CROSS SECTION DISTRIBUTION DYNAMICS

PhD Thesis
1996

Ana Rosa Lamo
London School of Economics

ACKNOWLEDGEMENTS

I am very grateful to Danny Quah for his excellent supervision, without his example and enthusiasm this thesis would not have been possible .

I want to thank Rafael Repullo and Juan J. Dolado for encouraging me to continue my studies abroad when I was at the CEMFI and for their support since.

A very warm acknowledgement to those friends who during the years in London have helped and supported me in many different ways, among them: Abigail Fallot , Chiara Betta, Denis Gromb, Fabian Burkart, Ignacio Lobato, Ignacio Soto, Javier Suarez, Jose-Manuel Mates, Lucas Burkart, Maite Fernandez, Mary Amiti, Mike Burkart, Reinout Koopmans, the Ratcliffes (Moira, David, Johana and Caroline) and all the Spanish connection at the LSE. And to Andreas Irmen.

Thanks to Javier Andrés for stimulating conversations, to my office-mates Gustavo Nombela and Marcus Rubin for their patience in dealing with me, to the CEP and all its members.

A very special thank you to Philip Macdowell for 'being always there' and for the hours he has spent reading my papers and unravelling my English.

Finally, thanks to Charlie Bean, Fabio Canova and Javier Hidalgo for kindly agreeing to be my PhD examiners.

SUMMARY

This thesis contains four chapters. Each chapter constitutes an empirical exercise in which I apply econometric ideas on studying the dynamics of large cross sections of data (*Random Fields).* Three of them concern the empirics of convergence and the fourth analyses business cycle fluctuations.

The first, "Notes on Convergence Empirics: Some Calculations for Spanish Regions," describes the econometric methods for studying the dynamics of the distributions and how to characterise convergence in this framework, explains why the standard cross-section regression analysis is misleading when testing for convergence and then performs some calculations for regions in Spain.

The second chapter, "Dynamics of the Income Distribution Across OECD Countries", considers its baseline hypotheses to be those generated by the Solow growth model. Using sequential conditioning, it studies whether the convergence hypothesis implications can be shown to hold for the OECD economies. It finds that neither absolute nor conditional convergence, in the sense of economies approaching the OECD average, has taken place.

The third chapter, "Cross Sectional Firm Dynamics: Theory and Empirical Results", extends ideas of distribution dynamics to a discrete choice setting, and extends the reasoning of Galton's Fallacy to the logit model. It provides evidence of the tendency of firm sizes to converge for the US chemicals sector by analysing dynamically evolving cross-section distributions.

Finally, the fourth chapter, "Unemployment in Europe and Regional Labour Fluctuations" applies distribution dynamics ideas to a business cycle setting. It analyses the dynamics of employment for 51 European regions from 1960 to 1990, addressing the issue of whether regional shocks have aggregate effects on unemployment or the opposite. It uses a model for non-stationary evolving distributions to identify idiosyncratic and aggregate disturbances.

NOTE:

The four chapters in this thesis have been conceived as independent papers, as a consequence of that and for completeness of the papers, there is some overlapping among chapters.

TABLE OF CONTENTS:

CHAPTER 2: **Dynamics of the Income Distribution Across OECD Countries**

CHAPTER 3: Cross Sectional Firm Dynamics: Theory and Empirical Results

CHAPTER 4 : Unemployment in Europe and Regional Labour Fluctuations

LIST OF TABLES

CHAPTER 1:

CHAPTER 2:

CHAPTER 3:

CHAPTER 4 :

LIST OF FIGURES

CHAPTER 1:

CHAPTER 2:

CHAPTER 3:

FIGURES7a to 7d: Stochastic Kernel, NMSH2 5 Year Transition

FIGURES 7e to 7h: Stochastic Kernel, NMSH2. Contour Plot

CHAPTER 1:

NOTES ON CONVERGENCE EMPIRICS :

Some calculations for Spanish Regions

1. INTRODUCTION

Whether initially poor countries or regions tend to became as wealthy as the initially richest ones; whether wages equalise across industries or regions, whether in an industry firms tend to reach the same level of capacity or whether they tend to an equilibrium size in terms of the number of workers etc. are economic questions which have been intensively studied from both a theoretical and empirical perspective. Although each one of them refers to different areas in economics, all of them involve the idea of convergence as a process of economic homogenisation rather than a persistence of the inequality.

This paper concentrates on convergence across economies i.e. states, countries, regions, provinces, etc.. This issue is important given that the implications for economic policy are very different depending on whether the income inequality is vanishing or not, in other words whether poor economies are catching up with the richest ones or not. Consequently the subject has generated a wide and controversial literature from both the theoretical and empirical point of view. From the Economic Theory perspective, there is a contest between the growth models which predict convergence across economies (generally those that take growth as an exogenous process) and those models that do not (endogenous growth models). Therefore, it is very important to guarantee a correct performance of the empirical analysis.

The numerous and varied empirical studies of convergence in the recent growth literature differ basically in two aspects: the way in which they formalise the convergence hypothesis and the econometrics tools or empirical approach they use to test convergence. These two aspects are linked together in the sense that not every econometrics approach is able to test for all the aspects involved in the convergence issue. A first basic idea of convergence as a process of homogenisation across economies over time already suggests that the formalisation and testing of the convergence hypothesis must involve the characterisation of the behaviour of a broad cross-section of economies over long periods of time. The usual or standard empirical approaches to convergence may not be adequate to give evidence about convergence.

The empirical approach which has generated a wider and more questionable literature is the so-called cross-section regression analysis. It examines the regression of (averaged) growth rates on initial levels of income across economies. The more elaborate studies use panel data techniques or pool data regression in an attempt to avoid part of the wastage of information in averaging. Additionally this approach looks at the dispersion of incomes across economies, and tries to account for convergence formulated in following way: each country eventually becomes as rich as all the others, the cross section dispersion diminishes over time. This approach, as will be argued below, is not very informative about convergence, in fact it might be misleading.

A second approach to convergence is time-series analysis. It tests for the convergence hypothesis formulated as a lack of persistence in income disparity between economies, studying persistence in the context of unit roots analysis. In other terms it tests whether disparities across economies have neither units roots nor diverging deterministic time trends. The econometrics tools are time series and units roots analysis .

A third approach has been recently suggested by Danny Quah.[1] It analyses the dynamics of large cross-sections by using econometrics ideas like mixing and ergodicity. It models directly the dynamics of the cross-section distributions as a more natural way to study convergence .Let us call it Cross-Section Distribution Dynamics. It is concerned with transitional characteristics and deals with the following formulations of the convergence hypothesis: (i) economies originally richer than average are more likely to fall below average eventually and vice versa, the cycle repeats and (ii) whether a country's income is eventually above or below the cross-section average is independent of that economy's original position.

Cross-Section regression and time-series analysis (call them standard empirical approaches) are not adequate to draw conclusions about convergence. In spite of its popularity regressing growth rates on initial levels of income across economies (growth equations) may be misleading when evaluating convergence. The time-series analysis does not use all the cross sectional information. Additionally, neither of them are able to account for transitional characteristics. On the other hand analysing the dynamics of the cross-section distributions embraces the other two approaches and it seems to be an adequate way to study large cross-sections.

The remainder of this paper is organised as follows: Section 2 describes critically the cross-section regression analysis and explains where and why standard approaches fail. Section 3 presents the cross-section distribution approach as an alternative empirical framework for studying the income or GDP dynamics of a broad cross-section of economies and offers some calculations for the regions in Spain. Section 4 studies the notions of *conditional convergence*. Section 5 concludes.

The empirical study of convergence in other areas of economics, like industrial organisation or labour markets, deserves similar comments to those we will make for growth theory. The cross section analysis may fail in evaluating convergence and the study of the dynamics of large cross section would be also adequate for these cases.[2]

2. THE STANDARD EMPIRICAL APPROACHES TO CONVERGENCE.

2.1. Cross-Section Regression Analysis

Standard cross-section regression analysis proposes two measures of convergence: β and σ convergence.

β convergence is a property of the classical Solow growth model. The β convergence analysis consists of estimating growth equations in which the (average) growth rate of income over time for each economy is regressed on the initial levels of income (and a bunch of steady variables). It interprets a negative correlation between the growth rate of income per capita and its initial level as economies converging toward a common steady-state. The results are surprisely uniform: a negative and significant estimate of the initial level of income coefficient (positive β) of about 2% for different samples and periods[3].

σ convergence is model free and has been proposed by the cross-section regression approach to be used together with β convergence. There is σ convergence if the dispersion of the real income across economies tends to fall over time.

Let us start by looking at the definition and empirical performance of β convergence, to subsequently analyse both together, β and σ measures. [4]

2.1.1.- β Convergence: Concept and Interpretation.

The studies on β convergence use as a framework the classical Solow growth model or the augmented Solow model. See for example Baumol (1986), Barro (1991) Barro & Sala-i-Martin (1991) & (1992), Mankiw, Romer & Weil(1992), and Holtz-Eakin(1992). Solow's model assumes a neo-classical production function with diminishing returns to capital.(CRS model). It takes rates of saving, population growth and technological progress as exogenous.

Let us assume a Cobb-Douglas production function for economy *i*.

$$Y(t) = k(t)^{\alpha} (A(t)\ L(t)))^{1-\alpha} \tag{1.1}$$

$0< \alpha <1$

Where: Y is output, K & L are inputs, capital and labour respectively, and A is the level of technology.[5]

Assume A and L to grow at the exogenous rates g and n respectively. $L(t) = L(0).e^{nt}$, $A(t) = A(0).e^{gt}$. Denote $y = Y/AL$ and $k = K/AL$.

Capital evolves according to:

$$\dot{k} = sy(t) - (n + g + \delta)\ k(t) \tag{1.2}$$

s = saving rate, δ = depreciation rate.

Therefore the capital converges to a steady-state defined by:

$$k^* = [s/ (n + g + \delta)]^{1/(1-\alpha)} \tag{1.3}$$

The expression for the steady-state income per capita is found by substitution into the production function and taking log.

$$log(Y/L) = a + gt + (\alpha/ 1 - \alpha)\ log(s) -(\alpha/ 1 - \alpha)\ log(n + g + \delta) + \varepsilon \tag{1.4}$$

Where: $log\ (A(0))=a + \varepsilon$, s and ε independents.

Equation (1.4) shows how the steady-state level of income per capita in each country depends on its population growth and its accumulation of capital (physical in the original model and physical and human in the augmented version). Therefore different economies can reach different steady-states.

The empirical literature in estimating the steady-state growth path, includes variables such as government expenditure, population, investment, migrations, market distortions or political system as proxies of population growth and accumulation of physical capital. School enrolment rates, fertility, indicators of differences in the quality of the education, etc. are proxies of the human capital accumulation in the augmented version.

The Solow model predicts convergence of each economy to its steady-state. Let y^* be the steady-state in each economy and y the actual value, by approximating (1.4) around y^*, it can be proved that :

$$\dot{y} = \beta\, [\log(y^*) - \log(y(t))] \tag{1.5}$$

Where : $\beta = (n+g+\delta)(1-\alpha)$

Solving this differential equation and rearranging we find that the average growth rate of y over the interval between 0 and T is :

$$(1/T)\log[y(T)/y(0)] = g + [(1-.e^{-\beta T})/T]\log(y^*/y(0)) \tag{1.6}$$

The convergence in equation (1.6) is a *conditional convergence*. i.e. given the steady-state y^* the growth rate is higher the lower the initial income level $y(0)$. The steady-state might differ across economies and in consequence in the empirical analysis it is necessary to hold fixed this variation. β is the convergence coefficient, which governs the speed of adjustment. The greater β the higher the responsiveness of the average growth to the gap between the steady-state and the initial income.

The empirical analysis considers a version of equation (1.6) which applies for discrete periods to the economy i, and checks for two kinds of convergence: *absolute* and *conditional convergence*. It tests for *absolute convergence* by regressing across economies the average growth rate of income over time for each economy on the initial levels of income, a negative initial income level coefficient is interpreted as economies moving towards a common steady-state. It test for *conditional convergence* by regressing on initial income and on the variables hypothesised as determinants of the steady-state, in other words conditioning on the determinant of the steady-state. In this regression the conditioning variables (government expenditure, investment, schooling etc.) determine the long run growth or the permanent component and the initial income level controls the transitory dynamics. A negative initial income level coefficient is interpreted as convergence of each economy to its own estimated steady-state.

2.1.2.- β Convergence: Criticisms.

Some ideas have proposed that the above-described convergence test fails.

The cross-section regression analysis and in particular the concept of β convergence is based on a model for a single "representative" economy and extracts conclusions about the cross-section. This is not straightforward. In other words the adjustment process of the Solow model tells us whether or not each economy, after being perturbed from its steady-state path returns to it, approaching monotonically. This is a single-country implication and consequently has nothing to do with economies approaching each other[6]. Quah (1995), makes an enlightening distinction between the growth and convergence mechanism. He argues that the conventional analysis (cross-section regression) may be revealing in the growth mechanism or productivity performance for an economy but is uninformative on the convergence issue.

In addition to the fundamental question of whether this *conditional convergence* prediction of the Solow model has any practical interest, the steady-state in each economy y^* is not observable and it has to be estimated from observed values of the explanatory variables. In claiming that economies converge toward their steady-state it is been assumed that the estimated steady-state is the true one, this adds new serious problems to the concept of conditional convergence apart from the one of interpretation. For example, Mankiw, Romer & Weil (1992) estimates the steady-state by the observed values of investment in human and physical capital, consequently the economies will converge to their steady-state only if both human and physical capital have already reached their steady-state values, if not they will converge to a pseudo steady-state that does not characterises the true one (if it exists). Cohen (1992) makes this point and finds that the correlation between growth and human capital accumulation vanishes once time-invariant country-specific effects are included in the standard regression. In general, some authors have shown that the causality among growth and accumulation rates is weaker than what the classical growth model would predict. Blomstrom, Lipsey and Zejan (1993) show that causality runs from growth to investment rather than the other way around. The theoretical model in Quah(1995b) predicts inverse causality. Quah studies the formation of clubs of convergence and coalitions, such that when different convergence-clubs form variables that have been used by the conditional convergence literature as explanatory variables are endogenous, for example high human capital is only found among the rich countries. Finally, many studies have found a significant correlation between growth and short-term macroeconomic indicators (inflation, exports, and so on) which nonetheless is not always robust to changes in the model specification (Levine and Renelt (1992)).

Other criticisms (which apply to both *conditional* and *absolute convergence*) refer to the implicit assumption that the permanent component of the income for each economy is well described by a linear or log-linear deterministic time trend. This implicit hypothesis is required in order to justify the usual interpretation of these regressions and in order that the average growth rate (LHS of the regression) makes sense. In fact the averaged growth rate is simply the slope of the deterministic trend. Pooling data or even estimating economy by economy, there is still a deterministic trend.

For example, Graphic 1 represents per capita income for 50 regions in Spain over 35 year (i.e. 50 time-series with 35 time observations each). What cross-section regression analysis does is to take the slope of each time series (i.e. 50 observations) and regress it on initial income (i.e. the first observation of each time series).

Suppose that each economy's income (in log), $\{Y_j(t),\ _{j=1,2,..J,\ t=\ 0,...,T}\}$ could be decomposed into a permanent and transitory components, as follows:

$$Y_j(t) = X_{j1}(t) + X_{j0}(t) \quad {}_{j=1,2,..J, t=0,...,T} \tag{1.7}$$

$X_{j1}(t)$ is the permanent component, which is implicitly considered by the mentioned literature as a time trend: $X_{j1}(t) = \alpha_j + \lambda_j t$, where: α_j and λ_j are coefficients that do not depend on t. $X_{j0}(t)$ is the transitory component such that, $\mathrm{E}\, X_{j0}(t) = 0$.

Then:

$$Y_j(t) = \alpha_j + \lambda_j t + X_{j0}(t) \tag{1.8.1}$$

$$\Delta\, Y_j(t) = \lambda_j + \Delta X_{j0}(t) \qquad E\Delta\, X_{j0}(t) = 0 \tag{1.8.2}$$

Note that since $E\Delta\, X_{j0}(t) = 0$, $E\Delta\, Y_j(t) = E\Delta\, X_{j1}(t) = \lambda j$ i.e. λ_j is the growth rate of the economy j's income and of its unobservable permanent component. Then equation (1.8.2) says that the growth rate λ_j is the slope of the deterministic trend.

λ_j varies across economies according to :

$$\lambda_j = Z_j\, \beta_o + u_j \qquad E\, Z_j u_j = 0 \tag{1.9}$$

Z_j includes the conditioning (education, policy, etc.) as well as the initial income level. On the other hand λ_j is not observable and in the referred empirical work has been proxied by the averaged growth rate of Yj.

Summarising: these studies run the cross-economies regression of the averaged growth rate of income over time for each economy on the initial levels of income and on the conditioning variables hypothesised in

$$\hat{\lambda}_j = Z_j \beta_o + u_j \ , \qquad E\, Z_j u_j = 0 \tag{1.10}$$

Where $\hat{\lambda}_j$ (averaged growth rate) is the slope of an assumed time trend in the permanent component of the income of economy j.

However, under stochastic growth, imposing a deterministic trend structure can be very misleading. This remains true even for the case in which pooled data on shorter averages is used. The procedure still assumes the non-stationary component of per capita income to be trend deterministic, although allowing for a changing slope.

Quah (1993a) checks the validity of a smooth time approximation to the permanent movements in income by fitting linear time trends to the log of per capita income country by country (Summer & Heston data) and compares the slope for different periods of time. The data do not support the implicit hypothesis of a smooth time trend approximation. Even then, it can be a good approximation if significant economic shocks happen only at the beginning of the sample.

Besides all the limitations just mentioned, Quah (1993b) has shown that convergence tests based on regressing average growth rates on initial levels (and conditioning variables) are uninformative since a negative cross-section regression coefficient on the initial level is in fact compatible with cross-economies behaviours that are far from the idea of convergence, like overtaking, cycles etc..

The argument is the same as that which explains Galton's classic fallacy. Galton observed two facts that he could not reconcile: on the one hand taller than average fathers had sons who turned out to be not as much above the average as their fathers themselves, on the other, the observed population of male heights continued to display significant cross section dispersion.

The regression test of convergence suffers from Galton's fallacy in the sense that. economies with higher than average initial income (tall fathers) have income in the future that is not as large as the initial. But this does not imply a diminishing cross-section dispersion, a collapsing cross-section distribution of the income. Similarly currently richer economies might become poorer than average in the future without a significant narrowing in the cross-section dispersion of incomes. This argument applies to both, *absolute* and *conditional convergence*.

Danny Quah (1993b) shows how in estimating growth equations[7] a non-positive coefficient in the initial levels arises even if it is assumed that the cross-section distribution is time invariant (there is not real convergence). The same conclusion holds when estimating in the final or even in a middle condition. And it can be negative also if there is divergence. Exactly the same reasoning applies to the idea of *conditional convergence*, just by taking the variable of analysis to be the residuals of the output after conditioning on the exogenous variables.

In consequence to use cross-section regression is inadequate for extracting out dynamic implications.

2.1.3 σ Convergence

The standard approach introduces σ convergence as a measure of convergence in the light of the above criticism and tries to arrive at some information about the cross-section distribution. There is σ convergence if the dispersion of the real income across economies tends to fall over time. The dispersion is measured as the sample standard deviation (σ_t), and σ convergence is formulated $\sigma_t \leq \sigma_{t-1}$, $\forall t$, (equality in the case of the economy being already in steady-state).

σ convergence is model free and tries to contribute to the measurement of convergence with some kind of information about the dynamics of the cross-section distribution. The idea is that both a negative coefficient of the initial condition and a decreasing cross-section dispersion over time would be sufficient to show convergence[8].

There are cases in which the sample standard deviation (σ_t) says little about the cross-section dynamics and additionally cannot account for some intra-distribution mobility. It is only a point in time statistic and may not be sufficient to describe the cross-section distribution dynamics. Sometimes σ convergence does not give any information about the cross-section dynamics, for example observations from a bimodal[9] distribution may have the same sample variance as observations from a uni-modal one. The interpretation of those distributions in terms of convergence is obviously quite different. Also, it happens that σ convergence cannot account for some intra-distribution mobility, an illustrative example of this is Figure 2, where $\sigma_t \leq \sigma_{t+1}$, $\forall t$, i.e. σ convergence holds and the standard deviation is identical in both cases, but 2a shows a situation where economies are overtaking each other and 2b displays a case of economies approaching monotonically. σ convergence cannot distinguish between 2a and 2b.

The main limitation of the conventional approach is that it relies on two single statistics (mean and standard deviation) that are not always sufficient to describe the cross-section dynamics distribution. They might not be very informative in terms of convergence.

Another attempt to incorporate the time dynamics in the cross-section regression approach has been to use panel data techniques. It still does not give any information on intra-distribution mobility and in most of the cases leads to inconsistent estimators.[10]

2.2. Time Series

On the other hand, time series approach does not offer a better analysis. It is not possible to study the convergence of economies by looking at their univariate dynamics; for instance the income in each economy can be integrated, but jointly cointegrated across economies, therefore interpreting a unit root in each economy income as evidence against the convergence hypothesis would be incorrect. It is the relative behaviour and cross-section mobility or transactional properties which matter in analysing convergence. Looking at cointregration gives accurate information but still in large cross sections may be missing part of the dynamics and transition characteristics, it is not telling the whole story. An alternative may be to do vector regression, by defining a vector of cross sections, but then the dimension of the vector makes it impossible to carry out the calculations.

In consequence the convergence cannot be studied without considering both cross-section variation and dynamic behaviour over time.

3. CROSS-SECTION DISTRIBUTION DYNAMICS. A simple exercise for Spain. Absolute Convergence.

An alternative approach suggested by Danny Quah (1993) which encodes the traditional ones mentioned above and overcomes some of their difficulties. This approach takes into consideration the whole cross-section distribution and does not impose any structure, nature of convergence, trend etc. It is able to account for transitional properties of the data, to characterize convergence, polarization, stratification etc.. It has been designed to deal with large cross-sections of data, in probability theory these structures, where both dimensions have the same order of magnitude, are called *Random Fields.* The current section presents this approach and illustrates it with a simple exercise for regions in Spain.[11]

3.1. The data:

The data in the empirical analysis are those in Dolado,J., J.M. Gonzalez-Paramo and J.M.Roldan (1984)[12]. The sample cover 50 regions *(provinces*: traditional administrative division) over a period of 35 years (1955-1989). The basic variable is real GDP per capita (as a proportion of the labour force) in each individual region (*province*) relative to the same variable for the country[13]. Figure 1 is a three dimensional plot of that variable, the main message of this graphic is that both dimensions of variation in the data are very important, they have a interesting dynamics. It is these dynamics that the approach tries to account for.

3.2 The Random Element : Cross section Distribution

At each point in time there is an income cross-section distribution which is simply the realisation of a random element in a space of distributions. The idea is to model their dynamics directly. The dynamics of these distributions involve changes of the exterior shape and intra-distribution mobility which give us relevant information in characterising convergence. Whether the cross-section distribution evolves to a multi-modal or to a degenerated distribution etc. are features to look at in dealing with convergence. Furthermore, the dynamics of each country's relative position is also a crucial component of the notion of convergence.

Consequently, the random element which turns out to be a cross-section distribution function has to be estimated from data. There are two approaches to density estimation. The *parametric approach* which assumes that the data are drawn from one of the known parametric families of distributions with unknown parameters. The underlying distribution can then be estimated by estimating these parameters from the observed data set. The *nonparametric approach* which requires weaker assumptions. The data are allowed to speak for themselves in determining the estimator of the density function. This is the approach we are using in our analysis because it does not require imposing any assumption on the exterior shape or about the moments of the density function from which the data are drawn. Due to their flexibility, nonparametric methods are able to detect structures deviating from traditional parametric forms.

The technical appendix in this chapter gives details on nonparametric estimation . It concentrates on kernel estimation, since it is the predominant procedure applied in this thesis. The reason to use kernel estimation is that conceptually is quite straightforward and the asymptotic theory is well developed.

We have estimated the density functions of relative per capita GDP across the 50 Spanish regions year by year. In this context of random fields, looking at the estimation of this cross section distribution is equivalent to looking at the values of the variable, observation by observation, in time series analysis. A few conclusions may be extracted from this exercise. Firstly when examining up to the end of the 1960s our estimators present two modes, a group of regions tend to concentrate at around one and a half times the average income of the country. During the 1970s the distribution turns to be unimodal and the dispersion decreases only very slightly. It is mainly during the period 1955-60 and 1977-84 when the range of relative differences among regions' incomes diminishes as the poorest regions approach the average. From 1985 onwards two modes in the density function show up at 70% and 110% of the average. Figures 2a to 2d present some nonparametric estimated cross-section density functions of relative per capita GDP for each period of 3 or 4 years which summarise the evolution described. This first look at the density estimates sugests that the first and second moment do not entirely describe the behaviour of the distribution.

Another very useful and illustrative way of looking at our data before doing any modelling is the cross-profile graph, Figure 4. It ranks the regions according to the relative income per capita in the first year of the sample (1955) and shows the evolution of the ranking over time. Each line refers to a single year and sketches the relative income of the provinces ordered according to the initial ranking. The larger the income inequality the steeper these cross profile lines are. Notice that for 1955 the line is obviously monotonicaly increasing; for the richest regions it is much steeper (corresponding to the second mode). With the exception of the highest quantiles, the lines become flatter. The inequality does not seem to be persistent between the intermediate quantiles of income. With respect to the richest regions the inequality is persistent specially until 1970. A high mobility is observed with regard to changes in each region's relative position, note the number of peaks in each line.

Looking at the random elements is very intuitive but the distributions are just point estimates for the sample period and cannot be assumed to reflect out-of-sample patterns. This does not give any information about the long run steady-state nor about the intra-distribution mobility.

3.3 Modelling the Distribution Dynamics. Continuous case.

In order to solve these objections and make progress in the analysis it becomes necessary to develop a formal structure. In other words to develop a law of motion for the cross-section distribution of income as realisations of random element in the space of distributions. Then we need a model for a stochastic process that takes values which are probability measures associated with the cross-section distribution.

Let λ_j be the probability measures (one each year) associated with the cross-section distribution. The simplest probability model is as follows:

$$\lambda_t = T^*(\lambda_{t-1}, u_t) \qquad (1.13)$$

T^* maps probability measures together with a disturbance to probability measures. It encodes information on how for example, the income levels of economies grow apart. T^* must be estimated from the data.

The stochastic difference equation in expression (1.13) is unmanageable. By ignoring the disturbance and iterating it can be written as (1.14) :

$$\lambda_{t+s} = (T^*)^s \lambda_t \qquad (1.14)$$

So that as s goes towards infinity it is possible to characterise the long run distribution of income across economies. In other words, to characterise the existence and uniqueness of the steady-state. But it is still difficult to deal with equation (1.14).

In order to make (1.14) tractable one can use the concept of *Stochastic Kernel* (Stockey, Lucas & Prescott (1989)). Consider the measurable space (R, *R*). R is the real line where the realisations of income fall and R is its Borel sigma algebra. B(R, *R*) is the Banach space of finitely additive functions. Let λ_{t+1} and λ_t be elements of B that are probability measures in (R, *R*). A *Stochastic Kernel* is a mapping M:Rx*R* -> [0,1], satisfying:

(i) $\forall$ a $\in$ R , M(a,.) is a probability measure.

(ii) $\forall$ A in *R*, M(.,A) is a sigma measurable function.

Then M(a,A) is the probability that the next state period lies in the set A, given that in this period the state is a.

For any probability measure λ on (R, *R*), $\forall$ A in R:

$$\lambda t+1 = \int M(x, A)\, d\lambda t(x) \tag{1.15}$$

Where: $M(.,.)$ is a Stochastic Kernel. And $\lambda_{t+1}\ (A) = (T^* \lambda_t\ )A$. T^* is an operator associated with the Stochastic Kernel that maps the space of probabilities in itself, (adjoin of the Markov operator associated to M).

Equation (1.15) measures the probability that the next period state lies in the set A, if the current state is drawn according to the probability measure λ_t. $(T^* \lambda_t)$ is the probability measure over the next period state, if λ_t is the probability measure over the current period. So we will consider T^* in (1.13) and (1.14) as being generated in the differential equation (1.15).

Thus the *Stochastic Kernels* are a complete description of transitions, but they are simply point estimates, there is not a fitted model. Inference cannot be performed and it is not possible to calculate the long run. But there may be addressed some questions like whether convergence has taken place or whether it is taking place.

Figures 5 and 6 are non-parametric[14] estimated *stochastic kernels* for relative income of k-year transitions (k=1, 5). Figures 5 display three dimensional plots of the transitions probability function, while Figures 6 display the contours of the function in Figures 5. A slice parallel to the t+k axis gives a probability density that describes transitions from a part of the income distribution to another in k periods. The location of the probability mass will give us the information about persistence and mobility that allows us to extract conclusions in term of convergence. Concentration of the probability mass along the positive sloped diagonal indicates high persistence in the economies' relative position, and consequently low mobility. Concentration along the negative sloped diagonal would indicate that regions are overtaking each other in the income ranking. The transition probability describing horizontal lines (parallel to t+k) shows that there is very low persistence, the probability of being at any point in t+k is independent of the position in t. Finally, if the mass of probability is orthogonal to the t + k axis there is convergence.

According to Figures 5a and 6a, during the sample period (1955-89), the Spanish regions have a low probability of changing their relative position in one year (k=1), tending to remain where they are relative to each other, although the contour in Figure 6a shows a tendency of the poorest regions to concentrate around 50% of the average income.

Following the analysis in Dolado *et al* (1994), we divide the whole sample of 35 years into the three following[15] sub-periods: 1955-64, 1964-77 and 1977-88. Looking at these we can observe that this concentration took place in the period 1955-64.

Regarding longer horizons, say 5 year transitions (k=5), the contour, Figures 6e-h, in most of the cases shows the probability placed along the main diagonal but less concentrated than before, which indicates there is a higher intra-density mobility. The contour for the 1956-64 period (Figure 6f), again seemingly indicates that the poorest regions tend to approach the rest and concentrate at a level of income of around 50% the average. It shows a high persistence for the high-income regions. For the period 1964-77 the probability of transition is mainly over the diagonal, although there is much more mobility than before for the richest. The contour for the period 1977-89 is steeper than the diagonal only for the poorest regions.

3.4 Modelling the Distribution Dynamics. Discrete case.

There is still a question about convergence that cannot be addressed, this being: will economies converge in the long run? To address this long term calculations and inferences are required. The way to proceed will be to work out T* from (1.15) and do the calculations in (1.14). T* can be approximated by assuming a countable state-space for income levels $S_t = \{s_{1t}, s_{2t}, .., s_{rt}\}$. I.e. T^* is simply a *transition probability matrix* Q_t such that $\lambda_t = Q_t\,(\lambda_{t-1}, u_t)$

(1.16) is tractable. Now Q contains all the relevant information about convergence and long run calculations are possible.

Under some regularity conditions, Q_t defined for the fixed grid ($S_t = S$) is time-invariant and the long run calculation in (1.14) can be done in an explicit way. The sequence of powers of this matrix converges to a matrix whose rows (all of them identical) are the *ergodic distribution*, which allows us to talk about steady-state. In this setting convergence is characterised as the ergodic distribution tending towards a mass point.

Taking per capita income for each region, relative to the sample average, we discretize the space of possible values, S, in *r* states. For example the state i=(0.7,1.1) includes the regions which have an income between 0.7 times and 1.1 times the average for the country. The discretization defines a grid that can be thought of as an estimator of the initial unconditional probability distribution λ. Each element of the Q matrix indicates the probability of transition from one state to another in *k* periods: the (q_{ij}) entry is the probability that a region in state *i* transits to the state j. Each row is a conditional probability vector. Each row of the matrix (in terms of the *stochastic kernel* above) is analogous to the density probability defined for each point in S when cutting the figure at that point by a plane parallel to the t+k axis.

Table 1 presents some estimates of the transitions matrix Q. The length of the defined states are different for providing a uniform distribution for the first year of the sample. In fact they are quite different, much wider for low and high income than for the intermediate ones.

The first column is the total number of transitions over the whole time sample, starting at each state. The rest displays an estimator of the time-invariant transition probability matrix *rxr* for a single period, calculated as an average over the total sample. The values in the main diagonal are high for the poorest and richest regions, being around 88%, which indicates that the probability of a region moving from being on the lowest and highest income group in one year is around 0.10. For instance, the probability that a region with income between 0.4 and 0.7 times the average moves in one period to an income between 0.7 and 0.9 times the country average is 0.11. There is less persistence for the intermediate groups, the probability of being off the diagonal for those reaches 20% and is symmetrically distributed between the probability of moving up or down (to a higher and lower relative position). Finally, Table 1 also offers an estimator of the ergodic distribution, which is the closest concept to the steady-state in this setting. The ergodic distribution tells us the unconditional probability for an economy to end up in a particular income range.

Recall that the states are defined in a way such that the initial distribution is uniform. We end up with a distribution that is not degenerated at its mean value, but it gives a slightly higher probability of reaching the average state. Table 2 displays some estimates of the transitions matrix Q for the sub-periods defined above. For the second period (1964-77) the evidence is of less mobility than for the other two periods, except for the richest regions for which the probability of moving to a lower level remains about the same as in the rest of the periods.

These calculations require time-invariant transition probability, which is not always reasonable for long periods in which, for example, some economic structural changes may happen. And in our sample the previous analysis suggests different patterns of behaviour for different time sub-samples.

Alternatively it is possible to construct time-variant transition matrices by fixing the probability vectors to be uniform and identical for every time point ($\lambda_t = \lambda$) and define a time-variant grid $S_t = \{s_{1t}, s_{2t}, .., s_{rt}\}$. Associated to that there is a sequence of transition probability matrices, Q_t.

For example, let r be 4 , and define $\lambda_t = \lambda =$ (0.25, 0.25, 0.25, 0.25), then the set of quantiles defining the grid consequently determines the sequence of cross-section distribution. Associated to these grids there is a sequence of transition probability matrices. The change in the grid describes the evolution of the cross-section distribution, i.e. the dynamics that refer to the exterior shape. The sequence of transition probability matrices shows the intra-distribution mobility .

The characterisation of convergence in this setting focuses on the sequence of quantiles tending to approach and it allows characterisation of the long-run. For example the sequence of quantiles degenerating to the mean will indicate convergence.

Table 3 and Figure 7 show the sequence of quantiles. Again the picture is one of persistence during the second period and slight convergence during the first and third periods, especially the limits of the lowest quantil have been rising since 1977.

4. BRINGING CONDITIONAL INFORMATION

Over and above all the problems faced by the traditional approach, the obvious and essential question that arises when talking about *conditional convergence* is whether the fact of a single economy reaching its own steady-state, or more precisely tending to its estimated[16] steady-state, has any practical interest.

This section departs from Solow predictions[17] and from conditional convergence understood as the classical model predicts. We simply try to illustrate how to bring conditional information into the cross-section distribution dynamics approach. Conditioning in this context means analysing the residuals from a first stage regression in which the effects of the variable we are conditioning on are removed. In other terms the idea is to analyse the income disparities that are orthogonal to the conditioning variables. This also may be of little interest since a great deal of information is likely to be removed.

4.1 Conditioning variables[18]

For Spain, we choose as the conditioning variable the inter-regional migration flows as a percentage of the previous year population. The data are those in Bentolilla[19] and Dolado (1991), the sample covers the 50 provinces above, over the period 1962 to 1986.

Consequently, the whole exercise consists of studying the dynamics of the distribution of income disparities which cannot be explained by migration flows. The methodology in this exercise follows the one in Quah (1996). Firstly we perform a causality test for bivariate VARs in the growth rates of regional per capita GDP (relative to the country GDP growth rates) and migration rates. There is significant dynamic correlation between those variables: GDP growth rates help to predict future values of net migration flows and vice versa, net migration helps to predict GDP growth rates.

4.2 Conditioning regression

There is not a structural interpretation of the causality above and consequently, following Quah (1996), we estimate a two sided projection of GDP growth rates on migrations rates, (Table 4). Then we accumulate the residuals from this projection (recall that the depend variable is the growth rate) to get the corresponding residual components in per capita regional GDP[20], (relative to the country GDP), which is going to be our basic variable.

4.3 Some results

Given that the sample available for migrations is shorter than the one for GDP (raw data), we will compare the results after conditioning with the ones for raw data during the 2nd period analysed in section 3, i.e. 1964-1977, which was the one showing more persistence (or less convergence).

Figure 8 presents the cross profile plot for the raw data (before conditioning) and for the residuals after conditioning out migrations. The evidence from here seems to indicate that migration had a negative effect on the convergence process.

Table 5 presents the time-invariant transitions matrix, defined by taking as the initial distribution a uniform one. Comparing this with Table 2.b, Table 5 shows higher values in the diagonal for the regions with higher income per capita and lower for all the others. In other words the poor regions are more likely to move to higher levels, for example the probability of leaving the lowest state is now 0.76 while before conditioning it was 0.89, and the richest are more likely to stay in their income group. Looking at Figure 9, which displays an estimator of the stochastic kernel comparable with figures 6g. and 7g, the message is the same; after conditioning out the effect of migrations the contour presents a steeper slope in its middle and lower parts.

The effect of the migrations during the period analysed prevents the lower income regions catching up with the rest and on the other hand helps the richest to approach the average.

5. CONCLUSION

This paper describes some of the criticism concerning the traditional approaches to the hypothesis of convergence. It presents the econometric methods for studying the dynamics of the distributions as an alternative approach to characterising convergence and analyses the cross-section distribution of relative GDP per capita across regions in Spain. The main conclusions from this exercise are the following:

. The GDP disparities in Spain have narrowed (slightly) during the periods 1955-64 and 1977-89, and they showed persistence during the period 1964-77.

. During the years 1966 to 1977 the inter regional migration flows prevented the poor regions from catching up with the rest.

ENDNOTES

1 Bianchi (1995), Desdoigts (1994), Paap and Dijk (1995) etc. have also folowed this approach.

2 For an application of this approach in industrial organisation and in labour market empirical literature see Koopmans and Lamo (1995) and Kogning (1995) respectively.

3 Quah (1994b) illustrates how the uniformity of this 2% may arise from a unit root in the time series, instead of from reasons related to the dynamic of economic growth.

4 β convergence is a necessary but not sufficinet condition for σ convergence.

5 Mankiw, Romer & Weil (1992) augmented Solow's model by adding human capital accumulation. In this case the production function is as follows: $Y(t)=K(t)^{\alpha}H(t)^{\beta}(A(t)L(t))^{1-\alpha-\beta}$ where H is the stock of human capital. We can consider that in the regression of the original model, H is an omitted variable. For simplicity and without any loss of generality reference will be made to the equations for the original model.

6 Chapter 2 in this thesis looks at this point in more detail.

7 Same exercise for discrete choice models is done in the chapter 3 of this thesis.

8 Quah (1995) shows why no combination of β and σ convergence can provide a satisfactory solution.

9 See Quah(1996) for more about bimodal distributions.

10 See Pesaran and Smith (1995) and Canova and Marcet (1996).

11 Dolado, J., J.M Gonzalez-Paramo and J.M.Roldan (1994) study convergence across Spanish regions in the framework of the traditional approach.

12 I thank J. Dolado for kindly providing me with the data. Some interpolation was required for which I used the sectoral sructure of each region and the sectoral GDP growth rates.

13 This normalization is a way to abstract each individual region from the country growth and fluctuations.

14 Obtained using the squared of standard Epanechnikov kernel for estimating the joint density $f(X_{iT}, X_{iT+\tau})$ and then rescaling to obtain the conditional probability. X is the per capita income for each individual region relative to the same variable for the country, and τ= 1,5. The bandwidth is chosen by least square cross-validation (see Silverman(1986) section 3.4.3.). All the calculations were done with Quah's shell *tSrF*.

15 As in Dolado, Paramo and Roldan (1994).

16 Abstracting from the problems of estimating the steady-state.

17 Chapter 2 of this thesis take as the baseline hypothesis the Solow model predictions and performs a sequential conditioning exerciseThe objective is not to test whether there is convergence after conditioning, but to test the solow model.

18 The purpose of our exercise is merely to illustrate the technique. Levine and Renel (1992) found that the significance of hardly any of the conditioning variables (except saving rates) can be claimed robust. We use the variable migrations, in spite of the interpretation problems, due to its availibility.

19 I thank Samuel Bentolila for making the data available to me.

20 For more details in this procedure see the data appendix in Quah (1996a).

TABLE 1

Spain Relative Per-Capita Income
First Order Transition Matrix, Time Stationary 1956-89

Upper End of the States	0.756	0.906	1.061	1.879
(r)	(1)	(2)	(3)	(4)
412:	0.89	0.11	0.00	0.00
412:	0.10	0.79	0.11	0.00
415:	0.00	0.10	0.78	0.12
1411:	0.00	0.00	0.12	0.88
Ergodic Distribution	0.225	0.247	0.265	0.260

TABLE 2a
Spain Relative Per-Capita Income
First Order Transition Matrix, Time Stationary 1956-64

Upper End of the States	0.727	0.872	1.053	1.879
(r)	(1)	(2)	(3)	(4)
116:	0.84	0.16	0.00	0.00
111:	0.12	0.77	0.11	0.00
114:	0.00	0.11	0.75	0.15
109:	0.00	0.00	0.15	0.85

TABLE 2b
Spain Relative Per-Capita Income
First Order Transition Matrix, Time Stationary 1964-77

Upper End of the States	0.746	0.909	1.073	1.583
(r)	(1)	(2)	(3)	(4)
177:	0.91	0.08	0.01	0.00
177:	0.07	0.80	0.12	0.00
170:	0.00	0.09	0.81	0.10
176:	0.00	0.01	0.11	0.88

TABLE 2c
Spain Relative Per-Capita Income
First Order Transition Matrix, Time Stationary 1977-88

Upper End of the States	0.798	0.932	1.065	1.501
(r)	(1)	(2)	(3)	(4)
147:	0.86	0.14	0.00	0.00
152:	0.14	0.75	0.11	0.00
154:	0.00	0.13	0.73	0.14
147:	0.00	0.00	0.12	0.88

TABLE 3
Quantiles (0.25, 0.5, 0.75)
Spain Relative Per-Capita Income

1956	0:0.457	1:0.677	2:0.843	3:1.071	4:1.864
1957:	0:0.446	1:0.705	2:0.873	3:1.026	4:1.653
1958:	0:0.457	1:0.712	2:0.884	3:1.038	4:1.618
1959:	0:0.467	1:0.716	2:0.890	3:1.045	4:1.591
1960:	0:0.481	1:0.712	2:0.839	3:1.051	4:1.644
1961:	0:0.486	1:0.723	2:0.839	3:1.048	4:1.642
1962:	0:0.453	1:0.736	2:0.848	3:1.058	4:1.540
1963:	0:0.459	1:0.757	2:0.857	3:1.062	4:1.482
1964:	0:0.518	1:0.721	2:0.871	3:1.033	4:1.472
1965:	0:0.528	1:0.768	2:0.925	3:1.088	4:1.571
1966:	0:0.532	1:0.768	2:0.922	3:1.083	4:1.561
1967:	0:0.463	1:0.736	2:0.848	3:1.028	4:1.435
1968:	0:0.471	1:0.749	2:0.856	3:1.028	4:1.402
1969:	0:0.471	1:0.693	2:0.886	3:1.057	4:1.481
1970:	0:0.459	1:0.693	2:0.885	3:1.065	4:1.460
1971:	0:0.443	1:0.727	2:0.904	3:1.071	4:1.390
1972:	0:0.443	1:0.731	2:0.915	3:1.066	4:1.390
1973:	0:0.422	1:0.749	2:0.896	3:1.084	4:1.391
1974:	0:0.414	1:0.749	2:0.926	3:1.086	4:1.423
1975:	0:0.419	1:0.729	2:0.929	3:1.068	4:1.328
1976:	0:0.422	1:0.730	2:0.929	3:1.051	4:1.354
1977:	0:0.437	1:0.753	2:0.939	3:1.045	4:1.335
1978:	0:0.450	1:0.774	2:0.933	3:1.045	4:1.350
1979:	0:0.484	1:0.802	2:0.944	3:1.028	4:1.339
1980:	0:0.490	1:0.827	2:0.927	3:1.037	4:1.327
1981:	0:0.464	1:0.843	2:0.933	3:1.044	4:1.316
1982:	0:0.469	1:0.763	2:0.909	3:1.029	4:1.374
1983:	0:0.496	1:0.828	2:0.927	3:1.081	4:1.327
1984:	0:0.489	1:0.800	2:0.898	3:1.056	4:1.375
1985:	0:0.484	1:0.776	2:0.899	3:1.075	4:1.450
1986:	0:0.510	1:0.783	2:0.918	3:1.074	4:1.451
1987:	0:0.538	1:0.780	2:0.926	3:1.081	4:1.490
1988:	0:0.526	1:0.775	2:0.898	3:1.066	4:1.447
1989:	0:0.535	1:0.754	2:0.874	3:1.074	4:1.328

TABLE 4

Conditioning Regression

Dependent variable: DGP growth rates_Spanish regions(relative to the country one)

Conditioning variable: migration flows rates

two sided projection

	coefficiets-OLS	se(OLS)	se (HKC))
migration (3)	1.139	1.041	0.833
migration (2)	0.361	1.300	1.007
migration (1	4.054	0.990	0.822
migration	20166	0.750	0.652
migration (-1)	0.652	0.781	0.627
migration (-2)	0.066	0.795	0.673
migration (-3)	0.494	0.519	0.439
constant	0.961	0.007	0.007

R^2= 0.296 , OLS, Sample 50 Regions, 1966-82

TABLE 5
Spain Residuals First Stage Regression
Conditioning out migration flows
(unexplained by migration flows)
First Order Transition Matrix, Time Stationary 1966-77

(r)	(1)	(2)	(3)	(4)
146:	0.76	0.24	0.00	0.00
146:	0.02	0.71	0.27	0.00
139:	0.00	0.02	0.71	0. 27
119:	0.00	0.00	0.05	0.95

