.039)	(0.400)
1.00000	.000001	3412.8	0.8718	-0.16	1451	0.9296	-0.056
				(0.340)			(0.415)
1.00000	.100000	3419.4	0.8404	-0.17	1455	0.9211	-0.098
			(0.015)	(0.350)		(0.119)	(0.400)
1.00000	1.00000	3507.8	0.7762	-0.29	1471	0.8896	-0.098
			(0.207)	(0.870)		(0.085)	(0.400)
∞	∞	3485.9	0.7250	-0.31	1470	0.8805	-0.098
			(0.296)	(0.920)		(0.094)	(0.402)
1.0365	0.9841	3509.2	0.7699	-0.29	—	—	—
			(0.199)	(0.880)			
1.0060	0.9780	—	—	—	1472	0.8808	-0.098
						(0.097)	(0.400)

Notes: The basic model is given in equations (10) and (13)-(14)-(15) The sample is 1980-1992 for Regional data and 1950-1985 for Country data. σ_ν and σ_η are the standard deviations of the prior restrictions. ρ and ω are the average estimate of the speed of adjustment and of the steady state across 144 regions (Regio data) or 17 countries (Heston and Summers data) and Mode is the value of the log-posterior mode under the particular prior restriction. The dispersion (standard deviation) of the coefficients across units is in parenthesis. The row with $\sigma_\nu = \sigma_\eta = \infty$ corresponds to OLS estimates unit by unit and the one with $\sigma_\nu = \sigma_\eta = 0.00001$ corresponds to pooled estimates. The last two rows report ML estimates of σ_v and σ_η.

Table 3: **Equality of Steady States**

	Posterior Odds	π^*	Likelihood Ratio
	European Regions		
1980-1992 sample	-645.23	0.0000	0.0000
	European Countries		
1950-1985 sample	-92.78	0.0000	0.0000
1950-1965 sample	-99.54	0.0000	0.9651
1966-1985 sample	-25.71	0.0000	0.0000
1950-1979 sample	-79.73	0.0000	0.9495

Notes: The (Small Sample) Posterior Odds criteria is defined in equation (29), π^* is defined in equation (30). For the Posterior Odds ratio, the prior odds ratio is 1.0. In the column Likelihood Ratio we report the p-value of the test defined in (32).

Table 4: **Test of Persistence in Inequalities**

Regressors	**European Regions**	**European Countries**							
	80-92 Sample	**50-85 Sample**		**50-65 Sample**		**66-85 Sample**		**50-79 Sample**	
Constant	-0.04	-0.05	-0.36	-0.07	-0.10	-0.01	0.09	0.003	0.003
	(0.04)	(0.06)	(0.37)	(0.09)	(0.50)	(0.02)	(0.20)	(0.04)	(0.30)
Initial	0.78	0.51	0.47	0.33	0.29	0.77	0.76	0.06	0.05
Conditions	(0.11)	(0.15)	(0.14)	(0.15)	(0.11)	(0.07)	(0.07)	(0.09)	(0.10)
Secondary			0.45		0.63		0.20		0.16
Education			(0.40)		(0.86)		(0.24)		(0.44)
I/Y			0.003		0.01		-0.002		0.008
			(0.006)		(0.009)		(0.005)		(0.006)
Government			0.48		-3.29		-0.64		-1.58
Share			(1.54)		(1.36)		(1.36)		(0.87)
$\bar{R}^2$	0.33	0.47	0.40	0.10	0.13	0.86	0.89	-0.04	-0.09

Notes: The dependent variable of the regression is the estimated steady state computed as $SS^i = \frac{\hat{\alpha}}{1-\hat{\rho}}$. The I/Y variable is from the appendix of Mankiw, Romer and Weil (1992). The Secondary Education and the Government Share variables are from Barro (1991). Posterior standard errors are in parenthesis.

Table 5: Average Estimated Parameters

σ_ν	σ_η	ρ	ω	Mode
		European Countries, Sample 1950-1965		
.000001	.000001	0.9902	-0.101	517.0
.000001	.100000	0.9724 (0.029)	0.156 (0.413)	488.9
.000001	1.00000	0.9361 (0.118)	0.347 (1.019)	287.8
.001000	.000001	0.8871	-0.09 (0.478)	518.9
.001000	.100000	0.8762 (0.032)	-0.05 (0.449)	-207.3
.001000	1.00000	0.8268 (0.145)	-0.06 (0.449)	-1757.0
∞	∞	0.7875 (0.179)	-0.06 (0.451)	-6452
.000002	.000001	0.9886	-0.101 (0.00001)	519.9
		European Countries, Sample 1966-1985		
.000001	.000001	0.9915	0.245	844.6
.000001	.100000	0.9833 (0.019)	0.270 (0.156)	856.6
.000001	1.00000	0.9802 (0.026)	0.270 (0.156)	857.7
.100000	.000001	0.8856	-0.044 (0.407)	870.0
.100000	.100000	0.8830 (0.029)	-0.076 (0.409)	873.7
.100000	1.00000	0.8805 (0.092)	-0.076 (0.409)	877.1
1.00000	.000001	0.8482	-0.044 (0.407)	870.1
1.0000	.100000	0.8415 (0.035)	-0.076 (0.409)	875.2
1.0000	1.00000	0.8316 (0.108)	-0.076 (0.409)	879.8
∞	∞	0.8294 (0.106)	-0.076 (0.411)	877.2
1.3270	1.0290	0.8301 (0.102)	-0.075 (0.409)	880.2

