DISCUSSION PAPER SERIES

No. 1265

THE POOR STAY POOR: NON-CONVERGENCE ACROSS COUNTRIES AND REGIONS

Fabio Canova and Albert Marcet

INTERNATIONAL MACROECONOMICS

Centre for Economic Policy Research

ISSN 0265-8003

THE POOR STAY POOR: NON-CONVERGENCE ACROSS COUNTRIES AND REGIONS

Fabio Canova and Albert Marcet

Discussion Paper No. 1265
November 1995

Centre for Economic Policy Research
25–28 Old Burlington Street
London W1X 1LB
Tel: (44 171) 734 9110

CEPR Discussion Paper No. 1265

November 1995

ABSTRACT

The Poor Stay Poor: Non-Convergence Across Countries and Regions*

We study the issue of income convergence across countries and regions with a Bayesian model which allows us to use information in an efficient and flexible way. We argue that the very slow convergence rates to a common level of per-capita income found, for example, by Barro and Sala-i-Martin, is due to a 'fixed effect bias' that their cross-sectional analysis introduces in the results. Our approach permits the estimation of different convergence rates to different steady states for each cross-sectional unit. When this diversity is allowed, we find that convergence of each unit to (its own) steady-state income level is much faster than previously estimated, but that cross-sectional differences persist: inequalities will only be reduced by a small amount by the passage of time. The cross-country distribution of the steady state is largely explained by the cross-sectional distribution of initial conditions.

JEL Classification: C11, C23, D90, O47
Keywords: convergence, income inequalities, persistence, panel data, prior distribution

Fabio Canova and Albert Marcet
Department of Economics
Universitat Pompeu Fabra
Balmes 132
08008 Barcelona
SPAIN
Tel: (34 3) 542 2601/1700

*This paper is produced as part of a CEPR research programme on *Market Integration, Regionalism, and the Global Economy*, supported by a grant from the Ford Foundation (no. 920-1265). We would like to thank Steve Durlauf, Paul Evans, Angel de la Fuente, Ana Lamo, Robert Litterman, Teresa Garcia-Mila, Danny Quah, Xavier Sala-i-Martin and Fabrizio Zilibotti, the participants of seminars at Ohio State University, Boston College, European University

Institute, Universitat Pompeu Fabra, Università di Catania, University of Aarhus, CEPR's conference on 'Empirical Macroeconomics: What Methods and for What Problems' and the 1995 SEDC Meetings in Barcelona for very helpful comments and suggestions, and Rosa M Fernandez for help in organizing the data. An early version of this paper was prepared for the FEDER project 'The Effects of the Unified Market on Regional Development' administered by the Institut d'Anàlisi Econòmica (Universitat Autonoma de Barcelona). Financial support from DGICYT and CIRIT grants is gratefully acknowledged. Part of Marcet's research was undertaken while he was a visitor at the Federal Reserve Bank of Minneapolis.

Submitted 25 September 1995

NON-TECHNICAL SUMMARY

The issue of convergence of per-capita incomes across economic areas is an old one. Are income differences across countries and regions disappearing as time goes by? Do poor regions stay poor? This issue has been placed at the forefront of economic research in recent years; for example, by Barro and Sala-i-Martin (1991,1992). They analyse the available data with cross-section regressions to conclude that convergence occurs for a cross-section of countries, US states or even European regions, and at the very slow rate of approximately 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 by proxies for human capital and government policy. A large literature exploring these issues in different data sets and with different statistical methods has ensued, but the main results of Barro and Sala-i-Martin have, by and large, been confirmed. The main conclusions and estimates are even consistent for regions with very different political and economic systems to those in Europe or the United States. These results tend to 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 over time.

Typically, the convergence literature explains income growth for each unit (either country or region) by aggregating growth rates over the sample period, and then performing a cross-section regression with one observation per unit. This 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; and third, it forces the use of definitions of convergence that are not appropriate for discussing persistence of inequality.

In this paper we propose a Bayesian procedure to estimate convergence rates and steady states that uses the information available for all periods and all cross-sectional units. Our prior distribution is based on the belief that the parameters of the statistical model in different units have 'similar' but not necessarily identical values. This ensures efficient use of all information available without imposing unrealistic assumptions that may cause various types of bias. Once steady-state estimates are obtained for each unit we can test, in a second step, what variables determine the cross-sectional distribution of steady states. In particular, we can examine whether initial conditions or other factors explain the dispersion of estimated steady states and, in this manner, perform a meaningful test of persistence in inequality. This estimation strategy is applied to two data sets: yearly per-capita income from the Region data set of Eurostat, and from OECD European countries of the Summers and Heston data set.

Three major findings arise from our analysis. First, average estimates of the convergence rate are much higher than those found in the literature; approximately 11% for countries and 23% for regions, with each unit converging to its own steady state. These estimates imply a capital share in a neoclassical production function of around 0.20–0.35. Second, the hypothesis that the steady state is the same for all cross-sectional units is rejected by the data, both for regions and countries. Third, the initial income conditions are, by far, the most important determinant of the cross-sectional dispersion of steady states. Poorer regions and countries stay poor and, over time, differences are reduced by only a small amount.

We also find that when the prior forces all parameters to be exactly equal in all units (a case that most closely resembles the cross-sectional growth regressions), a systematic distortion emerges that causes the averaged estimated convergence rate to be biased downwards, and surprisingly, of the order of 2%. We argue that previous estimates at 2% are due to a fixed effects bias, well known in the panel data literature; such a value is mechanically obtained from the data when observations from heterogeneous units are pooled as if their data generating process were the same.

Our conclusions differ somewhat from existing convergence literature: a poor region can expect the gap between its initial level of income and the aggregate to be reduced by only 30–40% in the limit. Hence, current redistribution and development policies, such as the Regional and Cohesion Fund Policies carried out by EC governments, are only partially working; rich regions can be taxed more heavily in favour of poor regions 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 regions unless structural changes occur in the economic environment.

Era bella, di una dura bellezza bruna che nessunno 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 differences across countries and regions disappearing as time goes by? Do poor regions stay poor? 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 cross section of countries, US states or even European regions and that it happens 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 by proxies of 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 results 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.

Typically, the convergence literature attempts to explain income growth for each unit (either country or region) by aggregating growth rates over the sample period, and then performing a cross-section regression with *one* observation per unit. 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 definitions of convergence that do not capture closely the idea of persistence of inequality.

[1]See, for example, Mankiw, Romer and Weil (1992), Barro and Lee (1994), Sala-i-Martin (1995) or Durlauf and Johnson (1994). 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)).

In this paper we provide an alternative definition of convergence that allows us to analyzes the evolution across time of per capita income and propose a Bayesian procedure to estimate convergence rates and the steady states that uses the information available for all periods and all cross sectional units. Our prior distribution is based on the belief that the parameters of the statistical model in different units have "similar" but not necessarily identical values. This allows an efficient use of all information available without imposing unrealistic assumptions that may cause various types of biases. Once steady state estimates are obtained for each unit we can test, in a second step, what variables determine the cross sectional distribution of steady states. In particular, we can examine whether initial conditions or other factors explain the dispersion of estimated steady states and, in this manner, perform a meaningful test of persistence in inequality. This estimation strategy is applied to two data sets: yearly per capita income of European regions, and of OECD European countries.

Three major findings arise from our analysis:

- Average estimates of the convergence rate are much higher than those found in the literature: about 11% for countries and 23% for regions, with each unit converging to its own steady state. These estimates imply a capital share in a neoclassical production function of the order of 0.20-0.35.
- The hypothesis that the steady state is the same for all cross sectional-units is rejected by the data, both for regions and countries.
- The initial income conditions are, by far, the most important determinant of the cross sectional dispersion of steady states. Poorer regions and countries stay poor; over time, 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 averaged estimated convergence rate to be biased downward and, surprisingly, of the order of 2%. Then, we explain the previous estimates of 2% as arising from a fixed effects bias, well known in the panel data literature (see e.g. Hsiao (1985) and, more recently, with a different flavor, Pesaran and Smith (1995)); such a value is mechanically obtained from the data 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 present 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 hypothesis. Our definition of convergence is related to that of Bernard and Durlauf (1994) in that it focuses on the evolution across time of the expectation 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 convergence rate may have been underestimated by BS is also made by Evans (1995) 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 previously presented in the literature. Section 3 discusses the statistical model and the Bayesian estimation and testing strategy. Section 4 describes the data. Section 5 discusses estimates of the rate of convergence across specifications, tests of equality of estimated steady states and of persistent inequality. Section 6 examines possible sources of misspecification and econometric biases which may affect the essence of the results. Section 7 concludes.

2 A Definition of Persistence in Inequality

The issue we are interested in studying is: *is there a tendency for the income of initially poor units to become similar, on average, to the income of initially rich units as time passes? or is it the case that initially poor stay poorer than the rest?* In the former case we would say there is convergence, in the latter that there is persistence of inequality.

To properly state the issue at stake, we first provide a definition of persistence of inequality and of convergence that most closely formalizes the above ideas. We assume that observations 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 $t = 0, 1, \ldots$ indexes time. The set I can be the first n integers, the unit interval, etc. The initial values $\{Y_0^i\}_{i=1}^n$ are assumed to be random variables. 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; in section 3 we will argue that modelling this variable has advantages from both theoretical and econometric point of views.

Let $w^i = lim_{t\to\infty} E_0 y_t^i$, where the limit is assumed to exist. Notice that w^i is a random variable indexed by i alone.

Definition 1 $\{Y_t^i\}$ *displays unconditional persistence of inequality if the function f defined as*

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

is monotonically increasing.

Definition 2 $\{Y_t^i\}$ *displays persistence of inequality, conditional on variables X^i if the function f defined as*

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

is such that $f(., X_i)$ is monotonically increasing for all possible values of X_i.

Notice that the expectation in (1) and (2) is taken with respect the cross sectional distribution.

The first definition implies that initially rich units are expected to stay relatively rich, regardless of their specific characteristics, while conditional persistence in inequality allows for factors other than the initial conditions to affect income. The fact that in the definition of w_i we deal with expectations (as t gets large) disregards differences in units' income due to temporary shocks, and concentrates instead on differences that persist through time on average. Also, the fact that in (1) and (2) we deal with expectations across units allows for some units that started our poor to become rich (see Parente and Prescott (1993). Hence the presence of business cycles or of 'economic miracles' does not prevent persistence of inequality.