FIGURE 1
GDP Per Capita

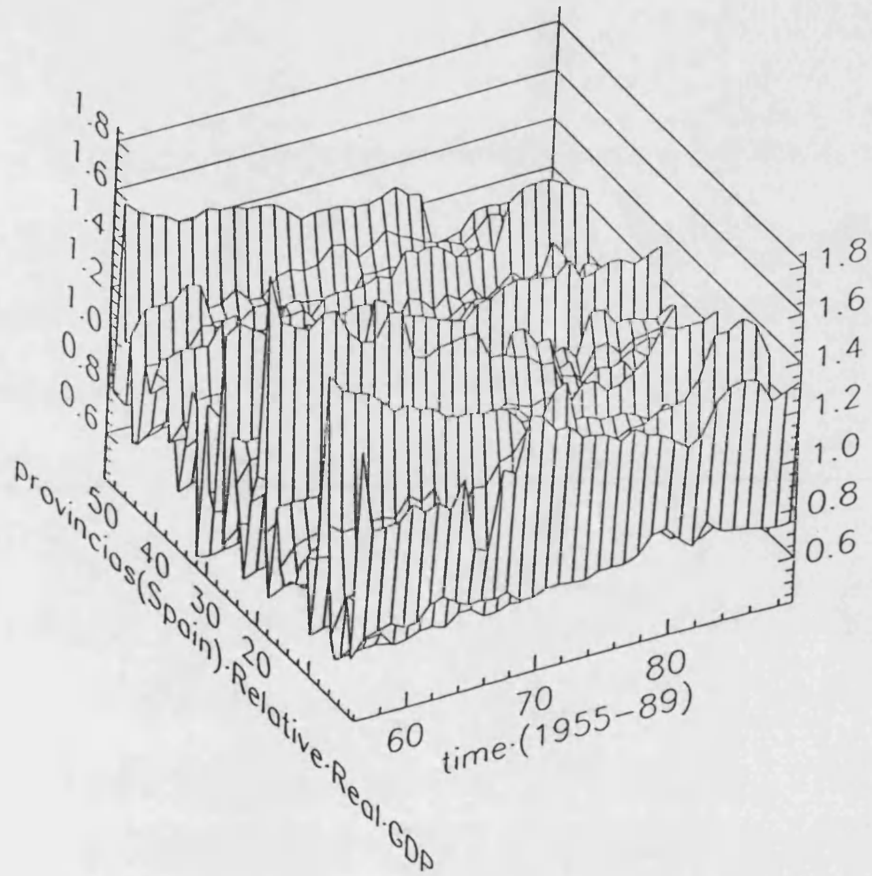
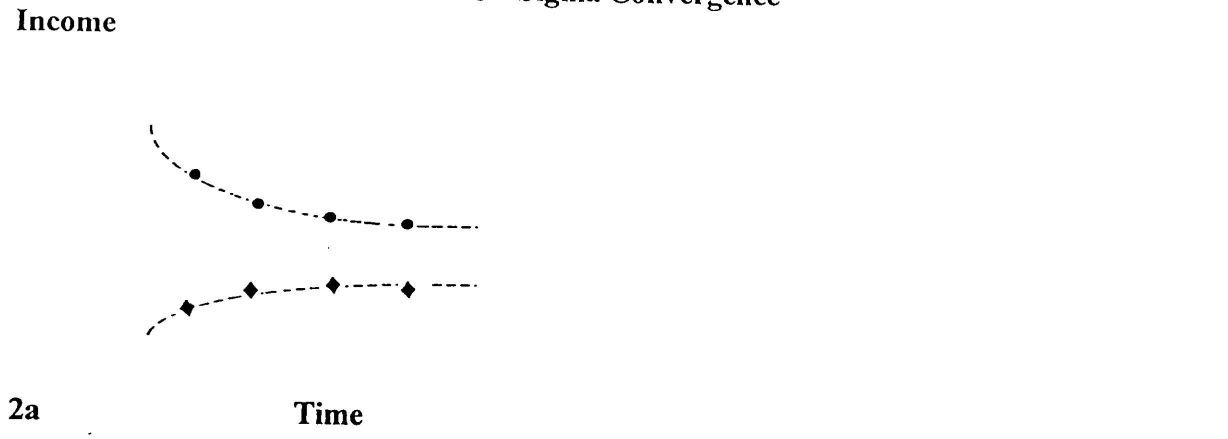
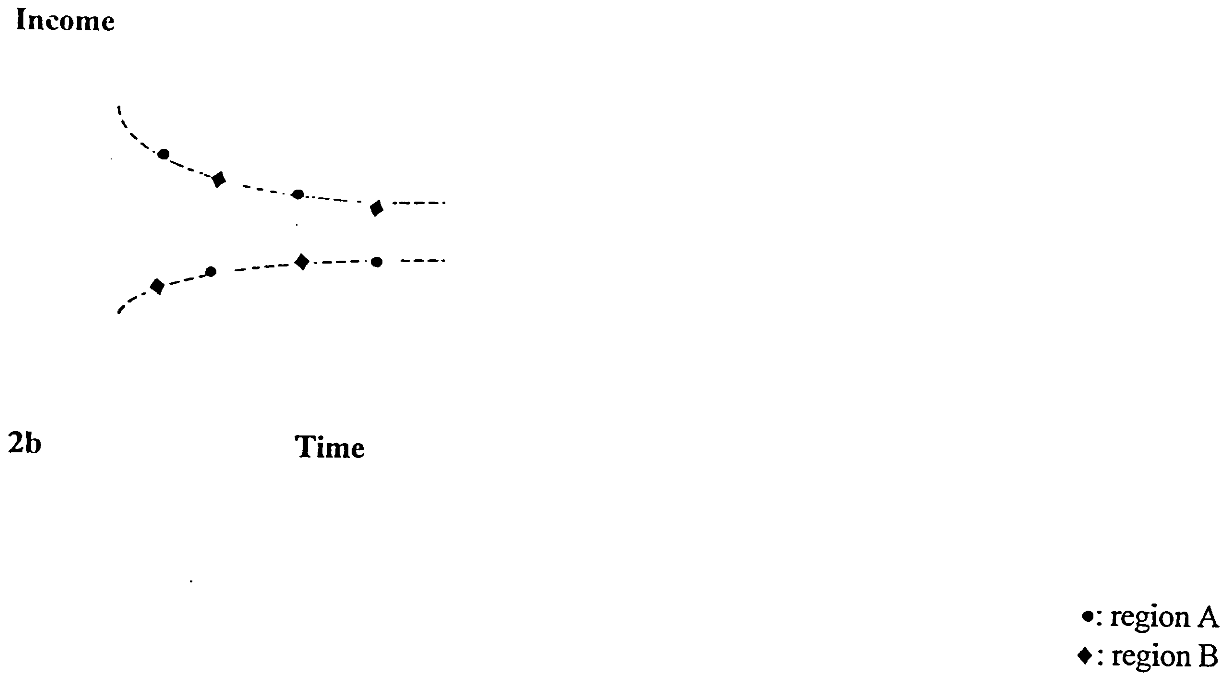


FIGURE 2
On Sigma Convergence



2a



2b

FIGURE 3
Estimated Density Functions
Spain.Relative GDP Per Capita

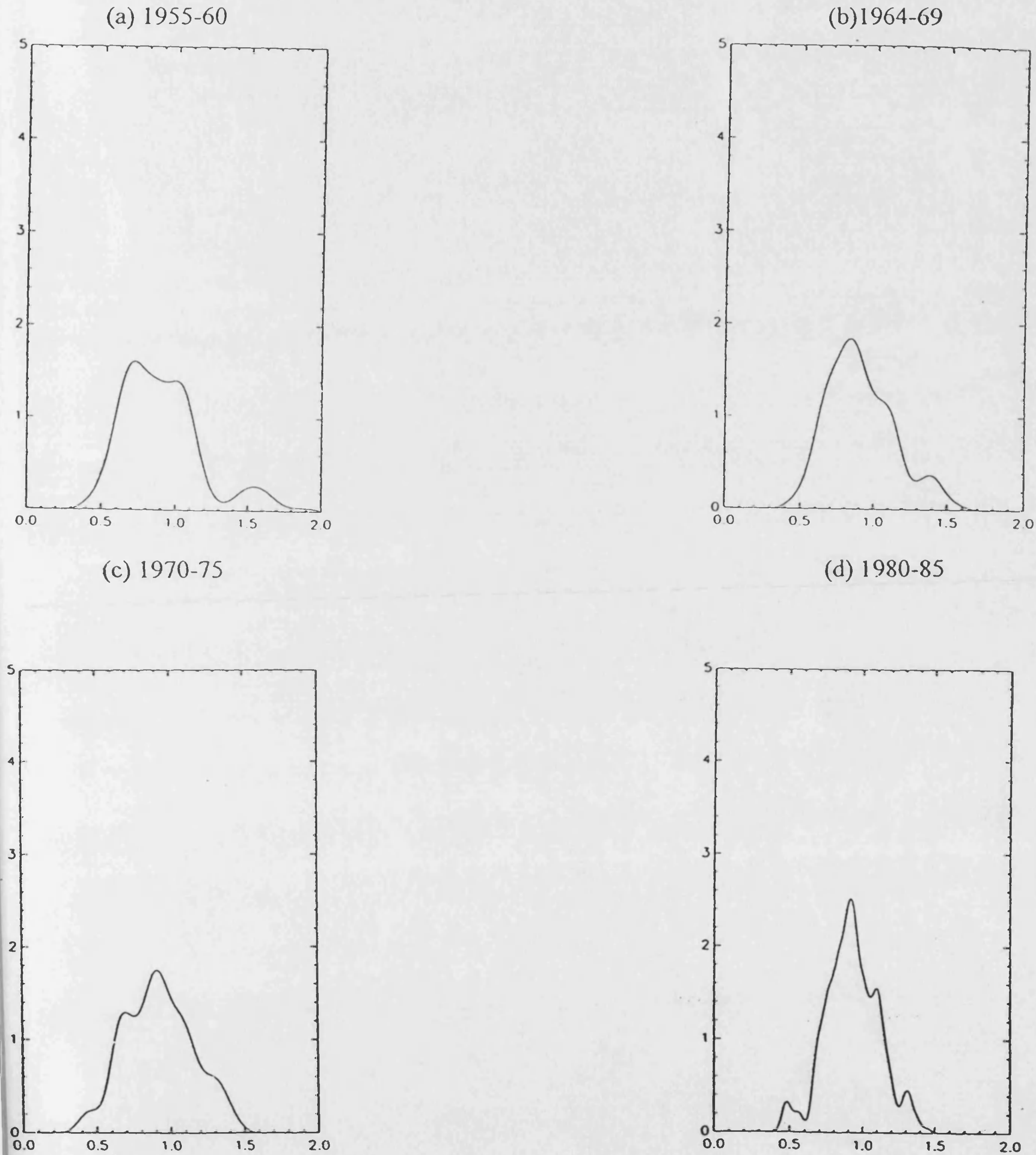


FIGURE 4
Spain.Relative GDP Per Capita, Cross Profile
1956, 1960, 1971, 1979, 1988

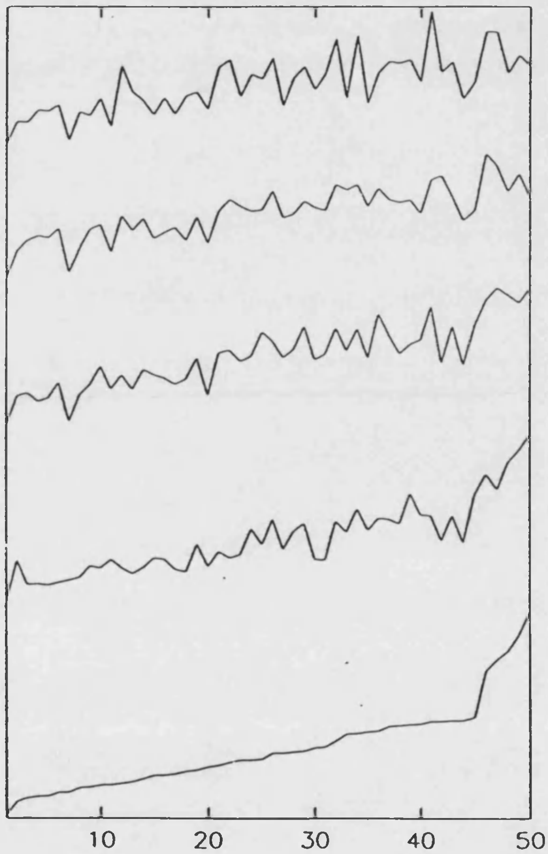
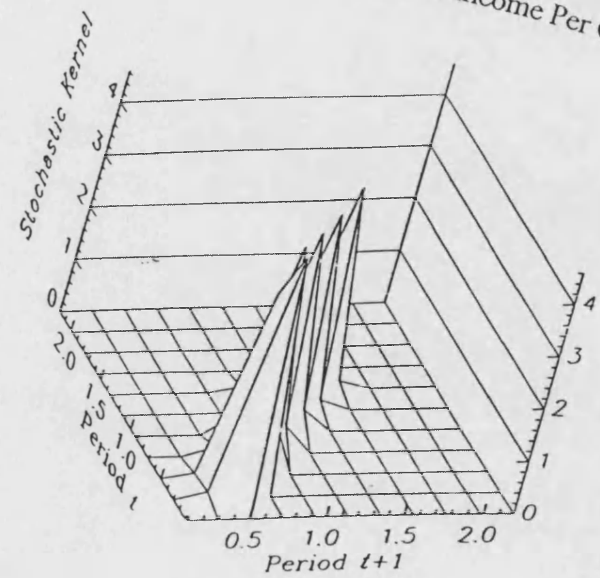
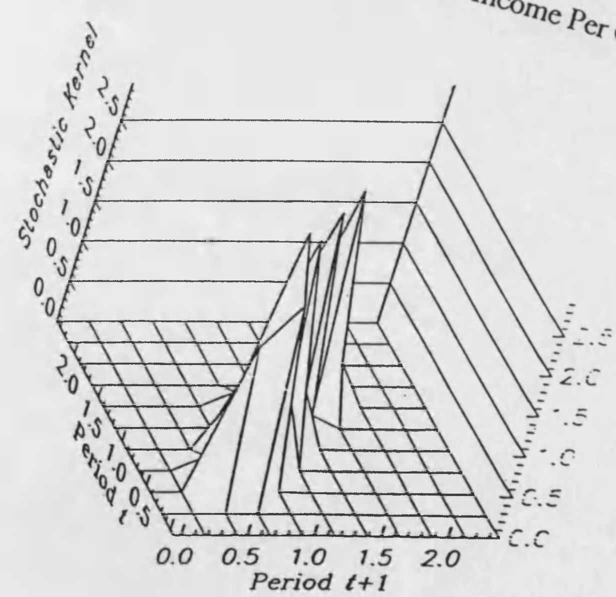


FIGURE 5a
SPAIN.Relative Income Per Capita



Stochastic Kernel. 1955-88
1 Year Transition

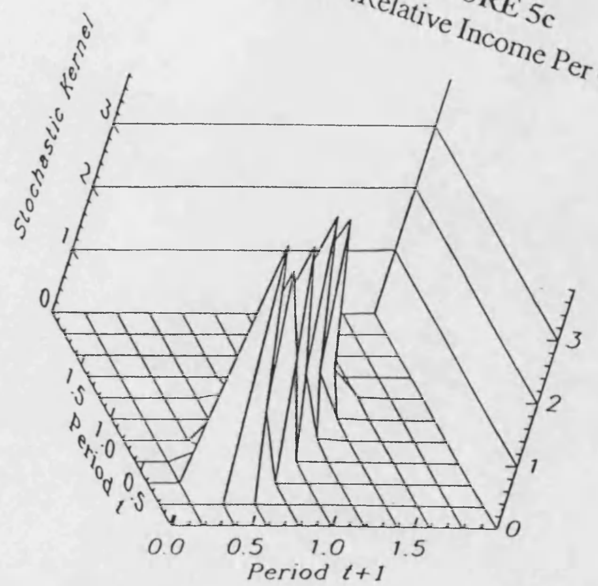
FIGURE 5b
SPAIN.Relative Income Per Capita



Stochastic Kernel
1 Year Transition 1955-64

FIGURE 5c

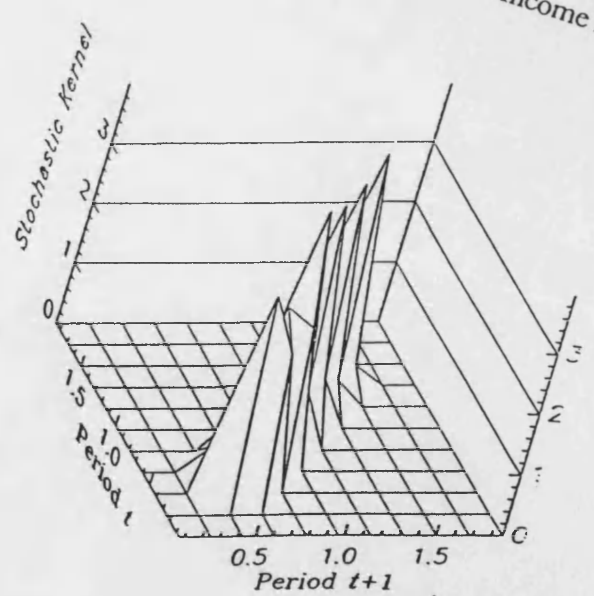
SPAIN.Relative Income Per Capita



Stochastic Kernel
1 Year Transition 1964-77

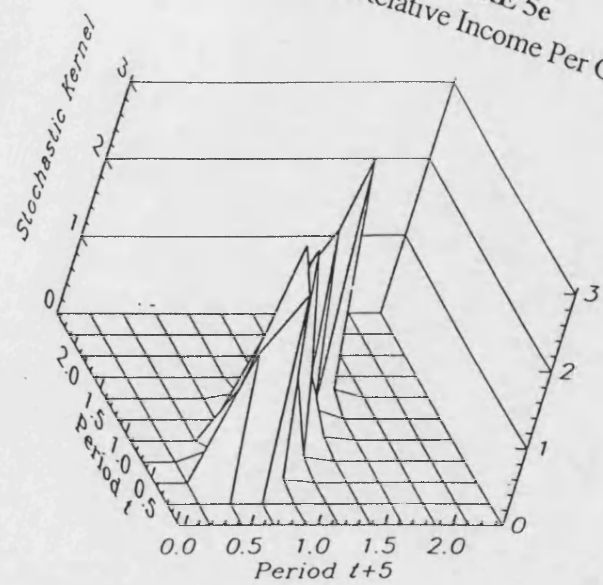
FIGURE 5d

SPAIN.Relative Income Per Capita



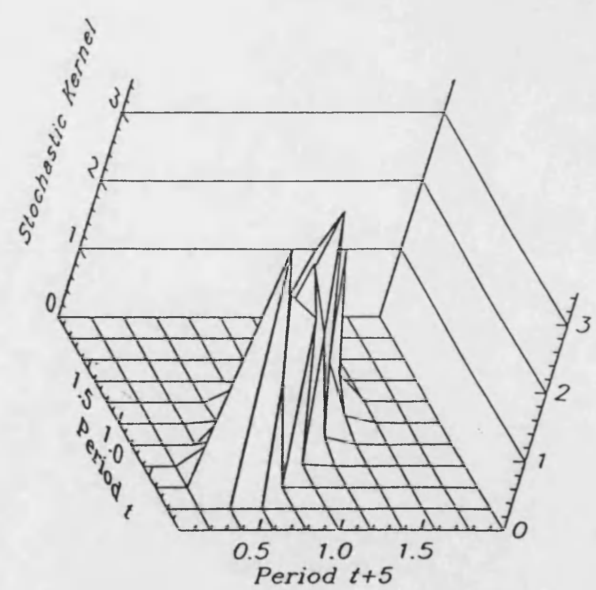
Stochastic Kernel
1 Year Transition 1977-88

FIGURE 5e
SPAIN.Relative Income Per Capita



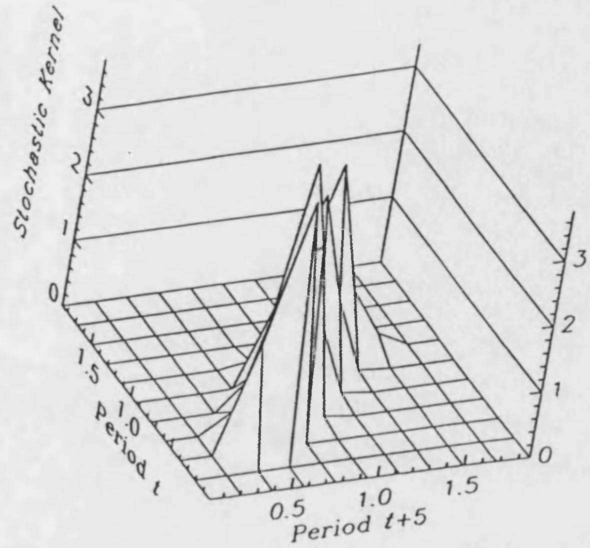
Stochastic Kernel
5 Year Transition 1956-64

FIGURE 5f
SPAIN.Relative Income Per Capita



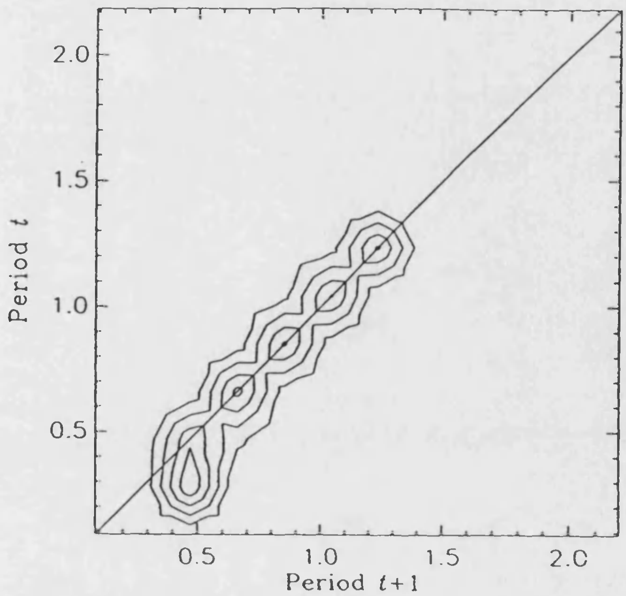
Stochastic Kernel
5 Year Transition 1964-77

FIGURE 5g
SPAIN.Relative Income Per Capita



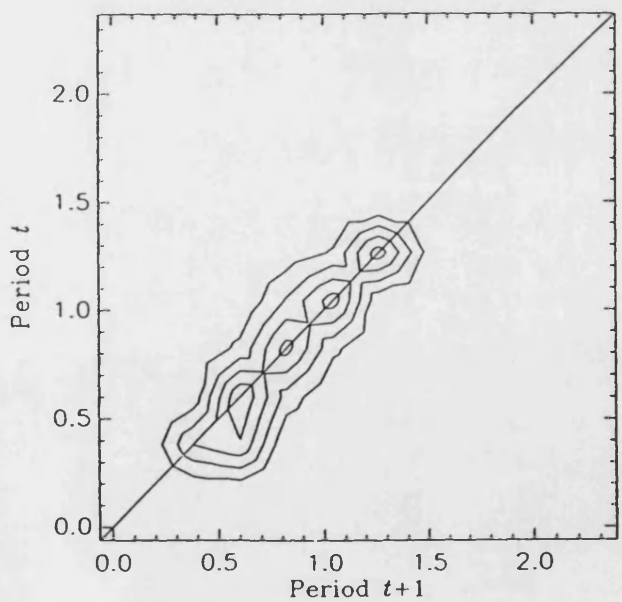
Stochastic Kernel
5 Year Transition 1977-85

FIGURE 6a
SPAIN.Relative Income Per Capita



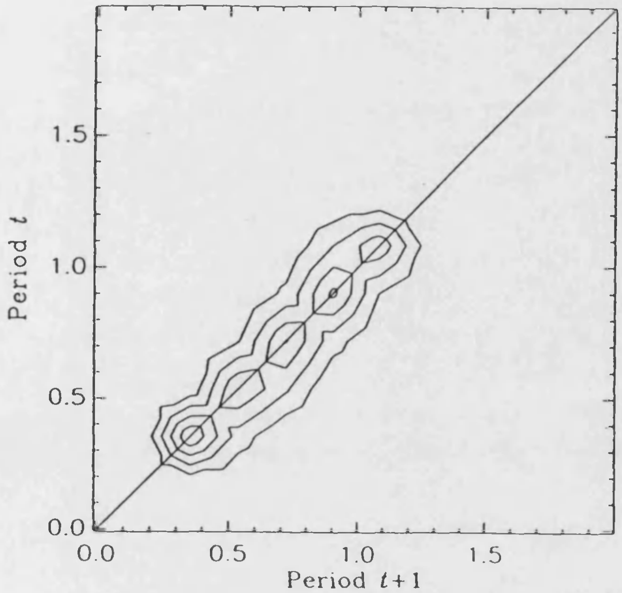
Stochastic Kernel
1 Year Transition. Contour Plot. 1955-88

FIGURE 6b
SPAIN.Relative Income Per Capita



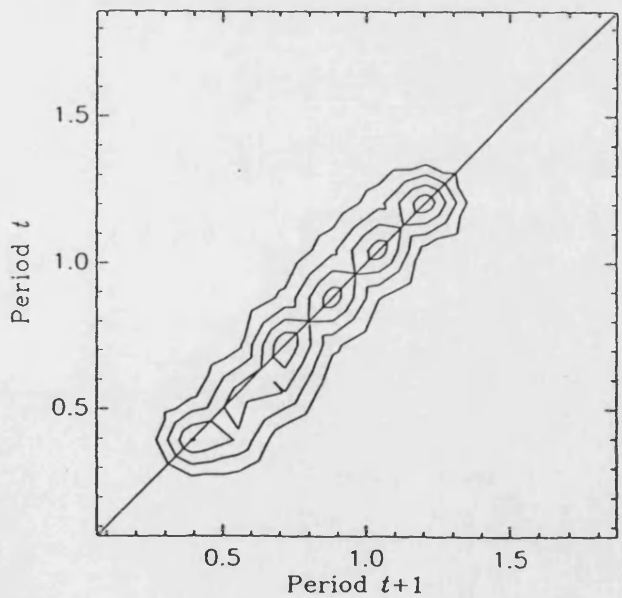
Stochastic Kernel
1 Year Transition 1955-64. Contour Plot.

FIGURE 6c
SPAIN.Relative Income Per Capita



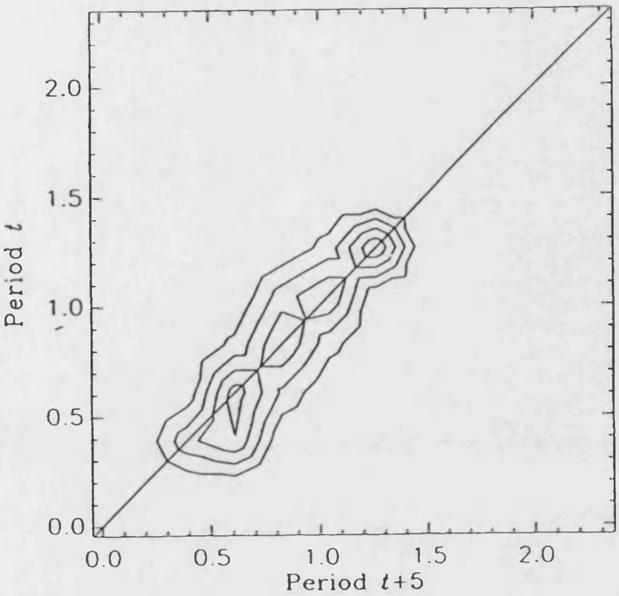
Stochastic Kernel
1 Year Transition 1964-77. Contour Plot.

FIGURE 6d
SPAIN.Relative Income Per Capita



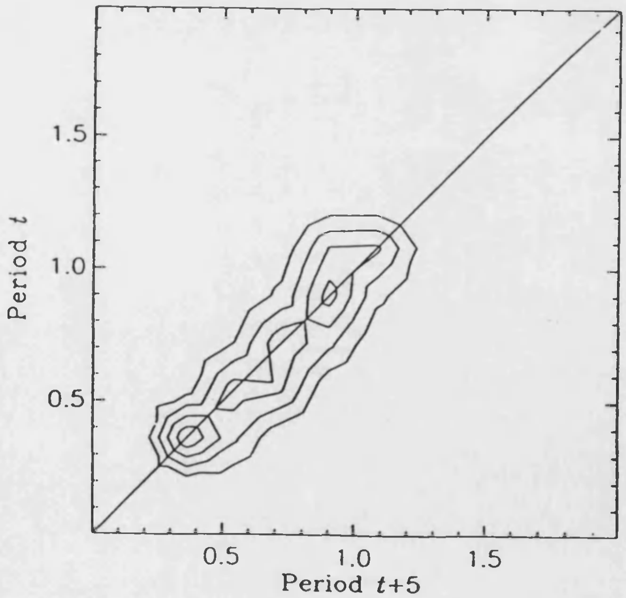
Stochastic Kernel
1 Year Transition 1977-88. Contour Plot.

FIGURE 6e
SPAIN.Relative Income Per Capita



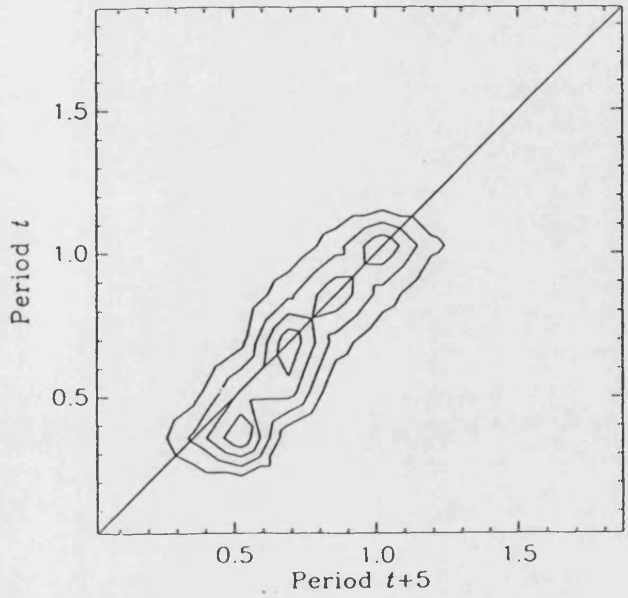
Stochastic Kernel Contour Plot.
5 Year Transition 1956-64

FIGURE 6f
SPAIN.Relative Income Per Capita



Stochastic Kernel contour
5 Year Transition 1964-77

FIGURE 6g
SPAIN.Relative Income Per Capita



Stochastic Kernel Contour Plot.
5 Year Transition 1977-85

FIGURE 7
Spain Relative Income Per Capita
Quantiles 1957-89

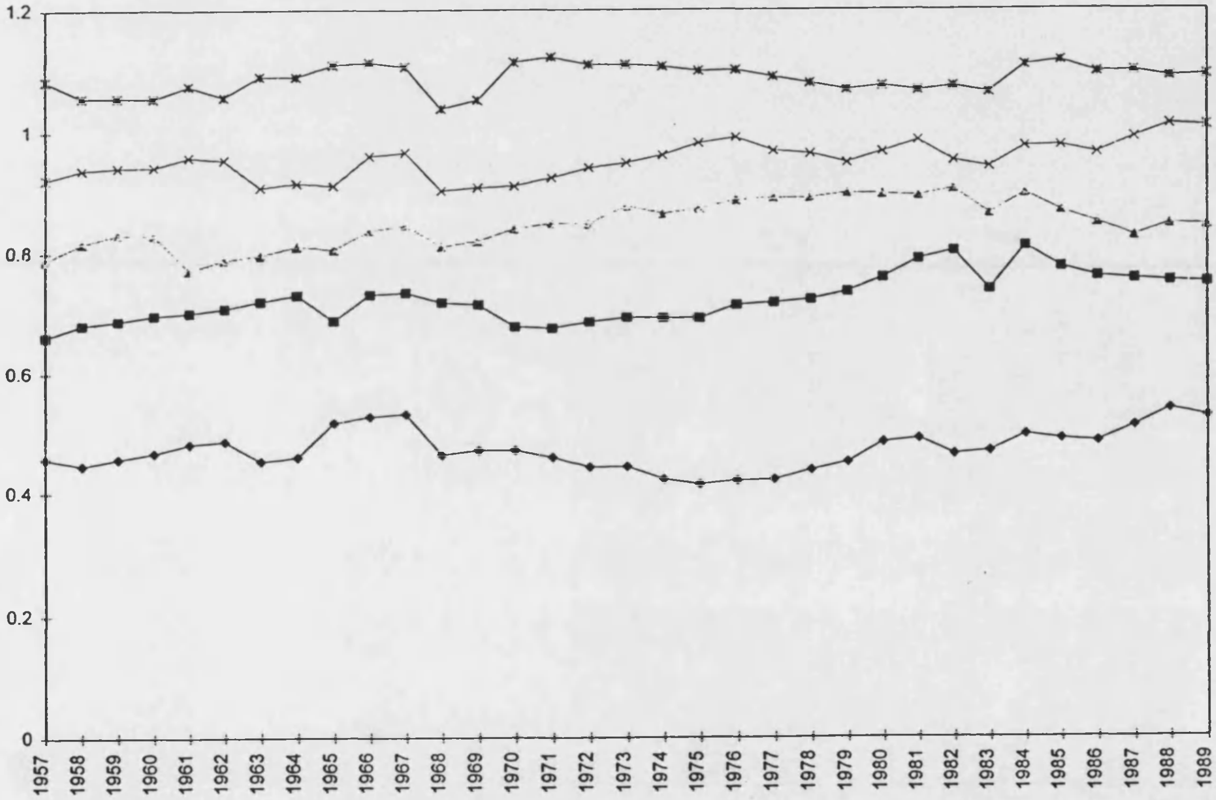


FIGURE 8a
Spain.Relative GDP Per Capita, Cross Profile
1966, 1970, 1975

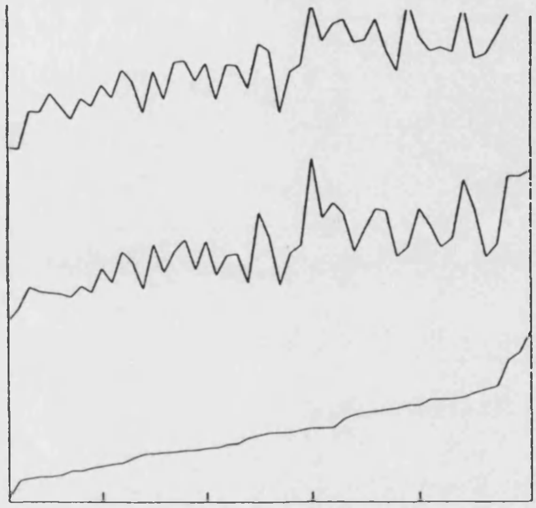


FIGURE 8b
Spain.Relative GDP Per Capita, Cross Profile
Uexplained by migration flows
1966 ,1970 1975

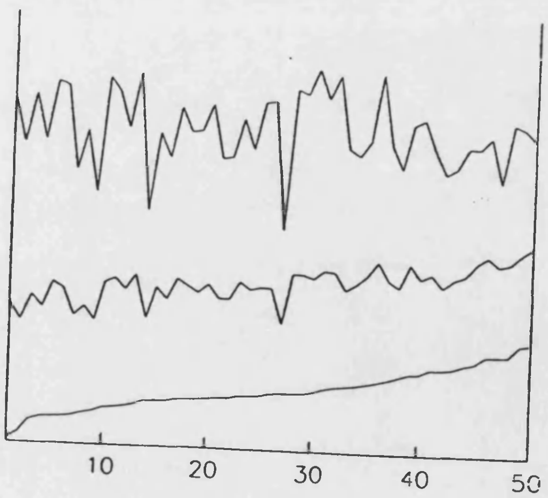
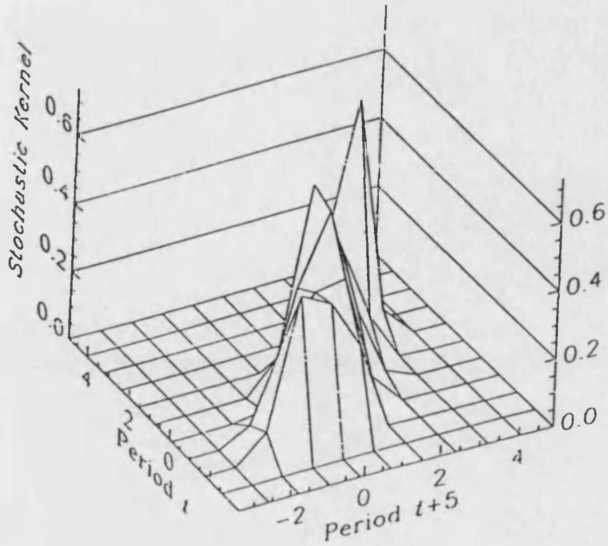
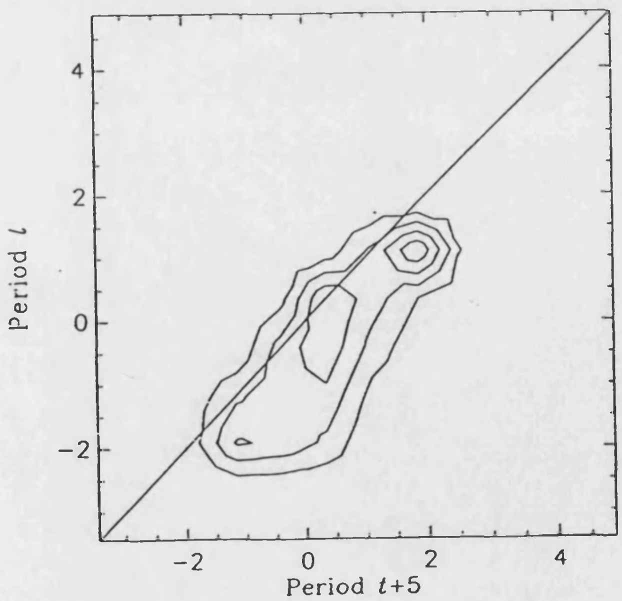


FIGURE 9a
SPAIN.Relative Income Per Capita
Unexplained by migration flows



Stochastic Kernel
5 Year Transition 1966-77

FIGURE 9b
SPAIN.Relative Income Per Capita
Unexpained by migration flows



Stochastic Kernel. Contour
5 Year Transition 1967-77

TECHNICAL APPENDIX

NONPARAMETRIC DENSITY ESTIMATION

Density estimation deals with the construction of an estimate of density functions from observed data. There is a rich literature on this topic (see among others the almost classical books by Tapia and Thompson (1978), Prakansa Rao (1983),
and Silverman (1986)). The ideas and techniques exposed here as well as the analysis in the paper follows Silverman (1986).

The remainder of this appendix consists of two sections. Section I concentrates on the estimation of a density underlying a set of *univariate* observations. Section II briefly considers the multivariate case.

I. UNIVARIATE DATA

This section is organised as follows. Paragraph I.1 sketches out some of the main nonparametric methods for univariate density estimation. Paragraph I.2 discusses the basic properties of the kernel estimate which is one of the best understood methods.

We consider a sample of n independent and identically distributed observations $\{X_1, X_2, \ldots, X_n\}$ from a continuous univariate distribution, whose underlying density function $f(x)$ is estimated from the data by $\hat{f}(x)$.

I.1. SOME METHODS

I.1.1. The histogram

The histogram is the traditional and most popular density estimator. In order to construct the histogram the data set is divided into a number of « bins ». Given a fixed origin x_0 we define these bins to be the interval $[x_0 + mh, x_0 + (m+1)h)$, where m is a positive or negative integer and h is a positive parameter called the *bin width*. The histogram is then defined by

$$\hat{f}(x) = (1/nh) \quad [\textit{number of } X_i \textit{ in the same bin as } x]. \qquad \text{(A.1)}$$

The parameter h controls the width of the partition and thus the smoothness of the histogram; as h decreases the number of peaks in the histogram tends to increase.

The popularity of the histogram is due to the fact that it is very easy to compute. However, it has some undesirable properties. Firstly, the asymptotic rate of convergence of the histogram to the true density is worse than for alternative density estimates. Secondly, the discontinuity of the histogram impedes the calculation of derivatives often needed as an intermediate tool or for their own sake. Finally, the choice of the origin x_0 affects substantially the final shape of the histogram. Thus, for the same data set the histogram can give a very different impression depending on the chosen origin. Additionally, the histogram requires the choice of the *bin width*. The latter difficulty is common among all methods presented in this appendix.

I.1.2. The Naive Estimator

Rosenblatt (1956) proposed the naive estimator (*NE*). It consists of using the histogram method with bins centred on the point to be estimated, i.e. the naive estimator can be seen as a histogram where every point is the centre of a bin. The *NE* can be derived directly from the definition of a density function

$$f(x) = \lim_{h\to 0} (1/2h)\, P(x-h<X<x+h). \tag{A.2}$$

If we estimate $P(x-h<X<x+h)$ by the proportion of the sample falling in the interval $(x-h, x+h)$, i.e. $[\hat{F}(x+h)-\hat{F}(x-h)]$, where $\hat{F}(x)$ is the empirical distribution function[1] then the density estimate is

$$\hat{f}(x) = 1/2nh(\text{number of } X_i \text{ falling in } (x-h, x+h)) = [\hat{F}(x+h)-\hat{F}(x-h)]\frac{1}{2h}. \tag{A.3}$$

It is proportional to the relative frequency of the data in an interval of length h centred around x.

The length of the interval is called *bandwidth* and has the same role for the *NE* as the *bin width* for the histogram. Since $\hat{F}(x)$ is an unbiased estimator of $F(x)$ with good statistic properties the density function $\hat{f}(x)$ is expected to be a good estimator of $f(x)= dF(x)/dx$ as $h\to 0$.

A more compact and transparent way of writing $\hat{f}(x)$ can be developed by defining the following weight function

$$w(x)\begin{cases} 1/2 & \text{if } |x|<1 \\ 0 & \text{otherwise} \end{cases}. \tag{A.4}$$

[1] $\hat{F}(x) = 1/n\sum I(X_i \le x)$ and I is the indicator function.

Consequently,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}w((x - X_i)/h). \tag{A.5}$$

Notice that

$$w(\frac{x-X_i}{h}) = \begin{cases} 1/2 & |(x-X_i)/h| < 1 \\ 0 & otherwise \end{cases} \quad \text{and} \quad |x\text{-}X_i|<h$$

$\Rightarrow X_i \in (x\text{-}h,\ x\text{+}h)$.

In other terms the *NE* consists of centring a « box » of width *2h* and height $(2nh)^{-1}$ around each observation and then summing them up. The *NE* solves most of the shortcomings of the histogram apart from the discontinuity problem.

I.1.3. The Kernel Estimator

The kernel estimator *(KE)* generalises the naive estimator and produces continuous density estimates. It replaces the rectangular weight function, which is a uniform density function, by a general function[2] *K(x)* which satisfies

$$\int_{-\infty}^{\infty} K(x)dx = 1 \qquad K(.): R \rightarrow R. \tag{A.6}$$

K(.) is called a kernel function. It will normally be continuous bounded, and symmetric.

The kernel estimator[3], is defined as

$$\hat{f}(x) = 1/nh \sum K((x - X_i)/h), \tag{A.7}$$

where *h* is the smoothing parameter known as *bandwidth* or *window width.*

The intuition is the following: for *x* fixed, if the realisation X_i is close to *x*, it follows that $(x - X_i)/h$ is approximately zero and $K((x - X_i)/h)$ will be large. The more realisations close to *x* the higher is the value of $\hat{f}(x)$. Obviously $(x - X_i)/h$ depends also on the value of *h* such that it can be close to zero because *h* is big. As mentioned before, the choice of *h* is a drawback shared by all nonparametric methods.[4]

The histogram and kernel methods may be misleading when the tails of the underlying density are large. This is because the smoothing parameter is fixed across the entire sample.

---

[2] It replaces « boxes » by « bumps ».

[3] It is immediate that if *K* is a continuous density function so is $\hat{f}(x)$.

[4] The next section of this appendix analyses in detail the bandwidth problem for the kernel estimate.

On the one hand if it is small, values in the tails of the data can generate spurious noise in the tail of the estimates. On the other hand if h is large enough to avoid this effects then the resulting estimate will hide details in the main part of the distribution. In order to deal with this difficulty several alternative methods have been proposed which allow the smoothing parameter to vary with the data.

I.1.4. Locally Adaptive Methods

These methods try to accommodate the level of smoothing to the local density of the data.

The alternative to the histogram is the *variable partition histogram* proposed by Anderson (1965) where each partition is of different length and contains a fixed number of data points k. Exactly the same idea is applied to kernel and naive estimates. For the naive estimate the locally adaptive alternative is called *kth nearest neighbour* density estimate. The boxes centred in each observation have the width needed to contain k observations. In other words, the smoothing parameter is the Euclidean distance between each observation and its *kth* nearest neighbour in the data set. Finally, the alternative to the kernel estimator is called the *generalised kth nearest neighbour estimate* defined as:

$$\hat{f}(x) = \frac{1}{n}\sum \frac{1}{hd(x)_k} K((x - X_i) / hd(x)_k), \tag{A.8}$$

where $d(x)_k$ is the Euclidean distance between x and its kth neighbour.

I.2. PROPERTIES OF THE KERNEL ESTIMATOR

Rosenblatt (1956) showed that most of the density estimates are biased. Consequently, there is need for another measure of the discrepancy between the estimate and the density in order to analyse how good the estimate is. The discrepancy measures are usually based on absolute values, quadratic distances etc. Here we focus on quadratic distance measures only.

Given x, when considering the estimation at a single point, the *MSE mean square error* is defined as

$$MSE_x \hat{f}(x) = E[\hat{f}(x) - f(x)]^2. \tag{A.9}$$

More interesting is the global discrepancy between the actual function and the estimated one. For that purpose define the *mean integrated squared error* as

$$MISE_x(\hat{f}) = \int MSE_x \hat{f}(x)d(x) = \int bias^2 \hat{f}(x)d(x) + \int Var f(x)d(x) \tag{A.10}$$

In order to evaluate the *MISE* we derive the expression of the bias and of the variance. For simplicity we assume that the underlying density function *f(x)* is continuous and has continuous derivatives and hat the kernel function satisfies

$$\int_{-\infty}^{\infty} K(x)dx = 1, \qquad \int_{-\infty}^{\infty} xK(x)dx = 0, \qquad \int_{-\infty}^{\infty} x^2 K(x)dx \neq 0. \tag{A.11}$$

Since the variables are independent the *bias*[5] can be written as follows:

$$bias\ \hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} E[K((x - X_i)/h)] - f(x) = \frac{n}{nh} E[K((x - X)/h)] - f(x) = \frac{1}{h}\int [K(x-y)/h] f(y)dy - f(x), \tag{A.12}$$

The expression (A.12) depends on *f(x)*. Consequently, the evaluation of the bias requires an assumption on the functional form of *f(x)*. This procedure is no very much in harmony with the nonparametric principles. We rather use an approximated expression of the bias. A change of variable y = *x-ht* gives the following formula

$$bias\ \hat{f}(x) = \int K(t) f(x - ht)dt - f(x), = \int K(t)[f(x - ht) - f(x)]d(t). \tag{A.13}$$

Using a second order Taylor approximation we write the bias as

$$bias\ \hat{f}(x) \cong -hf'(x) \int tK(t)dt + \frac{1}{2}h^2 f''(x) \int t^2 K(x)dt = \frac{1}{2}h^2 f''(x) \int t^2 K(x)dt. \tag{A.14}$$

We proceed in a similar way with the expression of the variance. Using the independence property it can be written as

$$Var\ \hat{f}(x) = \frac{1}{n^2h^2} \sum_{i=1}^{n} VarK((x - X_i)/h) = \frac{1}{nh^2} VarK((x - X_i)/h) = = \frac{1}{nh^2}\{E[K((x - X)/h)^2] - E[K((x - X)/h)]^2\} = \tag{A.15}$$

---

[5] The bias is : $bias\ \hat{f}(x) = E\hat{f}(x) - f(x)$

$$=\frac{1}{n}\left\{\int\frac{1}{h^2}[K(x-y)/h]^2 f(y)dy-\left[\frac{1}{h}\int[K(x-y)/h]f(y)dy\right]^2\right\}. \qquad (A.16)$$

The change of variable y = *x-ht* gives the following expression

$$Var\ \hat{f}(x)=\frac{1}{nh}\{\int f(x-th)K(t)^2dt-(1/n)[f(x)+bias\hat{f}(x)]^2\}. \qquad (A.17)$$

Approximating by a first order Taylor expansion results in

$$Var\,\hat{f}(x)\cong\frac{f(x)}{nh}\int K(t)^2dt\ . \qquad (A.18)$$

Substituting the expression of the bias and the variance into *MISE* gives

$$MISE_x\hat{f}(x)\cong(1/4)h^4(\int f''(x)^2dx)(\int t^2K(t)dt)^2+\frac{1}{nh}(\int K(t)^2dt). \qquad (A.19)$$

This measure depends on *K* and *h*. The first part of (A.19) is the systematic error (bias) which decreases as *h* goes to zero. The second is the random error (variance) which increases as *h* goes to zero. Consequently, the choice of the smoothing parameter implies a trade-off between random and systematic error.

I.2.1. Optimal kernel function

The choice of *K* is not crucial for the *MISE.* Using calculus of variations Epanechnikov (1969) proved that the optimal kernel is:

$$k(t)=\begin{cases}\frac{3}{4\sqrt{5}}(1-\frac{1}{5}t^2) & |t|<\sqrt{5}\\ 0 & otherwise\end{cases} \qquad (A.20)$$

This formula is known as the Epanechnikov kernel.

I.2.2. Optimal bandwidth

The *bandwidth* that minimises the approximated expression of *MISE is*

$$h_{op}\cong(\int t^2K(t)dt)^{-2/5}\left(\frac{\int K(t)^2dt}{n\int f''(x)^2}dx\right)^{1/5}. \qquad (A.21)$$

It depends on the density to be estimated. Several procedures have been proposed to calculate the optimal h.