European Countries, Sample 1950-1979

σ_ν	σ_η	ρ	ω	Mode
.000001	.000001	0.9870	0.011	781.37
.000001	.100000	0.9557 (0.436)	0.365 (0.342)	511.31
.000001	1.00000	0.9557 (0.437)	0.366 (0.344)	509.9
.100000	.000001	0.9274	-0.06 (0.407)	577.5
.100000	.100000	0.8685 (0.112)	0.06 (0.543)	-669.8
.100000	1.00000	0.8683 (0.121)	0.07 (0.545)	-672.5
1.00000	.000001	0.9274	-0.06 (0.407)	577.5
1.00000	.100000	0.8684 (0.121)	0.06 (0.543)	-671.4
1.00000	1.000000	0.8682 (0.121)	0.07 (0.545)	-674.1
∞	∞	0.8682 (0.120)	0.08 (0.545)	-675.7
0.00035	0.00032	0.9828 (0.001)	0.0001 (0.271)	784.36

Notes: The basic model is given in equations (10) and (13)-(14)-(15). σ_ν and σ_η are th standard deviations of the prior restrictions. ρ and ω are the average estimate of the speed c adjustment and of the steady state across 17 countries (Heston and Summers data) and Mod is the value of the log-posterior mode under the the particular prior restriction. The dispersio (standard deviation) of the coefficients across units is in parenthesis. The row with $\sigma_\nu = \sigma_\eta = \infty$ corresponds to OLS estimates unit by unit and the one with $\sigma_\nu = \sigma_\eta = 0.00001$ corresponds t pooled estimates. The last row of each panel reports ML estimates.

Table 6: Average Estimated Parameters
European Regions, Sample 1980-1992
Regional Income scaled by Country Means

σ_ν	σ_η	ρ	ω	Mode
.000001	.000001	0.9886	-0.19	3866.6
.000001	.100000	0.9900 (0.019)	-0.29 (0.17)	3908.5
.000001	1.00000	0.9420 (0.147)	-0.29 (0.50)	3936.7
.100000	.000001	0.9054	-0.07 (0.26)	4013.2
.100000	.100000	0.9007 (0.091)	-0.07 (0.27)	4017.4
.100000	1.00000	0.8684 (0.155)	-0.13 (0.29)	4057.6
1.00000	.000001	0.8611	-0.07 (0.24)	4039.5
1.00000	.100000	0.8224 (0.012)	-0.06 (0.24)	4056.5
1.00000	1.000000	0.6820 (0.169)	-0.08 (0.26)	4143.8
∞	∞	0.6334 (0.310)	-0.08 (0.27)	4190.4

Equality of Steady States

	Posterior Odds	π^*	Likelihood Ratio
Overall	-80.93	0.0000	0.00001
Belgium	-3.54	0.0000	
Germany	-29.26	0.0000	
Greece	-17.95	0.0000	
Spain	-21.08	0.0000	
France	-18.86	0.0000	
Italy	-18.95	0.0000	
Netherland	-13.86	0.0000	
Portugal	-4.76	0.0002	
UK	-17.00	0.0000	

Persistence in Inequalities

Constant	Initial Conditions	R^2
0.05 (0.04)	0.79 (0.23)	0.11

Notes: The basic model is given by (10) and (13)-(14)-(15). σ_ν and σ_η are the standard deviations ‹ the prior restrictions. ρ and ω are the average estimate of the speed of adjustment and of the steady stat across 144 regions (Regio data) and Mode is the value of the log-posterior mode under the particular pri restriction. The dispersion (standard deviation) of the coefficients across units is in parenthesis. Tł row with $\sigma_\nu = \sigma_\eta = \infty$ corresponds to OLS estimates unit by unit and the one with $\sigma_\nu = \sigma_\eta = 0.0000$ corresponds to pooled estimates. The tests for unconditional convergence with country names refer t the hypothesis that regions of the same country converge to the same steady state.

Table 7: Effects of Measurement Error in y_t^i

ρ	$\hat{\rho}$	μ	$\frac{\sigma_u}{\sigma_e}$
0.98	0.77	0.00	2.62
0.98	0.77	0.20	2.98
0.98	0.77	0.30	3.20
0.98	0.77	0.40	3.46
0.98	0.77	0.50	3.83
0.98	0.77	0.60	4.46
0.98	0.77	0.70	6.21
0.98	0.88	0.00	1.69
0.98	0.88	0.20	1.88
0.98	0.88	0.30	1.99
0.98	0.88	0.40	2.10
0.98	0.88	0.50	2.23
0.98	0.88	0.60	2.40
0.98	0.88	0.70	2.67

Notes: The model is $\hat{y}_t^i = y_t^i + \omega_t^i$ where $y_t^i = \rho y_{t-1}^i + u_t^i$, $\omega_t^i = \mu\omega_{t-1}^i + e_t^i$ and $E(e_t^i, y_t^i) = 0$ The table reports for a given ρ what is the variability of the innovation in y_{it} relative to the variability in the innovation in measurement error which is needed, for different values of μ, to get the $\hat{\rho}$ we obtain in the data.