A corresponding definition of unconditional (conditional) *convergence* states that the function f in (1) (the function $f(., X)$ in (2)) is equal to zero. Obviously, whenever we have persistence in inequality convergence fails and viceversa, but it is possible to find stochastic processes that display no convergence and no persistence of inequality.

2.1 An example and comparison of definitions.

There are many definitions of convergence available in the literature; each of them meaning different things and focusing on different aspects of the evolution of the distribution of income. Here we show by means of an example the relationship between previous definitions of convergence and our definition. We will see that those definitions do not allow for the kind of distinction that we want to study.

Consider the process

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

where ν, ρ are given constants, $\{\epsilon_t^i\} \sim i.i.d(0, \sigma_\epsilon^2)$ across time and units, and initial conditions $\{y_0^i\}_{i=1}^n$ are given. This model allows for initial conditions to influence the whole future through the parameter ν; the parameter ρ captures the dependence on the recent past, and allows for business cycles kind of variations.

It is easy to check that, if $|\rho| < 1$, the long run forecast of y_t^i is

$$E_0\left(y_t^i\right) \rightarrow \frac{\nu y_0^i}{1-\rho} \qquad \text{as } t \rightarrow \infty. \tag{4}$$

This formula already advances that, if $\nu > 0$, initial conditions affect the mean in the indefinite future and there is persistence of inequality. For the rest of this section, we determine for what parameter values we have convergence in the process (3) under alternative definitions of convergence.

As previously suggested, there are processes for which there is neither persistence of inequality nor convergence. This is the case for model (3) if $|\rho| < 1$ and $\nu < 0$ or if $\rho < -1$ [2]. In order to obtain an unambiguous answer, we assume that $\nu \geq 0$ and $\rho > -1$ for the rest of this section.

2.1.1 Our Definition

First, consider the case $|\rho| < 1$. It is clear from equation (4) that, if $\nu > 0$, there is persistence of inequality and if $\nu = 0$ there is convergence. The fraction $\frac{\nu}{1-\rho}$ is the proportion of initial income

[2] Notice that, if $|\rho| < 1$ and $\nu < 0$, equation (4) shows that we would have neither convergence nor persistence of inequality; for these parameter values, this is a model of the biblical prophecy 'the last will be first'. On the other hand, if $\rho < -1$, there is a deterministic cycle: if region i is such that $y_0^i > 0$, then $y_t^i \rightarrow \infty$ in odd periods and $y_t^i \rightarrow -\infty$ in even periods.

that is kept forever on average; inequality is reduced as time goes by if this fraction is less than one, otherwise inequality increases.

When $\rho = 1$, we have $E_0(y_t^i) = (t\nu + 1)y_0^i$. When $\rho > 1$, $E_0(y_t^i)$ goes to plus (minus) infinity if y_0^i is positive (negative). Therefore, if $\rho \geq 1$, there is persistence of inequality for all ν. Furthermore, inequality increases in all cases except when $\rho = 1$ and $\nu = 0$.

To summarize, convergence obtains only when $\nu = 0$ and $|\rho| < 1$ are both satisfied; in all other cases considered there is persistence of inequality. This situation is summarized in the first row of table 1.

2.1.2 σ-convergence

Let the dispersion of y be defined as $\Sigma_t \equiv \frac{1}{n}\sum_{i=1}^{n}(y_t^i)^2$, σ-convergence obtains if $\Sigma_t \leq \Sigma_s\ \forall 0 \leq s < t$. In the process defined by equation (3), for n arbitrarily large we have

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

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

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

where $\sigma_y^2 = \frac{\sigma_\epsilon^2}{1-\rho^2}$. This formula says that Σ_t is a weighted average of the initial dispersion Σ_0 and the limiting dispersion $\frac{\nu^2\Sigma_0}{(1-\rho)^2} + \sigma_y^2$. Hence, σ-convergence obtains if and only if

$$\Sigma_0 > \frac{\nu^2\Sigma_0}{(1-\rho)^2} + \sigma_y^2 \quad (7)$$

Clearly, if equation (7) is satisfied or not strongly related to persistence of inequality: (7) may be satisfied when $\nu > 0$ (as long as $|\frac{\nu}{1-\rho}| < 1$, and Σ_0 is sufficiently large), and it may fail when $\nu = 0$ (as long as σ_y^2 is sufficiently large)[3]. In addition, if $|\frac{\nu}{1-\rho}| > 1$ the above inequality fails, so that we do not have σ-convergence. Finally, (5) shows that σ-convergence fails when $\rho \geq 1$.

This situation is summarized in the second row of table 2.1.6.

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

2.1.3 β-convergence

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

$$y_T^i = \beta y_0^i + \eta^i \tag{8}$$

where $\eta^i \equiv \sum_{j=0}^{T-1} \rho^j \epsilon_{t-j}^i$ and $\beta \equiv \rho^T + \nu \sum_{j=0}^{T-1} \rho^j$. Equation (8) is the one estimated in the cross-section approach of BS.

Clearly, units that start out below average (i.e., $y_0^i < 0$) have a higher income relative to other units after T periods if $\beta < 1$; in this case, poorer units grow faster. Therefore, β-convergence obtains if $0 < \beta < 1$, and β-convergence fails if $\beta \geq 1$.

When $|\rho| < 1$, we have that $\beta < 1$ for T large enough if and only if $\frac{\nu}{1-\rho} < 1$. Therefore, we can have β-convergence coexisting with persistence of inequality when $\nu > 0$. On the other hand, it is clear from (8) that if $\rho \geq 1$, β-convergence and convergence in our sense both fail simultaneously. 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 Unit root convergence

Bernard and Durlauf (1995) 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, in the case when $\nu > 0$ and $0 < \rho < 1$ there is persistence of inequality and no unit roots. Furthermore, their definition only allows for pairwise comparisons. The fourth row of table 1 describes this case

2.1.5 Conditional β-convergence

Finally, we discuss the concept of conditional β-convergence. The idea is to test if poorer units grow faster after conditioning for certain observed variables X^i. This hypothesis is often tested, with a sample of T years, by running a cross-section regression of the form

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

for $i = 1, \ldots, n$. Notice that, with this approach, 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 $t = 1, \ldots, T-1$ (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 $\gamma = 0$ will be rejected even when it holds true because X_i are correlated with the residuals. Therefore, it is possible to accept conditional convergence ($\gamma \neq 0$ and $\beta < 1$) even though there is persistence in inequality.

2.1.6 Summary

Table 1 summarizes the cases for which different definitions would generate convergence for all possible values $\nu \geq 0$ and $\rho > -1$ in the example considered. The sign $\times$ indicates that convergence obtains; an empty box indicates no convergence; in the cases where conditions parameters other than ν and ρ affect convergence, this is indicated. The row for β-convergence is valid for T large enough.

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

	$\lvert\rho\rvert < 1$			$\rho = 1$	$\rho > 1$
	$\nu = 0$	$0 < \frac{\nu}{1-\rho} < 1$	$\frac{\nu}{1-\rho} > 1$		
our definition	$\times$				
σ-convergence	$\times$ if $\Sigma_0 < \sigma_y^2$	$\times$ if (7)			
β-convergence	$\times$	$\times$			
Unit root convergence	$\times$	$\times$	$\times$		$\times$

The table shows that various definitions are not strongly related to ours. Only in the fourth column there is complete agreement. The concept of $\beta-$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 tests found relevant. The conclusion is that, even though existing definitions focus on relevant aspects of the data, they are not appropriate for studying persistence in inequality.

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 the residual ϵ_t^i is assumed to have mean zero and to be independent across i's and t's. Using the *proportion* of per-capita income y_t^i as our basic variable, instead of plain per-capita income Y_t^i, alleviates problems of serial and residual cross-unit correlation: since recessions and expansions affect the world economy as a whole, 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) [4]. Appendix 1 provides a setup that formalizes this idea, and shows that the model is consistent with a business cycle shock and a trend that is common to all units; the business cycle variations within each unit are governed by ρ_i and an i.i.d. shock. BS (1992) also consider the possibility that regression residuals contain two components (one aggregate and one idiosyncratic). However, because they use $\log(Y_t^i)$ as the left hand side variable, they introduce proxies that hold constant the effect of aggregate shocks in their cross sectional regressions. Appendix 2 shows that (10) is consistent with the standard neoclassical growth model and the specification used by BS.

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 the per-capita income for all t's will be 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 is done in cross-section regressions. For example, BS (1992) showed that the theoretical rate of convergence depends on the parameters of preferences and technologies which may differ across units. However their empirical analysis constrains the rate of convergence 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 constant a function of observed characteristics of the unit. Because our approach allows the estimation of the steady states directly, we can

[4] We formally tested that these assumptions are satisfied for our data sets and our model specification. We rejected the existence of any correlation of the ϵ's both across time and units.

separately examine the issue of convergence to the steady state from questions concerning the features of the limiting distribution of per-capita-income.

It is straightforward to check that the steady state value of y_t^i (the long run forecast of y_t^i given information at time 0) is $a^i/(1-\rho^i)$ and that $1-\rho^i$ is the rate of convergence of each unit to its own steady state.

The main problem with our model specification is that, typically, there are too many parameters relative to the number of time series observations for each cross sectional unit. Then, if (a^i, ρ^i) are estimated separately using only the observations on unit i, the estimators will have very large standard errors and their small sample distribution may strongly deviate from the asymptotic one [5].

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 solves the small sample problem since Bayesian estimates are exact regardless of the sample size. Also, it 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 distribution we use assumes that the rate of convergence and the intercept of the model do not differ too much across units; more precisely, our prior distribution satisfies

$$(\rho^j - \rho^i) \sim N(0, \sigma_\rho^2) \qquad \forall i, j \tag{11}$$

$$(a^j - a^i) \sim N(0, \sigma_a^2) \qquad \forall i, j \tag{12}$$

Note that (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 on $\beta^i = \{a^i, \rho^i\}$ is left unspecified.