(i) -Reference to a standard distribution. In order to compute the h_{op} from (A.21) we need a value for f''. Assuming that $f(x)$ belongs to a particular family of densities (e.g. Gaussian) the parameters of the assumed density (e.g. mean and variance in the case of a Gaussian family) can be estimated using robust procedures from the data. Then f'' can be computed numerically. By plugging its value into (A.21) we obtain the optimal h.

(ii) -Subjective choice. This approach consists in choosing the parameter h which gives the density estimate $\hat{f}(x)$ most in accordance with one's prior ideas about the density.

(iii) -A very popular alternative is to treat h as a parameter which is estimated by *optimising a criterion function.*

a.-*Least-square cross validation.* The criterion function (to be minimised) is an estimator of the *MISE*. This is a data driven method which leads to the choice of a *bandwidth* that asymptotically minimises

$$MISE = \int (\hat{f}(x) - f(x))^2 d(x) = \int \hat{f}(x)^2 dx - \int \hat{f}(x) f(x) dx + \int f(x)^2 dx. \tag{A.22}$$

Note that the last term of the above expression does not depend on $\hat{f}(x)$. Thus the optimal h in the sense of minimising (A.23) is the result of minimising the first two terms

$$R(\hat{f}) = \int \hat{f}(x)^2 dx - \int \hat{f}(x) f(x) dx. \tag{A.23}$$

This gives an automatic method for choosing the smoothing parameter. The idea is to construct with the given data an estimator of $R(\hat{f})$:

$$\hat{R}(\hat{f}) = \int \hat{f}(x)^2 dx - 2n^{-1} \int \hat{f} \sum_i \hat{f}_{-i}(X_i), \tag{A.24}$$

where $\hat{f}_{-i}(x)$ is the density estimate constructed from all the data points except X_i

$$\hat{f}_{-i}(x) = (n-1)^{-1} h^{-1} \sum_{i \neq j} K[(x - X_j)/h] \tag{A.25}$$

Stone (1984) found that the h chosen by minimising this function is the best in the sense of minimising the *MISE*[6]. Let $I_{lscv}(X_1,...X_n)$ be the *MISE* of the density estimator constructed

[6] Unless otherwise is stated, least-square cross-validation is going to be used in this paper and in the rest of the thesis.

using the h parameter that minimises the function $\hat{R}(\hat{f})$. Let I_{opt} be the *MISE* if the h is chosen optimally. Under very mild conditions and with probability one

$$\frac{I_{lscv}(X_1,\dots X_n)}{I_{opt}(X_1,\dots X_n)} \to 1 \qquad \text{as } n\to\infty . \tag{A.26}$$

Thus, least-squares cross-validation achieves the asymptotically best possible choice of the smoothing parameter.

b.-*Likelihood cross-validation.* h is chosen such that it maximises the following function

$$LCV = \frac{1}{n}\sum_{i=1}^{n} \log \hat{f}_{-i}(X_i) \tag{A.27}$$

where $log\ \hat{f}_{-i}(X_i)$ is the cross-validated log likelihood function, $\hat{f}_{-i}(X_i)$ as in (A.25).

(iv) -Other procedures are described in Silverman (1986), sections (3.4.5) and (3.4.6.).

II. MULTIVARIATE DATA

Multivariate methods generalise the univariate methods in an immediate way. Instead of dealing with areas one deals with volumes in the *d*-dimensional space R^d. Here we define only the kernel estimator.

Consider the random sample $\{X_1, X_2, \dots, X_n\}$, where X is now a d-valued vector with unknown density function $f(x)$ to be estimated from the data. The kernel estimator is defined as

$$\hat{f}(x) = (1/nh^d)\sum K(\frac{1}{h}(x - X_i)), \quad x \in R^d, \tag{A.28}$$

with $K(x) \geq 0$, and $\int_{R^d} K(x)dx = 1$.

As for the univariate case approximated expressions of the bias and variance can be derived and then be used for the choice of the kernel function and bandwidth. In order to choose the parameter h least-square cross-validation is possible and again the Epanechnikov kernel[7] is optimal.

---

[7] The multivariate Epanechnikov kernel is $K(x) = \{1/2c_d^{-1}(d+2)(1-x'x)^2 \;\; if \;\; x'x < 1, \; 0 \text{ otherwise},$ where c_d is the volume of the unit d-dimensional sphere.

REFERENCES

Barro, Robert J., (1991), "Economic Growth in a Cross-Section of Countries", *Quarterly Journal of Economics,* Vol.106(2), May, pp.407-443.

Barro, R. and Sala i Martin, X., (1991), "Convergence Across States and Regions", Brookings Papers on Economic Activity, Vol.1, April, pp.107-182.

Barro, R. and Sala i Martin, X., (1992), "Convergence", *Journal of Political Economy,* Vol.100(2), April, pp.223-251.

Ben-David, D., (1994), "Convergence Clubs and Diverging Economies", Working Paper 922, CEPR, London W1X 1LB , February.

Baumol, W.J., (1986), "Productivity Growth, Convergence and Welfare", *American Economic Review,* Vol.76(5), December, pp.1138-55.

Bernard, A.B. and Durlauf, (1991), "Convergence of National Output Movements", NBER Working Paper, Cambridge M.A., May.

Bianchi, M., (1995),"Testing for Convergence: A Bootstrap Test for Multimodality", Working paper. Bank of England, May.

Blomström, M., Lypsey, R.E. and Zejan, M., (1993), "Is Fixed Investment the Key to Economic Growth", NBER WP 4436.

Canova , F and Marcet A.,(1996) "The Poor Stay Poor: Non-Convergence Across Countries and Regions " UPF Discussion Papers

Cohen, D., (1992), "Test of the Convergence Hypothesis. A Critical Note", CEPR Discussion Paper No. 691, August.

Cohen, D., (1992), "Economic Growth and the Solow Model. Some Further Empirical Results", CEPREMAP, August

De Long, J.B. and Summers, L., (1992), "Macroeconomic Policy and Long-Run Growth", Federal Reserve Bank of Kansas City, 77.

Desdoigts, A.. *Changes in the World Income Distribution: A Non-parametric approach to Challenge the Neo-classical Convergence Argument*. PhD thesis, European University Institute, Florence.June 1994.

Dolado, Juan.J, Gonzalez-Paramo, J.M. and Roldan, J.M, (1994) "Regional convergence in Spain", *Moneda y Credito.*

Durlauf, S. and Johnson, P., (1992), "Local versus Global Convergence across National Economies", NBER WP 3996.

Friedman, M., (1992), "Do Old Fallacies Ever Die?" *Journal of Political Literature*, Vol.30(4), December, pp.2129-2132.

Genewe, J, Marshall, R.C. and Zarking, G.A., (1986), "Mobility Indices in Continuous Time Markov Chains", *Econometrica*, Vol.54(6), November, pp.1407-1423.

Holtz-Eakin, D., (1992), "Solow and the States: Capital Accumulation, Productivity and Economic Growth", NBER Working Paper 4144, Cambridge, M.A.

Kelly, M., (1992), "On Endogenous Growth with Productivity Shocks", *Journal of Monetary Economics*, Vol.30, pp.47-56.

Knight, M., Loyza, N. and Villanueva, D., (1992), "Testing the Neoclassical Theory of Economic Growth: A Panel Data Approach", International Monetary Fund WP No.106.

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

Lucas, R.,(1988), "On the Mechanics of Economic Development", *Journal Of Monetary Economics*, Vol.22, July.

Lucas, R.,(1993), "Making a Miracle", *Econometrica*, Vol.61, pp.251-272.

Mankiw, N., Romer, D., and Weil, D., (1992), "A Contribution to the Empirics of Economic Growth", *Quarterly Journal of Economics*, Vol.107, pp.503-530.

Paap, R. and van Dijk, H. K. (1994) , " Distribution and Mobility of wealth of nations". Working paper, Tinbergen Institute, Erasmus University, October.

Parente, S. and Prescott, E., (1993), "Changes in the Wealth of Nations", *Quarterly Review*, Federal Reserve Bank of Minneapolis, Vol.17, pp.3-13.

Pesaran, H. and Smith, R., (1995), "Estimating Long-Run Relationships From Dynamic Heterogeneous Panels", Forthcoming in *Journal of Econometrics.*

Prakasa Rao, B. L. S., (1983) *Nonparametric Functional Estimation*, Orlando:Academic Press.

Quah, D., (1992), "National Patterns of Growth: II. Persistence, Path Dependence and Sustained Take-off in Growth transitions" Economics Department Working Paper, LSE, October. (First draft July 1990.)

Quah, D., (1993a), "Empirical Cross-Section Dynamics in Economic Growth", *European Economic Review*, Vol.37(2/3), April, pp.426-434.

Quah, D., (1993b), "Galton's Fallacy and Test of Convergence Hypothesis", *The Scandinavian Journal of Economics*, Vol.95(4), December, pp.427-443.

Quah, D., (1994), "Empirics for Economic Growth and Convergence" Economics Department Working Paper, LSE, September.

Quah, D., (1995a), "Coarse Distribution Dynamics, Divergence and Polarization", Economics Department Working Paper, LSE, London , July..

Quah, D., (1995b), "Ideas Determining Convergence ", Economics Department Working Paper, LSE, London , August..

Quah, D., (1996a), " Twin Peaks: growth and Convergence in Models of Distribution Dynamics ", *Economic Journal* . Forthcoming.

Quah, D., (1996b), "Convergence Empirics across Economies with (Some) Capital Mobility", *Journal of Economic Growth* . Forthcoming.

Romer, P., (1986), "Increasing Returns and Long-Run Growth", *Journal of Political Economy*, Vol.94.

Romer, P., (1989), "Capital Accumulation in the Theory of Long-Run Growth".

Rosenblatt, M., (1956) "Remarks on Some Nonparametric Estimates of a Density Function" *Annals of Mathematical Statistics*, 27,832-837.

Silverman, B., (1986) *Density Estimation for Statistical data Analysis,* London:Chapman and Hall

Shorrocks, A.F., (1978), "The Measurement of Mobility", *Econometrica,* Vol.46(5), September, pp.1013-1024.

Solow, R., (1956), "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics.*

Stokey, N., Lucas, R. and Prescott, E., (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press.

Tapia, R. A., and Thompson, J.R., (1978) *Nonparametric Probability Density Estimation,* Baltimore, MD: John Hopkins University Press.

CHAPTER 2 :

DYNAMICS OF THE INCOME DISTRIBUTION ACROSS OECD COUNTRIES

This chapter is joint work with Javier Andrés.

1. INTRODUCTION

Convergence across economies (states, countries, regions, etc) is an important economic issue. In recent years, it has been intensively studied from both a theoretical and an empirical perspective. If convergence is understood in terms of poor economies becoming as wealthy as the initially richer ones or equally as a process of economic homogenisation (non-persistent inequality) then a natural approach to the convergence proposition is to study the dynamics of the cross-economy income distribution along time. Nevertheless, most of the standard literature on convergence departs from this "natural approach" at least in two ways. Firstly, it collapses the time dimension by averaging and applies cross-section analysis. It summarizes the main features of the evolving distribution in a few sample statistics which might not always be very informative in terms of convergence. Secondly, it uses the framework of the "representative economy" model, (in particular, the classical or the augmented Solow growth model) to extract conclusions about the cross-section. It conceives the distribution of incomes as normally distributed around the mean representative economy. Under these assumptions, standard parametric methods yield consistent estimates of some basic moments of this distribution and make it possible to draw conclusions about its time evolution.

In a recent set of papers Danny Quah criticizes that approach, showing that regardless of how fancy it is, the standard cross-section regression over time averages delivers little information for the purposes at hand. While standard regressions come up with a positive and significant 2% of convergence across countries and regions, Quah finds that the evolution of the cross-section income distribution in the Summers and Heston data set shows a two camp world with very little upward mobility in the income ranking. The sequence of distributions displays an increasing concentration of economies at the two ends, as well as a few sharp upwards and downwards changes in the ranking[1].

In this paper we analyse the dynamics of the income per capita distribution across the OECD countries during the 1960-90 sample period. We use Quah's approach, which studies the dynamics as well as the interrelation among economies by using all the cross-section and time series information on the data. It allows us to see (independently of any theoretical model) whether convergence is taking place during the sample period and to examine convergence as a long-run property. Additionally, conditioning in this framework helps also to discriminate among theoretical models.

Firstly, we try to asses whether the behaviour of the distributions tells us the same story about convergence on income per capita as the standard literature does. The main results of this first exercise go against the widespread view that OECD economies are approaching each other at a positive rate, and also indicate that wealth differences across countries are more persistent than what the constant returns growth model would suggest. There is a significant number of countries with per capita income around 50% of the OECD average. Furthermore, the transition probability from the low income to the high income group is very low for each country.

Secondly, we analyse the cross-section dynamics of the residuals series from a first stage regression of per capita income on different information sets as the neoclassical model suggests. It is a simple conditioning exercise which tries to replicate what the standard literature does (from a new econometric point of view) and allows us to say something about to what extent the differences in accumulation rates account for permanent differences in per capita income across countries. The results remain essentially unchanged after removing the effect of differences in accumulation rates.

Finally we illustrate the Solow Model prediction for the OECD data, studying the cross-section dynamics after conditioning each economy on its individual steady-state. The residuals from the first stage regression display a very high intradistribution mobility along with a strong tendency towards the estimated steady state value and a very low inertia. We conclude that the convergence prediction of the Solow model might hold as a single country property. This does not indicate any tendency whatsoever for countries to approach each other over the long-run. The results must be interpreted cautiously given the substantial amount of conditioning that is needed to achieve them. Most of the cross-country long-run differences in wealth are accounted for by country specific factors, other than differences in accumulation rates. This suggests that the conventional cross-country regression models might suffer form severe misspecification problems. In other words, the estimated convergence rate in the literature is merely a poor estimate of the average first order autoregression coefficient in the OECD per capita income series.

The rest of the paper is organised as follows. Section 2 explains why the standard cross-section regression analysis approach is misleading and identifies the relevant issues to look at in the analysis of convergence. Section 3 applies Quah's approach[2] to study the dynamics of the cross-section distribution of per capita income across the OECD countries. Section 4 analyses the distribution of income after conditioning in the way that it has been done by the standard regression literature and illustrates Solow Model predictions. Section 5 concludes with some final remarks.

2. CONVERGENCE AND CROSS-SECTION REGRESSION ANALYSIS

Intensive growth takes place whenever the marginal return of further investment is higher than that of consumption. According to the constant returns model, this can only happen while the capital-labour ratio is below some critical ("steady state") level. In this setting growth ceases when the economy approaches the steady state. The empirical counterpart of this property is the well known convergence proposition that simply asserts that an economy will monotonically return to its steady state income level if for some reason (initial conditions, shocks, etc) it happens to be away from it.

Two measures of convergence have been proposed in this field (see for example Baumol, 1986; Barro, 1991; Barro and Sala-i-Martin 1991, 1992; Mankiw, Romer and Weil, 1992; and Holtz-Eakin, 1992). There is σ-convergence if the dispersion of the real per capita income across economies tends to fall over time. The σ measure is model free. On the other hand, a negative correlation between the growth rate of income per capita and its initial level is called absolute β-convergence. This is a property of the Solow growth model[3] which incorporates a neoclassical production function with diminishing returns to capital.

The main prediction of the Solow model is that for each economy the income per-capita converges to a steady state defined by

$$\log(y) = a + gt + \frac{\alpha}{1+\alpha}\log(s) + \frac{\alpha}{1+\alpha}\log(n+g+\delta) + \varepsilon \qquad (2.1)$$

Where:

$$\log(A(0)) = a + \varepsilon$$

where Y is output, L is labour, y=(Y/L), A is the level of technology and s, n, g and δ are the rates of saving, population growth and technological progress and depreciation[4].

Equation (2.1) shows that the steady state level of income per-capita in each country depends on its population growth and its accumulation of capital (physical in the original model and physical and human in the augmented version). Therefore different countries can reach different steady states. Let y^* be the steady state in each economy and y its current value, by approximating (2.1) around y^*, it can be shown that:

$$\dot{y} = \beta[\log(y^*) - \log(y(t))] \tag{2.2}$$

Solving this differential equation and rearranging we find that the average β growth rate of y over the interval between 0 and T is given by (2.3),

$$\left(\frac{1}{T}\right)\log\left(\frac{y(T)}{y(0)}\right) = g + \frac{1-e^{-\beta t}}{T}\log\left(\frac{y^*}{y(0)}\right) \tag{2.3}$$

The parameter β is the conditional rate of convergence, it is the rate at which the economy closes the gap between its current per capita income and its own y^*, and then it is expected to differ across economies. The greater β the higher the responsiveness of the average growth to the gap between the steady state and the initial income.

The empirical analysis has concentrated mainly in estimating the parameter β in a discrete time version of equation (2.3) in which the average growth rate of income over time for each economy is regressed on the initial levels of income and a bunch of steady state variables. The results for these regressions is a negative and significant estimate of the initial level of income coefficient, that is a positive and significant estimate of the parameter β, mainly after controlling for the determinants of the steady state (conditional convergence). What is noteworthy is that this empirical finding is about the same for very different samples, whether across states in the US, regions in Europe or across countries in the world and over different periods of time[5]. This, together with a weaker evidence of σ-convergence for some regions in the world, has been interpreted as strong and robust evidence of convergence across world economies caused by diminishing marginal returns to capital.

The extension of the stability property of the classical growth model for the "representative economy" to convergence in a multi-country setting is not straightforward. The adjustment process of the Solow model tells us whether or not each country when perturbed from its long-run pace gets back to it approaching monotonically, which is basically a single country implication. This property has nothing to do with countries truly approaching each other. In fact, growth theory teaches us that we should not expect that to happen unless the behaviour in some crucial aspects of economic and social life is alike across countries. Additionally, the traditional way of measuring the long-run, by taking time averages, is bound to leave us with very few degrees of freedom using single country series. To avoid this difficulty, the use (and indeed the production) of multi-country data sets has become very popular in recent times[6].

The convergence rate that comes up from these cross-country regressions must be understood as the speed at which a country returns to its long-run trend from the level of per capita income it has at a given date. Of course, this speed is country and time specific, but as long as it is drawn form a multi-country data set the particular value that is obtained is merely an average. However, this interpretation has recently been questioned. As Pesaran and Smith (1993) have pointed out, cross-section averages or even panel data methods do not yield consistent estimates of average parameters in *random fields* data sets in which time and cross-section dimensions are reasonably large and similarly sized.

Other criticisms refer to how the steady state in each economy, y^*, is estimated from the observed values of the explanatory variables. Some authors have shown that the causality among growth and accumulation rates is weaker than what the classical growth model would suggest. Cohen (1992) finds that the correlation among growth and human capital accumulation vanishes once time invariant country specific effects are included in the standard regression. Similarly

Blomström, Lypsey and Zejan (1993) show that causality runs from growth to investment rather than the other way around. Finally, many studies have found a significant correlation among growth and short term macroeconomic indicators (inflation, exports, and so on) which nonetheless are not always robust to changes in the model specification (Levine and Renelt, 1992).

In order to justify the usual interpretation of convergence regressions and for the notion of average growth rate (the independent variable in the regression) to make sense, it is implicitly assumed that the permanent component of the income for each economy is well described by a linear or log-linear deterministic time trend. In fact, the average growth rate is simply the slope of the deterministic trend whose cross-section variation is explained in the standard regression. However, under stochastic growth, imposing a deterministic trend structure can be very misleading. This remains true even for the case in which pooled data on shorter averages is used. The procedure still assumes the non stationary component of per capita income to be trend deterministic, although allowing for a changing slope.

Over and above all the limitations just mentioned, Quah (1993b) has shown that convergence tests based on regressing average growth rates on initial levels are uninformative, since a negative cross-section regression coefficient on the initial level is in fact consistent with absence of convergence. The argument follows the so-called Galton's classical fallacy, which tells us that "tall fathers" are not expected to have "taller than average sons". Similarly, currently richer economies might become poorer than average in the future without a significant narrowing in the cross-section dispersion of incomes. This applies to absolute and conditional convergence.

The evolution over time of the cross-country distribution of per capita income involves two kind of dynamics: changes in the shape (bimodal distribution, mass-point etc) and intra-distribution mobility. The main limitation of the conventional approach is that it relies on two single statistics that might not be a good description of the behaviour of the entire cross-section distribution and consequently may not be very informative in terms of convergence. A positive parameter β is compatible with cross-economies behaviours (overtaking, cycles etc) that are far from the idea of convergence. Sometimes σ convergence does not give any information about the cross-section dynamics, for example observations from a multi-modal distribution may have the same sample variance as observations from a uni-modal one. The interpretation of those distributions in terms of convergence is obviously quite different. Also, it happens that σ convergence cannot account for some intra-distribution mobility.

Consequently, it is necessary to develop an alternative econometric strategy to study what in probability theory are known as *random fields*. Danny Quah (1993a,b; 1994) suggests such a new and comprehensive approach that encodes the two kinds of convergence, β and σ measures. This approach takes into consideration the whole distribution (rather than a few statistics) and does not impose any structure about trends, nature of convergence and so on. The idea is to consider the cross-section distribution of the variable of interest, at each point in time as the realization of a random element in a space of distributions, and to analyse its dynamics. The variable of interest can be either income per capita or its residual after removing the effect of some variables trying to approximate the steady-state of each economy.

3. DYNAMICS OF THE EVOLVING CROSS-SECTION DISTRIBUTION OF INCOME ACROSS THE OECD COUNTRIES. ABSOLUTE CONVERGENCE.[7]

Contrary to what the constant returns to scale models would suggest, the cross-country distribution of per capita incomes in the world tends to concentrate around two extreme levels[8]. There are then two group of countries, one with a high level of per capita income and the other with a much lower level; with little mobility between these groups. This picture remains essentially unchanged after conditioning for some investment and human capital indicators, which possess a challenge to the widespread consensus about a 2% rate of ß convergence. Nevertheless, most economists would argue that very few countries in the world meet the explicit or implicit assumptions of the canonical growth model. If this is the case, lack of convergence should not be a surprise to anyone.

The OECD is a different matter for several reasons. First, the OECD and Asia are the only regions with substantial and steady growth from 1960 to 1990 (see Figure 1). Second, in this region the share of GDP devoted to increase the capital stock is higher and far more stable than in other regions in the world. Similarly, these countries are predominantly market economies and share many social, political and cultural features. Third, since 1960 there has been a sizeable reduction in the dispersion of income per capita levels within the OECD (Figure 2), while this has not happened in Africa, Latin America or Asia (except Japan). Figure 2 depicts the evolution of the sigma measure of convergence from 1960 to 1990 for these four economic regions. The dispersion of per capita income within the OECD has fallen by 33% since 1960, whereas it has gone up in Africa and, specially in Asia, in a significant proportion. In fact most of the reduction in the OECD took place from 1960 to 1975, and has remained roughly stable since then. Finally, growth rates in the OECD also present two specific features: a negative correlation among growth rates and initial conditions (whereby yesterday's poorer countries are expected to grow faster than the average) and a significant persistence[9] (so that today's fast growers are expected to enjoy a higher than average growth for some time). This has been usually presented as convincing evidence of the ability of the exogenous growth model to account for the long-run path of these economies. Actually, conventional regression analysis on the OECD sample yields a positive and significant estimate of the absolute convergence rate, although weaker than the conditional rate. If taken literally this would imply that OECD economies are effectively approaching each other, at least over the long-run. Since, the evidence of convergence within the OECD seems to be overwhelming according to the conventional regression analysis, the OECD countries become an adequate framework to confront the two econometric approaches. We start by analysing the distributional dynamics of the observed per capita income, without conditioning. The results in this section are directly comparable to the standard absolute convergence analysis.

Once the variable of interest is well defined, we proceed to estimate its cross-section distribution at each point in time as well as the intradistributional mobility from one period to another.

At this point we must define what we understand by convergence in this setting. A degenerate ergodic distribution would imply that all observations tend to concentrate around their sample mean, regardless of the initial conditions. Hence, all economies would be moving towards their steady state. The long-run inference requires some assumptions about stationarity time invariant transition probability) that might not be reasonable for long periods in which for example, some economic structural changes may happen. Alternatively we could have a non degenerate long-run distribution and still claim that convergence has taken or is taking place. This would be the case when the intradistributional mobility indicates that income disparities among countries are not persistent. Recall then that the property of high mobility or low inertia is the crucial one to look at when testing for convergence.

3.1. the variable of analysis - sample and characteristics

The data are taken from OECD statistics, in particular the GDP (and its components) is drawn from *National Accounts, 1960-1991* which uses the System of National Accounts (SNA) definitions. The sample covers 24 countries over a period of 31 years (1960-1990). Every nominal variable has been transformed in real terms by using its price index from National Accounts, then expressed in national dollars of 1985 by using estimated PPPs from 1990 for each aggregate. PPPs series are calculated involving only OECD countries[10]. The basic variable of analysis is the per-capita GDP in each individual country relative to the same variable for the entire OECD as a whole[11]. Figure 3 is a three dimensional plot of the variable, for the 24 economies over the 31 years of the sample. It is clear that both dimensions of variation in the data appear to be very important and the regression analysis is missing a significant amount of the dynamics. It is precisely this two dimensional dynamics that we are interested in, for which conventional convergence analysis is not well suited. For example, the cross-section analysis simply takes the slope of the linear time trend for each country over the time sample and studies the dynamics of that slope across economies. In this way most of the time dimension is lost. Alternatively estimating (2.3) for each individual country the cross-section dynamics is lost.

3.2. estimate the cross-section distribution of that variable at each point in time

The first step is to estimate the cross-section distribution of incomes at each point in time. The purpose of this exercise is to uncover any particular pattern in the time evolution of this distribution. In the context of random fields, the realization of the random element turns out to be a cross-section distribution function that has to be estimated from data. This requires the use of non-parametric and semi-parametric methods. Notice that it is the shape of the distribution that we are interested in, we must avoid imposing any prior assumption about it, or about the moments of the density function from which the data are drawn.

Figures 4a to 4d present some non-parametric estimated cross-section density functions of relative per capita income for each period of 3 or 4 years. The range of relative differences among countries' incomes diminishes mainly during the first decade of the '60s as the richest countries approach the average. The poorest countries also move towards the average but far more slowly. During the '70s and '80s a second mode in the density function shows up very clearly. A group of countries tends to concentrate around a 50% level of OECD average income. Rich economies still move slowly towards the mean. For the sample period as a whole there is a concentration (of probability mass) around two values. The cross-section distribution of income seems to be fluctuating over time. At first glance, these estimates suggest that the first and second moment do not entirely describe the behaviour of the distribution. In consequence by focusing only on the falling path of the sample variance (see Figure 5) we could conclude that there is σ convergence. However, σ is a statistic that gives us limited information about the dynamics of the countries relative cross-section position. The narrowing range gives an idea of diminishing dispersion, but the presence of two modes suggests that there are significant and persistent differences among countries that do not vanish, at least over the sample period.

3.3. dynamics of the (evolving) cross-section

This analysis is very intuitive but the distributions are just point estimates for the sample period and cannot be assumed to reflect out of sample patterns. Furthermore, the behaviour of the cross-section distribution refers not only to changes on the shape but also to the intra-distribution mobility. The dynamics of each country's relative position is a crucial component of the notion of convergence that the growth literature is concerned with. To make progress in the analysis requires a formal statistical structure. To develop a law of motion for the cross-section distribution of income, as realizations of random element in the space of distributions, we need a model for the stochastic process taking values which are probability measures associated with the cross-section distribution.

Let $\{\lambda_t\}$ be the sequence of probability measures associated with the cross-section distribution. The simplest probability model is as follows:

$$\lambda_t = T^*(\lambda_{t-1}, u_t) \qquad (2..4)$$

where T^* maps probability measures plus a disturbance term into probability measures. Hence, T^* encodes information on how economies transit from t to t+1, in particular about whether they get closer or grow apart. Unfortunately, this stochastic difference equation in (2.4) is unmanageable. By ignoring the disturbance and iterating it can be written as (2.5),

$$\lambda_{t+s} = (T^*)^s \lambda_t \tag{2.5}$$

such that as s goes towards infinity it is possible to characterize the long-run distribution of cross-country income. In other words, it allows us to make out of sample inference[12] and, eventually, to characterize the steady state.

In order to make (2.5) tractable one can use the concept of *Stochastic Kernel* (Stokey, Lucas and Prescott, 1989). Consider the measurable space (R, *R*), where R is the real line in which the realizations of income fall and *R* is its Borel sigma algebra. B(R, *R*) is the Banach space of finitely additive functions. Let λ_{t+1} and λ_t be elements of B that are probability measures in (R, *R*). A *Stochastic Kernel* is a mapping M (x,A) M:Rx*R* -> [0,1] satisfying:

(i) $\forall$ a $\in$R M(a,.) is a probability measure.
(ii) $\forall$ A in *R* M(.,A) is a sigma measurable function.

Then M(a,A) is the probability that the next state period lies in the set A, given that in this period the state is a. For any probability measure λ on (R, *R*), $\forall$ A in R:

$$\lambda_{t+1} = \int M(x, A)\, d\lambda_t(x) \tag{2.6}$$

Where M(.,.) is a *Stochastic Kernel*, $\lambda_{t+1}(A)=(T^*\lambda_t)A$, and T^* is an operator associated with the *Stochastic Kernel* that maps the space of probabilities into itself (the adjoin of the Markov operator associated to M). Equation (2.6) measures the probability that the next period state lies in the set A, if the current state is drawn according to the probability measure λ_t. $(T^*\lambda_t)$ is the probability measure over the next period state, if λ_t is the probability measure over the current period. We shall consider T^* in (2.4) and (2.5) as being generated in the differential equation (2.6).

Since the *Stochastic Kernel* is a complete description of transitions from a state into any other, it gives us information about the intra-distribution mobility. It describes the dynamics through the sample period.

However, before we can say anything about the steady state we must work out T^* from (2.6) and do the calculations in (2.5). Given that the *Stochastic Kernel* is infinite dimensional, some simplification is required to do empirical work. At the present, T^* can be approximated by assuming a countable state space for income levels $S = \{s_1, s_2, .. , s_r\}$. In this case T^* is simply a transition probability matrix Q such that the difference equation (2.7) is tractable.

$$\lambda_t = Q(\lambda_{t-1}, u_t) \tag{2.7}$$

The matrix Q encodes the relevant information on turnover in the distribution so that the long-run calculation in (2.5) can be performed. Under some regularity conditions, the sequence of powers of this matrix converges to a matrix whose rows (all of them identical) are the ergodic distribution, which allows us to talk about steady state.

Alternatively, by fixing the probability vectors to be uniform and identical for every time point, $\lambda_t=\lambda$, we define a time-variant grid (quantiles), and associated to that, a sequence of fractiles transition probability matrices, Q_t.

3.3.a. estimation of the Stochastic Kernel

Figures 6a to 6e and 7a to 7d show some non parametric[13] estimated *stochastic kernels* for relative income of k-year transitions (k=1, 5). Figures 6 display three dimensional plots of the transitions probability function, while Figures 7 display the contours of the function in Figures 6. A slice parallel to the t+k axis gives a probability density that describes transitions from a part of the income distribution to another in k periods. The concentration of the probability mass along the positive sloped diagonal indicates high persistence in the economies relative position, and implies low mobility. Concentration along the negative sloped diagonal, on the other hand, would indicate that economies are overtaking each other in the income ranking. The transition probability describing horizontal lines (parallel to t+k) shows that there is very low persistence, the probability of being at any point in t+k is independent of the position in t. Finally, if the mass of probability is parallel to the vertical axis we get convergence in the β terms, ie richer countries growing slower than the poorer ones.

According to this, Figure 6a shows how OECD economies tend to remain where they are relative to each other with a low probability of changing state in one year. Relative incomes in the OECD are highly persistent from one year to another. This can be better seen in the contours in Figure 7a, which surround the positively sloped diagonal. For longer horizons (Figures 6b to 6d and 7b to 7d), the highest probability of transitions is still along the main diagonal, but is not as concentrate as before, which suggests a somewhat higher intra-density mobility. The contour for the 1960-65 period (Figure 7b), ranging from 0.4 to 1.6, appears to be slightly steeper than the 45 diagonal, indicating that the poorer countries transit to a better state, whilst the opposite is true for the rich ones. For 1971-75 and 1981-85 (range 0.3 to 1.2) the probability of transition is mainly over the diagonal (Figures 7c & 7d). It seems that most of the convergence among the OECD countries took place in the first part of the sample period, in particular during the '60s, and was mainly due to the catching-up process of low income countries, which display substantial upward mobility.

3.3.b. estimation of the Transition Probability Matrix Q

These transition kernels are still point estimates. They deliver additional information about transition probabilities, but still inference cannot be drawn. To make the model operational we need to estimate a simple version of the Q transition matrix. Taking per-capita income for each economy, relative to the sample average, we discretize the space of possible values, S, in r states. For example the state i=(0.7,1.1) includes the economies which have an income between 0.7 times and 1.1 times the average for the total sample. The discretization defines a grid, that can be thought of as an estimator of the initial unconditional probability distribution λ_t. Each element of the Q matrix indicates the probability of transition from one state to another in *k* periods: the q_{ij} entry is the probability that a country in state *i* transits to the state j. (Each row is a conditional probability vector.) The matrix row is analogous to the density probability defined for each point in S (S is continuous) when cutting the figure at that point by a plane parallel to the t+k axis.

Table 1 presents some estimates of the transitions matrix Q. The grid divides the total observed sample into categories for providing a uniform distribution for the first year of the sample. Consequently the length of the defined states are different. These categories are very narrow around the value 1, which corresponds to countries having an income around the mean, and much wider for low income countries. The first column is the total number of transitions over the whole time sample, starting at each state. The rest displays an estimator of the time invariant transition probability matrix *rxr* for a single period, calculated as an average over the total sample. The values off the main diagonal are very low, which indicates that the probability of a country moving from one state to another in one year is very low. For instance, the probability that a country with income between 0.6 and 0.9 times the average, transits in one period to an income between 0.9 and 1.04 is 0.09. This persistence is even higher for the low income group, preserving the conclusions from the estimation of the *stochastic kernels*. Finally, Table 1 also offers an estimator of the ergodic distribution, which is the closest concept to the steady state in this setting. The ergodic distribution tells us the unconditional probability for an economy to end up in a particular income range. Recall that the states are defined in a way such that the initial distribution is uniform. We end up with a distribution that is not degenerated at its mean value, but it gives an approximately equal probability of reaching different states (although this probability is slightly higher for the average state). These calculations require time invariant transition probability, which is not always reasonable for long periods in which, for example, some economic structural changes may happen.

Let us define the grid in such a way that the set of quantiles determines the sequence of cross-section distribution, hence, the change in the grid describes the evolution of the cross-section distribution for one period to the next one this would allow us to study whether convergence is taking place and to characterise the long-run as the sequence of quantiles degenerating to the mean (approaching). Associated to these grids there is a sequence of fractile transition probability matrices that show the intra-distribution mobility. Table 2 and Figure 8 show the sequence of quantiles. The 25% of countries with the lowest incomes in 1960 fell in a range of 0.26 to 0.54 of the OECD average. The upper limit of this interval rose steadily until the mid-'80s (reaching a 90% of the average by 1983 and has been falling steadily since. The lower limit has not changed substantially. Today we still find a quarter of OECD countries with incomes between 0.3 and 0.8 of the OECD average[14]. On the other hand, the second and third quantil tend to concentrate around the mean while the richest quantil start with an upper limit of 1.56 and evolves until 1.25 times the average.

The associated sequence[15] of matrices has high values over the diagonal, which indicates persistence, hence there is a low probability of the countries move from one income quantil to another. This is especially so for the lowest income countries during the second part of the sample period. Unlike in Table 1, the transition matrix Q is time variant so that we can evaluate the mobility across quantiles as well as its behaviour over time. In Figure 9 we present a set of mobility indices suggested by different authors (Shorrocks, 1978; Genewe *et al*, 1986; and Quah, 1994) which summarize in one way or another the information contained in the transition matrices. The overall picture is one of very limited mobility. We shall return to this later to compare the mobility indices as well as the distributional dynamics and the ergodic distribution after conditioning for different information sets.

Figure 10[16] gives the same kind of information. It ranks the countries according to the relative income per capita in the first year of the sample and shows the evolution of the ranking over time. Each line represents, for a single year, the relative income of the OECD countries ordered according to the initial ranking. The larger is the income inequality the steeper these cross profile lines are. Notice that for 1960 the line is obviously monotonically increasing and from the end of the '70s becomes flatter for the middle quantiles. The inequality is persistent and even increasing with respect to the poorest 20%.

At this stage we can draw some conclusions about the growth process in the OECD. The results so far show a bimodal distribution of income per capita across OECD countries. The lowest income quantil is still quite far away from the OECD average, and has not been monotonically approaching it over the sample period. Furthermore, the intra-distributional mobility is low. Some countries seem to be stuck in lower than average income paths. All this militates against the notion of convergence, in the sense of countries approaching each other over the long-run. This exercise is comparable to the notion of absolute convergence in conventional regression analysis, and it can be argued that even within the OECD absolute β is low and weakly significant in a fully specified constant returns to scale model (Andrés *et al*, 1994).

Absolute convergence is interesting on its own, but in order to establish a comparable analysis with that of the conventional regression we must proceed to remove the variance in accumulation rates and to analyse convergence in conditional per capita income defined as the residuals series from a first stage conditioning regression. This is what we discuss in the next section.

4. CONDITIONING ON THE SOLOW MODEL STEADY STATE.

The debate in growth theory is mainly about which model best represents the long-run evolution of market economies, and what policy implications we may draw from it.

The whole purpose of the statistical model in the previous section was not to explain why economies converge or not, but to test whether they converge or they fail to do so. The results there indicate that differences in income per capita among OECD economies are rather persistent, and this is the most relevant conclusion on practical grounds. However, if this conclusion is to be taken as evidence against the convergence proposition of the exogenous growth model, as it is usually presented in the empirical growth literature, we must control for the variation of accumulation rates (ie the steady state) across OECD economies. Conditioning in this context means simply analysing the distributional dynamics of the residuals from a first stage regression, and there are different ways to do so (Quah (1994)). The analysis of the cross-section distribution of first stage residuals seems to be of little practical interest since it removes the structural differences among economies but to some extent it helps for the purpose of discriminating among rival models of growth.

Convergence in this setting would simply mean that each individual economy tends to revert to its steady state, regardless of whether this is approaching the OECD average or not. And this is precisely what the constant returns growth model is about. If the first stage residuals still do not collapse to their average value, we ought to conclude that either the convergence property does not hold or else that the various accumulation rates included in the regressors do not suffice to capture the steady state behaviour of each particular economy. This latter possibility cannot be denied, and suggests a straightforward strategy for sequential conditioning. Thus, conditioning here is used for the purpose of testing the constant returns to scale model, bridging the gap between the structural (regression) approach and the analysis of the dynamics of the cross-section distribution of income across OECD economies.

The evidence from the standard analysis[17], using alternative data sets, suggests that the Solow model, augmented to include human capital as a productive factor, explains rather well the evolution of growth rates in the OECD. According to the CRS model, growth is explained by two factors, the amount of resources devoted to accumulate human and physical capital (the steady state component), and the distance with respect to the long-run (sustainable) human and physical capital labour ratios (catching-up component). Strictly speaking, convergence only operates through the second component, hence it seems wise to remove the effect of the accumulation rates in order to asses whether or not it has effectively taken place. This implies that if the CRS model holds upon conditioning on these rates, we ought to be able to find a stronger tendency towards a degenerate distribution of incomes. The remaining differences in income across countries being explained by permanent differences in the way countries allocate their resources between savings and investment.

To analyse the evolution of the cross-section distribution of per capita income conditioned to its steady state, we have tried three different versions of the conditioning sets as suggested by the standard regression literature and by the basic Solow model.

The variable of analysis is defined as:

$$Y_c = Y - Y^{ss}$$

where Y and Y^{ss} are the logs of the GDP per capita and the (time varying) steady state GDP per capita respectively. According to the Solow model, the steady state is estimated as,

$$Y^{ss}{}_{it} = \pi_{i0} + \pi_{i1}\log(IY_{it}) + \pi_{i2}\log(TN) + g_i t \tag{2.8}$$

Where IY and TN are the basic accumulation rates of physical capital and population[18]. A linear deterministic trend is included to capture (although rather imperfectly) the exogenous increase of total factor productivity. We have tried different versions of (2.8).

4.1. conditioning on a common technology across countries

We first impose the assumption of a common technology across countries, so that the parameters of the steady state can be consistently recovered in a pooling estimation of the steady state equation. In other words, we restrict the parameter values to be the same for all countries in the sample ($\pi_{i0}=\pi_0$, $\pi_{i1}=\pi_1$, $\pi_{i2}=\pi_2$, $g_i=g$, all i) so that their long-run differences are only due to their different accumulation rates[19].

The dominant picture of the results in this first stage conditioning is quite similar to the one we got in the unconditional case. As Figures 11a to 11d make clear, conditioning does not make a huge difference to the estimated density functions. They are, if any, slightly smoother than the unconditioned ones, but still display very similar shapes and dynamics. Most of the observations tend to get closer to 1, although there is still a group (roughly 25% of the total) stuck at around 0.5. Quite similar features of persistence as in the unconditional case can be drawn from the picture of the *stochastic kernels* (Figures 12a-d and 13a-d).

The time-invariant matrix, defined by taking as starting point an empirically uniform unconditional distribution, shows a slightly higher mobility than before. Similarly, the ergodic distribution, far from being degenerate around 1, accumulates a significant mass of probability in other income ranges. Finally, the sequence of quantiles and mobility indices confirm that little is gained by conditioning in this simple way. The 25% smallest relative incomes in 1960 were between 0.35 and 0.70 of their estimated steady state, while the upper limit of this interval got closer to 1 by the mid '80s it went down again slowly. A group of countries remained below their expected long-run path all through the sample period. The second and third quantil tend to concentrate around 1 while the richest quantil starts with an upper limit of 1.98 and evolves until 1.43 times the steady state.

There is little change in the results after removing the effect of accumulation rates. It should be noticed that the interpretation of the bimodal distribution is slightly different to the one we made earlier. Now, the variable of interest is not income relative to the OECD average but income relative to its own steady state. The failure to obtain a degenerate distribution around 1 is not only evidence against conditional convergence but also about the way the steady state is defined. If a group of countries turns out to be persistently 50% of alleged steady state, we can conclude that either the saddle path property of the neoclassical model does not hold or that the steady state is wrongly measured, or both. In fact, as Figure 15 makes clear, part of the blame should be put on the conventional estimation of the steady state using multi-country data sets. Imposing the assumption of common technological parameters, we find that most countries do not even achieve their (estimated) steady state during the period 1960-1990. Even more worrying than this is the fact that over the whole sample period the richest countries are systematically above their steady state while the poorest ones tend to be below theirs.

4.2. allowing for country-specific and time-invariant effects

Controlling for accumulation rates is not enough to achieve the features of conditional convergence. One way to ascertain whether the saddle-path property of the neoclassical model does not hold or if we have not properly captured the long-run determinants of per capita income, is to improve our approximation to the steady state. In the second version of (2.8) we introduce a time-invariant country-specific effect to allow for differences in the initial level of accumulated technical knowledge. This means that π_{i0} is allowed to vary across countries. By doing so we are controlling for structural differences across countries that we do not try to explain.

This procedure is related to some work in the empirical growth literature that has explicitly considered the possibility of country-specific effects in one way or another[20]. The overall conclusion of these studies is that country specific effects are relevant and that convergence rates are higher once those effects have been controlled for. The analysis of the income distribution that we have performed is a sort of reduced form analysis which is not intended to uncover the economic factors behind the evolution of relative incomes. The specific effect does shift the estimated steady state for each country in such a way that the residuals are around zero. In this case, countries cross their steady state from time to time, but still the residuals are far from stationary and the level of income does not return to its steady state value even after long time periods.

Turning now to the main results of this exercise, we find, as expected, higher mobility and concentration around the newly defined steady-state. The estimated 1 year (average) transition kernel (Figure 16a) still displays a significant degree of persistence year by year, with the mass of probability concentrated around the main diagonal. Out of the initial state, transition probabilities are higher over a 5 year horizon (Figures 16b-d, 17b-d). The estimated contour is steeper than the main diagonal for the period 1961-65 indicating a strong tendency towards the steady-state during these years. It shows high mobility during the 1970-75 period and the persistence increases again for the 1981-85 period. The transition matrices give very much the same message although with, less persistence than in previous cases for the time invariant Q, but still an uniform ergodic distribution ie from a uniform initial distribution, the probability of each economy being above or below its steady state is the same. The states that define the sequence of quantiles approach 1 from above. The upper bound of the lowest quantil approaches 1 steadily until the mid-`80s and then falls again. The lower bound remains relatively unchanged at a low value (less than 0.3), so some countries remain quite far from their steady-state income level. The mobility indices are higher than in previous cases but there are still many zeroes in the off-diagonal entries in the transition matrices which imply substantial persistence that again increases during the `80s.

The transition matrices give very much the same message, less persistence than in previous cases for the time invariant Q and an uniform ergodic distribution i.e. the probability of each economy being above or below its steady state is the same.

4.3. a simple illustration of Solow Model's predictions

Given the structure of the multi-country data set, some authors have argued that the estimation of the steady-state should explicitly allow for cross-country differences in the parameter set, over and above differences in the constant term. In fact, Pesaran and Smith (1993) show how the relevant parameters in these kinds of models can only be recovered in a consistent manner by averaging the estimated parameters in individual country regressions. In related work, Andrés and Boscá (1993) have shown how the hypothesis of a common parameter set across the OECD economies can be rejected using standard econometric tests. This suggests another possibility of enlarging the conditioning set in a way that is consistent with the "representative economy" model. In the final part of this section we analyse the evolution of the distribution of relative incomes calculated as in (2.8), but allowing for country specific differences in all π_{0i}, π_{1i}, π_{2i} and π_{3i} parameters. This exercise ensures a meaningful steady state (one with respect to which the residuals seem more stationary) and it permits comparisons with the analysis carried out so far to see whether (as expected) permanent differences across OECD countries are explained not only by differences in the accumulation rates but also in the technological parameters.

This conditioning set yields a series of relative incomes such that for every single year), the distribution is unimodal and clearly degenerated around 1. The *stochastic kernels contours* are still around the main diagonal for 1 year transitions but become virtually vertical at the 1.0 level for five year periods (Figure 18a-d and 19a-d), indicating that income transits quickly to its steady-state value regardless of the initial level. For the 1960-65 period, the contours are even negatively sloped indicating a significant amount of overtaking among countries ie those countries below their potential income tend to be above it 5 years later. The invariant transition matrix is indicating great mobility, with no entry with a zero value and a similar probability of upwards and downwards mobility. The ergodic distribution is again uniform. The sequence of quantiles is very stable, but within very close range of 1[21]. Similarly, the time varying transition matrices have very few zero entries (some of them on the main diagonal) and suggest a substantial mobility which is confirmed by the very high mobility indices (Figure 9) which are twice as large as those found in previous exercises.

The features in the data on relative incomes, appear to be consistent with what we would expect if convergence is taking place. Nevertheless, we must bear in mind that this income is relative to a country specific steady state. Following these results one may or may not trust the ability of the CRS growth model to account for the long-run behaviour of each economy taken separately. It should be noticed, that high mobility and low persistence in the residuals from the first stage regression might simply be a feature of the autoregressive process of per capita income in each country. In particular this low persistence is also an indication of stationarity in the first stage residuals. However, this is precisely what convergence is about; convergence must be seen as the tendency of income per capita to return to its long-run path. What is clear, however, is that the model, at least in the way it has been extensively used in the literature in recent times, has a much harder job of explaining the evolution of the distribution of incomes across countries. In fact, nothing in our results suggests that OECD economies should get even closer in the long-run.

In order to render the first stage residuals stationary we have had to remove a great deal of country specific features that account for most of the differences across countries and indeed for most of their persistence. In other words, the conventional exogenous growth model might (or might not) be a good representation of the representative economy, but it certainly fails to capture the long-run evolution of the distribution of incomes across a group of economies interacting with each other. Hence, an alternative theoretical approach, which takes into account actions among economies is needed to take out conclusions about convergence.

5. CONCLUSION AND FINAL REMARKS

The main conclusions from the exercise carried out in this paper can be summarized as follows:

·The pattern of the cross-country distribution of per capita income within the OECD is not one of full convergence. Although the income gap has narrowed substantially over the sample period, there remain substantial differences that do not show a falling path in the last 15 years or so. Upward mobility in the income ranking is very limited, especially out of the 25% group of the poorest countries. In fact, the steady state cross-country distribution of income is not degenerate around the OECD average, but it reflects non vanishing sizable differences. These differences do not merely reflect variations in accumulation rates nor those in initial conditions (fixed effects). In fact, after conditioning we have found very much a similar picture, in which there is little tendency of countries to bounce around their steady state path, measured as is a common practice in most empirical studies. To obtain a picture compatible with the convergence proposition, we need to allow for time varying country specific effects as well, indicating that if the constant returns model would hold, it would only do so on a country by country basis, and that we should not use it to make inference about the long-run path of relative incomes, even within a homogeneous region such as the OECD.

·The conventional cross-country regression analysis and the study of the dynamics of the cross-country distribution, yield somewhat different conclusions that are not easily reconcilable. While the dispersion and regression towards the mean measures indicate that OECD countries converge in the σ and in the β sense, the dynamic approach in this paper suggests a more pessimistic view.

·Finally, in the paper we have also shown that the standard way of testing the propositions of growth theory using multi-country data sets can be very misleading. This evidence is damaging for the conventional analysis in two ways. Firstly, because it indicates that the saddle point property for the "representative economy" cannot be easily extended to make multi-country comparisons. Secondly, because one of the advantages of the cross-country regression analysis, namely that of collapsing the most relevant information about long-run patterns in a bunch of relevant statistics, does not hold. From the empirical point of view, these conclusions suggest the advantage of using an approach that imposes as little prior structure as possible on the data. From the theoretical one, it is necessary to develop in more detail the multi-country implications of existing growth literature, beyond the representative economy assumption, and set up models that explicitly take into account the kind of idiosyncratic factors that lie behind the persistent economic differences among countries.

ENDNOTES

1. "Miracles" and "disasters" in Parente and Prescott's (1993) terminology.

2. See Danny Quah (1993a,b; 1994).

3. Although it may also take place in an endogenous growth setting (Kelly, 1992).

4. Mankiw, Romer & Weil (1992) augment the Solow model by adding human capital accumulation. For simplicity and without any loss of generality, reference will be made to the equations of the original model.

5. Quah (1994b) illustrates how the uniformity of this 2% may arise from a unit root in the time series, instead of from reasons related to the dynamic of economic growth.

6. This approach implies imposition (without formal testing) of the assumption of a common technology across countries in the sample. However, whenever the assumption of a common technology has been tested it has been more often rejected than not, suggesting that the "best practice" technology is not available to all countries in the sample at a particular point in time, leading also to substantial differences in convergence rates. See for instance Durlauf and Johnson (1992). Andrés and Boscá (1993) formally test and reject the assumption of a common technology across OECD countries, obtaining parameter values which are inconsistent with the constant returns technology.

7. The calculations and graphics have been executed using Danny Quah's Time Series, Random-Fields shell *tSrF*.

8. D. Quah (1993a,b).

9. See Easterly *et al* (1993), among others.

10. See Andrés J., Domenech, R. and Molinas, C., (1994).

11. This normalization is a way to abstract each individual economy from the overall growth and fluctuations.

12. Although without any indication about the accuracy of the predictions, yet.

13. Obtained using the squared of standard Epanechnikov kernel for estimating the joint density $f(X_{iT,}X_{iT+\tau})$ and then re-scaling to obtain the conditional probability. X is the per capita GDP in each individual country relative t the same variable for the OECD as a whole, $\tau=1,5$. The *bandwidth* is chosen by least square cross-validation (see Silverman (1986), section 3.4.3). All the calculations were done with Quah's shell *tSrF*.

14. Hence the distance with the richest group average is still larger.

15. The entry estimates are not presented here to save space.

16. These cross-profile representations were used in Juan J. Dolado, Jose Manuel Gonzalez-Paramo and Jose Maria Roldan (1994) and in Quah (1994).

17. Mankiw *et al* (1992) and Andrés *et al* (1994) found evidence in favour of the constant returns to scale model which comes not only from the positive convergence rate, but also from the fact that the theoretical restrictions imposed by this technology seem to fit very well the OECD experience over the last thirty years.

18. We present here the results conditioning for investment rates, population growth, trend and dummies, conditioning also with respect to human capital does not change the results in a significant manner.

19. Although we present the results based in the linear version of (8) we have also estimated the coefficients in a non linear error correction format with very similar results.

20. See Durlauf and Johnson (1992), Knight, Loyza and Villanueva (1992), Andrés and Boscá (1993).

21. In 1960, for instance, the lowest quantil had bounds 0.918 and 0.954, whereas the highest ones were 0.991 and 1.030.

TABLE 1
OECD Relative Per-Capita Income
First Order Transition Matrix, Time Stationary 1960-89

Upper End of the States	0.778	0.940	1.026	1.578
(r)	(1)	(2)	(3)	(4)
175:	0.95	0.05	0.00	0.00
174:	0.03	0.90	0.07	0.00
172:	0.00	0.06	0.87	0.07
175:	0.00	0.00	0.09	0.91
Ergodic Distribution	0.166	0.265	0.314	0.255

TABLE 2
Quantiles (0.25, 0.5, 0.75)
OECD Relative Per-Capita Income

Cell Partition:					
1961:	0.265	0.538	0.853	1.064	1.565
1962:	0.264	0.561	0.860	1.042	1.540
1963:	0.272	0.584	0.860	1.042	1.526
1964:	0.263	0.585	0.868	1.042	1.514
1965:	0.254	0.593	0.858	1.045	1.520
1966:	0.265	0.610	0.861	1.030	1.523
1967:	0.263	0.653	0.848	1.001	1.506
1968:	0.262	0.698	0.842	1.004	1.485
1969:	0.258	0.715	0.859	1.022	1.451
1970:	0.258	0.745	0.893	1.012	1.442
1971:	0.268	0.775	0.903	1.033	1.449
1972:	0.267	0.776	0.909	1.017	1.423
1973:	0.260	0.766	0.917	0.998	1.392
1974:	0.275	0.802	0.914	1.019	1.409
1975:	0.295	0.831	0.920	1.024	1.351
1976:	0.305	0.824	0.928	1.021	1.352
1977:	0.302	0.824	0.914	1.005	1.358
1978:	0.295	0.814	0.915	0.992	1.363
1979:	0.279	0.849	0.925	0.997	1.348
1980:	0.270	0.865	0.954	1.010	1.325
1981:	0.273	0.866	0.939	1.018	1.327
1982:	0.282	0.889	0.954	1.036	1.296
1983:	0.280	0.903	0.947	1.020	1.302
1984:	0.278	0.891	0.941	1.013	1.322
1985:	0.279	0.897	0.942	1.008	1.315
1986:	0.290	0.899	0.940	1.022	1.309
1987:	0.297	0.873	0.934	1.017	1.302
1988:	0.291	0.828	0.949	1.000	1.294
1989:	0.281	0.807	0.942	0.993	1.284
1990:	0.296	0.785	0.946	1.004	1.268

TABLE 3
OECD Residuals First Stage Regression
Conditioning on a Common Technology across Countries
First Order Transition Matrix, Time Stationary 1960-89

Erreur! Signet non défini. Upper End of the States	0.866	1.102	1.250	2.142
(r)	(1)	(2)	(3)	(4)
169:	0.95	0.05	0.00	0.00
165:	0.02	0.90	0.07	0.00
169:	0.00	0.08	0.80	0.11
169:	0.00	0.00	0.13	0.87
Ergodic Distribution	0.163	0.318	0.279	0.241

TABLE 4
Quantiles (0.25, 0.5, 0.75)
OECD Residuals First Stage Regression
Conditioning on a Common Technology across Countries

Cell Partition:					
1961:	0.351	0.701	0.990	1.303	2.124
1962:	0.384	0.736	1.029	1.274	1.980
1963:	0.392	0.747	1.066	1.312	1.992
1964:	0.389	0.754	1.060	1.298	1.969
1965:	0.365	0.770	1.130	1.305	1.991
1966:	0.394	0.780	1.007	1.280	1.971
1967:	0.407	0.797	1.012	1.256	1.946
1968:	0.406	0.805	1.033	1.309	1.931
1969:	0.396	0.805	1.027	1.315	1.926
1970:	0.431	0.823	1.082	1.308	1.916
1971:	0.424	0.861	1.137	1.287	1.875
1972:	0.412	0.872	1.154	1.272	1.866
1973:	0.386	0.891	1.138	1.301	1.810
1974:	0.403	0.888	1.098	1.277	1.803
1975:	0.368	0.871	1.111	1.236	1.808
1976:	0.433	0.865	1.089	1.211	1.802
1977:	0.447	0.896	1.106	1.224	1.807
1978:	0.416	0.928	1.121	1.225	1.793
1979:	0.414	0.946	1.118	1.214	1.773
1980:	0.404	0.896	1.120	1.218	1.721
1981:	0.408	0.910	1.107	1.150	1.654
1982:	0.407	0.913	1.100	1.193	1.619
1983:	0.417	0.925	1.098	1.170	1.614
1984:	0.409	0.954	1.073	1.140	1.593
1985:	0.397	0.940	1.064	1.160	1.587
1986:	0.417	0.938	1.069	1.150	1.576
1987:	0.423	0.942	1.058	1.188	1.579
1988:	0.419	0.934	1.061	1.193	1.570
1989:	0.426	0.949	1.082	1.131	1.438

TABLE 5
OECD Residuals First Stage Regression
Country-Specific and Time-Invariant Effects
First Order Transition Matrix, Time Stationary 1960-89

Upper End of the States	0.960	1.009	1.060	1.217
(r)	(1)	(2)	(3)	(4)
160:	0.83	0.17	0.00	0.00
171:	0.17	0.61	0.21	0.01
170:	0.01	0.20	0.58	0.21
171:	0.00	0.01	0.21	0.78
Ergodic Distribution	0.259	0.249	0.248	0.243

TABLE 6
Quantiles (0.25, 0.5, 0.75)
OECD Residuals First Stage Regression
Conditioning on Country-Specific and Time-Invariant Effects

Cell Partition:					
1961:	0.694	0.893	0.961	1.016	1.161
1962:	0.720	0.898	0.960	1.009	1.175
1963:	0.766	0.910	0.972	1.025	1.160
1964:	0.779	0.932	0.970	1.008	1.188
1965:	0.829	0.930	0.985	1.017	1.202
1966:	0.871	0.927	0.984	1.010	1.129
1967:	0.840	0.945	1.001	1.031	1.130
1968:	0.886	0.989	1.016	1.049	1.176
1969:	0.915	1.009	1.035	1.058	1.139
1970:	0.931	1.009	1.029	1.055	1.116
1971:	0.959	1.022	1.046	1.074	1.132
1972:	0.948	1.032	1.066	1.083	1.212
1973:	0.964	1.040	1.061	1.090	1.158
1974:	0.954	1.000	1.029	1.060	1.124
1975:	0.959	1.014	1.051	1.078	1.130
1976:	0.992	1.007	1.046	1.078	1.104
1977:	0.988	1.011	1.048	1.072	1.139
1978:	0.974	1.004	1.052	1.082	1.127
1979:	0.955	0.994	1.028	1.085	1.141
1980:	0.938	0.958	1.004	1.051	1.132
1981:	0.899	0.928	0.978	1.045	1.119
1982:	0.872	0.950	0.969	1.031	1.115
1983:	0.873	0.954	0.993	1.026	1.124
1984:	0.873	0.941	0.993	1.018	1.139
1985:	0.853	0.923	0.981	1.016	1.125
1986:	0.829	0.919	0.965	1.024	1.137
1987:	0.806	0.898	0.957	1.028	1.149
1988:	0.768	0.886	0.945	1.026	1.156
1989:	0.726	0.881	0.930	1.013	1.172

TABLE 7
OECD Residuals First Stage Regression
Conditioning on Country-Specific and Time-Variant Effects
First Order Transition Matrix, Time Stationary 1960-89

Erreur! Signet non défini. Upper End of the States	0.976	1.001	1.025	1.167
(r)	(1)	(2)	(3)	(4)
159:	0.64	0.21	0.12	0.03
170:	0.22	0.41	0.26	0.11
171:	0.08	0.26	0.41	0.25
172:	0.04	0.10	0.23	0.63
Ergodic Distribution	0.238	0.247	0.256	0.259

TABLE 8

Quantiles (0.25, 0.5, 0.75)

OECD Residuals First Stage Regression

Conditioning on Country-Specific and Time-Variant Effects

1961:	0.918	0.954	0.972	0.991	1.030
1962:	0.896	0.964	0.976	1.004	1.043
1963:	0.926	0.973	0.982	1.000	1.038
1964:	0.943	0.965	0.980	1.003	1.059
1965:	0.888	0.963	0.981	1.006	1.057
1966:	0.919	0.960	0.973	1.001	1.026
1967:	0.898	0.966	0.991	1.007	1.089
1968:	0.940	0.998	1.011	1.037	1.049
1969:	0.963	0.988	1.016	1.054	1.100
1970:	0.967	1.003	1.017	1.045	1.115
1971:	0.984	0.999	1.013	1.056	1.095
1972:	0.914	1.015	1.033	1.047	1.163
1973:	0.965	0.999	1.054	1.068	1.156
1974:	0.958	0.987	1.006	1.031	1.137
1975:	0.960	0.996	1.007	1.039	1.106
1976:	0.988	1.001	1.017	1.038	1.087
1977:	0.985	1.003	1.019	1.037	1.096
1978:	1.001	1.009	1.025	1.039	1.079
1979:	0.983	1.008	1.028	1.039	1.095
1980:	0.953	0.987	1.020	1.029	1.077
1981:	0.948	0.976	0.996	1.020	1.047
1982:	0.936	0.969	0.985	1.006	1.047
1983:	0.950	0.971	0.997	1.012	1.041
1984:	0.954	0.979	0.988	1.005	1.041
1985:	0.949	0.973	0.987	1.006	1.035
1986:	0.953	0.960	0.978	1.007	1.028
1987:	0.902	0.961	0.980	0.997	1.017
1988:	0.901	0.954	0.962	0.989	1.029
1989:	0.770	0.940	0.958	0.996	1.079

FIGURE 1
GDP Per Capita

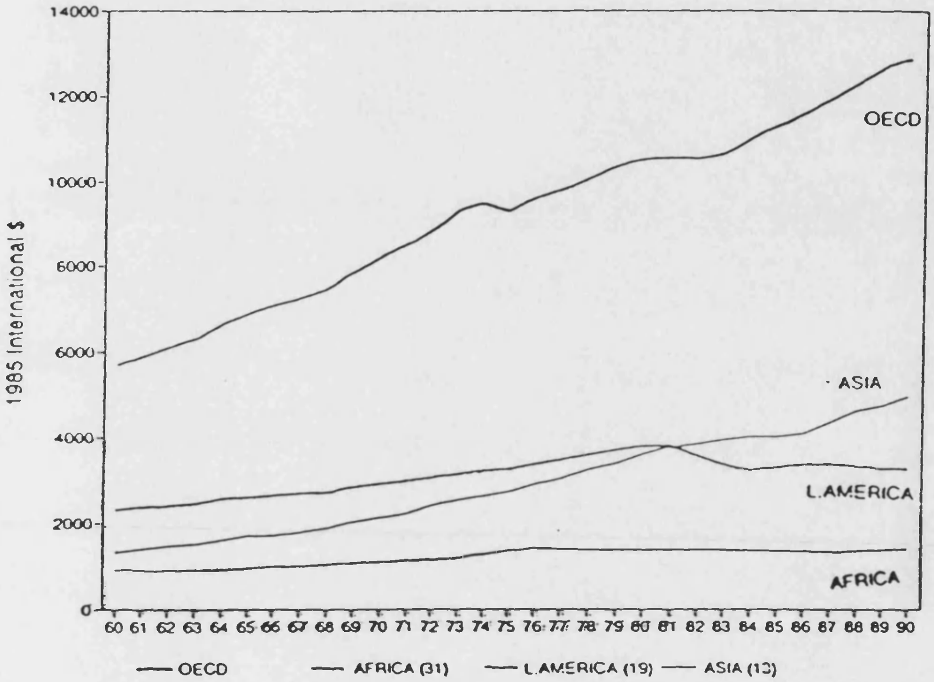


FIGURE 2
Sigma Convergence

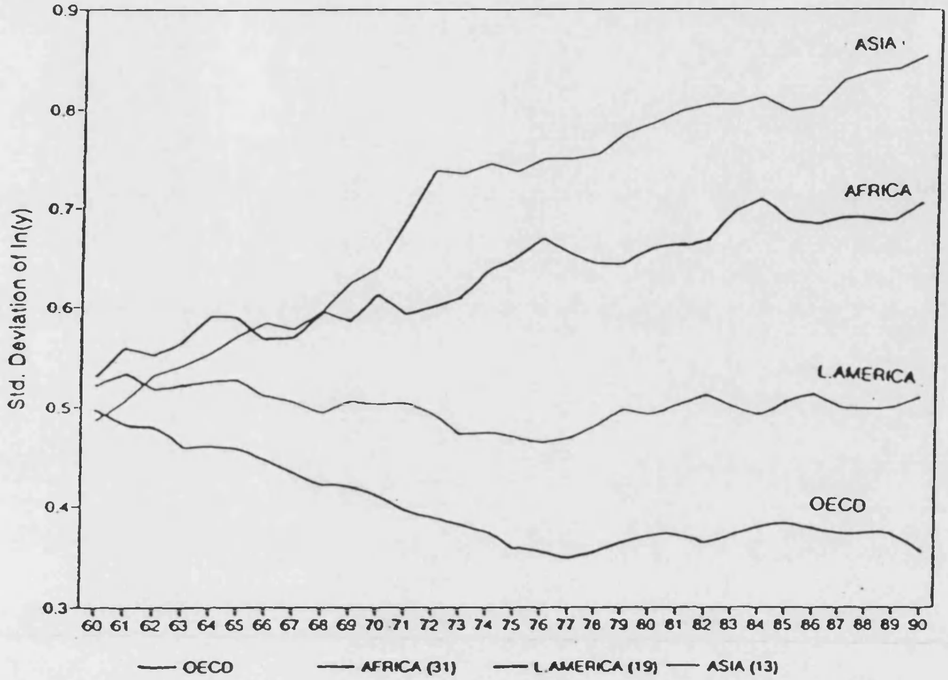


FIGURE 3
OECD, Relative GDP Per Capita

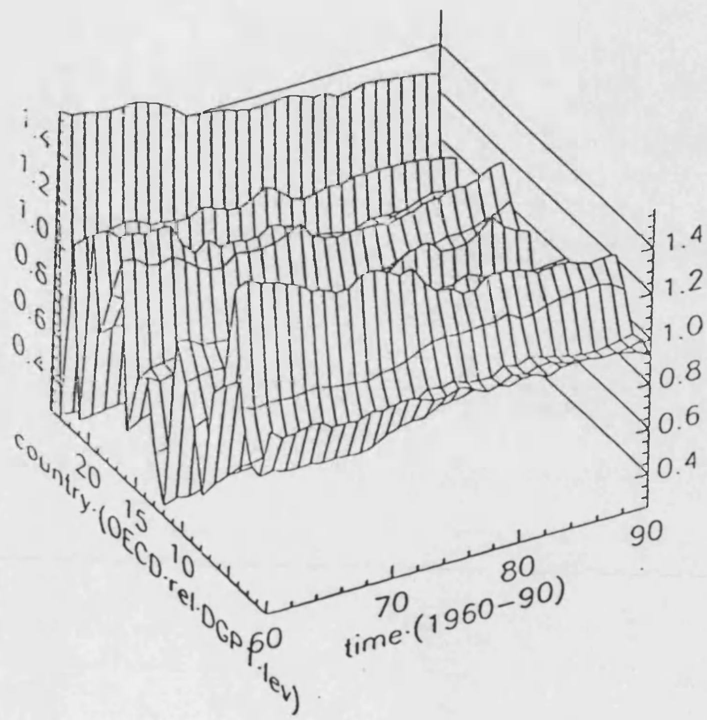


FIGURE 4

Estimated Density Functions
OECD Relative GDP Per Capita

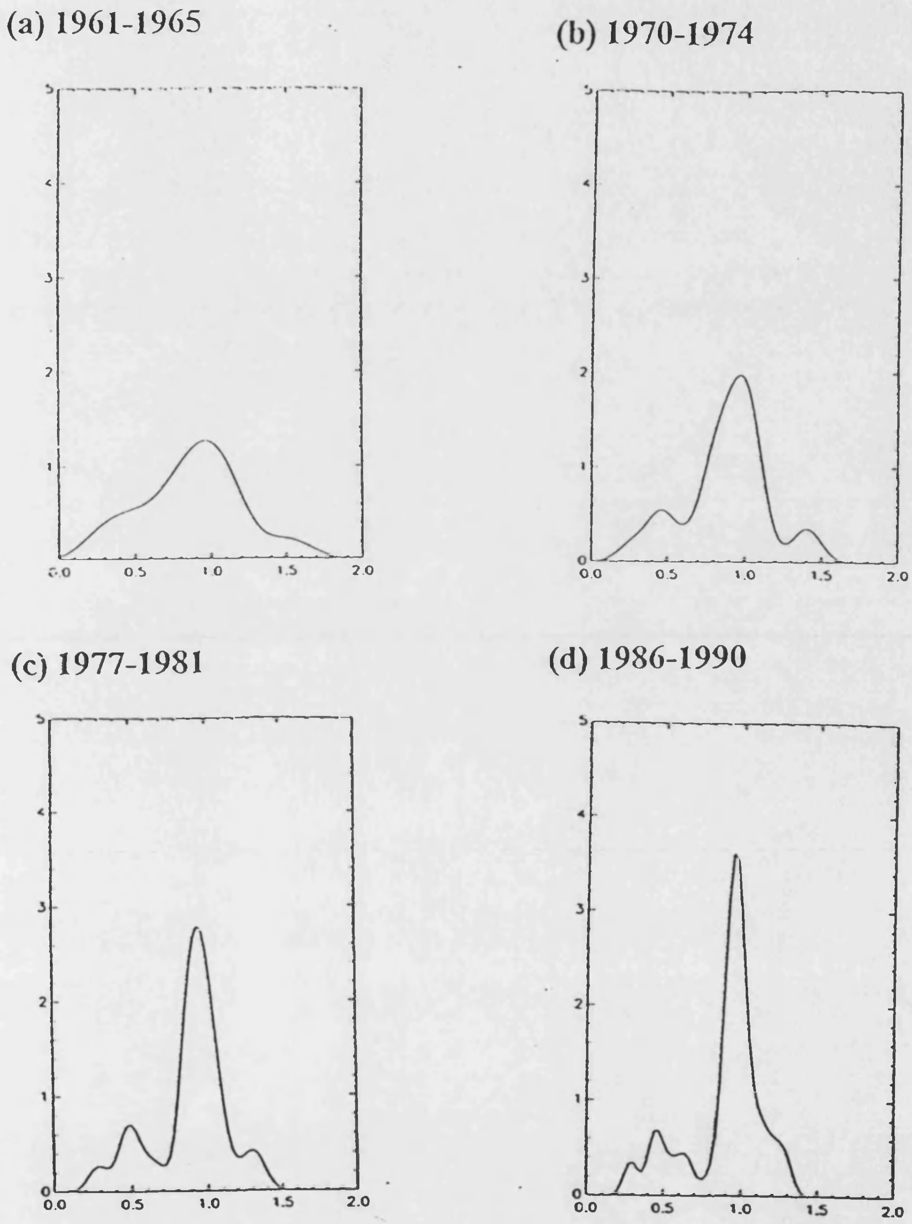


FIGURE 5
Empirical and Theoretical Sigma Convergence in OECD

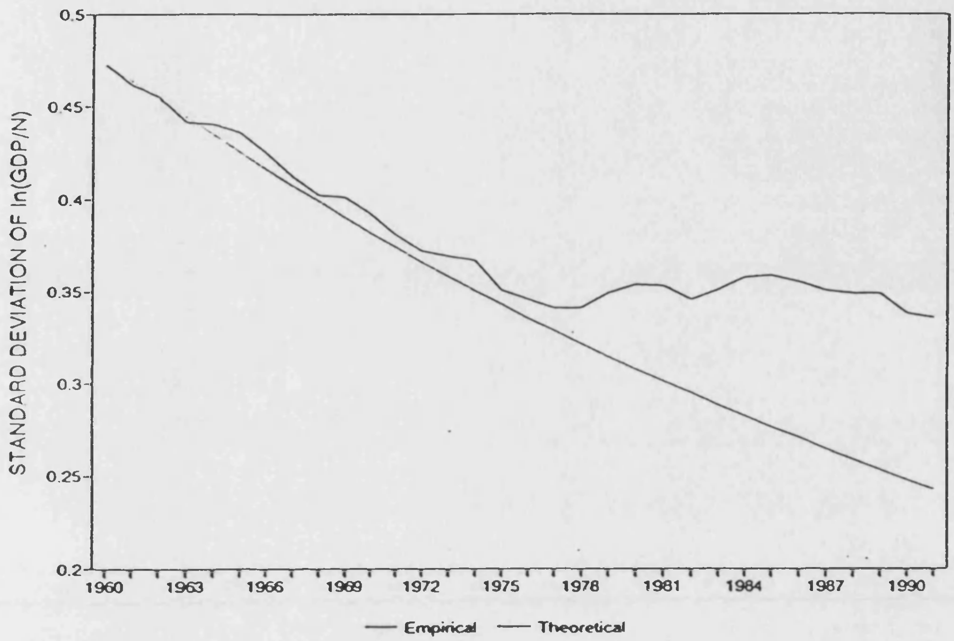
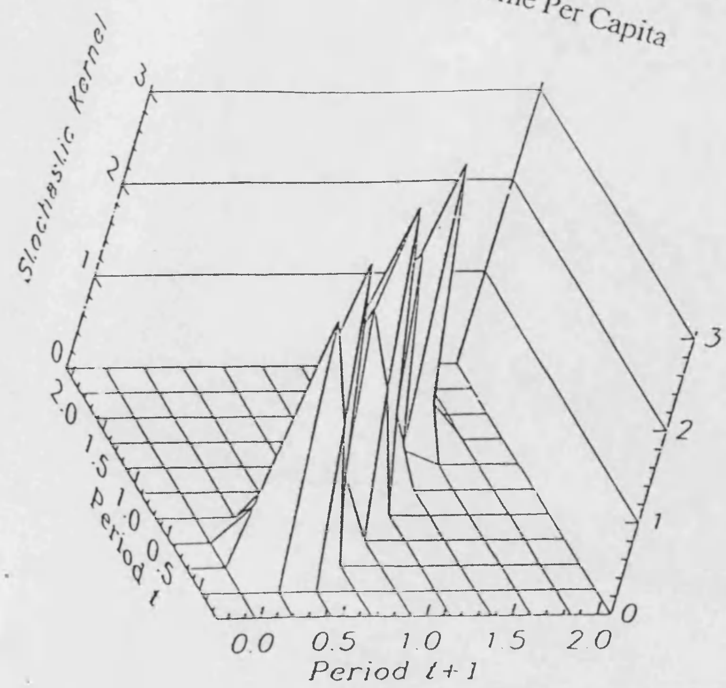
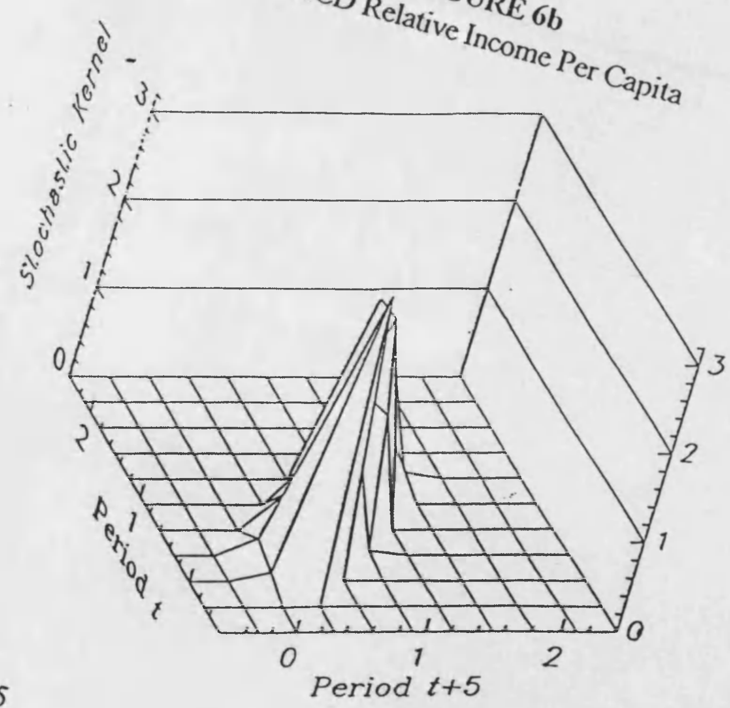


FIGURE 6a
OECD Relative Income Per Capita



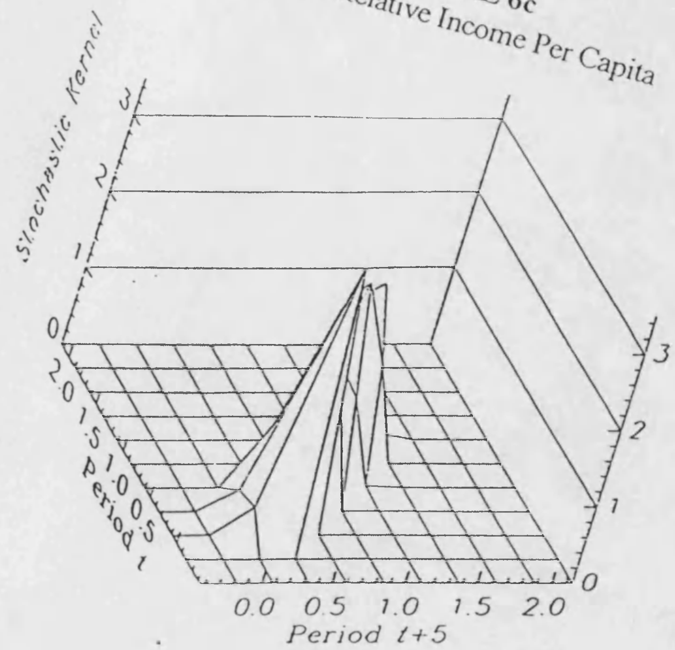
Stochastic Kernel
1 Year Transition

FIGURE 6b
OECD Relative Income Per Capita



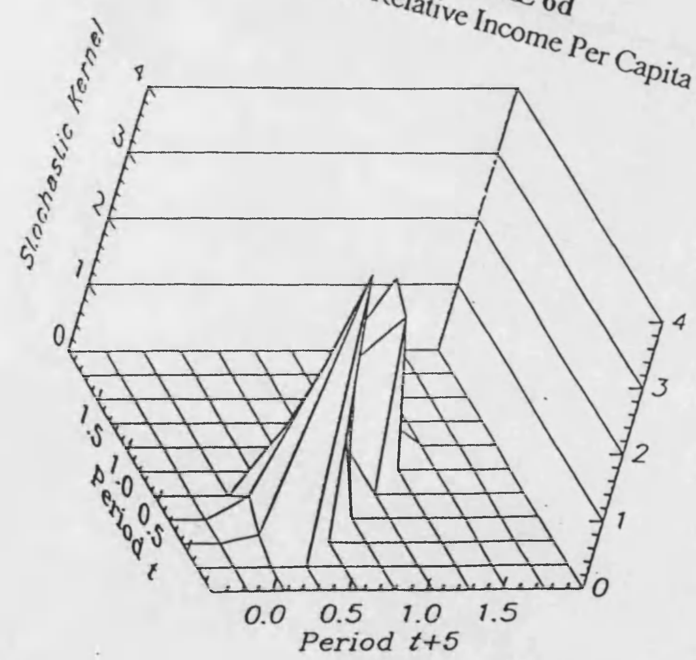
Stochastic Kernel
5 Year Transition 1961-65

FIGURE 6c
OECD Relative Income Per Capita



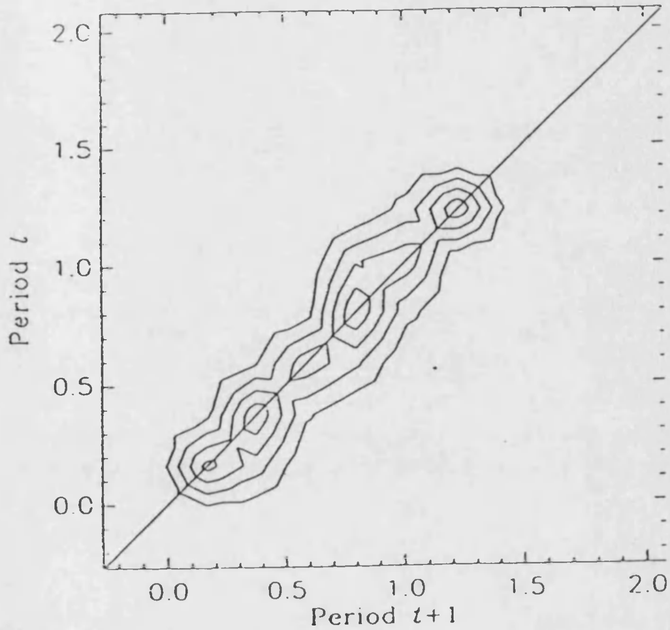
Stochastic Kernel
5 Year Transition 1971-75

FIGURE 6d
OECD Relative Income Per Capita



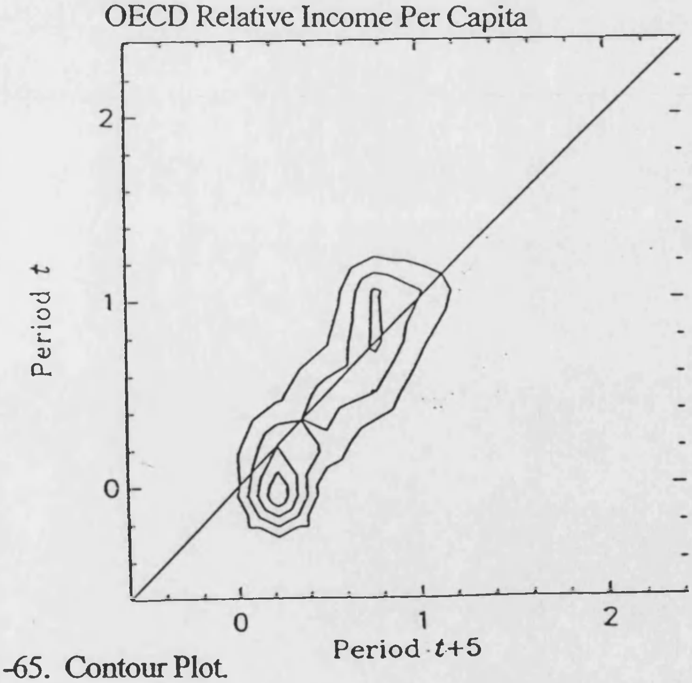
Stochastic Kernel
5 Year Transition 1981-85

FIGURE 7a
OECD Relative Income Per Capita



Stochastic Kernel
1 Year Transition. Contour Plot.

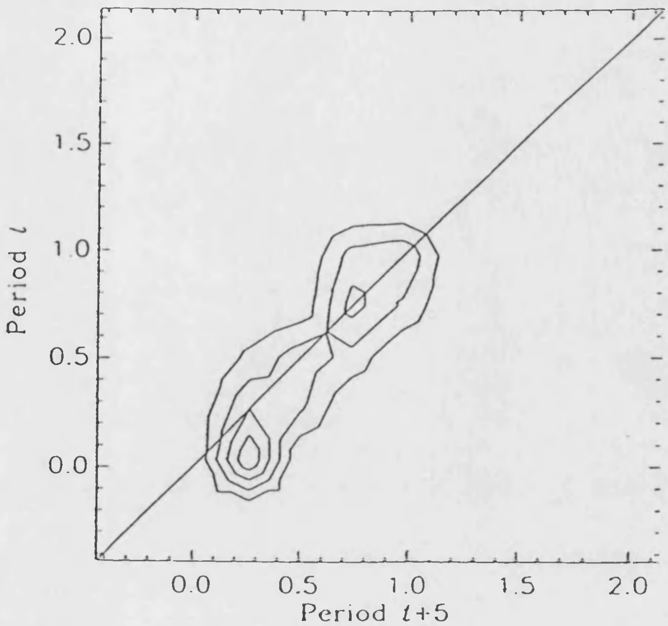
FIGURE 7b
OECD Relative Income Per Capita



Stochastic Kernel
5 Year Transition 1961-65. Contour Plot.

FIGURE 7c

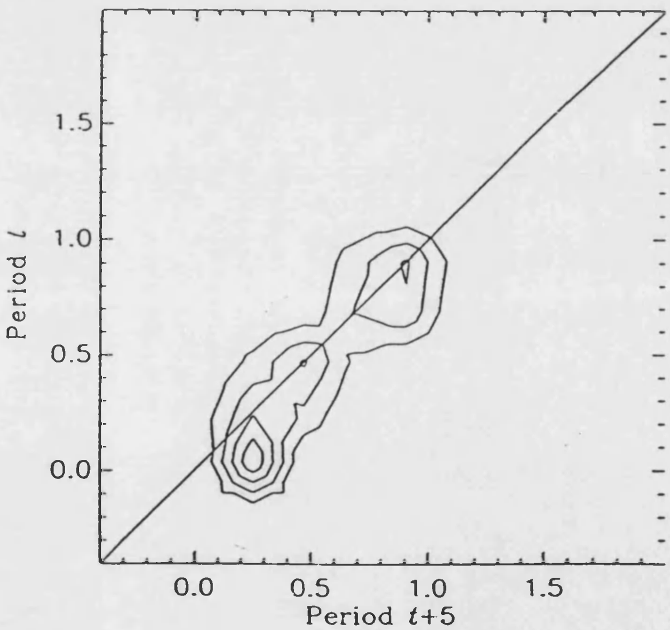
OECD Relative Income Per Capita



Stochastic Kernel
5 Year Transition 1971-75. Contour Plot.

FIGURE 7d

OECD Relative Income Per Capita



Stochastic Kernel
5 Year Transition 1981-85. Contour Plot.

FIGURE 8
OECD Relative Income Per Capita
Quantiles 1960-90

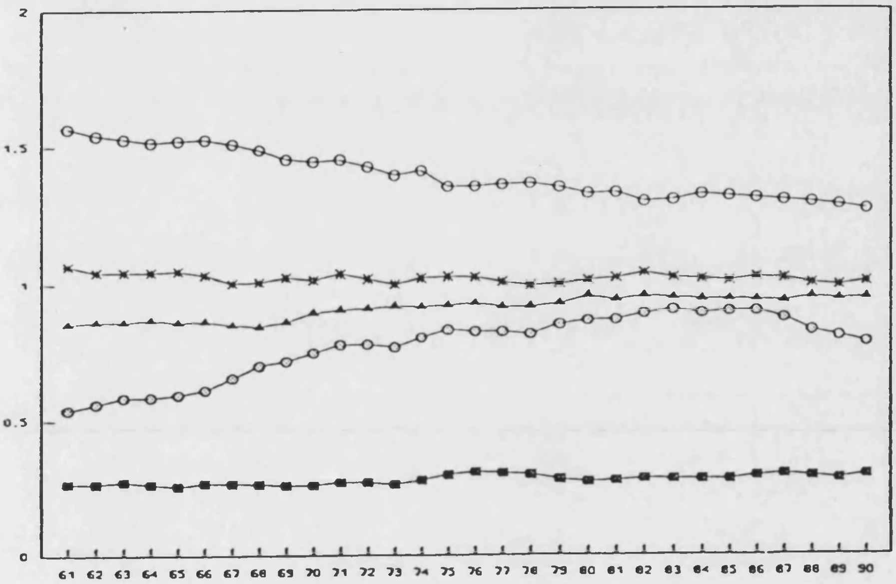


FIGURE 9

Mobility Indices, OECD 1960-1990

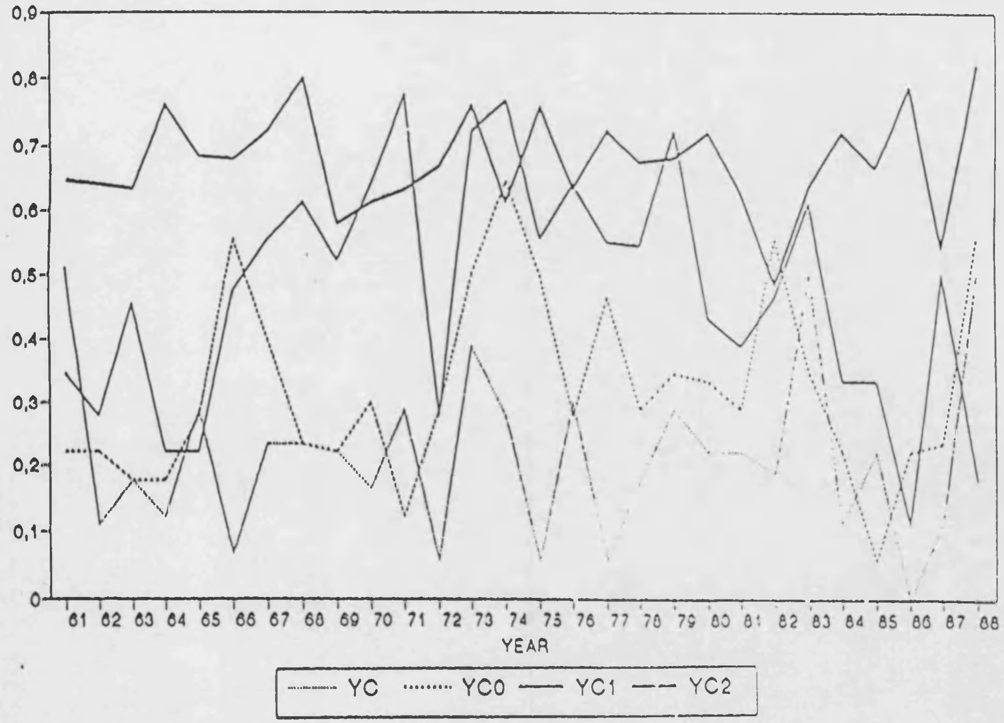


FIGURE 10
OECD Relative GDP Per Capita, Cross Profile
1960, 1969, 1973, 1980, 1985, 1989

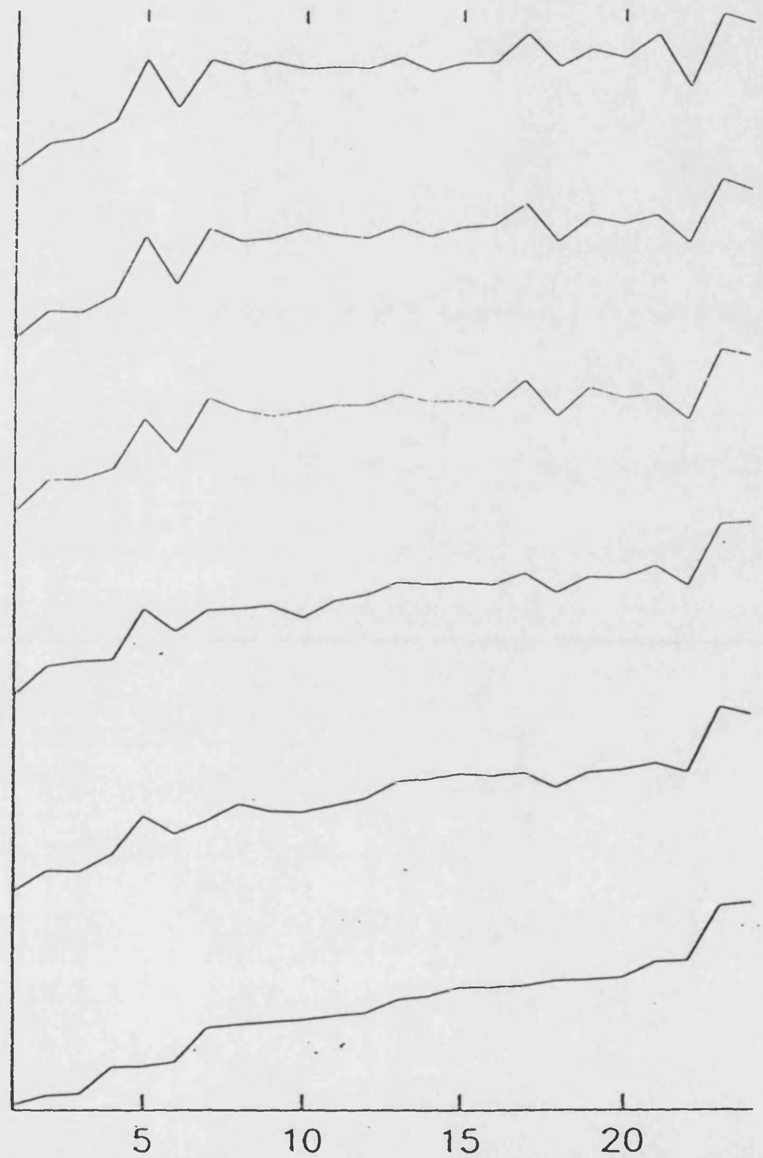


FIGURE 11

Estimated Density Functions, Residuals Fist Stage Regression
Conditioning on a Common Technology