If all σ's are set to zero we are a-priori imposing equality of coefficients across units therefore pooling estimates of the parameters towards their cross sectional mean. Hence, setting all σ's

[5] 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 Chamberlin (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 by IV or GMM procedures.

to zero in (11)-(12) would roughly replicate the cross sectional analysis performed by BS or Mankiw, Romer and Weil (1992) (MRW). On the other hand, if we let the σ's tend to infinity, the β_i are believed a-priori to bear no information for β_j so that 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 both from the cross-section and the time series dimensions of the panel.

The idea of constructing posterior parameter estimates trading-off the information contained in the cross-section and the time series dimension is tightly related to the literature on "exchangeability priors" discussed, e.g., in Linsday 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 of the model [6].

Posterior estimates of the coefficients are easily obtained with an augmented least square procedure after rewriting the model in order to mimic the setup of Theil mixed-type estimator (see e.g. Judge et. al. (1985)). To do so we treat the prior as an additional observation with explanatory variables that take the values 1 or -1 that multiply coefficients; that is, we add to (10) equations like

$$0 = \rho^i \cdot 1 + \rho^{i+1}(-1) + \eta^i \qquad \text{for } i = 1, I-1 \qquad (13)$$

$$0 = a^i \cdot 1 + a^{i+1}(-1) + \nu^i \qquad \text{for } i = 1, I-1 \qquad (14)$$

It turns out that the prior (13)-(14) is equivalent with (11)-(12) and the following structure

[6] The prior distribution used by Garcia-Ferrer et. al. (1987), Zellner and Hong (1989) assumes a correlation structure among η's that greatly simplifies the formula for the posterior mean. This assumption implies that the prior is empirical, and that it is itself determined by the data.

for covariances:

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

Notice that it is sufficient to write the restrictions for adjacent units only, and that it is not correct to assume independence of all η's.

The discussion so far leaves open the question of how to select the $\sigma = [\sigma_\eta, \sigma_\nu]$ parameters which regulate the trade-off between the information contained in the time-series and the cross-section dimensions of the panel. In a standard Bayesian approach one imposes an improper prior on σ and conducts posterior inference given these priors (see e.g. Judge, et. al. (1985)). Rather than taking the Bayesian approach literally, we start the empirical analysis in section 5 by exploring the likelihood generated by different σ's, and explore how different specifications match with the data. This is in the spirit of the specification searches of Leamer (1979) and Sims and Uhlig's (1991) 'helicopter tour'. As an alternative, one could take (13)-(14) as part of the model specification and use the likelihood function $L(y|\sigma)$ as a way to formally estimate σ.

In discussing the issue of unconditional convergence we will be interested in examining whether estimates of the steady state level of relative per-capita income differ across units. That is, we need to examine the null hypothesis that $\hat{a}^i/(1-\hat{\rho}^i) = \hat{a}^j/(1-\hat{\rho}^j)\ \forall i, j$ versus the alternative composite hypothesis that they are different. To test this hypothesis in a manner which is consistent with our Bayesian approach we employ the Posterior Odds ratio (PO) criteria and the Schwarz criteria (see e.g. Leamer (1979) or Sims (1988)). The reader interested in the technical details of the approach may consult appendix 3. Here it is sufficient to note that the PO criteria combines a-priori odds with the likelihood of the data under the null and the alternative and that the PO and the Schwarz criteria are asymptotically equivalent but that the PO is more appropriate for the size of our samples. The null is rejected if the statistics are positive. To provide an alternative point of view we also compute what is 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 ex-post α (denoted by α^*). Small values of this statistics indicate, that unless the alternative is a-priori impossible, the data would always reject the null. Finally, for those who feel uncomfortable with our Bayesian testing approach, 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 dispersion of estimated steady states. To examine whether there is persistence of inequality (either unconditional or conditional) we run cross sectional regressions of the type:

$$\widehat{SS}^i \equiv \frac{\hat{a}^i}{1-\hat{\rho}^i} = \delta + \gamma y_0^i + \omega X^i + u^i \tag{16}$$

where the vector X^i includes, as in BS or MRW, variables proxying for differences in technologies, government policies, and human capital, and δ is the cross sectional mean of the steady state distribution. If we accept the hypothesis that $\gamma > 0$, this indicates that initial levels of income matter for 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 significant $\hat{\gamma}$ suggests that the ordering in the cross sectional distribution is preserved as time goes by but that the gap is eventually very small. At the opposite end, a $\hat{\gamma}$ which is significant and close to one implies persistence in the ordering and in the magnitude of inequality. Finally, a negative $\hat{\gamma}$ indicates the realization of the 'biblical prophecy ', i.e. the steady state income of initially poor will be higher that the steady states of the initially rich. Significance of ω, in addition to significance of γ, suggests that factors other than the initial conditions explain the distribution of the steady states.

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 among economists and is used here 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. Some 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 in our study we use the ratio of regional to aggregate per-capita income there is no need to convert nominal income 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 [7]

Using data at Nuts-2 level is important because in 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 the approach to the per-capita real GDP measured in international prices from Summers and Heston (1991) data set. This data set is well known to students of economic growth. We limit our study 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 the institutions and economic structure, and it makes it more likely that the hypothesis of convergence will be accepted. [8].

[7] BS (1991) sample on European regions uses a longer time span, but it introduces fewer countries and less detail in the regions.

[8] To make sure that the results did not depend on our choice of countries, we also consider a sample composed

5 The Empirical Results

5.1 Rates of Convergence and Unconditional Convergence

Our first set of results is contained in tables 2 and 3. Table 2 reports, for different settings of σ, the average estimates and the cross sectional dispersion of the parameters of the model (10) and the value of the likelihood obtained with various settings of σ. The first panel reports the results obtained with Regio data, the second those obtained with Summers and Heston data. Table 3 reports the values of the statistics used to test the hypothesis that the steady states are the same and the p-value of the likelihood ratio test for the same hypothesis.

Several important facts stand out from the tables. First, by forcing the model to have the same coefficients for each unit (case $\sigma_\eta = \sigma_\nu = 0.000001$), we approximately obtain the eerily ubiquitous average convergence rate of 2% per year with both data sets. Therefore, under this particular set of restrictions, our model reproduces standard cross-sectional regression results.

Second, when we allow for heterogeneity in parameter estimates across units the average rate of convergence increases up to about 23% a year with the Regio data set and about 11% a year with the Summers and Heston data set. For similar OECD countries, using cross sectional regressions, BS and MRW estimated the rate of convergence to be of the order of 1.4-1.8% while Evans (1995), using panel data techniques, finds convergence rates of the order of 6-9% a year. Note that our estimates imply a capital share in the neoclassical production function of 0.20-0.35, a range which is more reasonable than the one obtained by BS or MRW.

It is instructive to provide an intuitive explanation for these results. Consider a situation where the "true" model has different steady states, but similar ρ's, and the steady state is positively correlated with the initial condition across units. Figure 1 represents equation (10) for three regions, under these assumptions, and a likely cloud of points in a finite sample. It is clear that 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 and, equivalently, the convergence rate will be much lower. This phenomenon is well known among microeconometricians as the 'fixed effect bias' and occurs whenever heterogeneity across units is not appropriately accounted for. This bias does not

of 22 OECD countries standardly used in the literature. No substantial changes in the conclusions emerged.

disappear as more time series or cross sectional observations are collected. The pervasiveness of convergence rates around 2% obtained in previous studies may therefore be the result of a biased estimation procedure that ignores fixed effects present in the data.

Overall, when we do not force parameters to be the same across units, the average rate of convergence increases uniformly. 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 dimensions of the panel. To investigate such a trade-off, we examine the results obtained in two intermediate cases ($\sigma_\nu = 1.0, \sigma_\eta = 0.000001$ and $\sigma_\nu = 0.000001, \sigma_\eta = 1.0$) where we are imposing either that the convergence rate is the same for each unit and differences in steady states are solely due to unit specific fixed effects or that there is no unit specific fixed effect and that steady states differ because of different convergent rates.

For the Regio data set, setting $\sigma_\eta = 0.000001$ and $\sigma_\nu = 1.0$ reduces the value of the likelihood relative to the most unrestricted case. However, the convergence rate is still on average about 13% and the dispersion of the posterior estimates of the steady states is still large. When we eliminate the individual effect but we allow the rate of convergence to differ the reduction in the likelihood relative to the most unrestricted case is much larger, the estimated value of a is pushed toward zero and the rate of convergence is only about 1.2% a year. Therefore, our results are consistent with figure 1: forcing the a's to be the same causes a larger distortion then setting the ρ to be the same.

For the 17 European countries the results are similar, although less spectacular quantitatively. Restricting the rate of convergence to be the same across regions causes a drop in value of the likelihood relative to the case $\sigma_\eta = \sigma_\nu = 1.0$ and a drop in the average estimated rate of convergence. However, leaving out individual effects while letting the rate of convergence be country specific makes the magnitude of these drops much larger.

One may wonder if there is a way to judge which choice of the σ vector offers the best fit to the data. If one writes the likelihood of the data conditional on the σ vector, then the higher is the value of the likelihood the more probable is the choice of σ, given the data. Using this criteria, we find that the specification $\sigma_\nu = \sigma_\eta = 1.0$ is to be preferred among those considered in the table with both data sets and the implied values of a and ρ can be considered approximate maximum likelihood estimates of the true parameters.

As mentioned, such estimates of σ_ν and σ_η imply *average* convergence rates which are larger than those previously found in the literature. Panel A in figures 2 and 3 plots convergence rates for each unit in the two data sets under the preferred choice of σ against the initial conditions. Rates of convergence 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. In both data sets there are units for which $\hat{\rho} > 1$, implying divergence of per-capita income. For the Regio data they are primarily regions from France and Germany while in the Summers and Heston data the only case of divergence is represented by Norway. Note also that panel A of both figures also suggests that there is very little relationship between the initial conditions and the rate of convergence.

Third, with the best choice of σ, the dispersion of estimated steady states is substantial. Panel B of figures 2 and 3 provides a histogram of the estimates of the 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 from the pictures that the estimated steady state distribution for the 17 European countries is almost bimodal, while the one for the 144 European regions tends toward normality.