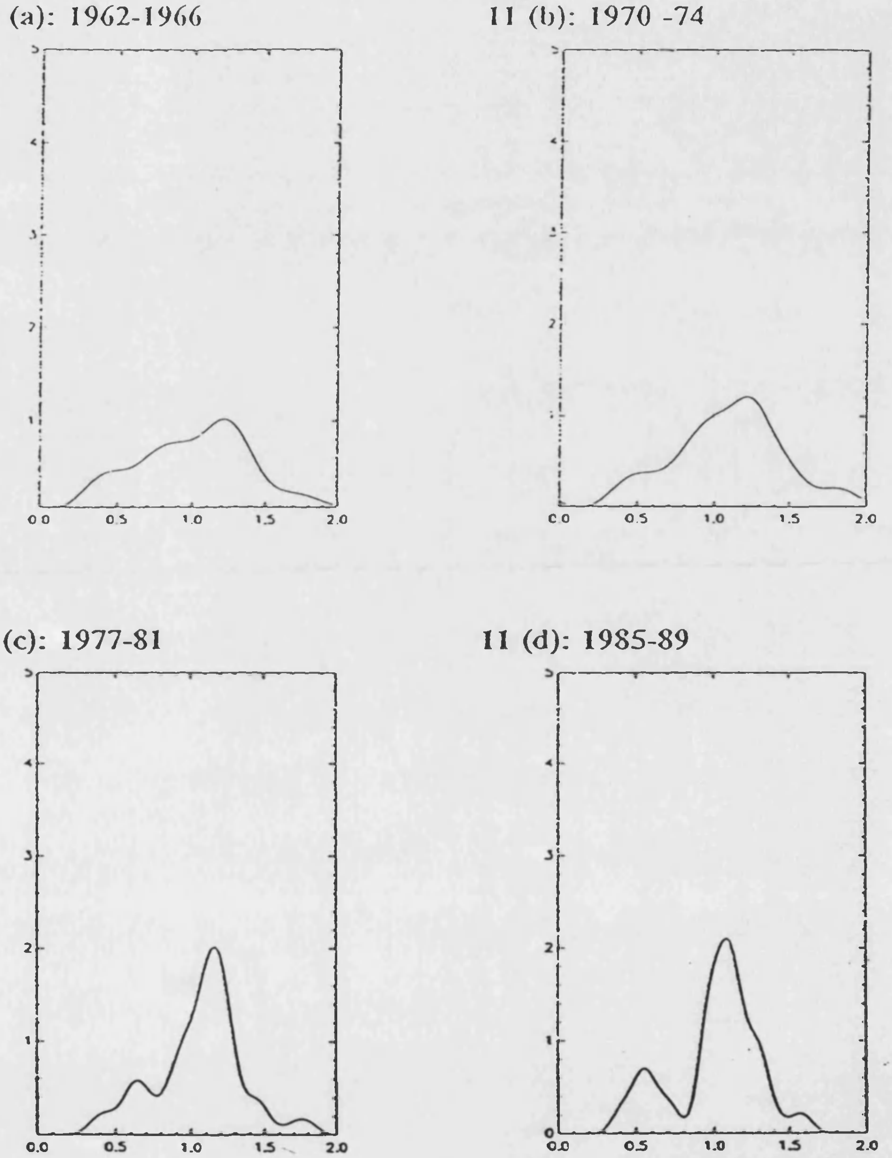
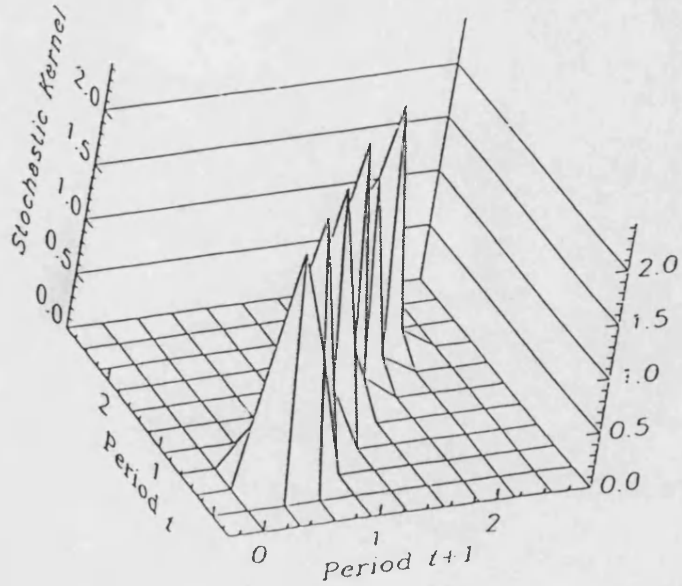
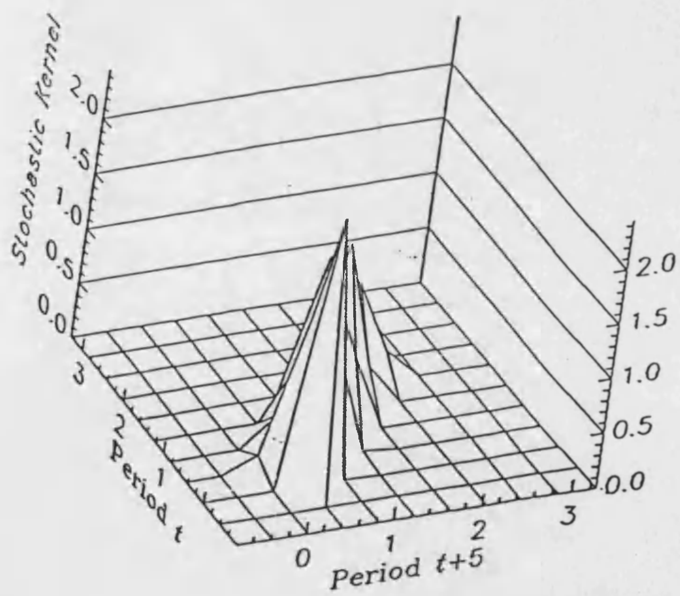


FIGURE 12a
Residuals, Conditioning on a Common Technology



Stochastic Kernel
1 Year Transition

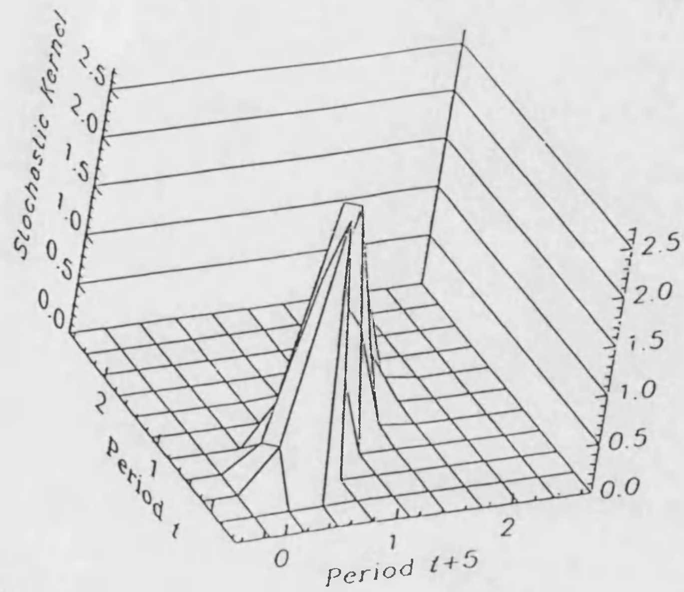
FIGURE 12b
Residuals, Conditioning on a Common Technology



Stochastic Kernel
5 Year Transition 1961-65

FIGURE 12c

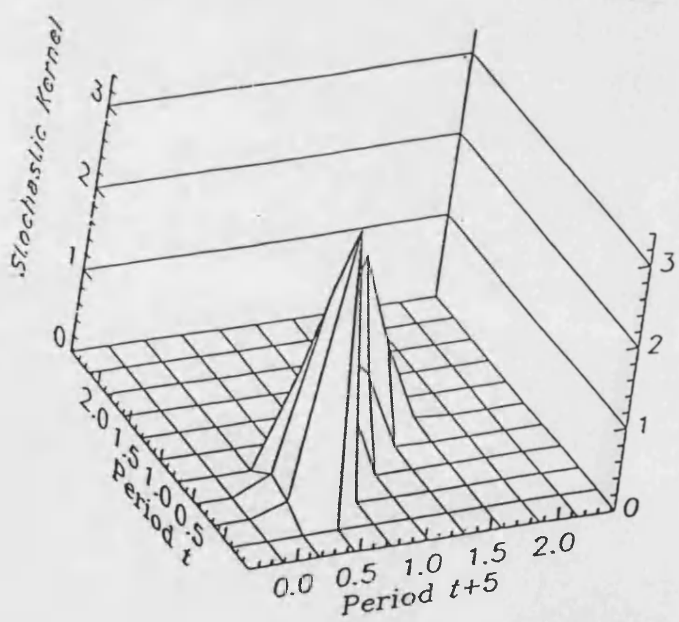
Residuals, Conditioning on a Common Technology



Stochastic Kernel
5 Year Transition 1971-75

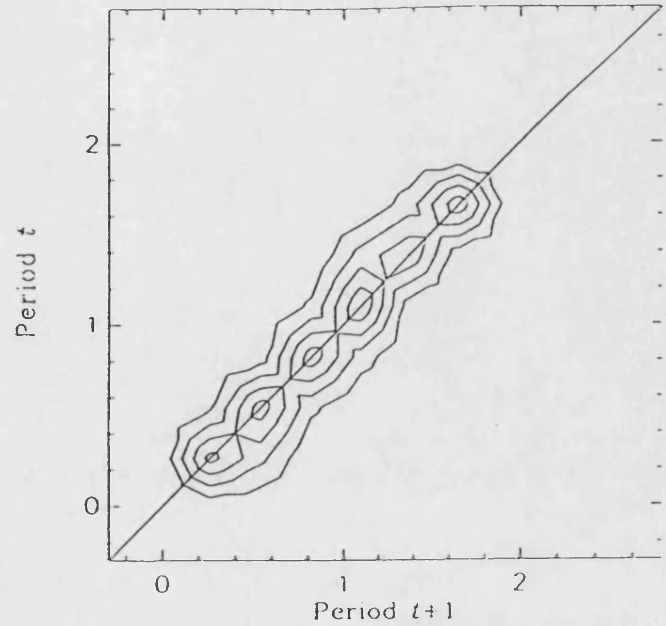
FIGURE 12d

Residuals, Conditioning on a Common Technology



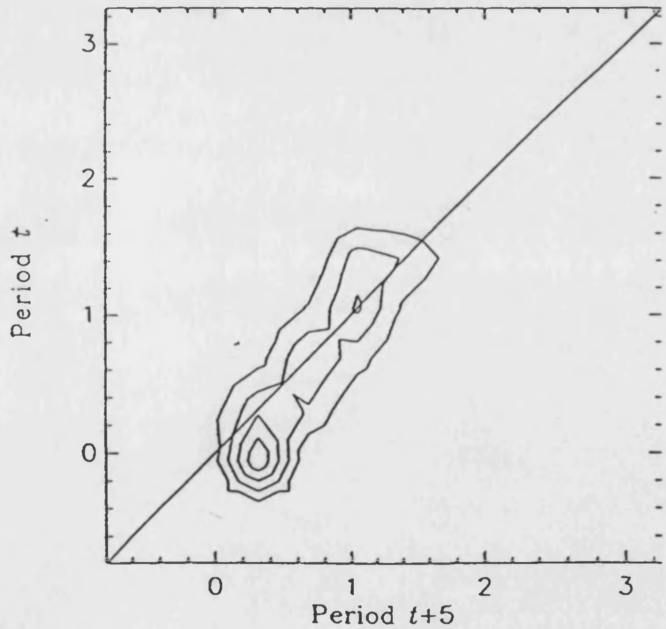
Stochastic Kernel
5 Year Transition 1981-85

FIGURE 13a
Residuals, Conditioning on a Common Technology



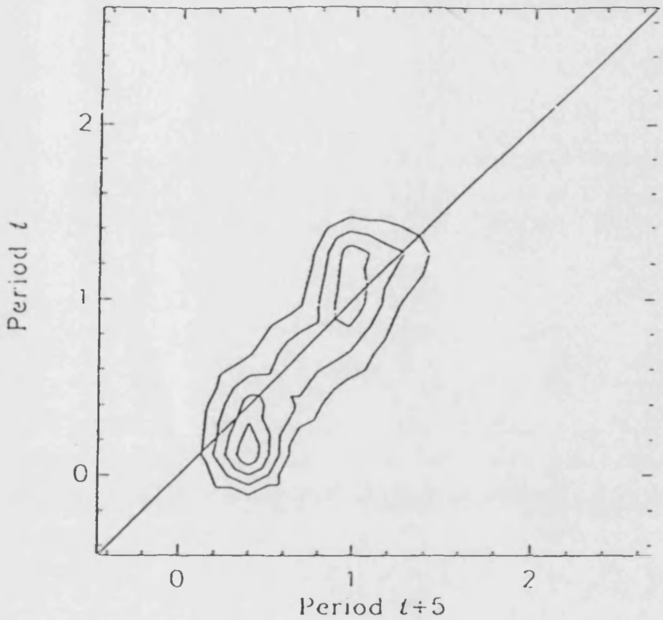
Stochastic Kernel
1 Year Transition. Contour Plot.

FIGURE 13b
Residuals, Conditioning on a Common Technology



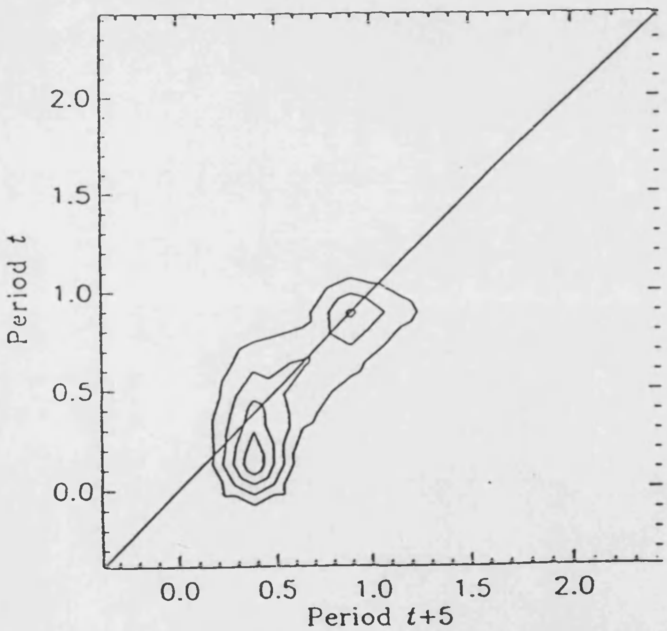
Stochastic Kernel
5 Year Transition 1961-65. Contour Plot.

FIGURE 13c
Residuals, Conditioning on a Common Technology



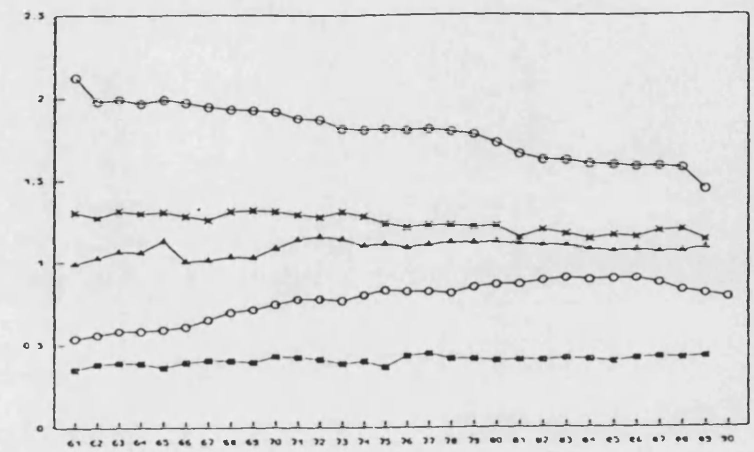
Stochastic Kernel
5 Year Transition 1971-75. Contour Plot.

FIGURE 13d
Residuals, Conditioning on a Common Technology



Stochastic Kernel
5 Year Transition 1981-85. Contour Plot.

FIGURE 14
OECD Residuals
Quantiles 1960-89



Conditioning on a Common Technology across Countries

FIGURE 15
Y Deviation from YSS
OECD Countries 1960-1990

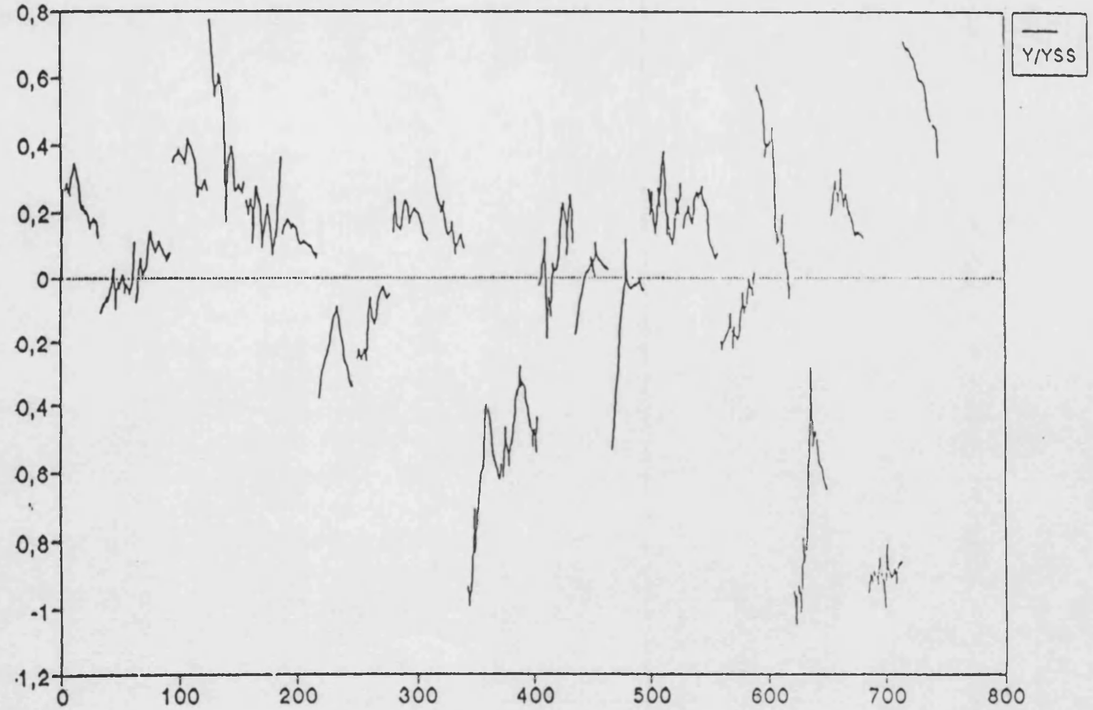
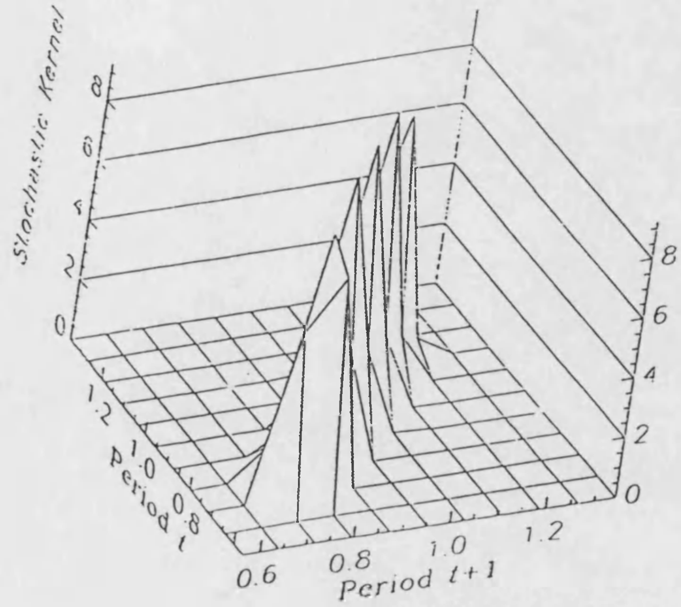
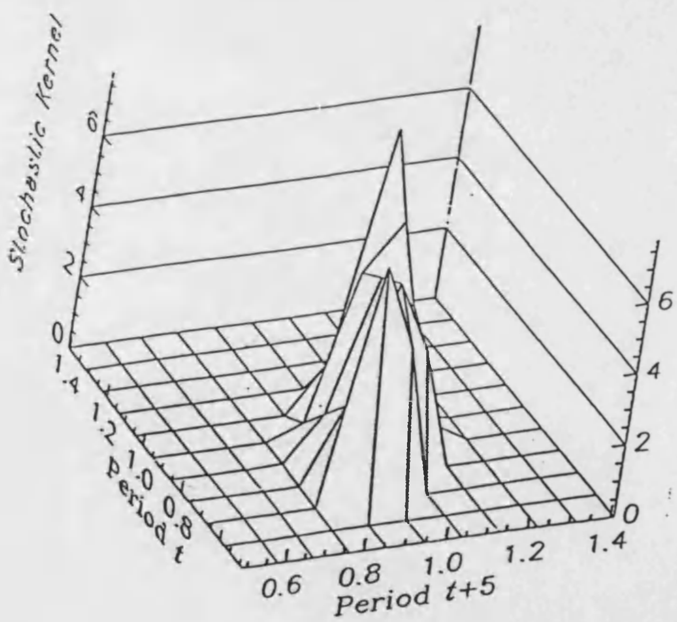


FIGURE 16a
Residuals,
Conditioning on Country-Specific and Time-Invariant Effects



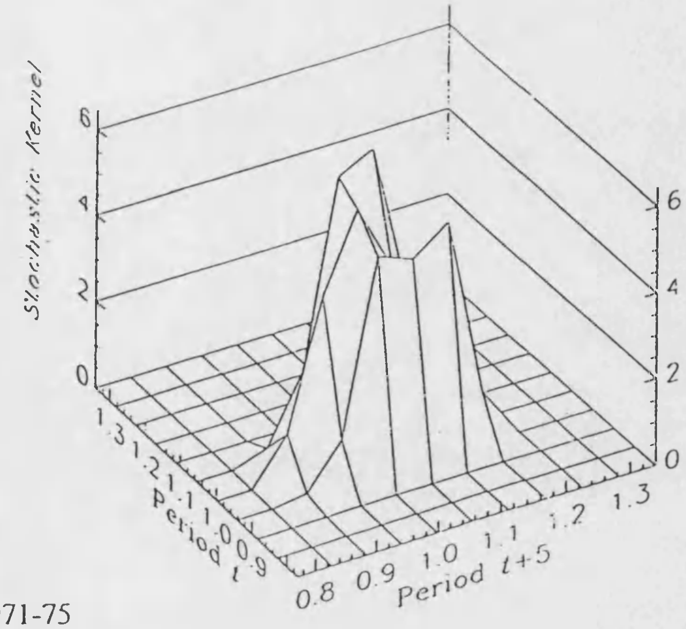
Stochastic Kernel
1 Year Transition

FIGURE 16b
Residuals,
Conditioning on Country-Specific and Time-Invariant Effects



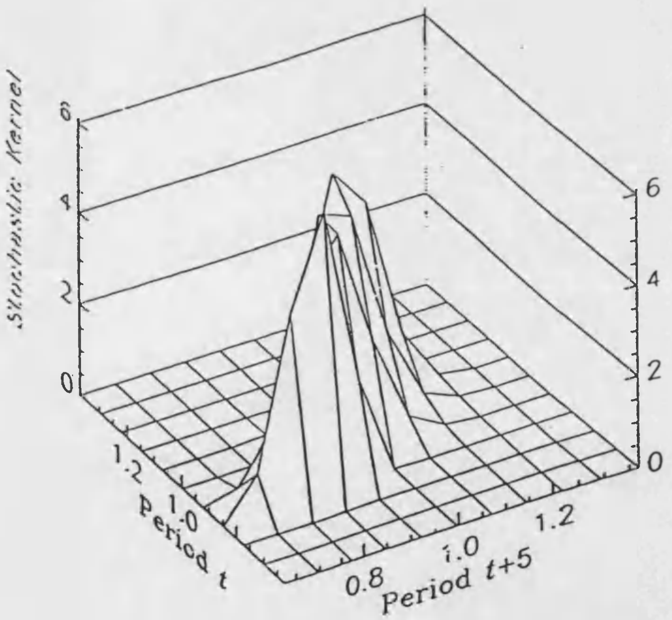
Stochastic Kernel
5 Year Transition 1961-65

FIGURE 16c
Residuals,
Conditioning on Country-Specific and Time-Invariant Effects



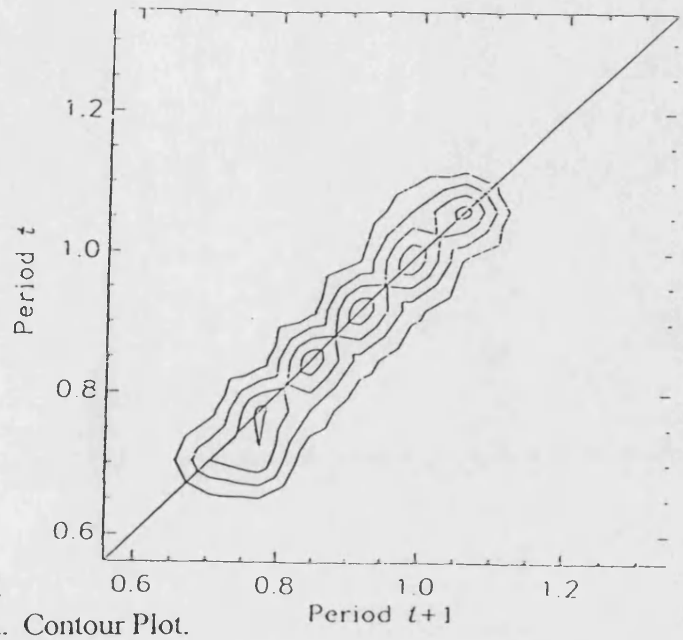
Stochastic Kernel
5 Year Transition 1971-75

FIGURE 16d
Residuals,
Conditioning on Country-Specific and Time-Invariant Effects



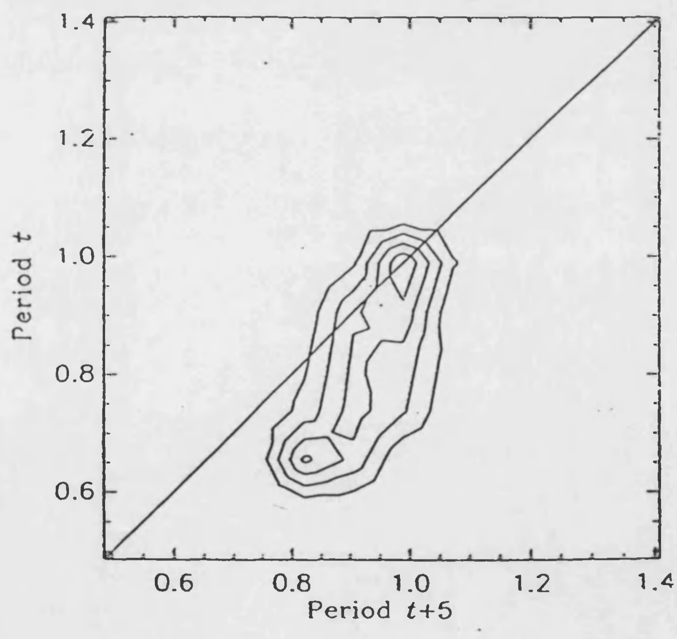
Stochastic Kernel
5 Year Transition 1981-85

FIGURE 17a
Residuals,
Conditioning on Country-Specific and Time-Invariant Effects



Stochastic Kernel
1 Year Transition. Contour Plot.

FIGURE 17b
Residuals,
Conditioning on Country-Specific and Time-Invariant Effects

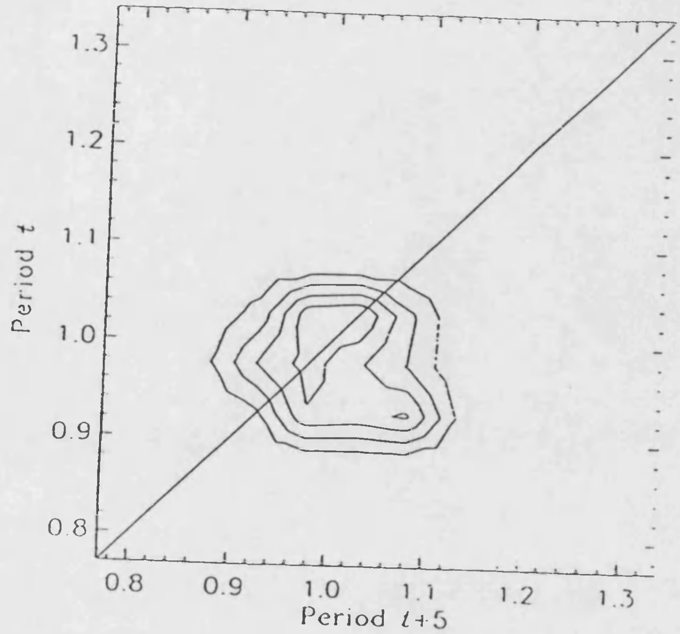


Stochastic Kernel
5 Year Transition 1961-65. Contour Plot.

FIGURE 17c

Residuals,

Conditioning on Country-Specific and Time-Invariant Effects



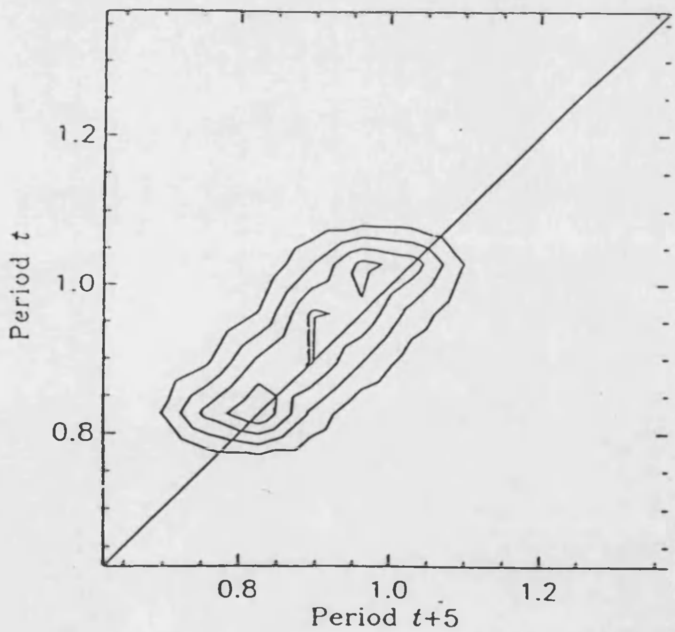
Stochastic Kernel

5 Year Transition 1971-75. Contour Plot.

FIGURE 17d

Residuals,

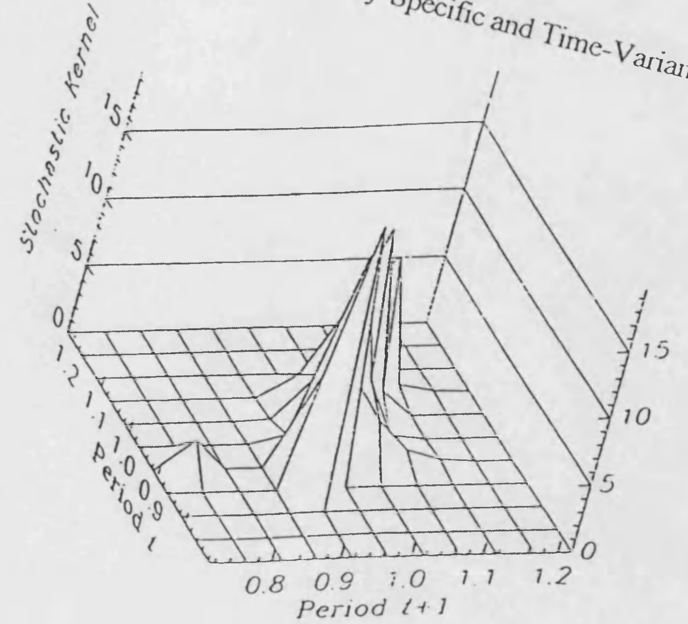
Conditioning on Country-Specific and Time-Invariant Effects



Stochastic Kernel

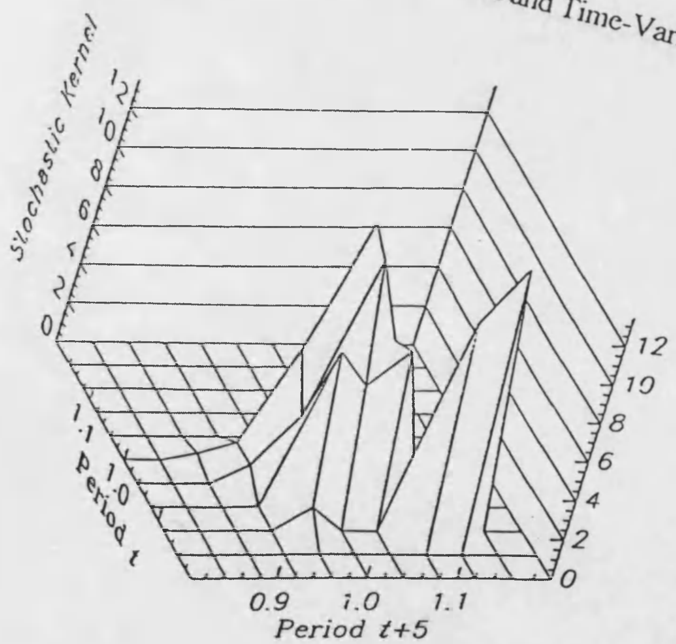
5 Year Transition 1981-85. Contour Plot.

FIGURE 18a
Residuals,
Conditioning on Country-Specific and Time-Variant Effects



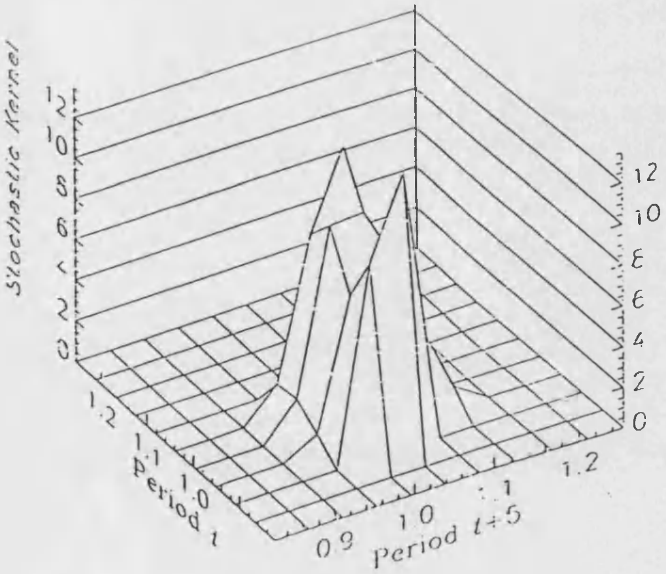
Stochastic Kernel
1 Year Transition

FIGURE 18b
Residuals,
Conditioning on Country-Specific and Time-Variant Effects



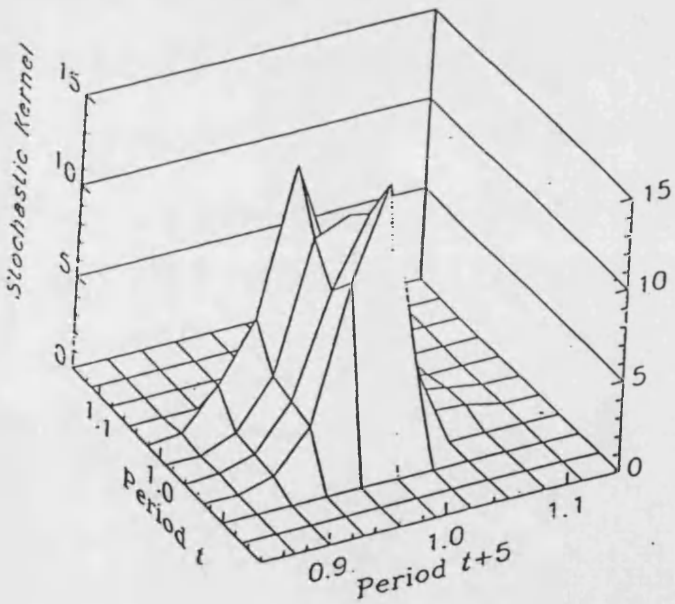
Stochastic Kernel
5 Year Transition 1961-65

FIGURE 18c
Residuals,
Conditioning on Country-Specific and Time-Variant Effects



Stochastic Kernel
5 Year Transition 1972-76

FIGURE 18d
Residuals,
Conditioning on Country-Specific and Time-Variant Effects

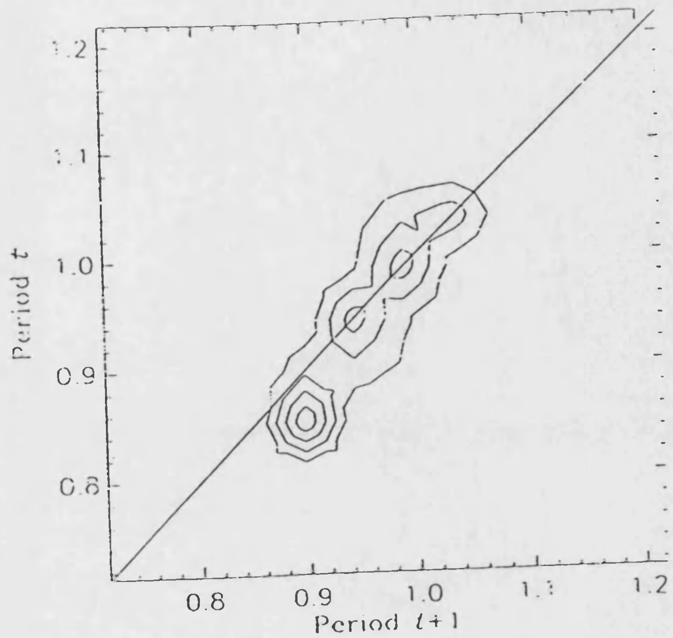


Stochastic Kernel
5 Year Transition 1981-85

FIGURE 19a

Residuals,

Conditioning on Country-Specific and Time-Variant Effects



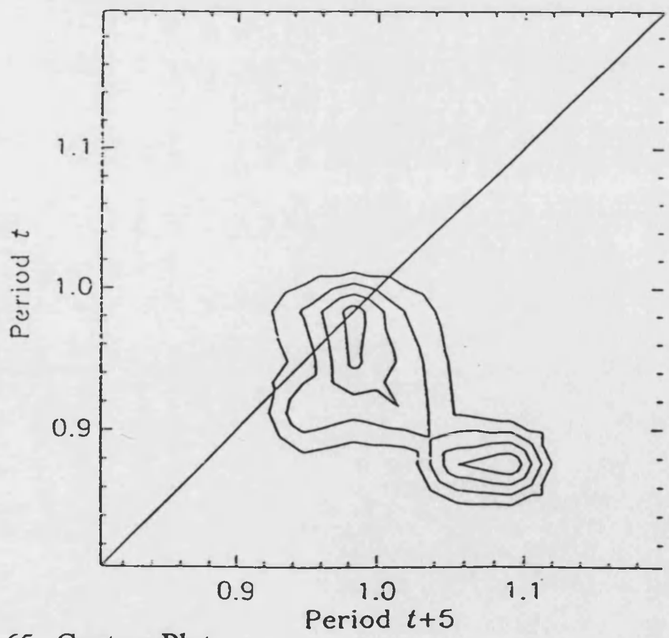
Stochastic Kernel

1 Year Transition. Contour Plot.

FIGURE 19b

Residuals,

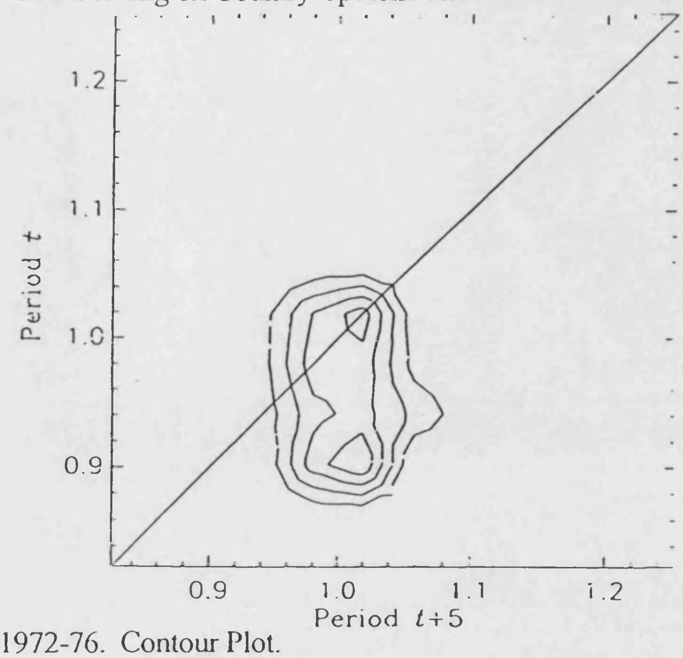
Conditioning on Country-Specific and Time-Variant Effects



Stochastic Kernel

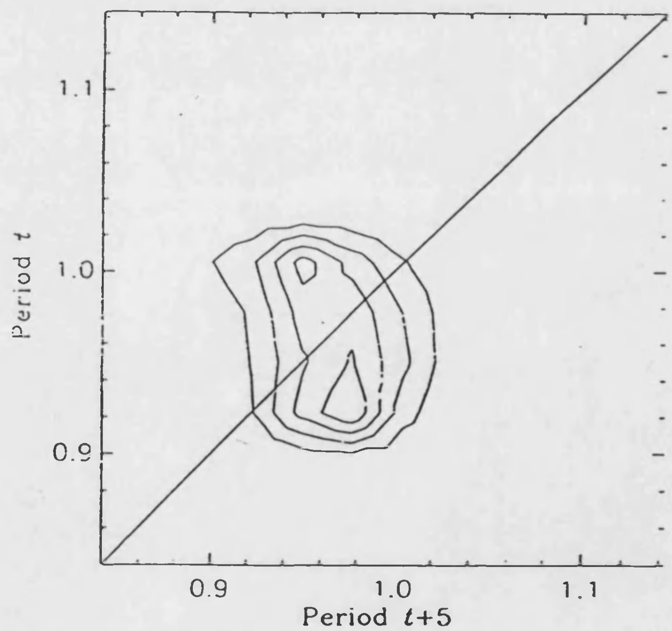
5 Year Transition 1961-65. Contour Plot.

FIGURE 19c
Residuals,
Conditioning on Country-Specific and Time-Variant Effects



Stochastic Kernel
5 Year Transition 1972-76. Contour Plot.

FIGURE 19d
Residuals,
Conditioning on Country-Specific and Time-Variant Effects



Stochastic Kernel
5 Year Transition 1981-85. Contour Plot.

REFERENCES

Andrés, J., Domenech, R. and Molinas, C., (1994), "Growth and Convergence in OECD Countries, A Closer Look", in B. van Ark and N. Crafts, (eds), *Catch Up and Convergence in Postwar Europe: Quantitative Aspects*, CEPR, Cambridge University Press.

Andrés, J. and Boscá, J., (1993), "Technological Differences and Convergence in OECD Countries", Mimeo.

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

Barro, Robert J., (1991), "Economic Growth in a Cross-Section of Countries", *Quarterly Journal of Economics,* Vol.106(2), May, pp.407-443.

Barro, R. and Sala i Martin, X., (1991), "Convergence Across States and Regions", Brookings Papers on Economic Activity, Vol.1, April, pp.107-182.

Barro, R. and Sala i Martin, X., (1992), "Convergence", *Journal of Political Economy*, Vol.100(2), April, pp.223-251.

Baumol, W.J., (1986), "Productivity Growth, Convergence and Welfare", *American Economic Review,* Vol.76(5), December, pp.1138-55.

Bernard, A.B. and Durlauf, (1991), "Convergence of National Output Movements", NBER Working Paper, Cambridge M.A., May.

Blomström, M., Lypsey, R.E. and Zejan, M., (1993), "Is Fixed Investment the Key to Economic Growth", NBER WP 4436.

Cohen, D., (1992), "Test of the Convergence Hypothesis. A Critical Note", CEPR Discussion Paper No. 691, August.

Cohen, D., (1992), "Economic Growth and the Solow Model. Some Further Empirical Results", CEPREMAP, August.

De Long, J.B. and Summers, L., (1992), "Macroeconomic Policy and Long-Run Growth", Federal Reserve Bank of Kansas City, 77.

Dolado, Juan.J, Gonzalez-Paramo, J.M. and Roldan, J.M, (1994) "Regional convergence in Spain", *Moneda y Credito.*

Durlauf, S. and Johnson, P., (1992), "Local versus Global Convergence across National Economies", NBER WP 3996.

Easterly, W., Kremer, M., Pritchett, L. and Summers, L., (1993), "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks", *Journal of Monetary Economics*, Vol.32, pp.459-483.

Easterly, W. and Rebelo, S., (1993), "Fiscal Policy and Economic Growth: An Empirical Investigation", *Journal of Monetary Economics*, Vol.32, pp.917-958.

Friedman, M., (1992), "Do Old Fallacies Ever Die?" *Journal of Political Literature*, Vol.30(4), December, pp.2129-2132.

Genewe, J, Marshall, R.C. and Zarking, G.A., (1986), "Mobility Indices in Continuous Time Markov Chains", *Econometrica,* Vol.54(6), November, pp.1407-1423.

Holtz-Eakin, D., (1992), "Solow and the States: Capital Accumulation, Productivity and Economic Growth", NBER Working Paper 4144, Cambridge, M.A.

Kelly, M., (1992), "On Endogenous Growth with Productivity Shocks", *Journal of Monetary Economics*, Vol.30, pp.47-56.

Levine, R. and Renelt, D., (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions", American Economic Review, Vol.82.

Lucas, R., (1988), "On the Mechanics of Economic Development", *Journal Of Monetary Economics,* Vol.22, July.

Knight, M., Loyza, N. and Villanueva, D., (1992), "Testing the Neoclassical Theory of Economic Growth: A Panel Data Approach", International Monetary Fund WP No.106.

Lucas, R., (1993), "Making a Miracle", *Econometrica*, Vol.61, pp.251-272.

Mankiw, N., Romer, D., and Weil, D., (1992), "A Contribution to the Empirics of Economic Growth", *Quarterly Journal of Economics*, Vol.107, pp.503-530.

Pesaran, H. and Smith, R., (1993), "Estimating Long-Run Relationships From Dynamic Heterogeneous Panels", Mimeo.

Parente, S. and Prescott, E., (1993), "Changes in the Wealth of Nations", *Quarterly Review*, Federal Reserve Bank of Minneapolis, Vol.17, pp.3-13.

Quah, D., (1992), "National Patterns of Growth: II. Persistence, Path Dependence and Sustained Take-off in Growth transitions" Economics Department Working Paper, LSE, October. (First draft July 1990.)

Quah, D., (1993a), "Empirical Cross-Section Dynamics in Economic Growth", *European Economic Review,* Vol.37(2/3), April, pp.426-434.

Quah, D., (1993b), "Galton's Fallacy and Test of Convergence Hypothesis", *The Scandinavian Journal of Economics,* Vol.95(4), December, pp.427-443.

Quah, D., (1994b), "Empirics for Economic Growth and Convergence" Economics Department Working Paper, LSE, September.

Quah, D., (1996), "Convergence Empirics across Economies with (Some) Capital Mobility", *Journal of Economic Growth.* Forthcoming

Romer, P., (1986), "Increasing Returns and Long-Run Growth", *Journal of Political Economy,* Vol.94.

Romer, P., (1989), "Capital Accumulation in the Theory of Long-Run Growth".

Shorrocks, A.F., (1978), "The Measurement of Mobility", *Econometrica,* Vol.46(5), September, pp.1013-1024.

Solow, R., (1956), "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics.*

Stokey, N., Lucas, R. and Prescott, E., (1989), *Recursive Methods in Economic Dynamics,* Harvard University Press.

CHAPTER 3

CROSS-SECTIONAL FIRM DYNAMICS: THEORY AND EMPIRICAL RESULTS FROM THE CHEMICAL SECTOR

This chapter is part of my joint research with Reinout Koopmans.

1. INTRODUCTION

Do firm characteristics converge or diverge? Do small firms grow faster than large firms, hence catching up, or do large firms have an inherent advantage in capturing new investment opportunities, dominating the industry in the long run? The classic papers by Simon and Bonini (1958) and Hart and Prais (1956) were among the first to study the dynamic nature of the firm size distribution. Major contributions have been made by Jovanovic (1982) and Cabral and Riordan (1994) on industry dynamics due to `passive' learning, by Gilbert and Harris (1984) on the evolution of industry structure in a growing (Cournot) market, by Pakes and Ericson (1987) on the strategic investment in dynamic context, known as `active' learning, and by Hopenhayen (1989) on dynamic competition between firms that face idiosyncratic cost shocks. A parallel literature on R&D and industry structure addresses similar issues although it is more specifically focused on the introduction of new products or technologies. A central question in the latter literature is which firm has the highest incentives to invest in R&D. Is it an entrant or the incumbent monopolist, the efficient or the inefficient firm? Two approaches have been used to address these issues[1]. One is an auction, as in Gilbert and Newbery (1982), Vickers (1986) and Katz and Shapiro (1987), the other a stochastic race, as in Lee and Wilde (1980) and Reinganum (1983).

The main problem in this literature is the lack of robust empirical implications. In many models, any outcome can be supported in equilibrium, depending on the details of the game or on specific parameter constellations which are empirically difficult, if not impossible, to observe[2]. This paper shows that if one specific assumption is made about the pay off structure, the indeterminacy can be resolved to some extent. To illustrate this resolution, three different games that allocate an investment project among firms will be analyzed. These are the auction and the stochastic race mentioned earlier, and one based on the `Grab the Dollar' game. The assumptions on the pay off structure are (i) that the profit of the project, if won, is identical for all candidates, (ii) the firms' existing operations are affected in the sense that winning changes the price cost margin (PCM) the firm earns on its existing capacity, and (iii) the project changes the PCM all competitors earn on their existing capacity (the market externality).

The theoretical result is that if this pay off structure is assumed, the three allocation mechanisms imply the same unique relationship between empirically observable industry characteristics and the evolutionary process of firm characteristics. In particular, it is the effect of the project on the firm's *own* PCM that determines if small firms tend to catch up with larger ones or large firms dominating the industry in the long run. If it is positive, then the large firm will generally win the project, if it is negative then the firm with the small initial capacity is in a better position to win[3]. This mechanism drives the result in a number of existing models[4]. What is pointed out here, though, is the generality of the result.

Although the results are general and not restricted to any industry in particular, the implications for homogeneous goods industries are particularly clear cut. Assuming firms competing in quantities, `projects' are opening or closing down capacity. If the market price is decreasing in the market capacity, then the implication of the theory is that the small firm is more likely to open new capacity than an already large one.

In the second part of the paper, the predictions of the model are tested .We use the population data of firms, of 24 product markets of the chemicals sector, covering the period between 1952 to 1983. The conventional way of testing this hypothesis is to estimate the probability of a firm being the next one to open a plant or capacity, which, according to the theory, should be decreasing in the firm's initial size. For illustrative purposes this is done for our dataset in Appendix 2. Although this is intuitively appealing, it can be shown that the negative sign of the initial size does not necessarily imply that differences in firm sizes tend to become smaller. We therefore use a methodology initially developed by Quah (1993a,b; 1994) to analyze convergence of *per capita* income across countries. Rather than looking at the negative coefficient of the initial condition in a cross- section growth equation or at the dynamic behaviour of the moments of the cross-section distribution, this approach analyzes the dynamics of the *entire* cross-section distribution, exploiting the time series and cross-section information more fully. *Stochastic Kernels* and transition matrices characterise the intra-distributional mobility of firms and we analyze the long term behaviour of the size distribution of firms. We find that in the period between 1952 and 1983 small firms have been more likely to increase capacity than larger ones, leading to a more fragmented industry structure. These tendencies have been particularly strong during those years in which there have been only a relatively small number of firms in the industry and they have seen strong growth.

The set up of the paper is as follows. Section 2 exposes the basic theoretical results. First a `Grab the Dollar' game is analyzed, augmenting it with the earlier assumption about the pay off structure. Then a similar procedure will be repeated to analyze the cross-sectional allocation in an auction and in a stochastic race. Section 3 describes the empirical implications of the theoretical analysis for the chemical sector and points out why the standard cross-section analysis can be misleading. In Section 4 we use a novel approach to study firm dynamics. Section 5 concludes.

2. THE MODEL

Two firms operate in a market. The large firm (1) has a historically given capacity q_1, the small one (2) has a historically given capacity q_2, where $q_1 > q_2 \geq 0$. If $q_2 = 0$, firm 2 is a (potential) entrant. Firm i earns PCM_i per unit of capacity. Assume that a profit opportunity arrives, which can be taken up by either firm. It can be opening another plant, a R&D project or an advertising campaign. We will say a firm `wins' the project if it is allocated to that firm. The other firm is the `loser'. The general features of the `project' are:

Assumption A1. It is unique in the sense that it can only be realized by one firm, though both are potential candidates.

All firms are potential candidates for winning the project implying that its arrival is independent of firm characteristics.

Assumption A2. The winner receives (pays) a net `fixed profit' (cost) equal to $\pi \in \mathbb{R}$ *which is identical for both firms.*

Assumption A3. Externalities. The project, if won, changes the *PCM* of the winning firm by $\Delta PCM \geq -PCM_i$ $i = 1,2$ (the externality) and the *PCM* of the loser by $\max\{\gamma\Delta PCM, -PCM_i\}$ $i=1,2$ (the market externality). The market externality is assumed to be smaller in absolute value than the externality: $\gamma \in [-1,1]$.

The fixed profit π, which can be interpreted as the profit an entrant would earn, is independent of the firm size, and so are the externalities $(\gamma, \Delta PCM)$. The *status quo*

PCMs might differ among the firms, i.e. $PCM_1 \neq PCM_2$. The validity of the assumption of the market externality being smaller in absolute value than the externality is an empirical issue and depends on the precise characteristics of the project. This will be discussed below.

Assumption A4. The firm that wins implements the project (`no shelving')[5].

The results presented here do not depend on intrinsic differences between the firms, only on differences in initial conditions.

If $\Delta PCM > 0$ and $\gamma > 0$, the project can be interpreted as an unique advertising project which generates positive spillovers for competitors. Alternatively, the project might be a product innovation or an improvement of existing technology, that is licensed to competitors or imitated by them[6]. The positive market externality might be due to diffusion of experience in the industry[7]. The project might be closing a plant in an industry, producing imperfect substitutes $(\gamma \in]0,1[)$ or homogeneous goods $(\gamma = 1)$.

If $\Delta PCM > 0$ and $\gamma < 0$, the project can again be interpreted as an advertising or (product or process) innovation project, but with negative spillovers. An innovation that is patented might give the holder a competitive advantage and undermine the competitive position of other firms. Even if it would be licensed, the competitors can face a lower PCM due to the royalty fees. An extreme example is where an innovation is `drastic' in the sense that the innovator can monopolize the market[8]. Then $\gamma \Delta PCM = - PCM$.

The constellation $\Delta PCM < 0$ and $\gamma > 0$ occurs for example if one firm adds capacity in a market of products that are imperfect substitutes, or homogeneous goods if $\gamma = 1$. Alternatively, it can be what Katz and Shapiro (1987) call a `major' innovation, an innovation that replaces the existing technology and production capacity based on that technology cannot be operated economically any longer. Then $\Delta PCM = - PCM$.

Finally $\Delta PCM < 0$ and $\gamma < 0$. An extreme version of this case is a firm exiting from an industry producing imperfect substitutes or homogeneous goods $(\Delta PCM = - PCM)$. If the production facilities are dismantled, the capacity reduction will increase the price cost margin of the firms that remain in the industry.

There are two examples where $\gamma \in [-1,1]$ is unlikely to hold. One is where goods are vertically differentiated and the introduction of a new variety can be located close to the competitor's variety in product space and far from the winner's own existing varieties, hence hurting the profitability of the competitor's existing varieties more than his own. The other is in the context of horizontal product differentiation. The project is opening a new plant, but there are transport costs. The plant can be located close to a competitor's existing plant, hence immunizing itself from the externality on its existing capacity. In these cases the market externality is higher in absolute value than the externality $(\gamma > 1)$.

The focus of the first part of the paper will be to determine which firm wins the project under which circumstances, i.e. for which parameter constellations $(\Delta PCM, \pi, \gamma, q_1, q_2)$ does either firm end up with the project. This will be analyzed in the context of three different selection mechanisms. First a `Grab the Dollar' game, in which firms choose whether or not to grab the project. The second is an auction in which the firms bid for the project. The third mechanism is a stochastic race, in which firms make strategic investments that increase the probability of winning.

2a. Grab the Dollar

Consider a one shot game in which a project is to be allocated (a dollar on the table). The strategy space of the two asymmetric agents is to `grab' the project (G) or to `pass' (P). The rules of the game are that if only one firm grabs that firm wins the project; if both firms bid, firm 1 will win with probability F. If neither grabs, the opportunity is lost and the PCMs are unchanged. The pay off firm i is $\pi + (PCM_i + \Delta PCM)q_i$ if it wins, and $(PCM_i + \gamma \Delta PCM)q_i$ if i looses. The pay off is $PCM_i q_i$ if none of the firms bid for the project. Grabbing itself is costless.

Firm 1 \ Firm 2	Grab $G2$	Pass $P2$
Grab $G1$	$F\pi+\big(F(1-\gamma)+\gamma\big)(PCM_1+\Delta PCM)q_1$ $(1-F)\pi+\big(1+F(\gamma-1)\big)(PCM_2+\Delta PCM)q_2$	$\pi+(PCM_1+\Delta PCM)q_1$ $\gamma(PCM_2+\Delta PCM)q_2$
Pass $P1$	$\gamma(PCM_1+\Delta PCM)q_1$ $\pi+(PCM_2+\Delta PCM)q_2$	$PCM_1 q_1$ $PCM_2 q_2$

(Expected) Pay Offs in the `Grab the Dollar' Game

A symmetric pure strategy Nash equilibrium in which both firms grab {G1, G2} exists if grabbing has a higher expected pay off than passing, given that the competitor grabs:

$$\gamma \Delta PCM q_i \leq H(\pi + \gamma \Delta PCM q_i) + (1-H)(\gamma \Delta PCM q_i) \quad i = 1,2$$

$$\Rightarrow \pi \geq (\gamma - 1)\Delta PCM\, q_i \tag{3.1}$$

where $H = F \quad if\ i = 1$
$\quad\quad H = 1-F \ if\ i = 2.$

An equilibrium in which neither firm bids {P1, P2} exists if for both firms grabbing yields a negative pay off, given that the competitor passes:

$$\pi + \Delta PCM\, q_i \leq 0 \quad i = 1,2. \tag{3.2}$$

The conditions for asymmetric equilibria in which firm i bids and j passes {Gi, Pj} are:
- for firm i it is optimal to bid, given j passes:

$$\pi + \Delta PCMq_i \geq 0 \Rightarrow \pi \geq -\Delta PCM\, q_i \tag{3.3}$$

- for firm j it is optimal to pass, given that i bids:

$$\gamma \Delta PCMq_i \geq H(\pi + \gamma \Delta PCMq_i) + (1 - H)(\gamma \Delta PCMq_i)$$

$$\Rightarrow \pi \leq (\gamma - 1)\Delta PCM\, q_j \tag{3.4}$$

where $H = F \quad if j = 2$
$H = 1 - F \; if j = 1.$

Figures 1 and 2 show the full set of pure strategy equilibria for $\Delta PCM > 0$ and $\Delta PCM < 0$ respectively. If $\Delta PCM > 0$ the large firm is more willing to incur a high cost $(\pi << 0)$ than the small firm, given γ. Hence, there exist parameter constellations for which the unique asymmetric equilibrium is {G1, P2}. More importantly, if $\Delta PCM > 0$ there are no parameter constellations for which there is a unique equilibrium in which only the small firm bids. If $\Delta PCM < 0$ then there do not exist parameter constellations for which the unique equilibrium is one in which only the large firm bids. The small firm is less affected by the negative externality and hence is willing to grab for lower levels of the level of π, given γ.

We can divide the space of outcomes into a part in which either firm or none wins in equilibrium and its complement, in which there is a unique equilibrium winner. We consider the latter first. A unique equilibrium winner exists if there is a unique asymmetric (pure strategy) equilibrium.

Proposition 1. If $\Delta PCM > 0$ and there exists a unique asymmetric pure strategy equilibrium, then the large firm wins. If $\Delta PCM < 0$ and there exists a unique asymmetric pure strategy equilibrium, then the small firm wins.[9]

The figures show that larger differences in initial firm sizes amplify this effect in the sense that the areas for which there exists a unique asymmetric pure strategy equilibrium are larger.

Proposition 1 holds if the 'no shelving' assumption (A4) is relaxed. If winning firms have the discretion not to implement the project, the optimal response of a passing firm in an asymmetric equilibrium becomes:

$$\gamma \Delta PCM q_i \geq H\, Max\{\, \pi + \Delta PCM q_i , 0\} + (\, 1 - H)(\, \gamma \Delta PCM q_i) \tag{3.5}$$

where $H = F \quad if\, j = 2$
$H = 1 - F \; if\, j = 1.$

Firm j only implements the project if the pay off is non-negative. If $\pi + \Delta PCM\, q_i \geq 0,$ then (3.5) reduces to (3.4). If $\pi + \Delta PCM\, q_i \leq 0,$ (3.5) becomes:

$$\gamma \Delta PCM < 0.$$

The consequence is that the areas defined by (3.3) and (3.4) are no longer unique asymmetric equilibria, since the passing firms are better off to grab the project and shelve it if won.

Now relax the assumption that the market externality is smaller than the externality, i.e. $\gamma \notin [-1,1]$. Proposition 1 goes through for $\Delta PCM < 0,$ but not for $\Delta PCM > 0.$ If $\gamma > 1$ and $\Delta PCM > 0,$ there exist parameter constellations for which {P1, G2} is an equilibrium, but {G1, P2} is not. This occurs if:

$$(\gamma - 1)\Delta PCM\, q_2 \leq \pi \leq (\gamma - 1)\Delta PCM\, q_1 \tag{3.6}$$

The area defined by relation (3.4) is not any longer a subset of the area defined by relation (3.3).

We now return to the original game and consider the outcomes in which either firm can win in equilibrium. These can be symmetric equilibria, multiple asymmetric equilibria or mixed strategy equilibria. Which firm is more likely to win the project in the symmetric equilibrium {G1, G2} depends on F. A weak auxiliary assumption would be that the outcome that leads to a lower industry profit is the less likely one. If firm i wins the project, the industry profit is $\Pi_i = \pi + \Delta PCM(q_i + \gamma q_j).$

Proposition 2. Assume that $F \geq \frac{1}{2}$ iff $\Pi_1 \geq \Pi_2$. In symmetric equilibria {G1, G2} the large firm is more likely to win the project if $\Delta PCM > 0$ and the small firm is more likely to win the project if $\Delta PCM < 0$.[10]

The intuition is as before. The equilibrium that generates the highest industry profits is - here by assumption - the most likely in symmetric equilibria. As long as the market externality is smaller in absolute value than the externality, the equilibrium industry profits are the highest if the large firm wins if $\Delta PCM > 0$ and if the small firm wins if $\Delta PCM < 0$.

There exist parameter constellations for which there exist multiple asymmetric pure strategy equilibria. One possible way around this indeterminacy would be to assume that with probability F the equilibrium is played in which firm 1 wins. Assuming that $F \geq \frac{1}{2}$ iff $\Pi_1 \geq \Pi_2$, this would imply that the large firm is more likely to end up with the project if $\Delta PCM > 0$, whereas the small firm is more likely to end up with the project if $\Delta PCM < 0$. Note that in the auction, that will be described in the next section, the indeterminacy is fully resolved in favour of the outcome obtained here using an `ad hoc' argument.

There is a caveat. Mixed strategy equilibria do not generate any clear cut results concerning which firm is more likely to win the project under which circumstances. The probability of firm i grabbing in a mixed strategy equilibrium if $F = 1/2$, is $p_i = \frac{2(\pi + \Delta PCM q_j)}{\pi + (\gamma+1)\Delta PCM q_j}$, which is increasing or decreasing in q_j, depending on the sign of $\Delta PCM \gamma \pi$.

2b. Selection in an Auction

We now turn to a second selection mechanism, which is essentially a (Dutch) auction. The dynamic nature of the game allows for a considerably richer way of describing the incentives for the firms to grab the project, thereby refining the intuition of the result in the last section. The project is won by the large firm if $\Delta PCM > 0$, and by the small one if $\Delta PCM < 0$. What will be shown is that, with identical discount rates, the large firm is more eager to implement it if $\Delta PCM > 0$, since his opportunity cost of waiting is higher. Secondly the opportunity cost of losing is higher for the large firm if $\gamma \in [-1,1]$, hence giving him an incentive to preempt. A similar reasoning holds *mutatis mutandis* for the small firm if $\Delta PCM < 0$.

The set up of the game follows Katz and Shapiro (1987). However, their analysis is augmented by Assumptions A2-A4.

Technically, the game is a stopping game. As before, there are two firms with historically given capacity $q_1 > q_2 \geq 0$. Decisions are made at discrete dates, at t = $0, \delta, 2\delta, \ldots,$ where $\delta \to 0$. The strategies of the firms are to `grab', given that no firm has yet grabbed the project, or to wait. Grabbing means winning the project, developing it and realizing the pay off[11]. Waiting means not grabbing at t and deciding again at $t+\delta$. The game ends as soon as one firm grabs the project. If neither of the firms grabs at any finite time, firm i earns PCM_i per unit of existing capacity, generating a continuous stream of $PCM_i q_i$. The incremental profit for the winner consists of a fixed stream of profits (cost), with initial present value equal to $\pi(t) > 0\,(< 0)$ if the project is grabbed in t. Its current value is continuously differentiable and increasing over time, though at a decreasing rate $\left(\frac{d\pi e^{rt}}{dt} > 0,\ \frac{d^2\pi e^{rt}}{dt^2} < 0\right)$ with finite limit $\lim_{T\to\infty}\pi(T)e^{rT} = \pi^{\infty}$. For example, because development costs fall at a decreasing rate over time, or opening a plant becomes more and more profitable due to growing demand. The PCM the winner earns on its existing capacity is changed by $\Delta PCM\ (> -PCM)$. If firm i wins, then, as before, there is a market externality on firm j's profits (the loser). Firm j's PCM changes by $\gamma\Delta PCM$, where $\gamma \in [-1,1]$.

Both firms have an identical discount rate r. The present value of the pay off of winning at time T is:

$$
\begin{aligned}
W_i(T) &= \pi(T) + \int_T^{\infty} e^{-rt}\,\Delta PCM\, q_i\, dt \\
&= \pi(T) + \frac{e^{-rT}}{r}\Delta PCM\, q_i \quad i = 1,2
\end{aligned}
\tag{3.7}
$$

Assume that winning the project initially is not profitable for either firm, i.e. $\pi(0) + \frac{\Delta PCM q_i}{r} < 0\ i = 1,2.$

The present value from losing at time T is:

$$
L_j(T) = \frac{e^{-rT}}{r}\gamma\Delta PCM\, q_j \quad j = 1,2 \tag{3.8}
$$

The equilibrium concept is a subgame perfect equilibrium, confining the analysis to pure strategy equilibria.

Two basic incentives determine the outcome. If firm j never grabs, then firm i's incentive to grab at any date depends upon the pay off from winning only. Following Katz and Shapiro (1987), we call the incremental profit of the winner $W_i(T)$ the `stand alone' incentive. Firm i is willing to grab at T or any time after, if $W_i(T) \geq 0$, since the actual value of winning, $W_i(T)e^{rT}$, is increasing in the grabbing date T. The optimal date to grab if j will never grab, the `stand alone date' $\hat{T}_i$ is the solution of:

$$\Delta PCM\, q_i = e^{r\hat{T}_i}\pi'(\hat{T}_i) \tag{3.9}$$

The right hand side (RHS) of (3.9) is decreasing in $\hat{T}$[12]. Therefore, if $\Delta PCM > 0$ firm 1 has an earlier stand alone date than firm 2 and *vice versa* if $\Delta PCM < 0$.

Lemma 1. If $\Delta PCM\, q_i > -r\pi^{\infty}$ for $i = 1,2$, then $\hat{T}_1 < \hat{T}_2$ if $\Delta PCM > 0$ and $\hat{T}_2 < \hat{T}_1$ if $\Delta PCM < 0$.

If $\Delta PCM > 0$ the large firm is more eager to implement the project than the small firm, since the large firm (1) is losing more by a further delay of the implementation of the project, although the firms have identical discount rates. It is this impatience that induces the large firm to implement the project earlier than the small firm would. A similar argument holds for the small firm if $\Delta PCM < 0$.

The other incentive is what Katz and Shapiro (1987) call the `incentive to preempt', which is the difference in profits from existing capacity between winning and losing, $(1 - \gamma)\Delta PCM\, q_i$. That is, firm i is willing to preempt at T or any date thereafter if for $t \geq T$, $W_i(t) \geq L_i(t), t > T$ even if it is before its stand alone date $(t < \tilde{T}_i)$. The `earliest preemption date' of firm i, $\tilde{T}_i$ is the solution of $W_i(t) = L_i(t)$:

$$-(1 - \gamma)\Delta PCM\, q_i = r\pi(\tilde{T}_i)e^{r\tilde{T}_i} \tag{3.10}$$

There exists a unique earliest preemption date if $(1 - \gamma)\Delta PCM\, q_i > -r\Pi$, since

$W_i(t)$ - $L_i(t)$ is increasing in t. Firm 1 has higher preemption incentive whenever $\Delta PCM > 0$, and because the RHS of (3.10) is increasing in t, firm 1's earliest preemption date is before firm 2's. If $\Delta PCM < 0$, by the same token, firm 2 has the higher preemption incentives and is the first to reach its earliest preemption date.

Lemma 2. If $(1 - \gamma)\Delta PCMq_i > -r\pi^{\infty}$, $i = 1,2$ then $\tilde{T}_1 < \tilde{T}_2$ if $\Delta PCM > 0$ and $\tilde{T}_2 < \tilde{T}_1$ if $\Delta PCM < 0$.

This result is driven by the prospect of the firm being worse off if it loses, than it would have been had it won the project. If $\Delta PCM > 0$ and $\gamma \in [-1,1]$ the large firm has higher preemption incentives than the small firm because the opportunity loss from not wining the project is higher for the large firm. Similarly for the small firm if $\Delta PCM < 0$.

Summarizing, if $\Delta PCM > 0$, then firm 1 has both an earlier stand alone and preemption date. Katz and Shapiro (1987) show that in this case firm 2 cannot win in equilibrium, since firm 1 would always preempt. Similarly, firm 2 will always preempt if $\Delta PCM < 0$ and firm 1 cannot win. Hence, Proposition 3 is a corollary of their result:

Proposition 3. If the selection mechanism is the above stopping game, then:

(i) the large firm is the equilibrium winner if $\Delta PCM > 0$,

(ii) the small firm is the equilibrium winner if $\Delta PCM < 0$.

Proof. The result follows from Lemmas 1 and 2, and from Katz and Shapiro's (1987) necessary condition for equilibrium no.5 (p.407).

If neither firm has a finite stand alone date (i.e. $\Delta PCM\, q_i < -r\pi^{\infty}$, $i = 1,2$), there is always an equilibrium without grabbing. If in addition both firms have a finite earliest preemption date (i.e. $(1 - \gamma)\Delta PCM\, q_i > -r\Pi \geq \Delta PCM\, q_i$, $i = 1,2$), there is a second equilibrium outcome in which the firm with the higher preemption incentives preempts before the earliest preemption date of its competitor (see Katz and Shapiro (1987) - Theorem 1b). These are the 'self defence' equilibria. Both firms would prefer not to grab, and only do so because the other one does. It can be shown that they can only occur if the market externality is negative, i.e.

$\gamma \Delta PCM < 0$[13].

Before turning to the analysis of the stochastic race, we compare our results to two related models that have been described in the literature. One is Gilbert and Newbery's (1982) model of preemptive patenting, which is essentially a second-price auction with an incumbent monopolist and an entrant bidding for a substitute product. If the monopolist wins, he remains the sole firm in the market. If the entrant wins, the market will become a duopoly, which reduces the profit of the monopolist. They find that the monopolist will win if entry results in any reduction of total profits below the joint maximizing level. In the unlikely case that the introduction of the substitute by the monopolist increases the profit margin of the existing variety, the outcome is consistent with our results. The more plausible case is, however, that the substitute will decrease the profit margin of the existing variety, though to a lesser extent than if the entrant would have introduced the variety. In our model this would mean that $\gamma > 1$, which is ruled out by assumption. This example shows, however, that by assuming $\gamma \in [-1,1]$ economically interesting cases might have been ignored. Furthermore, they show that `sleeping patents' might occur if the monopolist reduces his overall profit as a consequence of the introduction of the patent. In our model the implementation of the project was assumed.

Katz and Shapiro (1987) consider an R&D project, with a time dependent development cost. The loser faces no development costs, but earns a (different) profit flow due to imitation or licensing. The profit flows from winning and losing differ per firm. By choosing the appropriate parameter constellations all possible rankings of stand alone date and earliest preemption date can be generated. Hence, either firm winning the project can be an equilibrium outcome. Restricting the pay offs by adding assumptions A1-A4, effectively rules out the case in which firm i has greater stand alone incentives, but firm j has greater preemption incentives. The introduction of the additional assumption reduces the number of equilibrium outcomes. Moreover, the interpretation of the model as given here is more general than theirs in the sense that the fixed part of the profits can be either a cost or a profit and the externality on the profitability of existing capacity can be either positive or negative. The firms' pay off of winning is nevertheless concave in the grabbing date throughout, because it consists of either a time-increasing fixed profit with a given negative externality or a time-decreasing fixed cost with a given positive externality.

2c. Selection in a Model of a Stochastic Race

In this section we analyze the stochastic racing model of Loury (1979), Lee and Wilde (1980) and Reinganum (1983). At any point in time either firm can win. The probability density of firm i winning the project is determined by the level of its strategic investment. In the R&D literature the flow of investment is interpreted as research intensity, which determines the probability of developing the new technology at any point in time. What will be shown is that the results derived earlier hold qualitatively in this framework. If the pay off structure satisfies A1-A4, the large firm has a higher investment rate if $\Delta PCM > 0$ and the smaller firm has a higher investment rate if $\Delta PCM < 0$.

Consider an industry with two firms. As before, firm 1 has a large historically given capacity, firm 2 a small one, so $q_1 > q_2 \geq 0$. They are competing to be the first to win the project. Once one firm wins, the game ends.

The strategy of firm i is to select an investment rate z_i, that determines the probability density that firm i will win at any t. The success date of firm i is a random variable T_i, distributed according to:

$$Pr(T_i \leq t) = G_i(t) = 1 - e^{-h(z_i)t} \tag{3.11}$$

where $z_i \geq 0$ and $h(x_i)$ is the hazard rate. Assume that the hazard rate is twice continuously differentiable, with $h'(.) > 0$, $h'' < 0$, $h(0) = 0$ and $\lim_{z \to 0} h'(z) = \infty$, $\lim_{z \to \infty} h'(z) = 0$. The firm commits to a particular level at the start of the race, and pays z_i until one of the firms wins. Until the first success date the flow of profits of firm i is $PCM_i q_i - z_i$. If firm i wins, its flow changes by $\pi + \Delta PCM\, q_i + z_i$ and firm j's profit flow by $\gamma \Delta PCM\, q_j + z_j$. Let r be the common discount rate. The expected profit of firm i as a function of its own and the rival's investment rate is:

$$V_i(z_i,z_j) = \int_0^{\infty} e^{-rt} e^{-(h(z_i)+h(z_j))t}$$

$$\left[h(z_i)\left(\frac{\pi + \Delta PCM q_i}{r}\right) + h(z_j)\left(\frac{\gamma \Delta PCM\, q_i}{r}\right) - z_i \right] dt \tag{3.12}$$

$$= \left[h(z_i)\left(\frac{\pi + \Delta PCM\, q_i}{r}\right) + h(z_j)\left(\frac{\gamma \Delta PCM\, q_i}{r}\right) - z_i \right] \frac{1}{r + h(z_i) + h(z_j)}$$

The probability density of firm i winning at t is $h(z_i)e^{-(h(z_i)+h(z_j))t}$, generating a pay off of $\frac{\pi + \Delta PCM q_i}{r}$. The probability density of firm j winning is $h(z_j)e^{-(h(z_i)+h(z_j))t}$, generating a pay off for firm i of $\frac{\gamma \Delta PCM q_i}{r}$. With probability $e^{-(h(z_i)+h(z_j))t}$ neither firm has won before t and firm i pays z_i.

Proposition 4. If the selection mechanism is a stochastic race then in equilibrium

(i) the success date of the large firm stochastically dominates the grabbing date of the small firm in the sense of first order stochastic dominance if $\Delta PCM > 0$.

(ii) the success date of the small firm stochastically dominates the grabbing date of the large firm in the sense of first order stochastic dominance if $\Delta PCM < 0$.[14]

As before, if $\Delta PCM > 0$, both the stand alone and the preemption incentives are higher for the large firm. If the small firm is indifferent between winning and losing, the large firm strictly prefers to win. Hence, the large firm has always an incentive to preempt. Moreover, the opportunity cost of waiting is higher for the large firm than for the small one, hence the larger is more impatient. The reverse argument holds if $\Delta PCM < 0$.

In the auction model the incentive to preempt dominates the firm's decision as long as the stand alone incentive is non-negative. The main difference here is that preemption is stochastic. Consequently, the impatience of firms to implement the project also becomes relevant for the outcome. But since the larger firm has uniformly higher incentives to invest if $\Delta PCM > 0$ it has a higher probability of winning. By assuming this particular pay off structure, both incentives are always aligned.

This is of interest since the Gilbert and Newbery (1982) deterministic auction model and the Reinganum (1983) stochastic race model give opposite outcomes. In the former an incumbent firm preempts, whereas in the latter the potential entrant has a higher R&D effort. This can only occur if one firm has higher preemption incentives and the other has higher stand alone incentives.

3. APPLICATION TO THE CHEMICAL SECTOR

In this section we will test the theoretical implications that were derived in the last section, using data from the chemicals sector. The principal reason for taking this sector is that investment projects can be unequivocally defined as opening or closing a plant or production capacity within a plant. The industries in the dataset (see Table 1) are typically bulk chemicals, relative homogeneous by nature. Most are intermediate or final petrochemical products. The advantage that homogeneous products have over differentiated products in testing the theory is, that the effect on the price cost margin can be determined under very weak conditions. If the market price is decreasing in the total market capacity, then opening capacity will decrease the market price and *vice versa*. Hence, in terms of the terminology of the last section, it must be that $\Delta PCM < 0$ if capacity is increased and $\Delta PCM > 0$ if capacity is reduced, with $\gamma > 0$ throughout. In an industry with differentiated products a typical 'project' might be a combination of increased advertising and increased production. If the former increases the price cost margin and the latter decreases it, the net effect is typically indeterminate, and hence the empirical implication of the theory unclear.

However, within the chemical industry capacity is certainly not the only strategic choice firms face. R&D programmes are essential in the strategic interaction among firms (Quintella, 1993). Research in manufacturing technology has resulted in less expensive raw material, such as in the production Acrylonitrile and Vinyl Acetate. In Phenol, a more efficient process based on Cumene Hydroxide has been developed. In some cases the feedstock has changed. For example, Phtalic Anhydride used to be produced from Naphtalene, which then changed to ortho-Xylene. Research has focused on increases in size of existing plants, primarily by de-bottlenecking, and on 'scaling up' of entire production processes. For example, in the early 1950s the largest Ethylene plants had a capacity of about 100 million pounds per year. In the 1970s the newly constructed Ethylene plants produced well over 1 billion pounds per year (Spitz, 1988, Ch.11). Similar developments have occurred for Ammonia, Vinyl Chloride, Styrene and Methanol. Hence, the `project' of opening a plant or increasing capacity changed over time. This allows for the possibility that firms, depending on their R&D programme, faced different sets of opportunities in terms of opening additional capacity. In terms of the model the `fixed pay off' π might be firm specific, due to firms using different technologies. However, most production technologies are non-proprietary, particular in petrochemicals. Anecdotal evidence suggests that the diffusion of new production processes is quick, which is according to Spitz (1988) due to engineering contractors learning how to build the scaled-up plants or how to apply new production techniques, which makes the technology available to whoever is willing to pay for it (see p.424). Any firm that opened a new plant, was it an entrant, a small incumbent or a large incumbent, seemed to use the state of the art technology at any time, suggesting that new production technologies were widely available. Mansfield (1985) found that in the petroleum sector, 60% of the process technology was available to competitors within 18 months of a firm's decision to develop a major new process. This effect was even more pronounced in primary metals, though less in other chemicals. Spitz (1988) described the effect of new technologies on petrochemical industries as follows (p.393):

> *"If the new route represented a substantial economic improvement, but was not judged to be able to provide a dominant position, the company making the invention usually embarked on a licensing program, settling for the income provided by royalties and catalyst sales, as well as the presumed benefits of becoming a reasonably low cost producer. In other cases the company could not establish a controlling position, because it could not obtain broad patent protection to keep competitors from developing relatively similar process routes. In still other cases, such as Badger-Sherwin Williams' fluid-bed Phtalic Anhydride process, the new technology was not so much better that it forced a wave of shutdowns. Here, the new technology just added one or two new competitors and upgraded the economics of some of the existing producers, who switched to the new process."*

Well known exceptions include Du Pont in Titanium Dioxide (Ghemawhat, 1984), where it achieved a dominant position through its proprietary Chloride technology and its position in nylon, which it achieved through selective licensing. Sohio achieved a dominant position in Acrylonitrile, based on a revolutionary propylene technology (Stobough, 1987).

Related to this is that π or ΔPCM might be firm specific due to `increasing returns to scale'. Although it is well known that there are increasing returns in the chemical sector, they seem to occur primarily on plant level rather than on firm level. Both Spitz (1988) and Stobough (1987) show significant gains from increasing plant sizes in terms of reducing per unit production costs, but firm sizes do only seem to play a role in the availability of capital to finance large scale production facilities. Learning is frequently mentioned as an alternative source of increasing returns. However, in petrochemicals these gains are not firm specific according to Spitz (1988, Ch.10) and Stobough (1987, Ch.5), since again rapid diffusion of experience throughout the industry undermines any competitive advantage. Although there is evidence of significant `industry wide' learning (Lieberman, 1984), individual companies do not seem to be able to maintain an advantage through more experience.

We therefore claim that it is a reasonable first approximation to assume that in the capacity game all firms face an equal investment opportunity (`project'), and hence that firms only differ in their existing capacity.

Another assumption that has to be satisfied for the theoretical results to hold is that the market externality is smaller in absolute value than the externality on the winning firm's existing capacity $(-1 < \gamma < 1)$. Spitz (1988, p.540) describes how prices are often cut as new capacity comes on stream, due to firms giving discounts in order to fill new capacity. The effect on competitors is likely to be less than the full discount due to transportation costs, which are significant even though most of the US petrochemical capacity is located in the Gulf Coast region (Chapman, 1991, Ch.6). However, location is still relatively dispersed, due to the dependence of the US petrochemical industries on natural gas liquids as feedstock, rather than oil based raw materials as in Europe. An extensive network of pipelines gives firms considerable freedom in their locational choice without giving up nearness to raw material sources. Oil based feedstock would instead require the location close to refinery complexes and hence a higher degree of geographical concentration.

A priori it is not clear which game is the most appropriate description of the cross-sectional allocation of `projects'. However, all three that were described in the last section have a qualitatively identical empirical implication. Hence, the ***Empirical Hypothesis*** can be formulated independently of the allocation mechanism.

Empirical Hypothesis: In growing and declining chemical industries, firm sizes tend to converge.

In growing industries there is a sequence of arrivals of new investment projects, each of them being the opening of another plant. This decreases the PCM earned on existing capacity by the winning firm, and by Proposition 1-4 it follows that the initially small firm is more likely to win the project. Closing down a plant reduces capacity, increasing the PCM the winning firm earns on its remaining capacity. By Propositions 1-4 the firm that is larger *ex-post* is more likely to implement the project[15].

3a. The Data

The dataset used in this section is Gilbert and Lieberman's (1987) sample of 24 growing chemical industries, see Table 1. The demand for all products increased from the earliest observation until at least 1975. The sample includes industries with more than three but less than twenty competitors, all producing an homogeneous good. The sampling period starts, depending on the industry, between 1953 and 1965 and ends in 1983. The capacity data are from annual issues of the Directory of Chemical Producers (SRI International), reporting firm and plant capacities by product[16].

3b. The Standard Cross-Section Analysis of Convergence

The ***Empirical Hypothesis*** is tested for growing industries, using two different approaches. The first one, presented in Appendix 1, is a conventional logit analysis of the probability of firm i opening a plant. We find indeed a negative coefficient for the initial condition, though the result is quite unstable. It sounds intuitively appealing to interpret this negative sign as firms within one industry converging towards a common size. In this section we will show that this interpretation is not generally true and can be misleading in many cases. We then perform what we think is a more natural way of analyzing convergence.

A negative relation between the firm is probability of opening another plant and its initial size can be consistent with a growing dispersion in the cross-section distribution of firm sizes. This is due to Galton's Fallacy or Regression to the Mean[17]. The proof of the Fallacy for discrete choice models is in Appendix 2, showing that a negative cross-section coefficient for the initial level is consistent with absence of convergence.

Consequently, it has been argued[18] that the standard deviation of the cross-section distribution should also be considered, suggesting that both a negative coefficient of the initial condition and a decreasing cross-section dispersion over time would be sufficient to show convergence. But those are single statistics (mean and standard deviation) that summarize the information in the cross-section distribution, and there are cases in which they say little about the distribution dynamics, in particular catching-up. The cross-section distribution dynamics involves changes in the shape of the distribution and intra-distribution mobility, which can only be imperfectly captured by points in time statistics.

4. THE EMPIRICAL ANALYSIS OF CROSS-SECTION DYNAMICS

A more natural way of dealing with convergence is therefore to consider the dynamic behaviour and the cross-section variation of the entire size distribution. For that, it is necessary to develop an alternative econometric strategy, suggested by Quah (1993a,b; 1994), which deals with the dynamics and cross-section dimensions, based on what in probability theory is called *Random Fields*. These are data structures that have variation of the same order of magnitude in both dimensions. At each point in time there is a cross-section distribution of firm sizes, which is simply the realization of a random element in the space of distributions. The idea is to describe their evolution over time, which will allow us to analyze intra-distribution mobility, persistence of the firms' relative position, and to characterise the long run behaviour[19]. In this framework convergence is understood as the sequence of distributions tending towards a mass-point in the long run.

4a. The Variable of Analysis

The central conclusion of the theory in Section 2 is that small firms are more likely to install new capacity than their larger competitors. A somewhat stronger implication is that firms within one industry will ultimately reach the same size, i.e.the industry average. The difference between the size of a firm and the industry average should shrink over time, and possibly go to zero in the long run.

The theory suggests as the basic variables of analysis the firm's market share minus the industry average market share: ADMSH = MSH - 1/N, if there are N firms in the industry. Alternatively, firm sizes can be measured in terms of the number of plants a firm operates, in which case the variable of analysis is the firm's relative number of plants minus the average market share: ADPLS = PLSH - 1/N. The dynamics of the cross-section distribution of these variables is informative about convergence, since both go to zero as the firm sizes go to the industry average. The normalisation guarantees that this is independent of the actual number of firms in the market, consequently allowing to pool the industries.

In the literature a widely used variable to measure convergence has been the variable relative to the mean[20]. In our terms that would be the firm's market share relative to the industry average share (NMSH = MSH $\div$ 1/N) and the firm's relative number of plants normalised by the average market share (NPLS = PLSH $\div$ 1/N)[21]. For these variables convergence is understood as the sequence of their cross-section distribution tending towards a mass point at unity. The normalization is a way to control for overall growth and aggregate fluctuations of the industry, heterokedasticity, and again, it allows pooling of industries. In our case it can be shown that convergence in terms of NMSH and NPLS implies convergence in terms of ADMSH and ADPLS, since the normalization variable (1/N) goes to zero as the market becomes more fragmented[22]. We confine the analysis here to the former set of variables[23].

There is an important issue of potential entry. We have no indication for how long firms have been around, waiting in the wings, before they enter. We make alternative assumptions to test the robustness of the empirical results. We assume for example that an entrant has been around for seven years before opening its first plant. Given a construction lag of 2 years, we include the firm in the sample as a potential entrant (with 0 plants) for five years, obviously taking into account the starting year of the sample for that industry[24]. These are NMSH5 and NPLS5. Similarly, though assuming that entrants could have opened a plant for two years only, are NMSH2 or NPLS2.

The empirical analysis has been performed for both NMSH and NPLS under the alternative entry assumption. Figures 3a-b, show three dimensional plots of NPLS2. It is clear that in all cases both dynamic and cross-section variations are large, illustrating the importance of studying both dimensions if analyzing firm dynamics.

4b. Cross-Section Distribution of the Variable

In the context of random fields the realization of the random element is a cross-section distribution function that can be estimated from the data. Figures 4a-d present the cross-section density functions of the plant share for each period of 3 or 4 years. They have been estimated by non-parametric methods[25] for the available sample. No assumption has been made about the shape nor about the moments of the density function from which the data were drawn[26].

During the 32 years of the sample there is a tendency of NPLS to concentrate around the average industry size. However, there are two limitations of the distribution functions in this context. One is that convergence is generally a limit concept and the cross-section distributions are points in time estimates, available only for 1952-83. Hence, we cannot say anything about the long run behaviour of the size distributions. Further, the graphs do not give any information about the firm's relative situation and its movement over time. To deal with these limitations, it is necessary to derive a law of motion for the cross-section distribution in a more formal structure.

4c. Modelling Dynamics of the Cross-Section Distribution

Let λ_t be the probability measure (one for each year) associated with the cross-section distribution. The simplest probability model that can describe its dynamic behaviour is:

$$\lambda_t = T^*(\lambda_{t-1}, u_t) \tag{3.13}$$

T^* maps the probability measures and a disturbance into another probability measure. T^* encodes information on how the firms move over time relative to each other. By ignoring the disturbance and iterating, (3.13) can be written as:

$$\lambda_{t+s} = (T^*)^s \lambda_t \tag{3.14}$$

As s goes to infinity, the long run (ergodic) distribution of firm sizes can be characterised.

The stochastic difference equation in expression (3.13) is unmanageable, but so is (3.14). Given the impossibility of analytic solutions for T^*, we will assume T^* is being generated by the following differential equation:

$$\lambda_{t+1} = \int M(x, A) d\,\lambda_t(x) \qquad (3.15)$$

For any probability measure λ on the measurable space (R, *R*), R is the real line and *R* is the Borel sigma algebra, $\forall$ A in R. M is a *Stochastic Kernel*[27], that is, M(x,A) is the probability that the next state period lies in A given that in this period the state is x. T^* is an operator associated with the *Stochastic Kernel* that maps the space of probabilities into itself, and $\lambda_{t+1}(A) = (T^*\lambda_t)A$.

Equation (3.15) measures the probability that the next period's state lies in set A, if the current state is drawn according to the probability measure λ_t. And $(T^*\lambda_t)$ is the probability measure over the next period's state, if λ_t is the probability measure over the current period.

The *Stochastic Kernel* allows us to analyze the intra-distribution movements of firms, solving one of the limitations pointed out, but leaves the problem of the analysis of the long run behaviour unresolved, because the *Stochastic Kernel* is infinite dimensional. We can, however, simplify the problem by approximating T^* assuming a countable state space for firms sizes $S = \{s_1\ s_2 .. s_r\}$. In that case T^* is simply a transition probability matrix Q, which makes the difference equation (3.13) tractable.

$$\lambda_t = Q(\lambda_{t-1}, u_t) \qquad (3.16)$$

Q encodes the relevant information about mobility within the cross-section distribution. But the ergodic distribution of (3.16) can be calculated explicitly. Under some regularity conditions the sequence of powers of matrix Q converges to a matrix which has identical rows describing the ergodic cross-section distribution. This allows us to analyze the long run behaviour of the size distribution.

4d. Estimation of the *Stochastic Kernel*

Figures 5a-d, 6a-d and 7a-d are three dimensional plots of some *Stochastic Kernels* for NPLS5, NPLS2 and NMSH2, estimated non-parametrically[28]. They describe the transitions from one state to any other in 1 and 5 years respectively. Figures 5e-h, 6e-h and 7e-h present the contours of the functions in 5a-d, 6a-d and 7a-d respectively.

A slice orthogonal to the plane (t,t+k) and parallel to the t+k axis, represents the probability density that describes the transitions from one part of distribution to another in k periods. The probability mass concentrated along the positive sloped diagonal, indicates a high persistence in a firms' relative position. A concentration of probability mass along the negatively sloped diagonal implies that firms overtake each other in size rank. The transition probability describing horizontal lines (parallel to t+k) indicates that there is very low persistence, the probability of being at any point in t+k is independent of the position in t. Finally, the mass of probabilities located along a vertical line in size 1 (the industry average) implies convergence in the sense that small firms grow faster than large ones.

The theoretical results are consistent with the probability mass being both along the negative diagonal, implying an `action - reaction' pattern of opening a plant by alternating firms and along a vertical line around 1, in which case there is convergence of sizes in a stricter sense.

The graphs show persistence year by year, indicating that the firms remain in their relative position. Particularly for firms in size class zero, which represents the potential entry state. Not surprisingly, this effect is more pronounced for NPLS5 than for NPLS2, NMSH2.

The results are much more striking for the larger (5 years) horizon, where the probability of transition is no longer clustered along the positive diagonal but along the vertical line in 1, indicating convergence to the industry average. After 5 years, firms that were potential entrants initially, will have a positive share of the market. A large probability mass is concentrated under the positive diagonal at zero in period t, indicating that entrants in period t reach the average size (1) in period t+5.

The contours show that in the first decade (between 1955 and 1965) the tendency of the firm sizes to converge is the strongest. The estimated kernels are consistently steeper than later estimates, indicating more persistence in subsequent decades. It is worth noting that this corresponds to the decade in which the industries saw their largest market growth and the fewest number of firms in the industry[29]. This is encouraging for the theory since two essential features are that (i) capacity in the market is increasing and (ii) there is a negative effect of additional capacity on the market price, that decreases as markets become more competitive.

Note that these transition kernels are simply point in time estimates, describing what actually happened over the sample period. They are not fitted models. Hence, we cannot derive a law of motion, or make any inferences about the long run dynamic behaviour.

4e. Estimation of the Transition Matrix Q

Q is analogous to the *Stochastic Kernel* but in a discrete space such that inferences can be made and its long run behaviour described. Divide the space of possible values of the firm sizes into *r* states. For example, firms that have a plant share of 0.2 times the industry average to 0.6 times the industry average, are in state i = (0.2, 0.6). This defines a grid that can be thought of as an estimator of the initial unconditional probability distribution λ_{t-k}. Each element of the matrix indicates the probability of transition from one state to another: the entry (i,j) is the probability that a firm in state i moves to the state j in t periods. Hence, every row is a conditional probability vector, the discrete analogy of the distribution of the transitions in the figures above (*Stochastic Kernels*), when cutting the figure at a point by a plane parallel to t+k axis.

Tables 2, 3 and 4 present some estimators of the transition matrix Q. The grid divides the initial year total observed sample into approximately equal categories, i.e.a uniform initial distribution results by construction. Consequently the length of the defined states varies. Note that they are very narrow around the mean.

The first column in the tables is the total number of transitions over the whole time sample, starting at each state. An estimator of the time invariant transition probability matrix Q is presented in the remaining columns, for a single period and for differing number of states *r*. Q is calculated as a time average over the total sample.

Most of the entries of the matrices are different from zero implying that a transition to almost any state in the distribution can occur within one year. Hence, there is substantial mobility. The conclusions from the estimation of the *Stochastic Kernels* go through unchanged.

The last row in Tables 2, 3 and 4 show estimates of the ergodic distribution. Independently of the initial position of a firm, the ergodic distribution gives the probability of that firm being in a particular state. Recall the states were defined in such a way that the initial distribution is uniform. Though the ergodic distributions are not degenerate at 1, they are unimodal with a peak around this value. Whatever of the position of a firm in the initial uniform distribution, the probability of ending close to the average firm size is higher than the probability of ending up anywhere else.

Table 5 shows an estimators of Q for the subsample 1955 to 1965. Comparison of the ergodic distribution for this subsample with the one estimated using the full sample confirms the earlier finding that the tendencies to converge are strongest during the years of strong industry growth in industries with a small number of firms[30].

5. CONCLUSION

This paper theoretically shows that in the chemicals sector there is a tendency for firms sizes to converge. The appendix 2 of the paper demonstrates that the conventional empirical approach to test for convergence using discrete choice models is inappropriate in this context.

Analysing dynamically evolving cross section distributions the paper provides evidence of that tendency of firms sizes to converge for the USA chemical sector.

ENDNOTES

1. See Reinganum (1989) for a more complete description.

2. Although Budd, Harris and Vickers (1992) find that competition tends to evolve in the direction where joint profits are higher, the implication for `catching up' versus `increasing dominance' depends on the exact nature of the pay offs. For example in Vickers (1986), the outcome is reversed if price competition is Bertrand rather than Cournot. Reinganum (1983) shows how Gilbert and Newbery's (1982) results that a monopolist will spend more on R&D than an entrant is reversed if the very same question is analyzed in a stochastic race, rather than a bidding game.

3. In this paper a firm is defined by its presence in a particular market, its size by the capacity it operates in that market. Inter-industry ownership structures and issues relating to the boundaries of the firm are not addressed.

4. Gilbert and Harris (1984), for example, find in a dynamic Cournot-Nash oligopoly with increasing demand and indivisibilities in installing new capacity, that market forces tend to push the industry towards equal market shares as smaller firms invest to catch up with larger ones. This is consistent with our results, since installing additional capacity implies a decreasing PCM for existing capacity. Farrell and Saloner (1988) and Beggs and Klemperer (1992) find both that in the presence of consumer switching costs and new customers arriving in each period, market shares converge to equality, since the larger firm gains relatively more from charging a high price to exploit its current customer base than charging a low price to gain a greater share of the new customers. The smaller firm can attract the new customers at a lower cost, since the loss on its existing customer base from charging a lower price to attract new customers is smaller. These models solve the full dynamic game with a stream of profit opportunities arriving over time. In this paper the analysis is confined to a single project. Katz and Shapiro (1987) find in a dynamic model of R&D that the leader tends to develop an innovation that, if licensed or imitated, reduces cost by an equal amount for all firms (add-on technology).