Are differences in the estimated steady states statistically significant? Table 3 indicates that the hypothesis that the estimated steady states are the same for all units is rejected using all testing criteria. Particularly informative is the reported value of α^* (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). In both cases, unless we assume a-priori that the alternative is impossible, the null hypothesis will always be overturned by the data.

Finally, because there are 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 dètected that convergence is less prevalent in the decade of the 80's. 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 down the average rate of convergence and for the specification which maximizes the likelihood, the average estimated rate of convergence is substantially larger then the one found in the literature. 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) also display interesting differences. First, for these two subsamples the "best" specification is one where the rate of convergence across units 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 across units. Second, and as consequence of the above, the estimated distribution of the steady states is non-degenerate so that three of our tests favor the alternative hypothesis that the estimated steady states are different across units. The likelihood ratio test is however unable to reject the null hypothesis as the likelihood is somewhat insensitive to the choice of σ_{ν}. In other words, although the fit of the model, as measured by the peak of the likelihood, improves when there are heterogeneities in the constant, the improvement for these two data set are small. 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 the heterogeneities across countries become more marked. This result is in agreement with those of Blanchard and Katz (1992). More importantly, these heterogeneities turned out to emerge more strongly in convergence rates which, consistent with the results of our Regio data set, became very dissimilar across countries.

In conclusion, our first set of results can be summarized as follows: (i) the vast majority of countries and regions converge to their **own** steady states. Divergence is an important feature of the data only during the 80's where some polarization emerged. (ii) The estimated average rate of convergence varies with the σ vector in a way that is consistent with the fixed-effect bias described in figure 1. For the best choice of σ, the estimated average convergence rate is significantly higher than previously estimated and there is considerable dispersion in the estimates across units. Significant differences in convergence rates emerge from the beginning of the 80's. (iii) Estimated steady states differ across units for both data sets and for subsamples.

5.2 Explaining the Distribution of Steady States

Our results so far indicate that the estimated distribution of steady states is non-degenerate. Next, we proceed to examine which variables account for the cross sectional dispersion in the estimated steady states.

Cross-section regression analyses of convergence allow for differences in the steady states through the effect of 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 argues that these effects are rather small for OECD countries and absent for regions (see e.g. BS (1992)). The results of section 5.1, however, indicate that *some* variable must be having an effect on the level of steady states.

One candidate for the determinant of the limiting distribution of steady states can be found by inspecting their cross sectional dispersion: cross sectional estimates of the steady states tend to be higher (lower) than zero for initially rich (poor) regions. Panel C of figures 2 and 3 plots the estimated steady state against the initial income level for the two data sets we analyze [9]. It is clear from the graphs that the estimated steady states appear to have a strong *positive* connection with the initial conditions.

Table 4 presents the results. Because of the lack of disaggregated data on X^i for European regions, we restrict the attention to the initial conditions with the Regio data set and test only unconditional persistence of inequality. As argued by BS (1992) and Sala-i-Martin (1995), the omission of region specific characteristics may not be crucial since the conditioning variables may be either very similar across regions or unimportant to describe the evolution of steady states. 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 additional variables changes the essence of our results. In this latter case, we consider proxies 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 from Barro (1991)).

[9] Whenever $\hat{\rho} > 1$ we compute steady states using the small sample formula $\hat{SS}^i = a * \frac{1-\hat{\rho}^{T+1}}{1-\hat{\rho}} + \hat{\rho}^T y_0^i$.

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 both data sets. For regional data, the slope of the cross section regression is close to 0.6 and 21% of the variations of the cross section distribution of steady states is explained solely by initial conditions. For European countries the initial conditions alone explain 47% of the cross sectional distribution of steady states for the sample 1950-1985 and the other conditioning variables add no significant explanatory power to the regressions. More importantly, none of the conditioning variables appears to be correlated with steady states once the effect of initial conditions is accounted for - eliminating this variable from the regression does not lead to significant estimates for the other conditioning variables - and nonlinear effects, capturing possible clusters of units around particular steady state values, appear to be of minor importance. Overall these results indicate, in the language of Levine and Renelt (1992), that only the initial conditions appear to be a robust determinant of the relative position of a unit in the distribution of steady states and are consistent with Easterly et. al. (1993)'s conclusion that policy variables play a small role in explaining the pattern of growth rates.

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. For the 1950-1965 subsample the initial conditions are similarly important but now government share in GNP has significant explanatory power in the regression. Finally, for the 1950-1979 sample, initial conditions are insignificant but government share in GDP has marginal explanatory power in the regression.

In sum, we find that in four of the five samples, the initial conditions are the most important determinant of the estimated cross sectional distribution of the steady state of per-capita income. Countries tend to converge to their own steady states sufficiently fast but income disparities disappear at a very slow rate: they were reduced to some extent from the 1950 to the 1970, they persisted intact for most of the 1970's and they increased over the 1980's. This is true even if we condition for government variables, human capital etc.. A country (region) which is initially below the average per-capita income will eventually expect the gap to narrow somewhat but not to improve its relative standing in the cross sectional distribution. Hence, with some exceptions, the poor stay about as poor as they were at the beginning.

6 What can go wrong? Misspecificion and Biases

Our results are substantially at odds with those commonly found in the literature. We have provided an explanation that makes consistent ours and previous results; this explanation is the fixed-effects bias depicted in Figure 1. It is important to challenge our results, however, to see whether there are possible misspecifications or econometric biases intrinsic in our estimation/testing procedure that would account for our 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. However, with this scaling, there are about 10% of the regions for which $\hat{\rho} > 1$ (diverge) and we noted that this phenomena appears, primarily, for French and German regions. It is therefore worth examining whether this tendency to diverge is reduced using an alternative normalization which preserves the AR(1) properties for the scaled variable. For this reason we repeat the estimation process for the sample of European Regions scaling each unit at each point in time by its country mean [10]. This exercise also allows us to test whether there is any tendency for the steady state of income per-capita of regions 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 substantially confirm previous conclusions. Few additional features are worth noting. First, the choice of the σ vector which maximize the likelihood is $\sigma = \infty$ so that estimation by OLS equation by equation provides the best possible fit to the data. This suggests that knowledge of the α and ρ for one region does not provide relevant information for the same variables in another region. Put it in another way, with this scaling, income per-capita at regional level behaves as if there were no regional (or country) interdependences. Second, for the best specification of σ, the average convergence rate increase to about 36%. Finally, the hypothesis that the estimated steady states are the same for all regions of one country is soundly rejected using the Posterior Odds ratio, with Portugal being a marginal exception. From figure 7 we see that now only

[10] Since data for Denmark, Luxemburg and Ireland is available only at country level, we exclude them from the sample for this experiment. In terms of the model of Appendix 1 this implies that there is one trend common for all the regions of each country

4 regions display a $\hat{\rho}$ which exceeds one, that the cross sectional distribution of steady states is more normal and that the relationship between the position in the initial and steady state distribution of per-capita income is strong with no tendency toward reducing inequalities (the slope is 1).

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 may be downward biased (i.e. the estimated convergence rate 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. Can measurement error explain why our average estimates of the convergence rate are much larger than those existing in the literature? To quantify the extent of the problem for our two data sets we ask the following question. Suppose that the true convergence rate is 2% per year. What properties should the measurement error have to obtain estimates of the convergence rate of 23% (Regio data) or 12% (Summer and Heston data)? Table 7 presents the results allowing for a measurement error that is serially correlated. 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 GNP in European countries. When the measurement error is i.i.d, the variability of innovations in the measurement error should be about 40% (70%) of the variability of innovations in y_t^i.

On the other hand, measurement error would bias the cross sectional regressions we present in table 4 *against* our finding that initial conditions determine steady states. If measurement errors are present in the initial conditions, the conditioning variables in X_i will be correlated with the error term, making it difficult to accept the hypothesis that the initial condition is the most important variable in those regressions.

To summarize, measurement error is unlikely to be the reason for both the high average estimate of the convergence rate and the strong persistence of inequality found in the data.

One additional potential problem with our cross sectional regressions is that we neglect the fact that steady states are estimated on a short sample. That is, we asymptotically extrapolate given the sample, disregarding the fact that estimates of the steady state should also include a term $\rho^T y_0^i$. Neglecting this implies that the error term and the regressor may be correlated and the significance of the initial conditions in explaining the estimated cross sectional distribution of the steady states spurious. There are several ways to check the extent of this problem. For example, one can used direct small sample estimates or sample averages of the steady states. Alternatively, one could test if the slope coefficient is really $\bar{\rho}^T$ where $1 - \bar{\rho}$ is the average estimate of the convergence rate. Finally, one could use an instrumental variable procedure, instrumenting y_0^i with y_{-q}^i for q large. In all cases we find that, if there is a problem, it is very minor. For example for $T \geq 12$, ρ^T never exceeds 0.1 while for the two full samples the slope coefficient in the cross sectional regression is of the order of 0.5-0.6.

Finally, even if measurement error is absent, OLS estimates of ρ are downward 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. Since our posterior estimates in many samples are close to OLS estimates, this problem may be serious. We are currently doing some work on the short sample properties of our estimators; preliminary Monte-Carlo results indicate, however, that our basic results are not due to short 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.

7 Conclusion

The modern literature on convergence has concluded that there is a strong tendency for regions and countries to converge to similar steady states, although convergence is very slow. Limiting steady states may be different because of differences in technologies or government policies but the effects of variables proxying for these differences are weak and not present 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 conclusion that this paper offers are somewhat different: we find fast convergence rates to a 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 aggregate to be reduced by only 30%-40% in the limit. The conclusion seems to be that current redistribution and development policies, such as the Regional and Cohesion Fund Policies carried out by the governments of EC, are not working; rich regions can be taxed more heavily in favor of poor regions for solidarity reasons but not in the hope that these transfers this will foster development of the poor regions. Poor regions cannot expect to become as well off as rich regions unless structural changes occur in the economic environment; controlling the conditioning variables is not sufficient for convergence.

We also argued that the traditional results are econometrically biased and that the restrictions that cross section regressions impose on the data are strongly rejected in formal testing; then, the previous results can be accounted for as an econometrics bias or, alternatively, as the fact that the definition of convergence that had been used is not strongly related to the issue of persistence of inequality. We have also shown that, by exploring the data with a Bayesian procedure, we can both find the best model specification, examine the features of the data in a systematic way and formally test various convergence propositions.

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 we used, comforts us about possible incongruities present in the Regio data set. Second, the time span of the Regio data set is short so the prior may have a substantial influence on posterior estimates. "Misspecification" of the prior may therefore cause distortions. Third, by its own nature, the issue of convergence is an exercise in asymptotic extrapolation, which is well known to be an unreliable exercise. Because this sin is committed by all studies, we let the reader decide what to do with the entire empirical literature on convergence.

Appendix 1

A Model of Aggregation over units.

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

$$\overline{y}_t^i = a^i + \rho^i\ \overline{y}_{t-1}^i + \epsilon_t^i \tag{17}$$

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 [11].

Assume there is a continuum of regions $i \in [0,1]$. Clearly, the process $\{Y_t^i\}$ is consistently determined from the above equation and *any* process for aggregate 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 make sure that the model is well defined, we have to show that $\int_0^1 y_t^i\ di = 1$ for all t.

Here, ϵ_t^i is the idiosyncratic shock. As long as the process (17) 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 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 they satisfy*

$$\int_0^1 \overline{y}_0^i di = \int_0^1 \frac{a^i}{1-\rho^i}\ di = 1. \tag{18}$$

To show that $\int_0^1 y_t^i\ di = 1$ for all t under this assumption, notice that (17) can be rewritten as

$$\overline{y}_t^i - \frac{a^i}{(1-\rho^i)} = \rho^i \left(\overline{y}_{t-1}^i - \frac{a^i}{(1-\rho^i)}\right) + \epsilon_t^i = \tag{19}$$

$$\sum_{j=0}^{t-1} \left(\rho^i\right)^j\ \epsilon_{t-j}^i + \left(\rho^i\right)^t \left(\overline{y}_0^i - \frac{a^i}{(1-\rho^i)}\right) \tag{20}$$

Taking integrals in (20) and using the Assumption 1, we see that

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

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

$$\sum_{j=0}^{t-1} \int_0^1 \left(\rho^i\right)^j \, di \; 0 + \int_0^1 \left(\rho^i\right)^t \, di \; \int_0^1 (1 \; - 1) = 0 \tag{21}$$

Using (18) and the previous equation, we have

$$\int_0^1 \overline{y}_t^i di = \int_0^1 \frac{a^i}{1-\rho^i} \, di = 1 \tag{22}$$

for all t, as desired.

This model, then, allows for a fairly rich pattern of regional cycles, with a (possibly correlated) aggregate shock shared by all units, different cycles in different units, and growth in all units. 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 $\overline{y}_t^i$ is not affected by the aggregate shock, so that no units are allowed to respond more strongly than others to aggregate shocks.

One possibility is to assume that aggregate output follows a random walk with drift:

$$Y_t = Y_{t-1} \; \delta \; \eta_t \tag{23}$$

where η_t is a stationary process that may be serially correlated, and its log has mean zero, and is independent of all other random variables. In this particular case, multiplying both sides of (17), by Y_t we have

$$Y_t^i = a^i Y_t + \rho^i \; Y_{t-1}^i + \left(\rho^i \; Y_{t-1}^i (1 - \delta\eta_t) + \epsilon_t^i Y_t\right) \tag{24}$$

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. It is because of this undesiderable property of the residuals of the equation for Y_t^i that many authors have avoided estimation of convergence regressions with levels and panel data (see e.g. Pesaran and Smith (1995)).

Appendix 2

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

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

$$\log[\hat{y}_i(t)] = \log[\hat{y}_i(0)]e^{-\beta_i t} + \log[\hat{y}_i^*](1 - e^{-\beta_i t}) \tag{25}$$

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

The estimable specification BS employ is:

$$\frac{1}{T}\log[\frac{y_i(t_0+T)}{y_i(t_0)}] = B_i - \log[y_i(t_0)]\frac{(1-e^{-\beta_i T})}{T} + u_{i,t_0,t_0+T} \tag{26}$$

where u_{i,t_0,t_0+T} is a distributed lag of u_i for times between 0 and T and 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 technological progress. Using discrete time notation the above equation can be written as:

$$\log(y_T^i) = \alpha + \rho^T \log(y_0^i) + \gamma X^i + \epsilon^i \tag{27}$$

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

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

Appendix 3

This appendix describes in details the statistics used to test the hypothesis that estimated steady states are the same.

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

$$PO = 2 \cdot \log\left(\frac{(1-\alpha)}{\alpha}\frac{\phi(\chi)}{\Phi(\chi)}|\Sigma|^{-\frac{1}{2}}\right) \tag{28}$$

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 evaluated at $\chi = \sqrt{A'R\Sigma^{-1}R'A}$ and $\Phi(\chi)$ is the cumulative standard normal distribution at χ, A is a $n \times 1$ vector containing (approximate) ML estimates of the steady states for each unit, R is a $(n-1) \times n$ matrix with ones on the main diagonal, -1 on the following upper diagonal and zero everywhere else and Σ is the covariance matrix of the estimates of 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 (infinitely) many possible alternative values. Second, since in this study we are dealing with small samples, we explicitly include Σ in the criterion function. Asymptotically, $\log|\Sigma|^{-\frac{1}{2}}$ 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 OLS estimates of the steady states may differ substantially from those obtained in large samples.

In standard Bayesian literature the PO ratio is used to test linear hypotheses. Because the hypothesis we are interested in involves nonlinear function of the parameters of the model, we linearize the restriction around the average (cross sectional) estimate of the parameters before applying the PO ratio criteria.

Asymptotically, the selection criteria used by the PO ratio is identical to the one of the Schwarz criterion (see Sims (1988)) which, in our case, can be written as

$$SW = -\log(|\Sigma|) - \chi^2 \tag{29}$$

By comparing the results obtained with PO and SW we can therefore analyze the extent of the small sample bias which is present in our sample.

An alternative way of examining the equality of estimated steady states is to ask what is the largest 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 α^*, can be computed from (29) as:

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

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

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 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 favour 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 the available data set. 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 from the point of view of the data than the nonlinear combination of a and ρ implied by the null.

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] Bernard, A. and Durlauf, S. (1994), "Interpreting Tests of the Convergence Hypothesis", forthcoming, *Journal of Econometrics*.

[7] Bernard, A. and Durlauf, S. (1995), "Convergence in International Outputs", *Journal of Applied Econometrics*, 10, 97-108.

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

[9] 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.

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

[11] Durlauf, S. and Johnson, P., (1994), "Multiple Regimes and Cross-Country Growth Behavior", *Journal of Applied Econometrics*, 10, .

[12] Evans, P. (1995), "How Fast Do Economies Converge", forthcoming, *Review of Economic and Statistics*.

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

[14] 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.

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

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

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

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

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

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

[21] 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.

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

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

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

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

[26] Quah, D. (1994), "Empirics for Economic Growth and Convergence", CEPR discussion paper 1140.

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

[28] Sala-i-Martin, X. (1995), "Regional Cohesion: Evidence and Theories of Regional Growth and Convergence", CEPR discussion paper 1075.

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

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

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

[32] 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.

[33] 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	a	ρ	Likelihood	a	ρ	Likelihood
.000001	.000001	-0.0790	0.9848	3337.7	0.0008	0.9910	1410
.000001	.100000	-0.0910	0.9914 (0.018)	3364.0	0.0028	0.9832 (0.012)	1433
.000001	1.00000	-0.0158	0.9840 (0.103)	3390.8	0.0057	0.9738 (0.024)	1439
.100000	.000001	-0.0133 (0.023)	0.9401	3402.1	-0.0016 (0.015)	0.9606	1447
.100000	.100000	-0.0135 (0.023)	0.9445 (0.010)	3404.1	-0.0007 (0.015)	0.9551 (0.007)	1450
.100000	1.00000	0.0235 (0.024)	0.9599 (0.114)	3438.0	0.0026 (0.017)	0.9398 (0.039)	1460
1.00000	.000001	-0.0216 (0.043)	0.8718	3412.8	-0.0039 (0.029)	0.9296	1451
1.00000	.100000	-0.0273 (0.056)	0.8404 (0.015)	3419.4	-0.0053 (0.036)	0.9211 (0.119)	1455
1.00000	1.00000	-0.0661 (0.149)	0.7762 (0.207)	3507.8	-0.0140 (0.084)	0.8896 (0.085)	1471
∞	∞	-0.0860 (0.059)	0.7250 (0.296)	3485.9	-0.0151 (0.092)	0.8805 (0.094)	1470

Notes: The basic model is given in equations (10) and (13)-(14). 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. a and ρ are the average estimates across 144 regions (Regio data) or 17 countries (Heston and Summers data) and Likelihood the value of the 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.

Table 3: Test of Unconditional Convergence

	Schwarz criterion	Posterior Odds	α^*	Likelihood Ratio
	European Regions			
1980-1992 sample	-380.41	-645.23	0.0000	0.0000
	European Countries			
1950-1985 sample	-61.37	-92.78	0.0000	0.0000
1950-1965 sample	-68.13	-99.54	0.0000	0.9651
1966-1985 sample	5.69	-25.71	0.0000	0.0000
1950-1979 sample	-48.33	-79.73	0.0000	0.9495

Notes: Schwarz Criterion is defined in equation (30), the (Small Sample) Posterior Odds criteria is defined in equation (29), α^* is the ex-post prior probability on the alternative defined in equation (31). For the Posterior Odds criteria the prior probability odds are set to 0.50. In the column likelihood ratio we report the p-value of the test (the statistics is distributed as $\chi^2(2)$).