5. This is primarily relevant for R&D projects, in particular those that can be patented. In the EU national laws typically provide for compulsory licenses in two situations. The first is where an invention has not been worked within the country to the extent to meet national demands. The second is where the working of a subsequent invention is prevented by the prior patent.

6. Foster (1985) describes a strategy "often used in the chemical industry. [BASF] developed a catalyst, which it then improved. The first generation of the catalyst went to its licensees, the second and improved generation into its own plants" (p.119). Katz and Shapiro (1987) quote a number of studies in which competitors imitate innovations at a lower cost than the innovator faced. Although the licensee might have to pay a fee, or imitation involves some R&D expenses, it is generally recognised that this is less than the development cost, and can be normalized to zero within the framework of this model.

7. Mansfield (1985) finds evidence of a high rate of information diffusion in several industries.

8. See Arrow (1962).

9 For the Proof. see Koopmans and Lamo (1995)

10 For the Proof. see Koopmans and Lamo (1995)

11. If both firms grab at the same time, either firm will win with equal probability.

12. This is the case because $d^2\pi e^{rt}/dt^2 = d\pi'(t)e^{rt}/dt + rd\pi e^{rt}/dt < 0,$ and since $d\pi e^{rt}/dt > 0,$ it must be the case that $d\pi'(t)e^{rt}/dt < 0.$ The SOC $\pi''(t) + re^{-rt}\Delta PCM < 0$ is satisfied for $t = \tilde{T},$ since by (3.9) it is equivalent to $\pi''(\tilde{T}) + r\pi'(\tilde{T}) < 0.$ This is implied by $d\pi' e^{rt}/dt < 0.$

13. The two conditions are: (i) $\Delta PCM\, q_i \leq -r\pi^{\infty}$ $i = 1,2$ (neither has a finite $\ddot{T}_i$) and (ii) $-r\pi^{\infty} < (1-\gamma)\Delta PCM\, q_i$ $i = 1,2$ (both have a finite $\tilde{T}_i$). If $\Delta PCM > 0,$ (i) and (ii) can only be satisfied simultaneously if $\gamma < 0,$ and if $\Delta PCM < 0$ they can only both be satisfied if $\gamma > 0.$

14 For the Proof. see Koopmans and Lamo (1995)

15. Ghemawat and Nalebuff (1990) find a qualitatively similar result. The mechanism is, however, somewhat different. In a declining industry the optimal size of firms declines over time. Large firms have reached this optimal size and reduce their capacity accordingly. However, their smaller competitors have a suboptimal size and only reduce their capacity once the optimal size is smaller than their actual size.

16. See Gilbert and Lieberman (1987) for a more detailed description of the dataset.

17. See Quah (1993b), Friedman (1992), Huigen *et al* (1991), Hall (1987) and Leonard (1986) Strictly speaking this Fallacy refers to growth equations, with the initial size as one of the explanatory variables.

18. In particular by some authors in growth theory, see for example Barro and Sala-i-Martin (1992).

19. All the calculations and graphics in this section have been made using Danny Quah's Time-Series Random-Field shell tSrF.

20. For example, in growth theory the variable of interest is the per-capita income or output per worker of each individual economy relative to the same variable for the total economy as a whole. This variable has been used in Desdoigts (1994), Quah (1993; 1994).

21. These measures are equivalent to the absolute firm sizes (in terms of capacity or plants resp) relative to the corresponding average firm size in the industry.

22. This result holds if the series is bounded.

23. The empirical analysis has been performed for both sets of variables, the results are consistent.

24. i.e.if the period between the beginning of the sample for that industry (see Table 1) and the opening of the first plant is shorter than five years, accordingly fewer zeros are included.

25. See Silverman (1986), Section 2.10.

26. These graphs show clearly the limitations of describing density functions by their first and second moment.

27. See Stokey and Lucas (1989).

28. They are obtained using the squared of standard Epanechnikov kernel for estimating the joint density and then re-scaling to obtain the conditional probability (tSrF).

29. Between 1955 and 1965 the average of annual industry growth rates of market capacity was 14.6%, the average number of firms 6.8. Between 1965 and 1975 the former was 9.3%, the latter rose to 8.01. Between 1975 and 1983 the growth rate dropped to 4.4%, whereas the average number of firms was 7.7.

30. We also estimated kernels and transition matrices for a subsample of industries, leaving out those industries for which there was a proprietary production technology (i.e.Acrylonitrile, Caprolactam and Titanium Dioxide). Qualitatively, the conclusions remained unchanged.

TABLE 1

Product	First Observation	Total Plant Openings	Max Number of Firms
Acrylic Fibers	1953	4	6
Acrylonitrile*	1956	6	6
Aluminum	1956	16	13
Aniline	1961	5	6
Bisphenol A	1959	4	5
Caprolactam	1962	3	4
Ethylene Glycol*	1960	11	14
Formaldehyde*	1962	34	18
Isopropyl Alcohol*	1964	2	4
Maleic Anhydride	1958	9	8
Methanol	1957	11	12
Pentaerythritol*	1952	6	7
Phenol	1959	8	12
Phthalic Anhydride	1955	15	12
Polyethylene-LD	1957	15	15
Polyethylene-HD	1957	20	14
Sodium Chlorate	1956	15	10
Sodium Hydrosulphide	1964	4	6
Sorbitol*	1955	3	5
Styrene	1958	12	13
Titanium Dioxide	1964	7	6
1,1,1-Trichloroethane*	1966	2	4
Vinyl Acetate*	1960	8	7
Vinyl Chloride*	1962	12	13
		233	

* The industries that were included in the specification of column (7), Table A1b.

TABLE 2
PLSH2 First Order Transition Matrix

Time-Stationary

Upper End of the States (r)	0.625 (1)	0.786 (2)	0.857 (3)	1.000 (4)	1.429 (5)	4.103 (6)
916:	0.78	0.10	0.05	0.05	0.02	0.00
963:	0.07	0.71	0.13	0.05	0.02	0.01
1032:	0.01	0.17	0.71	0.09	0.01	0.01
1206:	0.00	0.03	0.11	0.78	0.05	0.02
478:	0.01	0.01	0.02	0.17	0.67	0.12
869:	0.00	0.01	0.02	0.03	0.06	0.88
Ergodic Distribution	0.086	0.188	0.209	0.243	0.089	0.184

TABLE 3
PLSH5 First Order Transition Matrix

Time-Stationary

Upper End of the States (r)	0.333 (1)	0.800 (2)	0.917 (3)	1.000 (4)	1.500 (5)	5.050 (6)
981:	0.76	0.08	0.06	0.05	0.03	0.02
979:	0.06	0.77	0.11	0.03	0.02	0.00
989:	0.03	0.15	0.64	0.15	0.02	0.01
986:	0.03	0.01	0.20	0.61	0.13	0.02
899:	0.02	0.01	0.02	0.19	0.69	0.07
935:	0.00	0.00	0.01	0.02	0.08	0.88
Ergodic Distribution	0.111	0.179	0.195	0.190	0.158	0.167

TABLE 4
NMSH First Order Transition Matrix

Time-Stationary

Upper End of the States (r)	0.273 (1)	0.507 (2)	0.775 (3)	1.139 (4)	1.756 (5)	5.843 (6)
886:	0.74	0.10	0.06	0.06	0.04	0.01
874:	0.10	0.78	0.09	0.03	0.00	0.01
862:	0.01	0.14	0.70	0.12	0.02	0.01
920:	0.00	0.01	0.16	0.71	0.11	0.01
872:	0.01	0.00	0.01	0.16	0.73	0.09
870:	0.00	0.00	0.00	0.00	0.11	0.89
Ergodic Distribution	0.077	0.156	0.178	0.203	0.191	0.195

TABLE 5
NMSH First Order Transition Matrix

Time-stationary, 1955-60

Upper End of the States (r)	0.00 (1)	0.75 (2)	1.00 (3)	1.11 (4)	1.66 (5)	5.05 (6)
180:	0.79	0.02	0.11	0.01	0.07	0.01
52:	0.02	0.67	0.27	0.00	0.04	0.00
172:	0.01	0.04	0.80	0.03	0.09	0.02
41:	0.00	0.00	0.59	0.37	0.02	0.02
108:	0.00	0.00	0.17	0.11	0.63	0.09
100:	0.00	0.00	0.00	0.00	0.15	0.85
Ergodic Distribution	0.030	0.056	0.439	0.057	0.209	0.208

FIGURE 1

$\Delta PCM < 0$

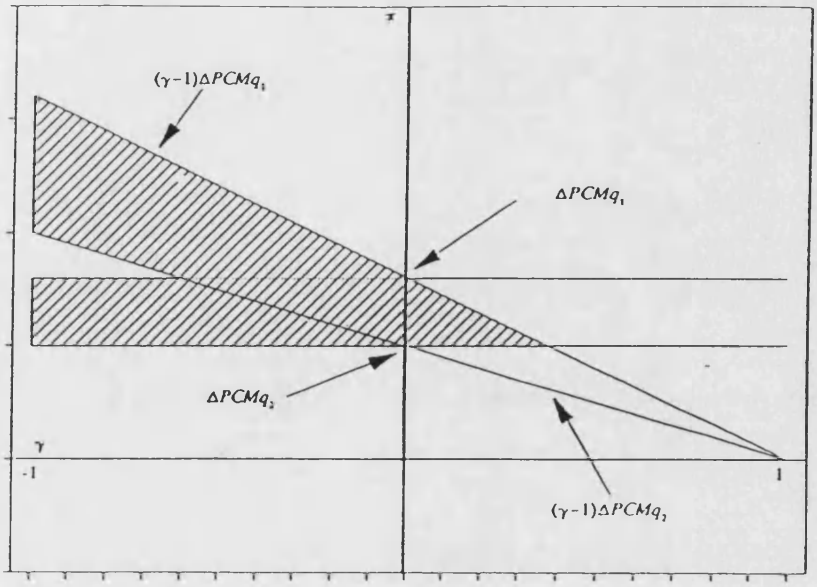


FIGURE 2

$\Delta PCM > 0$

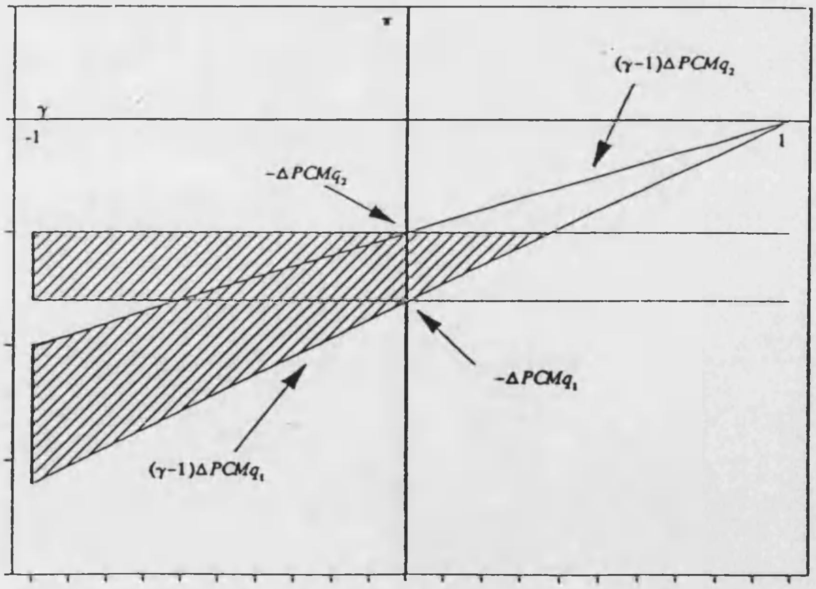


FIGURE 3a
NPLS2

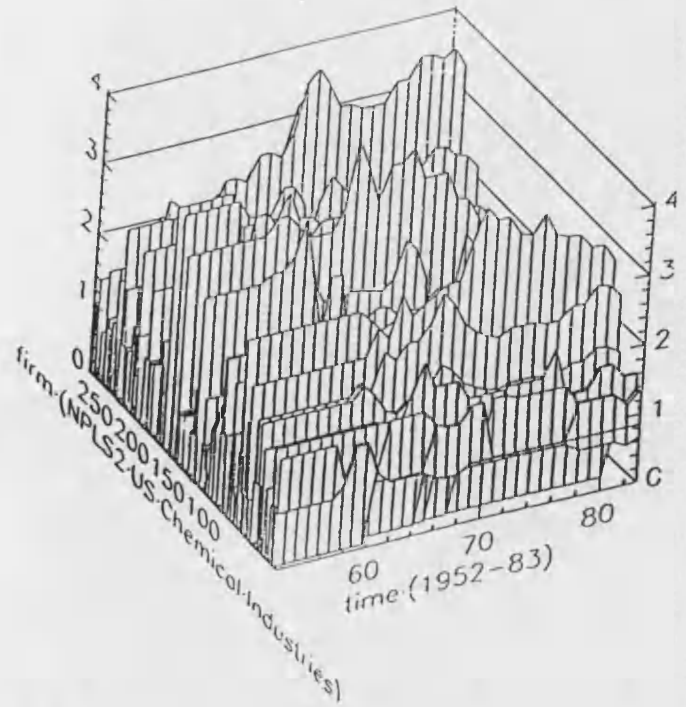


FIGURE 3b
NMSH2

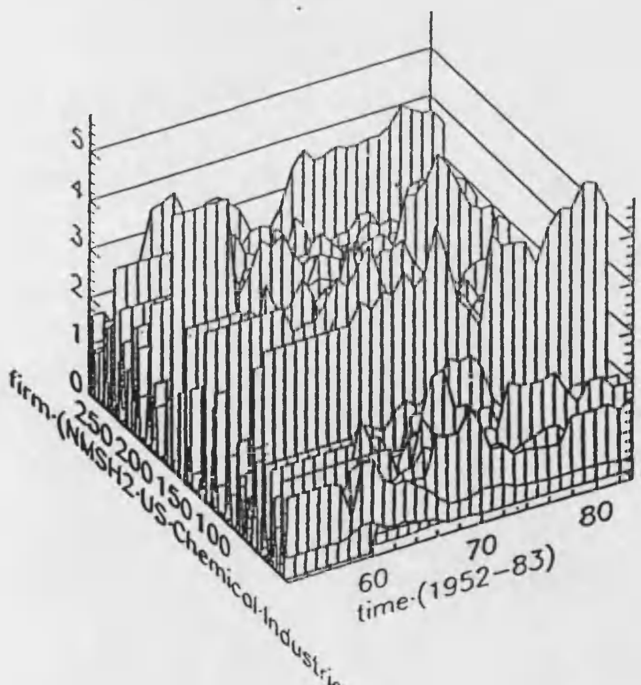
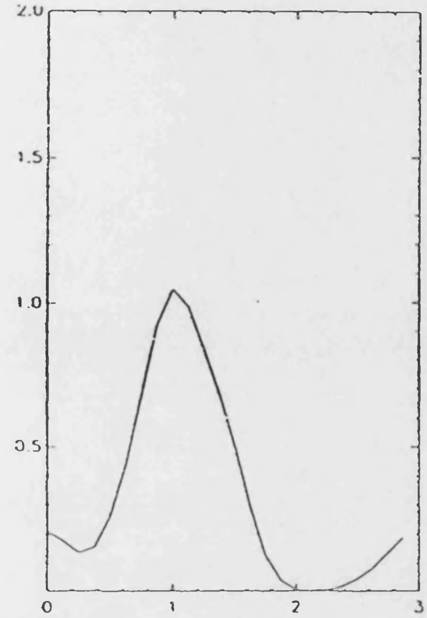
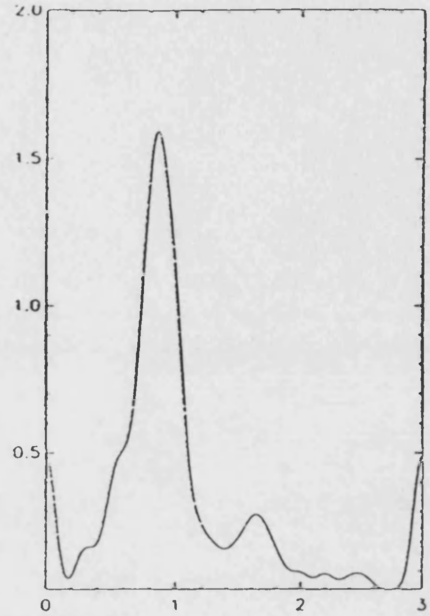


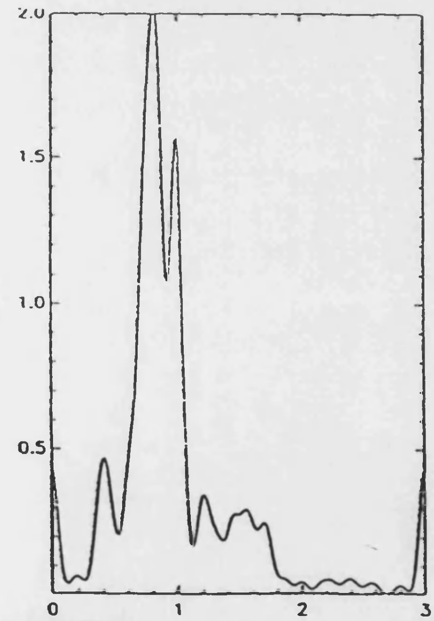
FIGURE 4
Estimated Density Functions NPL2