Table 4: Explaining the Cross-sectional Distribution of Steady States

Regressors	European Regions 80-92 Sample	European Countries 50-85 Sample		50-65 Sample		66-85 Sample		50-79 Sample	
Constant	-0.16	-0.05	-0.36	-0.07	-0.10	-0.01	0.09	0.003	0.003
	(-4.06)	(-0.76)	(-0.98)	(-0.73)	(-0.21)	(-0.42)	(0.45)	(0.07)	(0.01)
Initial	0.60	0.51	0.47	0.33	0.29	0.77	0.76	0.06	0.05
Conditions	(5.42)	(3.36)	(3.28)	(2.20)	(2.46)	(10.69)	(10.92)	(0.63)	(0.57)
Secondary			0.45		0.63		0.20		0.16
Education			(1.11)		(0.73)		(0.93)		(0.36)
I/Y			0.003		0.01		-0.002		0.008
			(0.50)		(1.11)		(-0.37)		(1.16)
Government			0.48		-3.29		-0.64		-1.58
Share			(0.31)		(-2.41)		(-0.98)		(-1.80)
$\bar{R}^2$	0.21	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 compute as $SS^i = \frac{a}{1-\rho}$ if $\rho < 1$ and $SS^i = a * \frac{1-\rho^{T+1}}{1-\rho} + \rho^T y_0$ if $\rho > 1$. 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). t-statistics for the hypothesis that the coefficient is zero are in parenthesis.

Table 5: Average Estimated Parameters

σ_ν	σ_η	α	ρ	Likelihood
		European Countries, Sample 1950-1965		
.000001	.000001	-0.0010	0.9902	517.0
.000001	.100000	0.0046	0.9724 (0.029)	488.9
.000001	1.00000	0.0134	0.9361 (0.118)	287.8
.001000	.000001	-0.0011 (0.052)	0.8871	518.9
.001000	.100000	0.0062 (0.051)	0.8762 (0.032)	-207.3
.001000	1.00000	0.0063 (0.061)	0.8268 (0.145)	-1757.0
.001000	.001000	-0.0006 (0.004)	0.9886 (0.002)	510.9
∞	∞	0.0051 (0.122)	0.7875 (0.179)	-6452
		European Countries, Sample 1966-1985		
.000001	.000001	0.0021	0.9915	844.6
.000001	.100000	0.0036	0.9833 (0.019)	856.6
.000001	1.00000	0.0046	0.9802 (0.026)	857.7
.100000	.000001	-0.0044 (0.045)	0.8856	870.0
.100000	.100000	-0.0093 (0.047)	0.8830 (0.029)	873.7
.100000	1.00000	-0.0173 (0.050)	0.8805 (0.092)	877.1
1.00000	.000001	-0.0067 (0.061)	0.8482	870.1
1.0000	.100000	-0.0170 (0.075)	0.8415 (0.035)	875.2
1.0000	1.00000	-0.0033 (0.097)	0.8316 (0.108)	879.8
∞	∞	-0.0041 (0.098)	0.8294 (0.106)	877.2

European Countries, Sample 1950-1979

σ_ν	σ_η	α	ρ	Likelihood
.000001	.000001	0.0009	0.9870	781.37
.000001	.100000	0.0009	0.9557 (0.436)	511.31
.000001	1.00000	0.0009	0.9557 (0.437)	509.9
.100000	.000001	-0.0165 (0.029)	0.9274	577.5
.100000	.100000	-0.0449 (0.122)	0.8685 (0.112)	-669.8
.100000	1.00000	-0.0450 (0.122)	0.8683 (0.121)	-672.5
1.00000	.000001	-0.0165 (0.029)	0.9274	577.5
1.00000	.100000	-0.0450 (0.122)	0.8684 (0.121)	-671.4
1.00000	1.000000	-0.0451 (0.127)	0.8682 (0.121)	-674.1
0.0003	0.0003	0.001 (0.003)	0.9828 (0.001)	784.36
∞	∞	-0.0451 (0.122)	0.8682 (0.120)	-675.7

Notes: The basic model is given in equations (10) and (13)-(14). σ_ν and σ_η are the standard deviations of the prior restrictions. α and ρ report average estimates across the 17 countries and Likelihood the value of the 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.

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

σ_ν	σ_η	α	ρ	Likelihood
.000001	.000001	-0.0023	0.9886	3866.6
.000001	.100000	-0.0030	0.9900 (0.019)	3908.5
.000001	1.00000	-0.0124	0.9420 (0.147)	3936.7
.100000	.000001	-0.0072 (0.025)	0.9054	4013.2
.100000	.100000	-0.0079 (0.025)	0.9007 (0.091)	4017.4
.100000	1.00000	-0.0168 (0.022)	0.8684 (0.155)	4057.6
1.00000	.000001	-0.0098 (0.033)	0.8611	4039.5
1.00000	.100000	-0.0127 (0.039)	0.8224 (0.012)	4056.5
1.00000	1.000000	-0.0285 (0.073)	0.6820 (0.169)	4143.8
∞	∞	-0.0405 (0.012)	0.6334 (0.310)	4190.4

Test of Unconditional Convergence

	Schwarz criterion	Posterior Odds	Ex-post α	Likelihood Ratio
Overall	178.3	-80.93	0.0000	0.00001
Belgium	13.06	-3.54	0.0000	
Germany	27.86	-29.26	0.0000	
Greece	6.09	-17.95	0.0000	
Spain	10.31	-21.08	0.0000	
France	23.56	-18.86	0.0000	
Italy	16.04	-18.95	0.0000	
Netherland	10.17	-13.86	0.0000	
Portugal	4.29	- 4.76	0.0002	
UK	5.20	-17.00	0.0000	

Test of Persistence in Inequalities

	Constant	Initial Conditions	R^2
	0.01 (0.25)	1.005 (4.27)	0.12

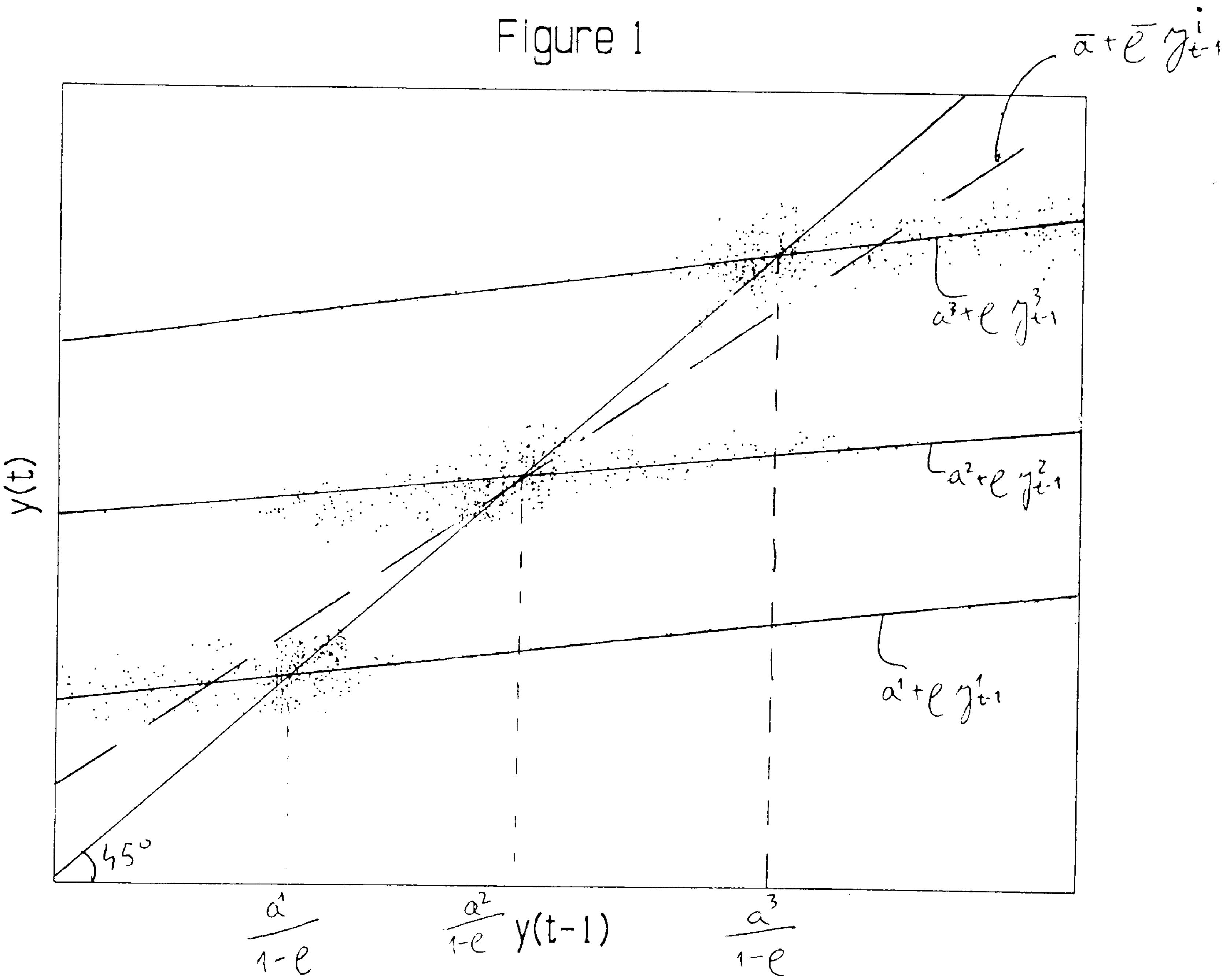
Notes: The basic model is given by (10) and (13)-(14). σ_ν and σ_η are the standard deviations of the prior restrictions. α and ρ are the average estimates across 144 regions (Regio data) and Likelihood the value of the posterior mode under the particular prior restriction. The dispersion (standard deviation) 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 tests for unconditional convergence with country names refer to 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	$\dot{\rho}$	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

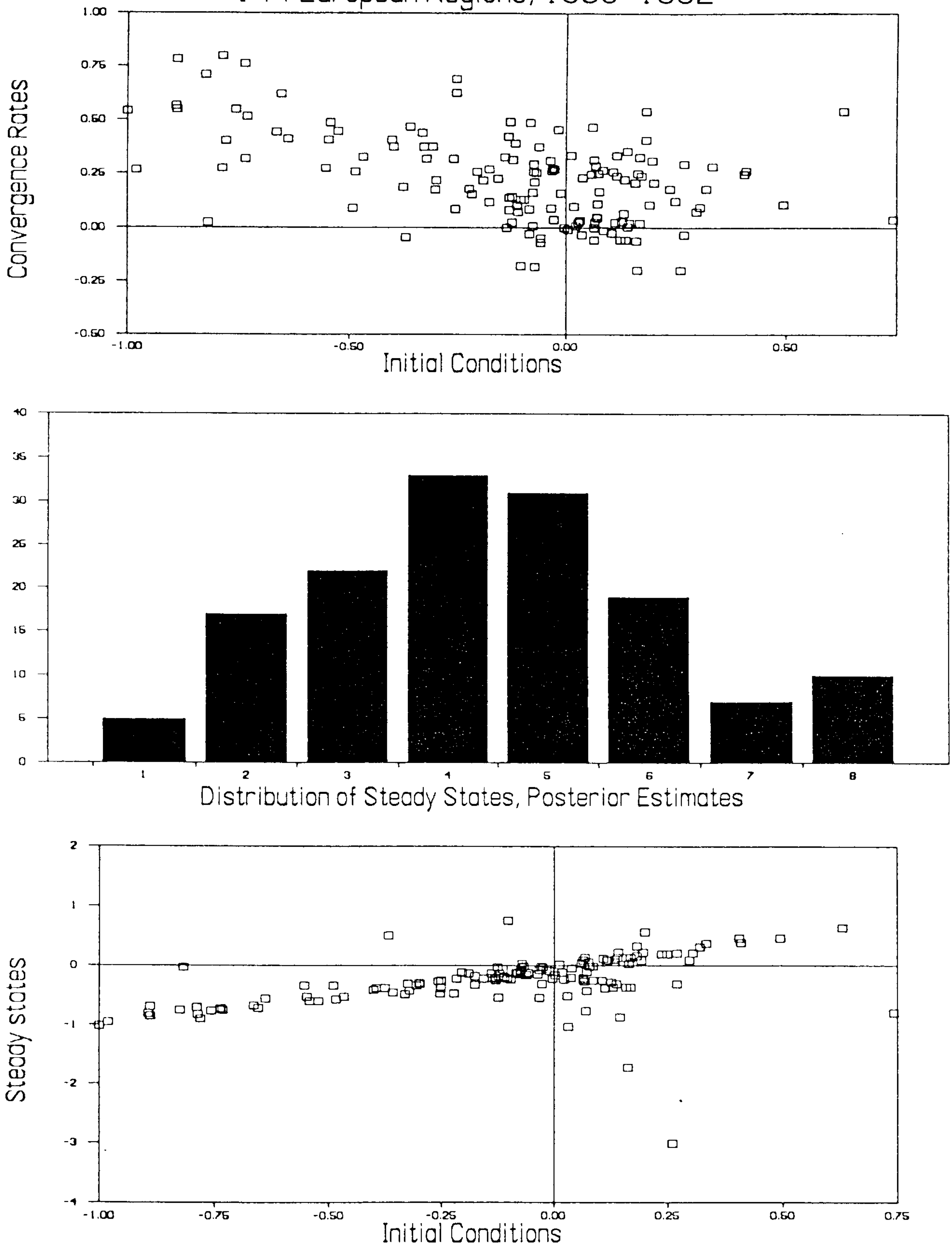
Notes: The model considered is $y_t^i = y_t^i + \omega_t^i$ where $y_t^i = \rho y_{t-1}^i + u_t^i$, $\omega_t^i = \mu \omega_{t-1}^i + e_t^i$ and $E(e_t^i, y_t^i) = 0$. The table reports for a given ρ what is the variability of the innovation in y_{it} relative to the variability in the innovation in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

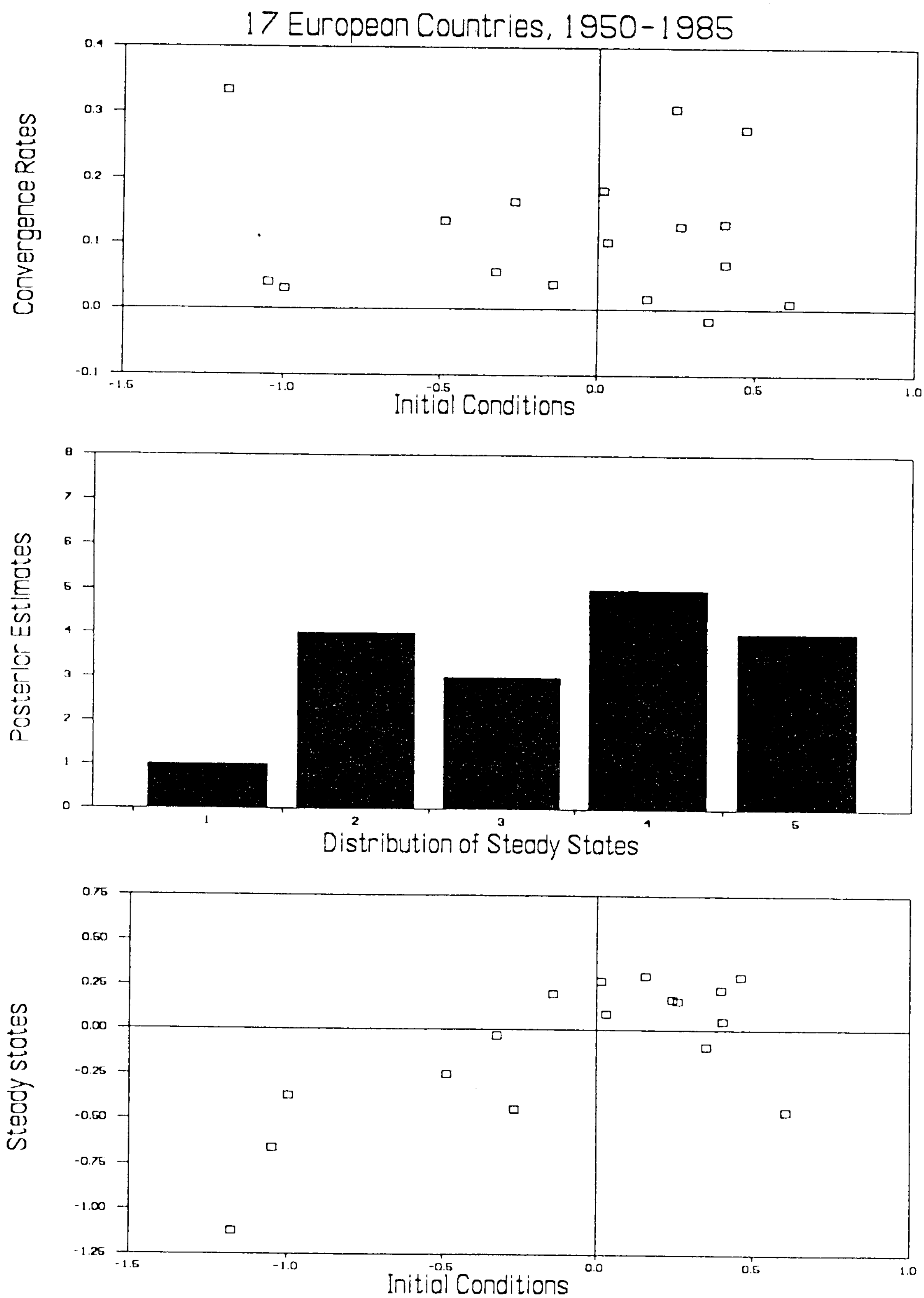
Figure 1

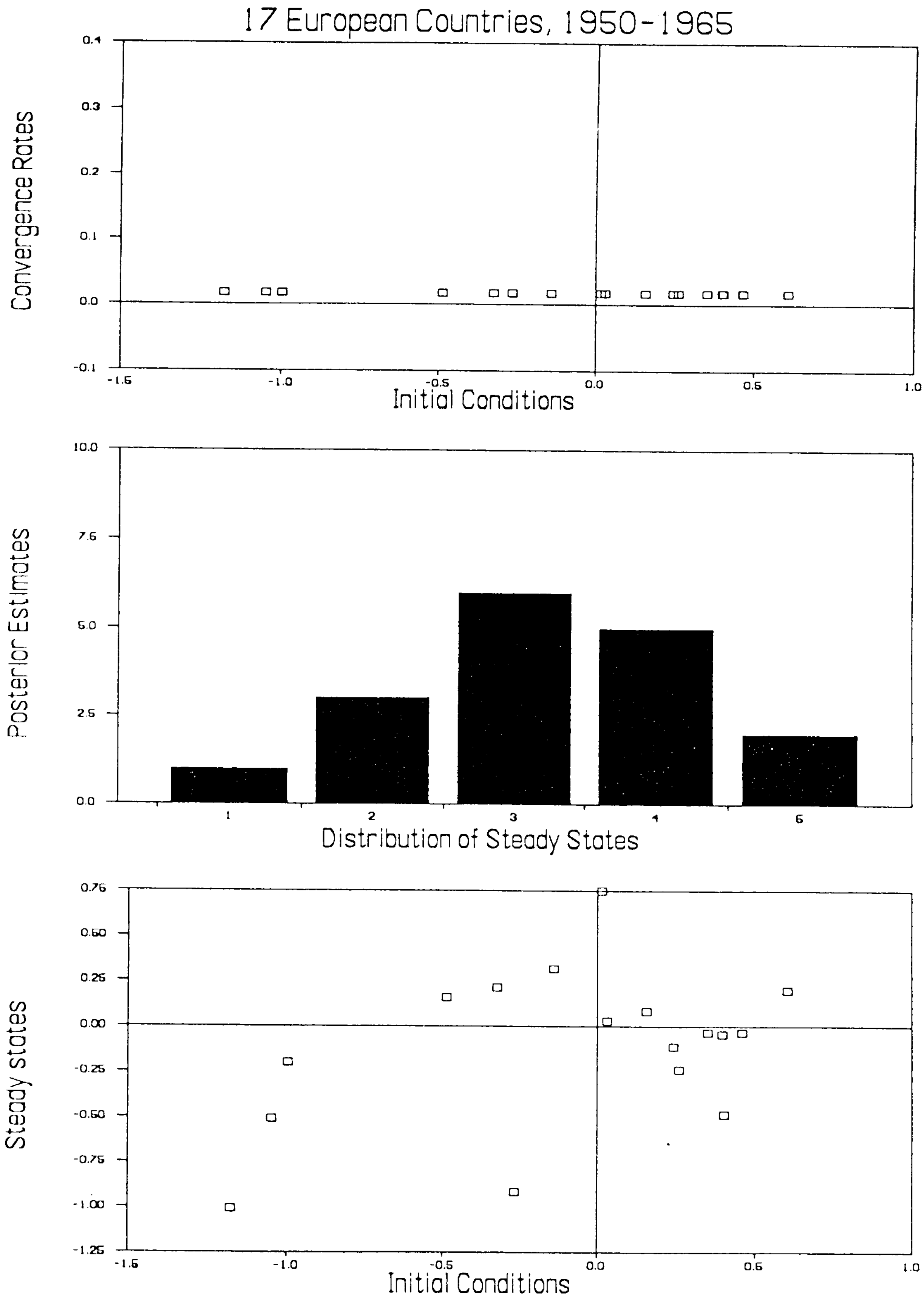


$\bar{a} + \bar{e}\, y^i_{t-1}$
$a^3 + e\, y^3_{t-1}$
$a^2 + e\, y^2_{t-1}$
$a^1 + e\, y^1_{t-1}$
y(t)
45°
$\frac{a^1}{1-e}$
$\frac{a^2}{1-e}$
y(t-1)
$\frac{a^3}{1-e}$

144 European Regions, 1980-1992

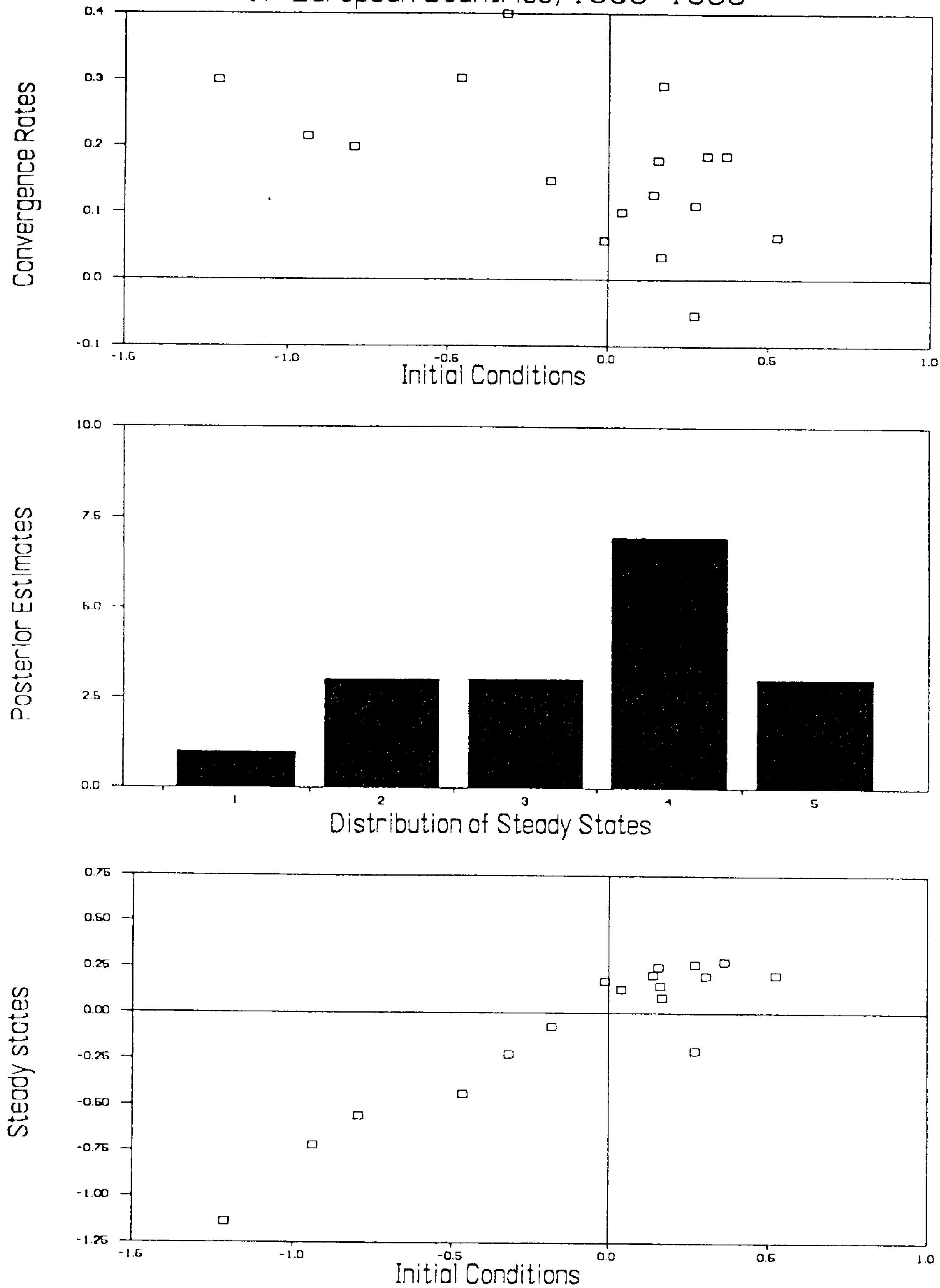


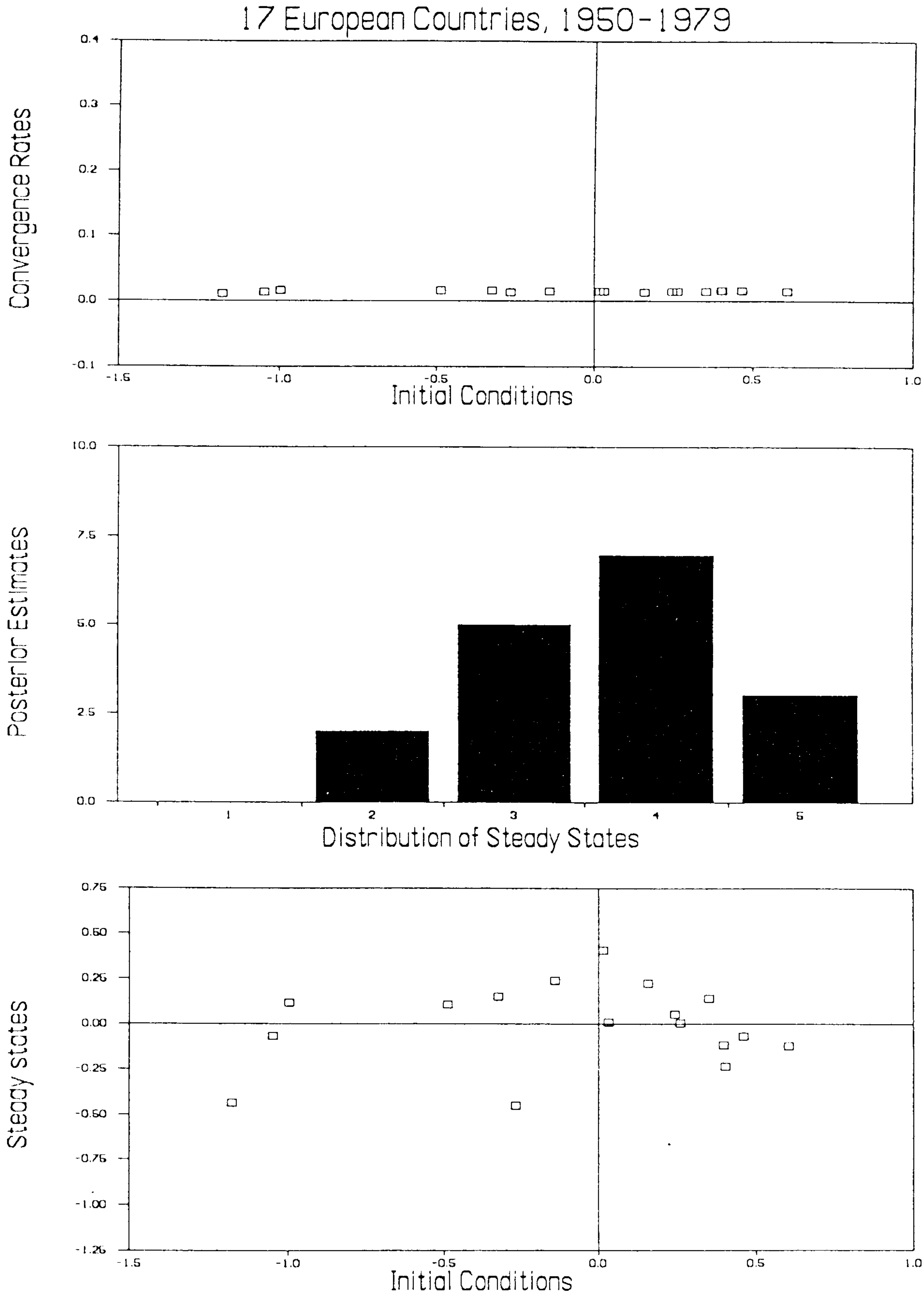

17 European Countries, 1950-1985
Convergence Rates
0.4
0.3
0.2
0.1
0.0
-0.1
-1.5
-1.0
-0.5
0.0
0.5
1.0
Initial Conditions
Posterior Estimates
8
7
6
5
4
3
2
1
0
1
2
3
4
5
Distribution of Steady States
Steady states
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
-1.5
-1.0
-0.5
0.0
0.5
1.0
Initial Conditions



17 European Countries, 1950-1965
Convergence Rates
0.4
0.3
0.2
0.1
0.0
-0.1
-1.5
-1.0
-0.5
0.0
0.5
1.0
Initial Conditions
Posterior Estimates
10.0
7.5
5.0
2.5
0.0
1
2
3
4
5
Distribution of Steady States
Steady states
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
-1.5
-1.0
-0.5
0.0
0.5
1.0
Initial Conditions

17 European Countries, 1966-1985




17 European Countries, 1950-1979
Convergence Rates
Initial Conditions
Posterior Estimates
Distribution of Steady States
Steady states
Initial Conditions

141 European Regions, 1980-1992