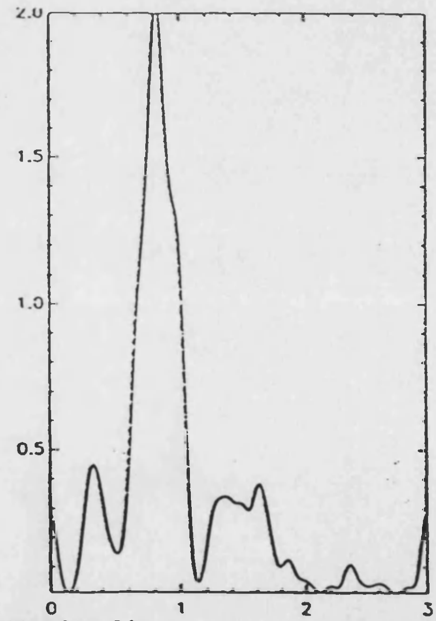
(a) 1954-58



(b) 1961-64



(c) 1969-72



(d) 1977-80

FIGURE 5a
Stochastic Kernel, 1 Year Transition

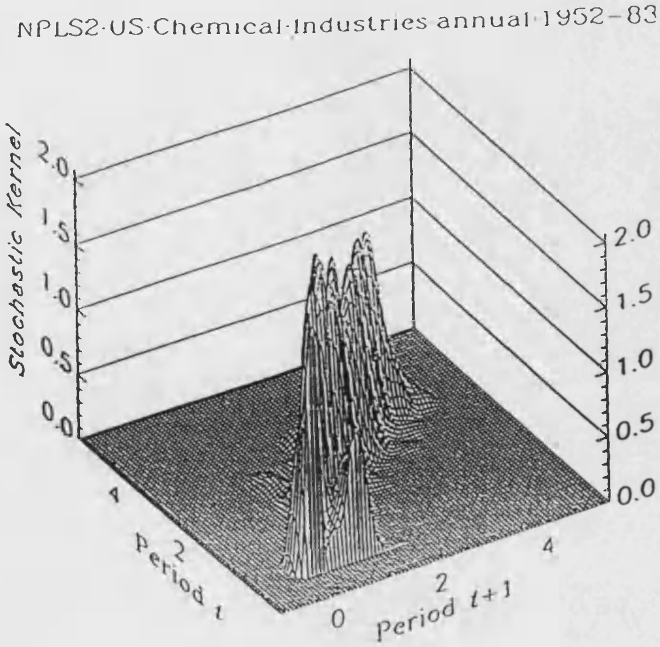


FIGURE 5b
Stochastic Kernel, 5 Year Transition

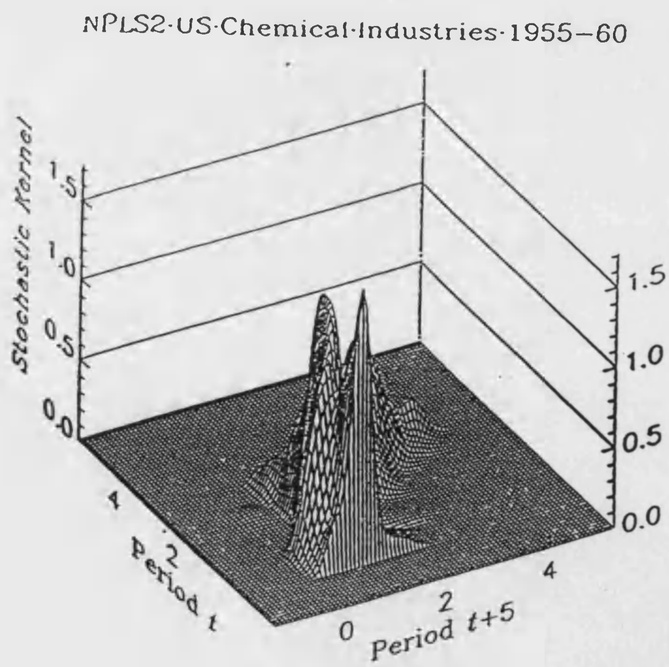


FIGURE 5c
Stochastic Kernel, 5 Year Transition

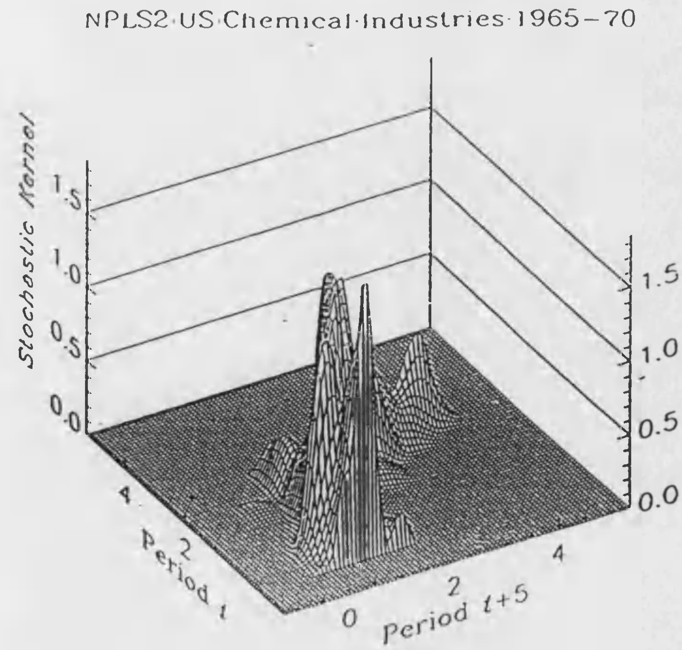


FIGURE 5d
Stochastic Kernel, 5 Year Transition

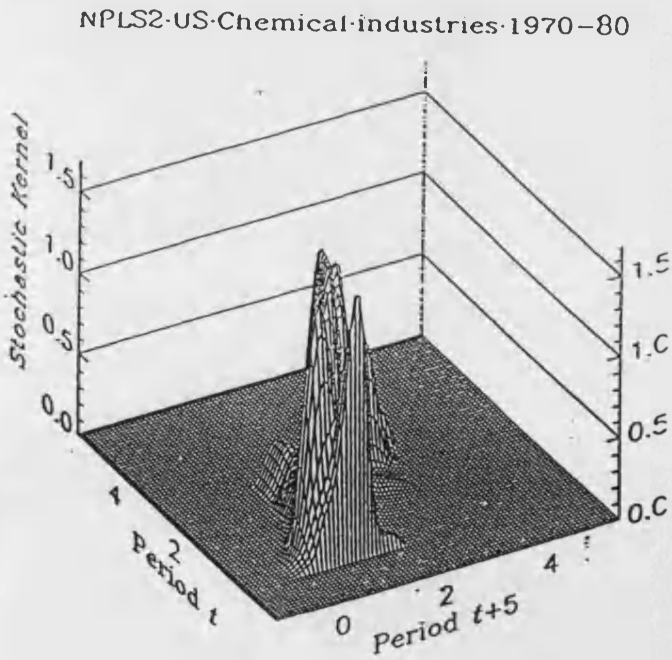


FIGURE 5e
Stochastic Kernel Contour Plot, 1 Year Transition

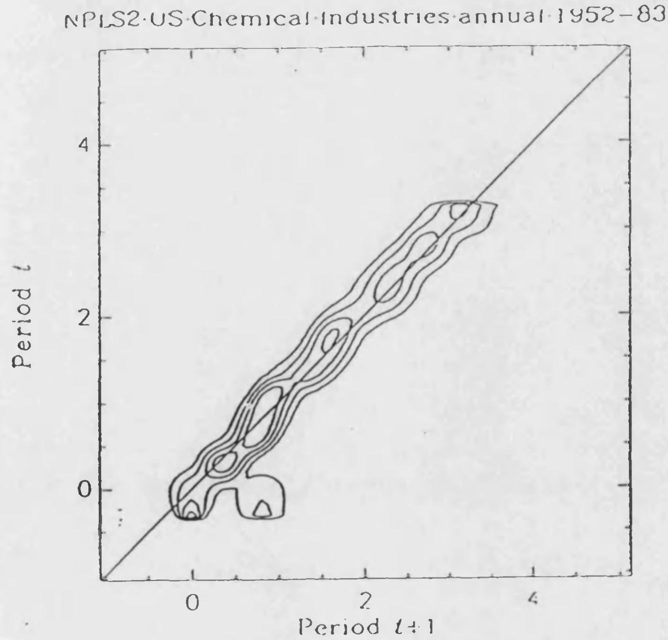


FIGURE 5f
Stochastic Kernel Contour Plot, 5 Year Transition

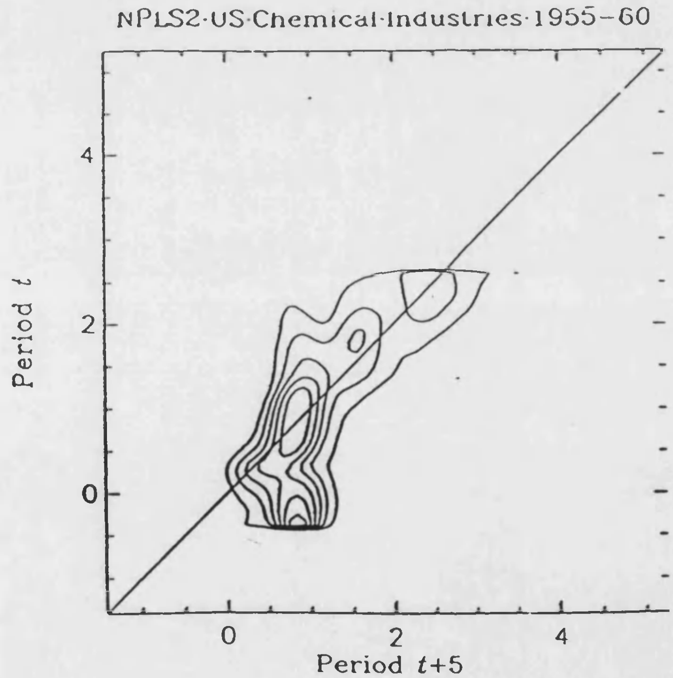


FIGURE 5g
Stochastic Kernel Contour Plot, 5 Year Transition

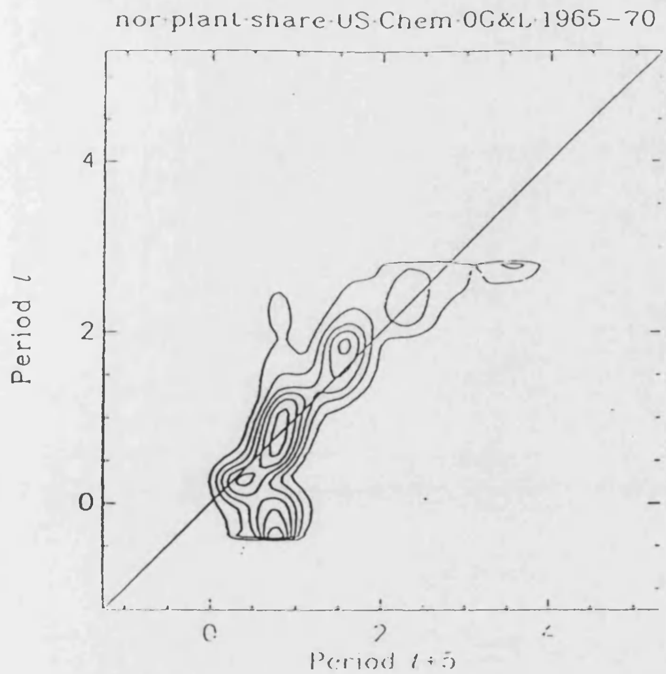


FIGURE 5h
Stochastic Kernel Contour Plot, 5 Year Transition

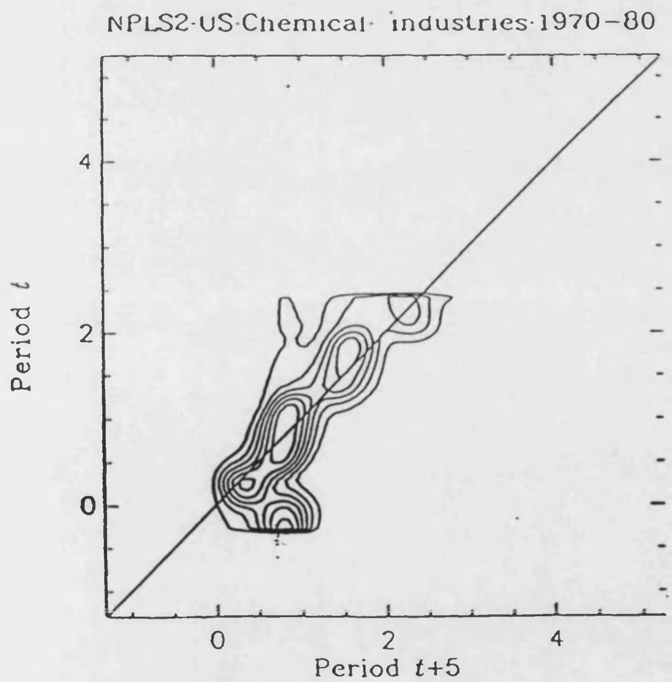


FIGURE 6a
Stochastic Kernel NPL5, 1 Year Transition

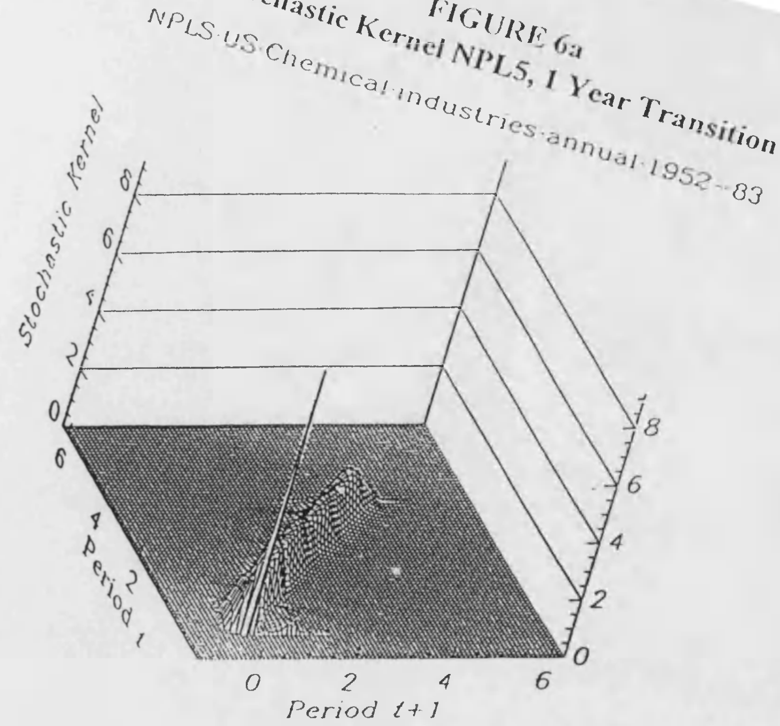


FIGURE 6b
Stochastic Kernel, 5 Year Transition

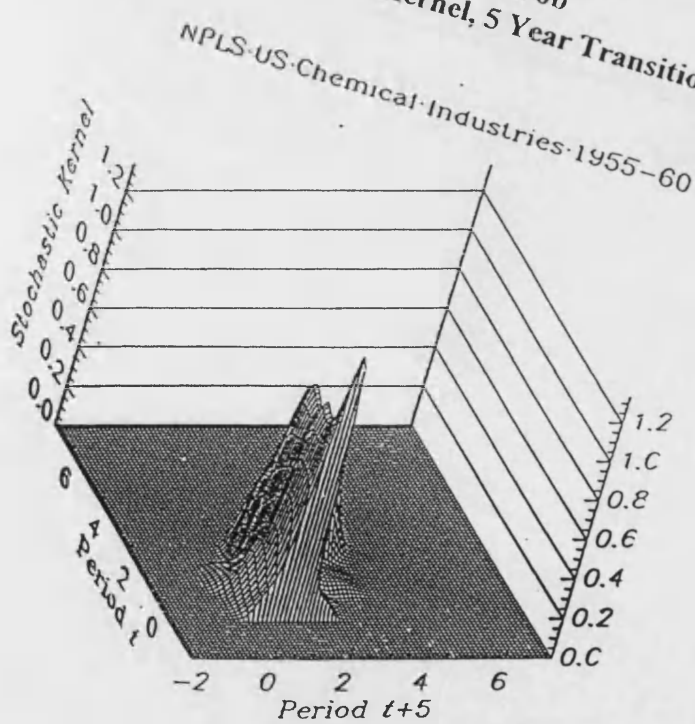


FIGURE 6c
Stochastic Kernel, 5 Year Transition

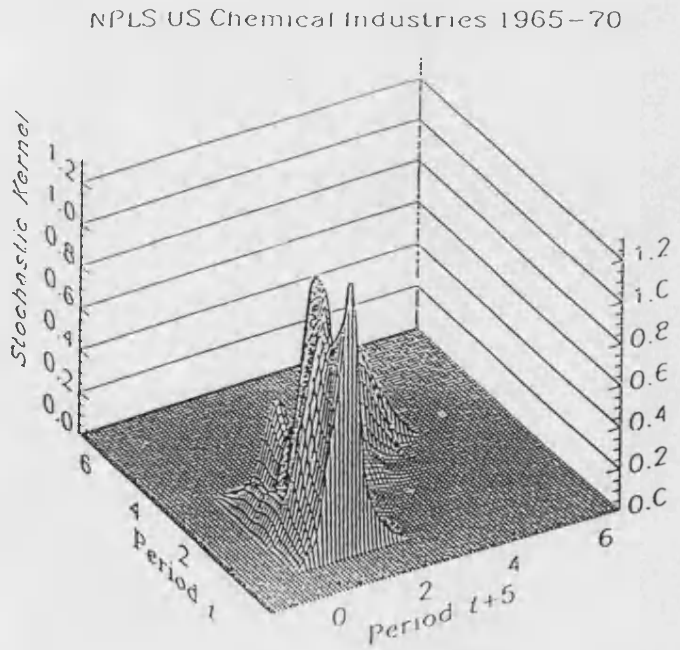


FIGURE 6d
Stochastic Kernel, 5 Year Transition

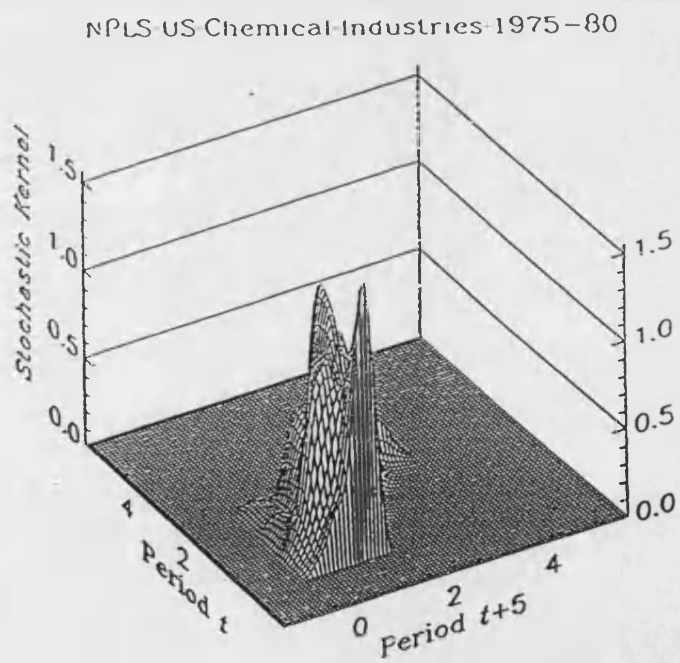


FIGURE 6e
Stochastic Kernel Contour Plot, 1 Year Transition

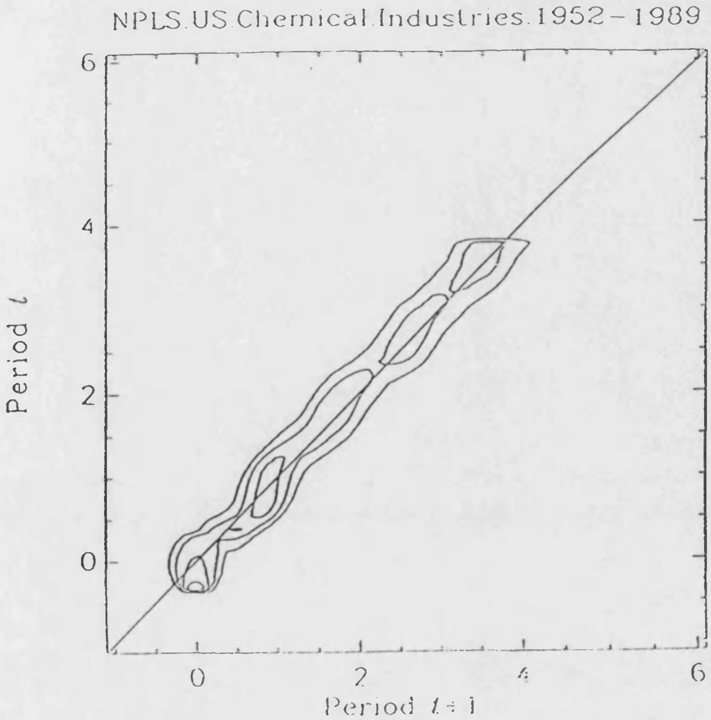


FIGURE 6f
Stochastic Kernel Contour Plot, 5 Year Transition

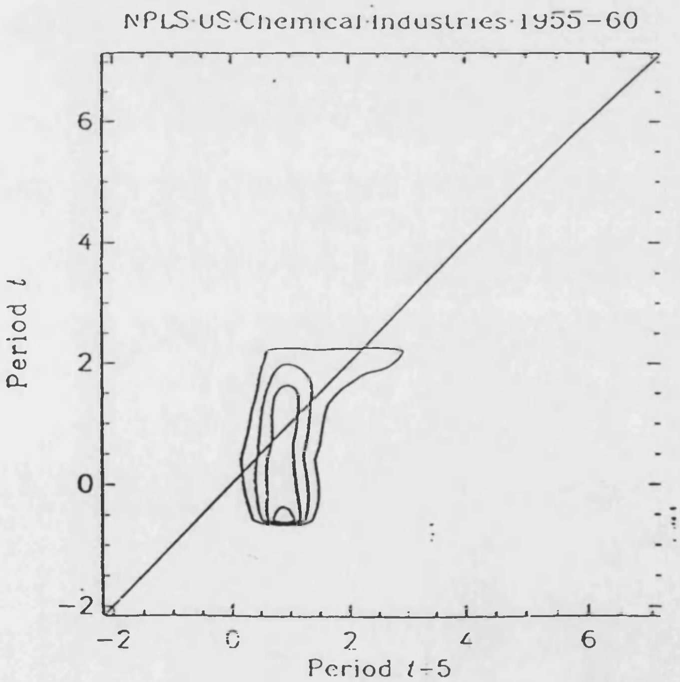


FIGURE 6g
Stochastic Kernel Contour Plot, 5 Year Transition

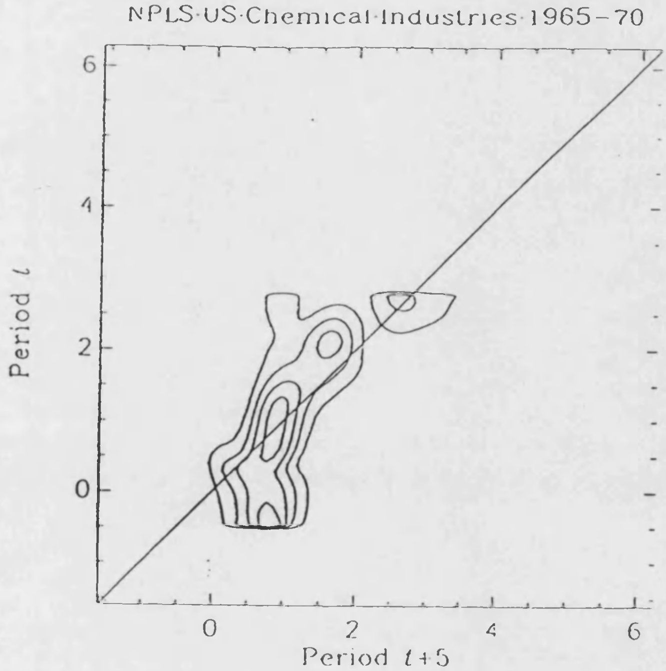


FIGURE 6h
Stochastic Kernel Contour Plot, 5 Year Transition

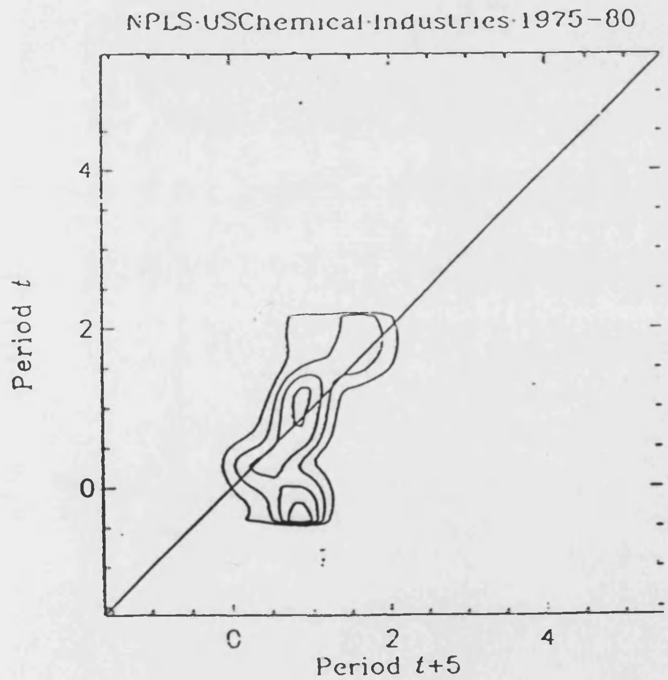


FIGURE 7a
Stochastic Kernel, 1 Year Transition

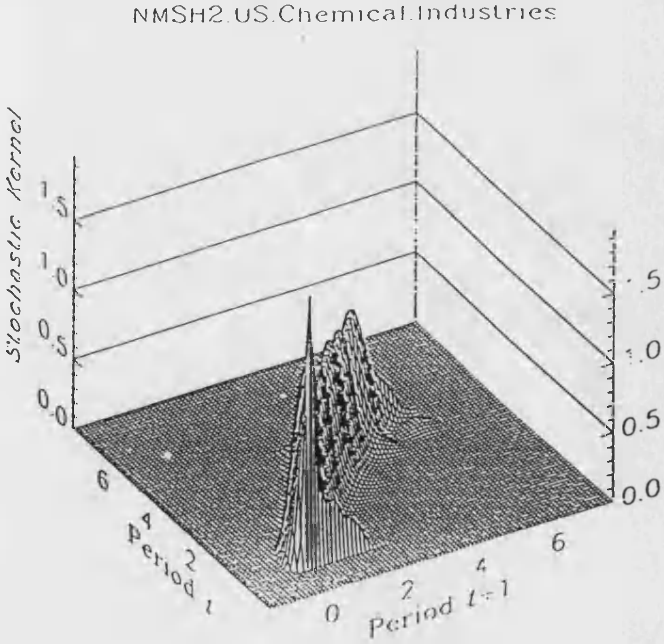


FIGURE 7b
Stochastic Kernel, 5 Year Transition

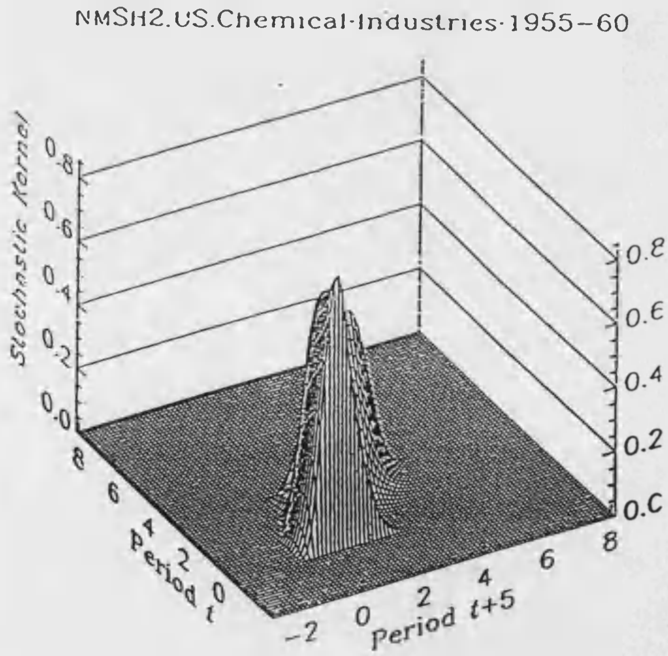


FIGURE 7c
Stochastic Kernel, 5 Year Transition

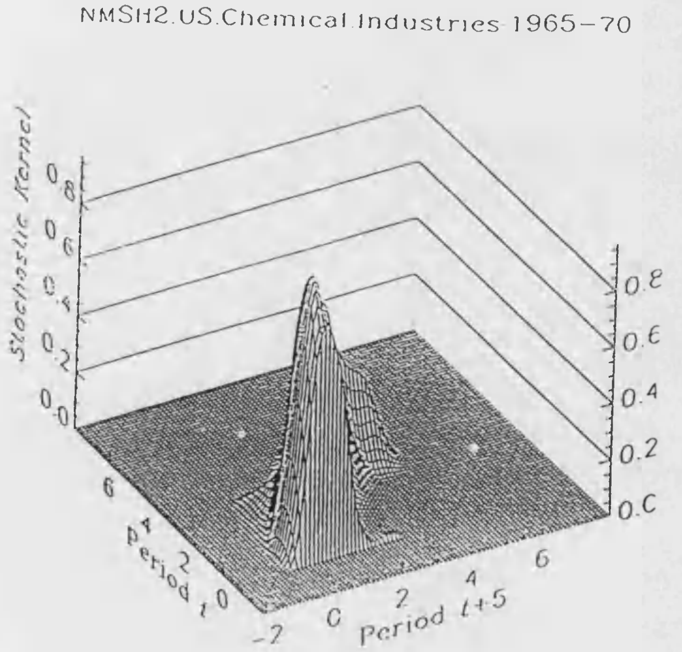


FIGURE 7d
Stochastic Kernel, 5 Year Transition

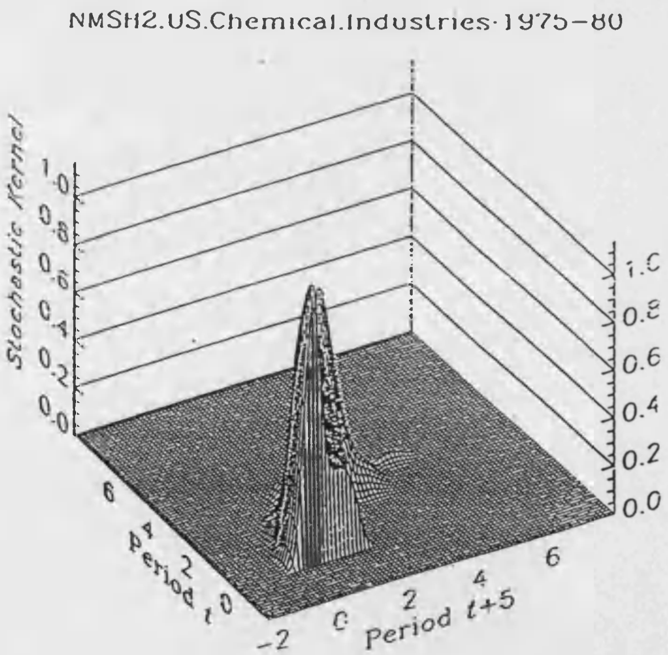


FIGURE 7e
Stochastic Kernel Contour Plot, 1 Year Transition

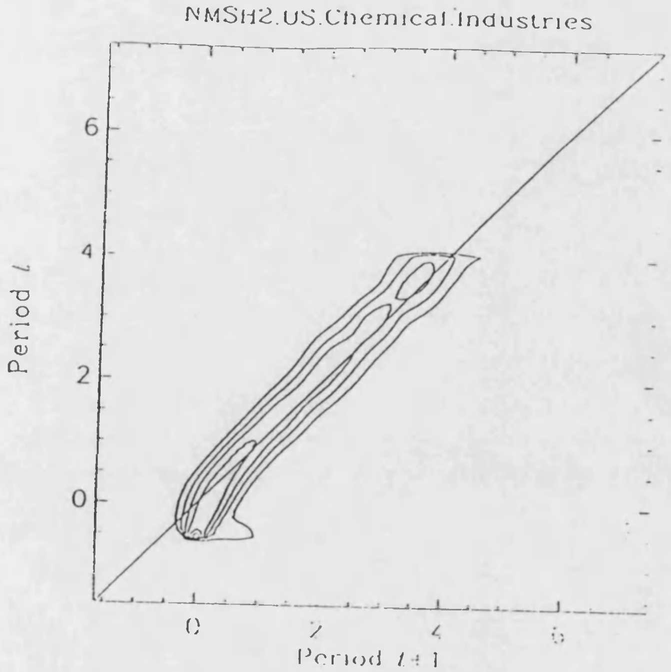


FIGURE 7f
Stochastic Kernel Contour Plot, 5 Year Transition

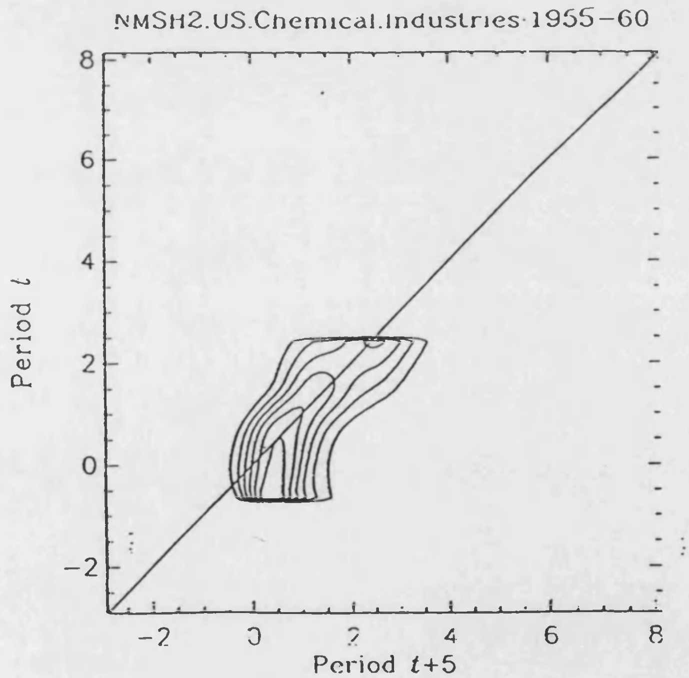


FIGURE 7g
Stochastic Kernel Contour Plot, 5 Year Transition

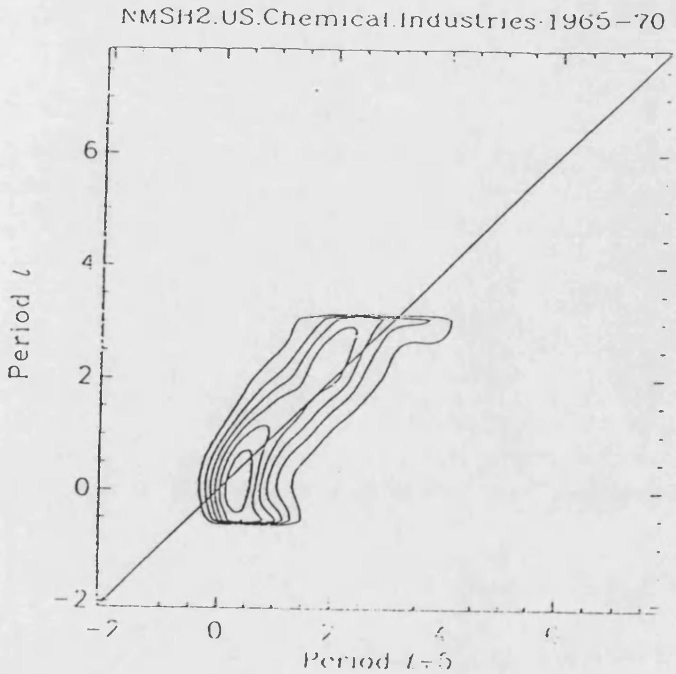
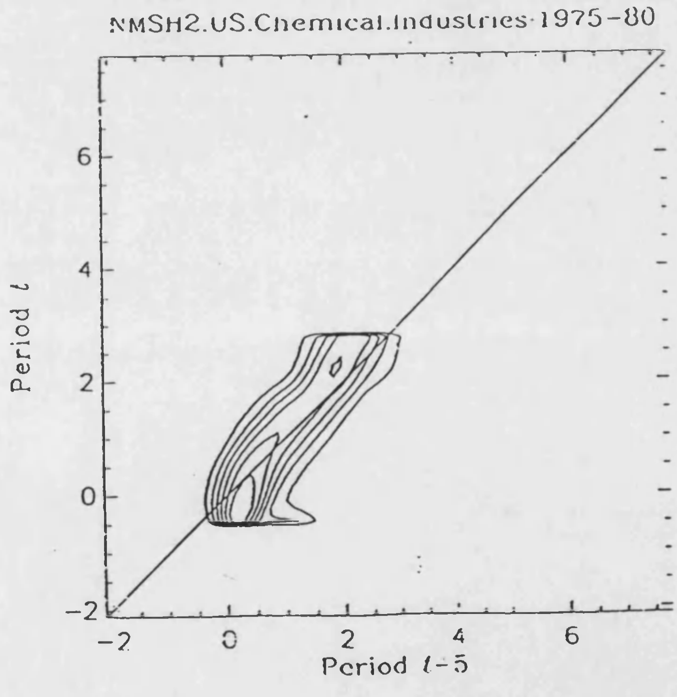


FIGURE 7h
Stochastic Kernel Contour Plot, 5 Year Transition



APPENDIX 1

The Logit Model

In this appendix we present the results of testing the ***Empirical Hypothesis*** using the conventional discrete choice approach.

The empirical model estimates the probability of firm i opening a plant, conditional on at least one firm in the industry opening one:

$$Prob(x_t = 1 | SUMOP > 0) = F(MSH_t,\ SUMOP_t,\ SUMF_t) \qquad \text{(A1.1)}$$

Where:

$SUMOP_{j,t}$:	total number of plants opened in industry j between t and t+1.
$x_{ij,t}$:	1 if firm i opens a plant in industry j between t and t+1, 0 otherwise.
F:	Logistic distribution function.
$MSH_{ij,t}$:	firm i's total capacity in industry $j(q_{ij,t})$ relative to the total industry capacity: $MSH_{ij,t} = q_{ij,t} / \Sigma_i\ q_{ij,t}$.
$SUMF_{j,t}$:	total number of firms in the industry and potential entrants.

We will test for alternative functional forms of the (log) odds ratio $h(.)$. Our basic specification, which reflects the outcome of the game, is linear:

$$h_t = a + b\,MSH_t + c\,SUMOP_t + d\,SUMF_t + u_t \qquad \text{(A1.2)}$$

The reason for conditioning on the number of openings (SUMOP) is that the theoretical model does not predict the time series dynamics, only which firm is more likely to win the project given that it is available.

In some of the estimated models MSH will be replaced by an alternative measure of relative firm size PLS, which is the number of plants of firm i in industry j relative to the total number of plants currently in operation in j. SUMOP is included as an explanatory variable to account for multiple arrivals of projects, as only the cross-sectional allocation is tested. The implicit assumption of this procedure is that the arrival process of new investment opportunities (projects) is independent of the firm characteristics. The validity of this is tested in Koopman and Lamo (1995)

A change in the number of firms competing for the projects changes the probability of firm i opening a plant, cp the relative sizes. But not only incumbent firms compete for opening new plants, potential entrants do too. This, as before, begs the question on how to deal with potential entry. In the dataset potential entry is not observable. If firms actually enter, it is not clear for how long a firm has been a potential entrant. All that can be said is that it has been for at least two years, since it takes on average two years to build a new plant[31]. We will make alternative assumptions on how long firms have been potential entrants before actually entering, to show that the results are robust in this respect. We will also assume that all potential entrants will enter over the course of the sampling period. Exiting firms remain potential entrants for some years by assumption. Hence, SUMF depends on the specific assumption that is made in this respect.

Large firms will open and close more plants than small firms due to replacement of old plants by new ones, effectively not changing the industry capacity. To control for this bias, a firm opening a plant t will be considered as replacement if the firm closes one between t-2 and t+2.

A testable hypothesis H_0 that follows from the ***Empirical Hypothesis*** is that $b < 0$, although in Section 3b we claimed that the interpretation of a negative coefficient of the initial condition as indicating convergence is not necessarily valid.

Table A1 shows the result of the ML estimators of (A1.2) for alternative specifications. Column (1) reports the estimated coefficients assuming that entrants have been around to grab a project for four years before opening their first plant, taking into account the two years it takes to build one. The coefficient of MSH is negative and significant. Hence, we cannot reject H_0 on first sight. The coefficients of both SUMF and SUMOP have their expected signs. Column (2) corresponds to a specification of the model that includes non-linear effects of MSH. The quadratic term is significant. It indicates that for MSH < 0.37 the relationship between the probability of opening and MSH is negative, though it is positive for higher values. It should be noted that 95% of the observations of MSH are below this critical value in this sample.

Some misspecification tests were performed. To test for industry specific effects we estimated the model (A1.2) for each industry, the estimates of b were instable, often insignificant and not always negative. Allowing b and the coefficient of MSH^2 to be industry specific, generated significant estimates only for a few industries. Adding industry dummies to the equation (A1.2) did not show significant industry effects[32].

Testing for time varying effects, time dummies were added to the original specification (A1.2). They were all insignificant and the estimation results did not change substantially. Splitting the sample into two sub-samples, pre- and post oil crisis (1973) shows that there is a structural break, but qualitatively the results are unchanged.

The coefficients in column (3) are estimated under the same entry assumption as (1) and (2), but replacing MSH by PLS. The conclusions are similar to those for (2). Also industry and time effects are as before. In (4) the sample is changed, assuming that all entrants could have opened a plant since the beginning of the sampling period and all exiting firms stay around as potential entrants until the end of the sampling period. The earlier conclusions remain unchanged.

The results of estimating (A1.2) using the sub-sample of incumbent firms are reported in column (5). In various alternative specifications, both MSH and PLSH become insignificant. Industry by industry, none of the estimates for b are significant. However, there might be a problem of sample selection[33] in the sense that we ignore the part of the sample with initial size being zero, since a firm does not enter the sample until it has a positive size. The estimation of the model for incumbents should, strictly speaking, take into account that the initial size being equal to zero is a truncation point in the considered sample[34]. On the other hand, to consider the whole sample as we did in the first place assumes that an entrant's decision of opening a plant can be described by the same model as the incumbent's. This is in line with the theoretical result since both respond to the same motivation, i.e. the fixed profit of the plant and the externality on the existing capacity.

Although these first results show some support for the theoretical results, they seem very sensitive to the exact empirical specification and the sample that is used.

Another testable implication of the theory are that the relative size of the firm is a sufficient statistic for determining which firm is most likely to win the project. Some exercises trying to test for this are performed in Koopman and Lamo (1995).

APPENDIX 2

Galton's Fallacy for Discrete Choice Probability Models (Probit)

Let the variable X_{it} be the size of the firm i at time t. Define:

$$Y_{it} = \begin{cases} = 1 & \text{if } X_{it} > X_{it-1} \\ = 0 & \text{if } X_{it} \leq X_{it-1} \end{cases} \quad \forall_{i,t} \tag{A2.1}$$

The probability of an individual firm increasing capacity given its current size, can be formulated as the following Discrete Choice Probability Model:

$$\text{Prob}\left[\, Y_{it}=1 \,/\, X_{it-1} \right] = \text{Prob}\left[\, X_{it} > X_{it-1} \,/\, X_{it-1} \right] = F(\,\beta' X_{it}) = 1 - F(-\,\beta' X_{it}) \tag{A2.2}$$

(A Probit Model if F is Φ).

We get (Appendix 2) a negative estimator of the parameter β. This negative relation between the probability of increasing size and the initial size has been understood in the literature as convergence in the firm's sizes. We claim that β is not informative in terms of convergence. In fact it can be shown that no convergence is compatible with $\beta < 0$.

Suppose that X_{it} follows the process:

$$X_{it} - \mu = \gamma(X_{it-1} - \mu) + \varepsilon_{it} \tag{A2.3}$$

Where μ is the mean and can be a function of exogenous variables $X'\phi$.

For simplicity redefine X_{it} as its deviation respect to the mean μ:

$$X_{it} = \gamma X_{it-1} + \varepsilon_{it} \tag{A2.4}$$

$\varepsilon \sim N(\,0, \sigma_\varepsilon)$ and $X_{it} \sim N(\,0, \sigma_\varepsilon/(1-\gamma)\,)$.

Let us assume that there is not convergence in the sense that the cross-section distribution of sizes remain unchanged along time. In other words X_{it} is stationary, i.e.:

$$|\gamma| < 1 \tag{A2.5}$$

Subtracting X_{it-1} from (A2.4),

$$X_{it}-X_{it-1} = -(1-\gamma)X_{it-1} + \varepsilon_{it} \tag{A2.6}$$

then (A2.2) is simply,

$$\text{Prob}\left[\,(X_{it}-X_{it-1}) > 0 \,/\, X_{it-1}\,\right] = \text{Prob}\left[-(1-\gamma)X_{it-1} + \varepsilon_{it} > 0\right] =$$

$$1 - \text{Prob}\left[\,\varepsilon \leq (1-\gamma)X_{it-1}\,\right] = 1 - F\left[\,(1-\gamma)X_{it-1}\right] \tag{A2.7}$$

So, our $-\beta = (1-\gamma)$ *i.e.* $\beta = (\gamma - 1) < 0$ even if the firm's sizes are stationary.

ENDNOTES TO APPENDICES

31. See Lieberman (1987).

32. The output of those specification tests are omitted for reasons of space. Available on request.

33. For sample selection problems in this context see Hall (1987).

34. This might be related to Gilbert and Lieberman (1987) finding a positive sign for MSH in their logit model of the probability of incremental capacity expansions. Their sample is also restricted to incumbents.

TABLE A1

Logit Analysis of Plant Openings.

(Estimated coefficients)

	(1)	(2)	(3)	(4)	(5)
MSH	-7.14* (1.53)	-15.65* (2.59)		-5.85* (1.53)	3.37* (0.98)
MSH^2		20.92* (5.52)		8.14 (2,81)	
PLSH			-28.75* (2.59)		
$PLSH^2$			50.50* (5.64)		
SUMOP	0.28* (0.04)	0.25* (0.04)	0.27* (0.05)	0.23* (0.03)	0.16 (0.04)
SUMF	-.18* (0.02)	-0.17* (0.02)	-0.21* (0.03)	-0.11* (0.03)	-0.05 (0.03)
CONST	0.27	0.27	-1.81	0.13	-2.13
No of Obs.	1642	1642	1642	2223	807
Log Likelihood	-606.8	-584.9	-527.6	-706.4	-291.2

Heteroscedasticity consistent (White) SE are in parentheses.
* Significant at 1% (One tailed test).

REFERENCES

Arrow, K., (1962), `Economic Welfare and the Allocation of Resources for Inventions', In Nelson, R., (ed), *The Rate and Direction of Inventive Activity*, Princeton University Press: Princeton, NJ.

Barro, R. and Sala-i-Martin, X. (1992), `Convergence', *Journal of Political Economy*, Vol.100, No.2, pp.223-251.

Beggs, A. and Klemperer, P.(1992), `Multi-Period Competition with Switching Costs', *Econometrica*, Vol.60, No.3, May , pp.651-666.

Budd, C., Harris, C. and Vickers, J.(1992), `A Model of the Evolution of Duopoly: Does the Asymmetry Between Firms Tend to Increase or Decrease?', *Review of Economic Studies*, Vol.60, No.3, pp.543-573.

Cabral, L. and Riordan, M. (1994), `The Learning Curve, Market Dominance and Predatory Pricing', *Econometrica*, Vol.62, No.5, September.

Chapman K., (1991), *The International Petrochemical Industry*, Blackwell: Oxford.

Desdoigts, A. (1994) , *Changes in the World Income Distribution: A Non-Parametric Approach to Challenge the Neoclassical Convergence Argument,* PhD thesis, European University Institute, Florence, June.

Dolado, J. and Gonzalez-Paramo, J.(1994), `Regional Convergence in Spain', *Moneda y Credito*, .

Farrell, J. and Shapiro, C.(1988), `Dynamic Competition with Switching Costs', *Rand Journal of Economics*, Vol.19, No.1, Spring, pp.123-137.

Foster, R., (1985), *Innovation: The Attacker's Advantage*, Summit Books.

Friedman, M. (1992), `Do Old Fallacies Ever Die?' *Journal of Political Literature*, Vol.30, No.4, December , pp.2129-32.

Ghemawat, P.(1984), `Capacity Expansion in the Titanium Dioxide Industry', *The Journal of Industrial Economics,* Vol.33, No.2, December, pp.145-163.

Ghemawat, P and Nalebuff, B.(1990), `The Devolution of Declining Industries', *Quarterly Journal of Economics,* Vol.105, 1990, pp.167-186.

Gilbert, R. and Harris, R. (1984), `Competition with Lumpy Investment', *The Rand Journal of Economics*, Vol.15, No.2, pp.197-212.

Gilbert, R and Lieberman, M.(1987), `Investment and Coordination in Oligopolistic Industries' *Journal of Economics*, Vol.18, No.1, Spring , pp.17-33.

Gilbert, R and Newbery, D.(1982), `Preemptive Patenting and the Persistence of Monopoly', *American Economic Review*, Vol.72, No.3, June, pp.514-26.

Hart, P. and Prais, S.(1956), `The Analysis of Business Concentration: A Statistical Approach', *Journal of the Royal Statistical Society*, Series A, Vol.119, pp.150-81.

Hall, B.(1987), `The Relationship between Firm Size and Firm Growth in the US Manufacturing Sector', *The Journal of Industrial_Organisation*, Vol.35, No.4, June, pp.583 - 606.

Huigen, R., Kleijweg, A. and Van Leeuwen, G.(1991), `The Relationship Between Firm Size and Firm Growth in Dutch Manufacturing Estimated on Panel Data', Research Paper 9105, Research Institute for Small and Medium-sized Business in the Netherlands, Zoetermeer October .

Hopenhayen H.(1989), `A Dynamic Stochastic Model of Entry and Exit to an Industry', mimeo, Graduate School of Business - Stanford University, Research Paper 1057.

Jovanovic, B.(1982), `Selection and Evolution of Industry', *Econometrica,* Vol.50, pp.649-70.

Katz, M. and Shapiro, C. (1987), `R&D Rivalry with Licensing or Imitation', *American Economic Review,* Vol.77, No.3, June , pp.402-20.

Koopman R. and Lamo A. (1995), ' Cross-Sectional Firm Dynamics: Theory and Empirical Results From The Chemical Sector' LSE, CEP Discussion Paper No.229, March .

Lee, T. and Wilde, L.(1980) , `Market Structure and Innovation: A Reformulation', *The Quarterly Journal of Economics*, Vol.94, March, pp.429-36.

Leonard, S.(1986), `On the Size Distribution of Employment and Establishments', NBER Working Paper No 1951, NBER: Cambridge, MA.

Lieberman, M.(1984), `The Learning Curve and Pricing in the Chemical Processing Industries', *Rand Journal of Economics*, Vol.15, No.2, Summer, pp.213-28.

Lieberman, M.(1987), - `Postentry Investment and Market Structure in the Chemical Processing Industries', *Rand Journal of Economics*, Vol.18, No.4, Winter.

Loury, G.(1979), `Market Structure and Innovation', *The Quarterly Journal of Economics*, Vol.93, August, pp.395-410.

Mansfield, E.(1985), `How Rapidly Does New Industrial Technology Leak Out?' *Journal of Industrial Economics,* Vol.34, No.2, December, pp.217-223.

Pakes, A. and Ericson, R.(1990), `Empirical Implications of Alternative Models of Firm Dynamics',mimeo,.

Quah, D.(1993a), `Empirical Cross-Sectional Dynamics in Economic Growth', *European Economic Review*, Vol.37, No.2/3, December pp.426-443.

Quah, D.(1993b) `Galton's Fallacy and Tests of the Convergence Hypothesis', *The Scandinavian Journal of Economics*, Vol.95, No.4, December, pp.427-443.

Quah, D.(1996)`Convergence Empirics Across Economies with (Some) Capital Mobility', *Journal of Economis Growth*. Forthcoming

Quintella, R., (1993), *The Strategic Management of Technology in the Chemical and Petrochemical Industries*, Pinter Publishers: London.

Reinganum, J.(1983), `Uncertain Innovation and the Persistence of Monopoly', *American Economic Review*, Vol.73, No.4, September , pp.741-48.

Reinganum, J (1989), `The Timing of Innovation: Research Development and Diffusion', in Schmalensee, R. and Willig, R., (eds), *Handbook of Industrial Organization, Vol.I*, Amsterdam, pp.849-908.

Silverman, B., (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall: New York, NY 10001.

Simon, H. and Bonini, C.(1958), `The Size Distribution of Business Firms', *American Economic Review*, Vol.48, September, pp.607-17.

Spitz, P., (1988), *Petrochemicals, The Rise of an Industry*, John Wiley & Sons: New York.

Stobaugh, R., (1987), *Innovation and Competition*, Harvard Business School Press: Boston, MA.

Stockey, N. and Lucas Jr, R., (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press: Cambridge, Mass.

Vickers, J.(1986), `The Evolution of Market Structure when there is a Sequence of Innovations', *The Journal of Industrial Economics*, Vol.35, No.1, September 1986, pp.1-12.

CHAPTER 4

UNEMPLOYMENT IN EUROPE AND REGIONAL LABOUR FLUCTUATIONS

1. INTRODUCTION

Whether idiosyncratic or disaggregate[1] (regional, sectors, etc.) shocks are responsible for aggregate business cycles or whether aggregate disturbances are the causes of macroeconomic fluctuations which then disseminate across regions or sectors is still an open question in the business cycles literature.

This paper attempts to be a contribution to the empirical business cycle literature in the labour market addressing the issue above for the European labour market. It studies unemployment and labour market fluctuations. Specifically it deals with the question of whether shocks in employment growth rates across regions in Europe have aggregate effects on European unemployment or on the contrary, the aggregate fluctuations are primary.

The standard theories of employment decompose unemployment into natural or structural rate and fluctuatuions around it, calling these fluctuations cyclical unemployment. The economy can be above or below the natural rate but in the long run equilibrium tends to it. The interest in cyclical unemployment appears during the 1980s, when macroeconomists try to explain the behaviour of European unemployment rates. The picture in the early 80s was one of rising unemployment together with a deflation process. The traditional approach explained the labour market behaviour by assuming that unemployment rates were over their natural or structural level. The puzzle emerges when in the second half of the decade the inflation levels stabilise and unemployment remains high in spite of the actions taken to reduce it. It seems that the natural rate of unemployment in Europe has risen[2] and arguments such as oil prices, high interest rates etc. fail to explain why. One of the possible explanations is that the cyclical unemployment in Europe is turning into structural unemployment. Consequently, business cycle literature has an important role to play, since understanding fluctuations in labour market becomes crucial to understanding the high unemployment rate in Europe and its persistence.

Traditionally, it has been argued (Barro (1977)) that the cause of unemployment fluctuations were aggregate demand shocks. More recent ideas (hysteresis theory) claimed that temporary labour demand shocks may have long-lasting effects on unemployment. Also it has been maintained that unemployment reacts imperfectly to permanent shocks and they have a delayed effect on unemployment. A controversial explanation of unemployment fluctuations was raised by David Lilien (1982). He claimed that an important part of the fluctuations in employment is due to shifts in demand across sectors or regions rather than aggregate disturbances.

There is an enormous amount of empirical work trying to relate aggregate and disaggregate (or idiosyncratic) fluctuations in labour markets, in other terms attempting to see to what extent idiosyncratic shocks can generate aggregate fluctuations in the labour market. This literature deals with aggregate unemployment, sectoral and regional labour imbalance and labour mobility. Methodologically it makes use of measures hardly useful to analyse the issue at hand and additionally it makes assumptions about stationarity. The current paper deals with the same kind of issues. The analysis performed here overcomes some of the methodological drawbacks of the existing literature. It uses a model for non-stationary evolving distributions to identify idiosyncratic and aggregate disturbances and then analyses their joint dynamics. It provides some evidence on whether regional shocks in the European labour market are responsible for the evolution of unemployment rates in Europe.

The rest of the paper is organised as follows. Section 2 describes the recent empirical literature in labour mobility and sectoral/regional imbalance and explains why the techniques and the measures used in the existing literature are not adequate to account for idiosyncratic shocks. Section 3 suggests a more natural approach in the context of cross-section dynamics analysis proposed by Quah (1994a, 1996). Section 4 studies the dynamics of employment for 51 European regions from 1960 to 1990, and provides some evidence on whether regional shocks have aggregate effects on unemployment or whether aggregate fluctuations spread across regions. Section 5 concludes.

2. CYCLICAL EMPLOYMENT AND LABOUR IMBALANCE LITERATURE

Lilien (1982) found a high positive correlation between the standard deviation of employment growth rates among sectors (σ_t) and the aggregate unemployment rate for the US during the post-war period. From this correlation he concluded that shocks in demand across sectors are responsible for an important part of the cyclical variation in unemployment. Conclusion that relies on two assumptions: the employment growth rates among sectors is a good proxy of labour reallocation and its standard deviation describes adequately the dynamics of the cross-section distribution. Further it interprets correlation as causality.

Lilien's argument generates a wide response, given that the implications for economic policy are very different depending on whether the driving force of cyclical unemployment is sectoral shifts or aggregate disturbances. Idiosyncratic shocks as the main cause of unemployment fluctuations suggest that an efficient policy would be that conceived to smooth the adjustment process of the labour force across categories and consequently it would discard the aggregate demand policies[3]. In this respect a very influential paper is Abraham & Katz (1986). They showed that a pure aggregate demand shock could produce a positive correlation between σ_t and the employment rate if some categories (regions, sectors, etc.) are cyclically more sensitive than others. They also gave evidence on how Lilien's measure may be affected by aggregate variation influences. They understand the correlation found by Lilien as reverse causality, in other words, aggregate fluctuations generate the dynamics in Lilien's measure and not the opposite.

Abraham & Katz (1986) suggest using information on job vacancy rates in order to indicate whether a pure idiosyncratic shift or a pure aggregate demand has been the more important cause of the correlation. This is based on the negative relationship between unemployment and vacancy rates. Holding structural characteristics fixed, the plot of unemployment rates versus vacancy rates describes a negatively sloped curve which is known as Beveridge or UV curve. Changes in aggregate demand lead to movements along this curve, then the response of unemployment and vacancies would go in the opposite directions. A pure idiosyncratic shock shifts the curve generating higher unemployment rate at each vacancy rate. This is compatible with movements of vacancies and unemployment rate in the same direction for example a negative shock across regions increase the unemployment rate but it also increase vacancies. In other terms there is a matching problem between demand and supply for labour. Excess and deficiency for demand and supply coexist.

From that, several studies in the literature have pointed out the concept of deficient matching between labour supply and demand for labour. They define some measures of the imbalance between unemployment and vacancies across different labour-market categories (sectors, regions, skills, etc.), which are called *mismatch indices* and study their evolution over time and their correlation with aggregate unemployment. There is a big variety of indices which correspond to different concepts of mismatch. In fact there is no unified view of the mismatch concept.

The most popular measures of mismatch arise from the equilibrium models. Mismatch is defined as the distance between the actual and the optimal unemployment rate derived from an equilibrium model. If the model is such that the optimal unemployment rate is the one at which the unemployment and vacancies ratio coincides across categories, then the empirical measure of mismatch is the following [4]

$$MM1 = \tfrac{1}{2} \sum | u_i - v_i | , \tag{4.1}$$

where u_i and v_i are respectively the share of unemployed persons and the share of job vacancies in category i= 1....N.

If the equilibrium unemployment rate is the NAIRU,[5] the empirical measure of mismatch is the following index[6]:

$$MM2 = \frac{1}{2} var[\frac{\frac{U_i}{N_i}}{\frac{\Sigma U_i}{\Sigma N_i}}]. \tag{4.2}$$

Where U_i and N_i are unemployment and employment in group i respectively.[7] In fact this index uses the idea of relative dispersion of regional unemployment as an indicator of mismatch.

There are several studies of mismatch based on the Dréze and Bean disequilibrium model.[8] The Unemployment can be constrained by lack of demand (Keynesian regime), lack of capacity (classical regime) or lack of labour (repressed inflation regime). Mismatch here is identified with regime disparity across regions, sectors, etc.. In each micro market the short side, determines the unemployment, the existence of rationing implies that there are unfilled vacancies or unemployment. Finally there is an approach which understands mismatch as a short term phenomenon. The index that better reflects this short-term approach to mismatch is a turbulence[9] index of the type of Lilien's measure, i.e. the sum of absolutes changes in regional /sectoral/etc. shares of employment.

The mismatch literature is inconclusive. The evidence from *MM1* indicates that mismatch increased in Germany and Japan but did not in the UK and Sweden during the post-war period. According to the *MM2* index mismatch falls over time in the majority of the categories (skill, occupation, region, etc.) and countries, but nevertheless it seems to explain more than one third of the total unemployment.[10] Lilien's measure has been included in the estimation of some labour market equations and is basically not significant. Few other studies replicate Lilien's work and find what they call evidence in support of Lilien's argument, for example Neeling (1987) for Canada or more recently Kazamaki (1994) for Sweden.

All the studies in this labour imbalance/idiosyncratic shock literature, suffer from the same kind of general but critical problem: the empirical measures used do not capture the economic phenomenon that they are trying to reflect.

Lilien simply takes the time series of cross-section variances of changes in the employment rate (in logs, i.e. employment growth rates) and examines its correlation with aggregate unemployment time series. Firstly, he measures labour reallocation due to sectoral shifts as rates of change in employment. Abraham and Katz (1986) argued that there is evidence to believe that Lilien's variable includes labour reallocation due to sectoral shocks and to aggregated fluctuations. There are a few attempts in the literature to construct proxies that distinguish aggregate shocks from the idiosyncratic ones. See for example Neeling (1987) or Kazamaki (1994)[11]. Secondly, the cross-section standard deviation σ_t is a point-in-time statistic of the cross-section distribution.

These studies assume that it is a good summary of all the relevant information about the dynamics of the variable in question (employment growth rates). An assumption that is quite questionable. In order to accept it, it would be necessary to test whether σ_t describes adequately the cross-section distribution of the employment growth rates.

The mismatch indices approach deserves similar comments. The study of labour imbalances requires the characterisation of the behaviour of employment and/or vacancies of a cross-section (regions, sectors, etc.) over time. Instead what this literature does is the following. Given a definition of imbalance between labour demand and labour supply across categories, it uses an index to measure that imbalance. Again, the mismatch indices are simply summary-measures, statistics of the cross-section distribution of that imbalance at each point in time. The indices collapse all the cross-section information in a single statistic. The studies of mismatch look at the evolution of these indices over time. In other words they are using data in the distribution across categories of employment and/or unemployment but they do not exploit efficiently the information contained in those data.

Entorf (1993) analyses the performance of the above defined *MM1* and *MM2* and shows that they can easily fail when unemployment shows upward additive shifts. He proves analytically that both measures can give spurious results arising from stochastic trends and changes in aggregate unemployment.

Consequently, the mismatch analysis which initially appeared as a promising alternative way to analyse disaggregate and aggregate fluctuations in the labour market, turns out to be misleading and requires alternative measures that take into account the dynamics of the imbalance phenomenon. Additionally most of the existing measures of mismatch are derived from stationary and more precisely static equilibrium models .

The current paper, goes back to the original problem in Lilien (1982) and analyses directly the dynamics of labour reallocation. It attempts to obtain some evidence on whether the evolution of European unemployment during the last 30 years is attributable in part to the dynamics of regional shocks. It models the dynamics of changes in employment, after conditioning out the aggregate component, across categories (in this case regions) using a natural approach that exploits all the cross-section information available and models its dynamics over time. The study of the dynamics of vacancies and unemployment imbalances is in the agenda.

3. REGIONAL EMPLOYMENT DYNAMICS AND AGGREGATE UNEMPLOYMENT FLUCTUATIONS

This section suggests an approach to study the dynamics of aggregates and disaggregates.[12] Our question is whether idiosyncratic regional shocks to employment (labour reallocation due to regional shocks) explain the dynamics of aggregate unemployment fluctuations. The empirical analysis of this question requires two steps. Initially there is need for an adequate proxy of regional shocks in labour. Once the variable is well defined, what this approach suggests is simply to characterise the dynamics of its cross-section distribution, instead of focusing on one statistic of that distribution as previous studies do. The idea is to model the dynamics of aggregates, whose fluctuations we are interested in explaining, together[13] with the dynamics of the disaggregates This proposed analysis is a natural way of looking at the information, it is model free and does not make any assumption about stationarity or similar, exploiting all the cross-section information .

While the aggregate unemployment rate in Europe, is a time series structure, the disaggregate (idiosyncratic regional shocks to employment) has the structure of a *Random Field.*[14] At *each* moment *t* there is one observation for *each* region, i.e. at each point in time is a cross-section distribution. The dynamics of these cross-section distributions involves: (a) changes on the exterior shape and (b) intra-distribution mobility. The way to proceed is to characterise that dynamics and to relate it to the unemployment dynamics.

Formally, let u be a vector of aggregates with a fixed finite dimension and let y be the cross-section of disaggregates. The hypothesis is that aggregates and disaggregates (u,y) evolve together over time. We are interested in their joint dynamics. Let us start by modelling the dynamics of the disaggregates. Let $\{\Phi_{yt}$, integer $t \geq 1\}$ be the measure (one for each year) describing the distribution of y. More precisely $\Phi_{y,t}$ is the dynamically evolving probability measure of the distribution of y. It is defined on the measurable space (R, R), (where R is the real line and R is the sigma algebra.

Dynamically evolving distributions ($\Phi_{y,t+1}$) can always be written in terms of the following stochastic kernel equation:

$$\Phi_{y,t+1} = \int \Xi_t(y, A)\Phi_{y,t}(dy) \ , \tag{4.3}$$

for every A in R, where Ξ_t is a Stochastic Kernel.[15] That is, $\Xi_t(y, A)$ is the probability that the next state period lies in A given that this period the state is y.

The $\{\Xi_t\}$ sequence of stochastic kernels, encodes all the dynamics of Φ_t (the cross-section distribution of the disaggregates. However Ξ_t is infinite dimensional, for the discrete case the stochastic kernel equation describes a Markov chain sequence. The latter may be parameterized by two elements: a sequence of transition matrices which indicates intra-distribution mobility and a sequence of grids which are an estimator of the cross section distribution shape[16]. Hence, $\{\Xi_t\}$ turns up to be a sequence of transition probability matrices and Φ_t are the corresponding marginal distributions (grids). Since we are interested in the relation between the dynamics of the disaggregates (regional fluctuations) and the aggregate fluctuations in unemployment, we must study correlation and causality between the aggregate and the sequence of transition matrices and the sequence of grids, which parameterize the disaggregate.

Let us define the grid in such a way that the set of quantiles determine the sequence of cross-section distributions, hence the change in the grid describes the evolution of the cross-section distribution Denote it by $q_t=\{ q_{1,t}, q_{2,t} \ldots q_{n,t} \}$, n= number of cells in the grid. The sequence of fractile transition probability matrices, associated with these grids, shows the intra-distribution mobility. Let us denote this sequence of fractiles transition probability matrices for regional disaggregates by $\{M_t\}$.[17]

Each transition probability matrix $\{M_t\}$ includes n x n cells. Therefore it is difficult to extract information about intra-distribution mobility. In order to do that we can use the notion of *Mobility Index*. A mobility index is a continuous scalar function defined over the set of transition matrices. Each index collapses the information about mobility contained in the $n(n-1)$ independent numbers of the matrix into a single number. From each time series of matrices $\{M_t\}$, each index defines a time series of mobility measures.

In this paper we are using three of these indices,[18] and a fourth one which additionally includes information on the quantile location. *(i)* Shorrocks (1978) proposes a measure of mobility the following form:

$$hm = \frac{n - tr(M)}{n-1} = \tag{4.4a}$$

$$= (\frac{n}{n-1})n^{-1}\sum_j (1 - M_{jj}), \tag{4.4b}$$

where M_{jj} is the probability of remaining in the state j and (1- M_{jj}) is the probability of exiting state j (non persistence). This index *(hm)*can be interpreted (see expression (4.4b)) as the inverse of the harmonic mean of the expected duration of remaining in a given part of the distribution. The higher *hm* the less persistence is in the transition matrix. *(ii)* A second index frequently used in the literature is the following :

$$e_2 = 1 - |\lambda_2|, \tag{4.5}$$

where λ_2 is the second largest eigenvalue of the transition matrix. To understand the intuition behind *e2* as an index of mobility notice that every stochastic matrix M has an eigenvalue equal to unity and the modulus of the others is smaller than one. If M implies a unique ergodic (long run) distribution, the sequence of matrices converges to this long-run at a speed given by the powers of the eigenvalues. In particular the rate of convergence is driven by the second largest eigenvalue. Consequently the second largest eigenvalue module is often used as a measure of the convergence speed. The higher the index *e2* the faster the convergence. *(iii)* Based on the same intuition there is another index, called it *ev*

$$ev = \frac{n - \sum_j |\lambda_j|}{n-1}, \tag{4.6}$$

where λ_j are the eigenvalues of M . The index *ev* relates positively to the average (not only the leading term) rate of convergence of the transition matrix towards the ergodic limit. Normally *ev* and *hm* are not related but when all the λ_j are real and positive *ev* coincides with *hm.* To see this notice that the trace of a matrix equals the sum of the eigenvalues, hence, *hm* can be written as:

$$hm = \frac{n - \sum_j \lambda_j}{n-1} \quad . \tag{4.4c}$$

These three indice (*hm, ev* and *e2)* are bounded in the interval [0, 1]. *(iv)* In addition we use the index of mobility proposed by Quah (1996), this index exploits simultaneously information on M and q. From each time-series of pairs $\{M_t, q_t\}$, each index defines a time series of mobility measures. Quah (1996) argues that the information on the quantiles sets is also relevant since it makes a difference moving from the lowest to the highest quantile when the latter are close or far away from each other. Hence, not only moving from one state (quantil) to another matters but also the location of those quantiles.

The index is derived from the autoregressive stochastic process corresponding to the evolution of the transition matrices. It is defined as the unity minus the correlation coefficient in that process. Quah (1996) defines the index as

$$ar = 1 - \rho_t \ , \tag{4.7}$$

where ρ_t is the correlation coefficient. Notice that a correlation coefficient is an indicator of predictability, i.e. of immobility.

The next section utilises the methodology above described to study regional employment dynamics and aggregate unemployment fluctuations in Europe.

4.- SOME RESULTS FOR EUROPE [19]

After questioning the validity of the existing literature on mismatch and proposing a new approach to analyse the relation between aggregate and disaggregate fluctuations, this section performs an empirical exercise which illustrates the proposed approach using data for regions of the European Union.

The choice of Europe,[20] in spite of the difficulties in data availability, does not need much justification. As it has been mentioned in previous sections understanding European unemployment in the past two decades is a challenge faced by macroeconomists, not only is its level high and persistent but also its behaviour differs from the OECD countries.

We do not try to explain European unemployment, we simply attempt to find some evidence on one of the multiple explanations that are present in the literature, that is regional shocks as a cause of unemployment fluctuations. Three different considerations have motivated this specific analysis: firstly, the regional shocks hypothesis seems to be compatible with the movement upward of the UV curve in Europe and the high unemployment rates correspond to certain regions, those with intensive heavy industry or agriculture. The second consideration is methodological; the existing empirical analysis on this question suffers from a few problems, discussed in section 3, which may prejudice the results. Thirdly the evidence from the traditional analysis is inconclusive.

An additional reason to choose European regions is that as a by-product of the [21] analysis it is possible to form an idea of whether the regional shocks in employment are symmetrically distributed, an issue that has been argued as relevant for the viability of the EMU. The reason for this is that a flexible exchange rate can balance the labour market shocks. Although the current paper will not consider this aspect, it will be treated in future work.

In this section the disaggregate refers to European regions and aggregate refers to Europe as a whole. The available data include 51 regions of similar population size[22] for a period of 31 years (1960-1990) (see data appendix). The basic variable for the aggregate is the European unemployment rate (u_t) and for disaggregates it is the log of changes in employment for each region after conditioning out the components which are common to all the regions.

The regional variable is meant to reflect the regional (idiosyncratic) shocks, however the log of changes in employment for each region is affected by aggregate and country-specific changes.

To substitute out all or a part of the aggregate influence we first choose a very simple variable: the growth rate of regional employment as a proportion of European employment, $h_{it} = \Delta\ log\ (N_{it}|N_{et}) = \Delta\ [log\ (N_{it}))- log\ (N_{et})\ ] = g_{it} - g_{et}$, where N_{it} is employment in region i at time t and N_{et} is European-wide employment at time t. $g_{it} = \Delta log\ N_{it}$ is employment growth rate in the region i, time t and $g_{et} = \Delta log\ N_{et}$ the employment growth rate in Europe.

Table 1b gives the contemporaneous correlation between the aggregate unemployment growth rates and the mobility indices for the cross-section dynamics of the disaggregates. The contemporaneous behaviour of the unemployment rates in Europe seems to be strongly related to the intra-distribution mobility of relative employment growth rates across regions. There is a positive and very high correlation in every case, ranking from 0.74 to 0.8.

Regarding the behaviour of the aggregate together with the quantile location (shape of the cross-section distribution), the contemporaneous correlation is negative for the lower quantiles, it is -0.7 for the 20th percentile, -0.75 for the 40th and only -0.30 in the case of the 60th percentile. A higher relative employment growth rate for the regions in the lower quantiles (i.e. as the lower quantiles approach the average) corresponds to a lower aggregate unemployment rate. The correlation of the aggregate with the 80th percentile and the maximum is positive and quite high: 0.64 and 0.79 respectively. Movements towards the average correspond to a lower aggregate unemployment rate.

The causality evidence, [23] is more complicated. We perform an exclusion restriction test in bivariate VARs (unemployment rates and, one by one, the measures that characterise the disaggregate dynamics). These kind of tests consist of testing the joint significance of the lags of a group of variables. The estimated VAR coefficient of unemployment suggests that it is an integrated variable. It is well known that if the variables in the VAR are integrated the exclusion tests in this context may have non-standard asymptotic properties. We follow a very simple alternative proposed in Dolado and Lütkepohl (1994), such that the test may be done directly on the coefficients (least squares estimators) of the VAR process in levels. It consists of fitting a VAR the order of which exceeds the true one. It does not require unit root test and is robust to the integration process properties.

Table 2b shows the marginal significance levels for the exclusion restriction test, in a bivariate VAR which includes our measures of intra-distribution mobility and the unemployment rate.

Testing in two lags systems shows no causality relations. The three lags analysis indicates that the indices *hm*, *ev* and *e2* help to predict aggregate unemployment rate and the opposite is also true for *e2*. The index that we will say is more efficient in the sense that incorporates more information, i.e. *ar* does not show any power to predict unemployment or vice versa.

Table 2a, for the quantile element, suggests that the 40th and 80th, percentiles cause aggregate unemployment rates. The msl for exclusion of the quantiles in the equation of the unemployment are 3% and 0.4% for 40th percentile, 10% and 2.3% for the 80th. Unemployment helps to predict the maximum (msl 1% and 0.5%.). The 60th percentile does not show any pattern of causality.

The previous normalisation assumes that the elasticity of labour reallocation across regions with respect to aggregate fluctuations is the same for each region and equals the unit. Relaxing this hypothesis implies to repeat the analysis above taking as the basic variable $h'_{it} = g_it - \hat{\beta}_i\, g_{et}$ [24] where $\hat{\beta}_i$ is the elasticity parameter estimated by fitting the following univariate process for each region:

$$g_{it}= \alpha_i + \beta_i\, \Delta\, log\, N_{et} + \varepsilon_{it} \tag{4.8}$$

How far to go with conditioning depends on what is understood by idiosyncratic or region-specific fluctuations. Part of the fluctuations in h'_{it} still may be not region-specific but common to all the regions in the same country, think of country-specific economic policy etc.. The sample analysed in this exercise, includes eleven countries that during the considered period have had different macroeconomic policies; five of these are divided in regions the rest are considered as a unique region.

To condition out the country-specific effect we take h'_{it} as the basic variable if the country is not divided into regions. For each one of the other countries, which are divided in regions, we fit [25]the following model.

$$g_{it}= \alpha_i + \beta_i^{`}\, \Delta\, log\, N_{et} + \gamma\, \Delta\, log\, N_{ct} + \varepsilon_{it} \tag{4.9}$$

where N_{ct} is country-wide employment, g_{ct} is country growth rate. g_{it} and $\Delta log\, N_{et}$ as before. Now the basic variable is $h'_{it} = g_{it^-}\ \hat{\beta}_i \Delta\, log\, N_{et} - \hat{\gamma}\ \Delta\, log\, N_{ct}$.

Tables 3a,b and 4a,b show the results of performing similar analysis to the one before after conditioning out the country-specific effects and allowing for different elasticity of labour reallocation across regions with respect to aggregate fluctuations. The contemporaneous correlation gives us the same picture as in the previous exercise. There is negative contemporaneous correlation for the lower quantiles and positive for the higher. The contemporaneous behaviour of unemployment rates and the intra-distribution mobility of regionally specific employment growth rate is very high and positive.

In regard to whether aggregate and disaggregate fluctuations are dynamically correlated we test causality as above. The evidence is that *e2* and *ev* cause aggregate unemployment. For *hm* and *ar* there is no evidence of causality (except for *hm* in the 4 lags system where msl is 10%). Unemployment does not help to predict the indices.

For the quantile location the causality evidence is that unemployment causes the maximum of the distribution. The 40th quantile causes unemployment and so does the 60th, for the 20th and 80th there is no causality evidence.

5. CONCLUSIONS AND REMARKS

1. There is an interesting dynamic relation between European unemployment and labour reallocation across regions that cannot be summarised with a simple (point in time) statistic in the way that the empirical literature does.

2. The contemporaneous behaviour of the unemployment rates in Europe seems to be strongly related to the dynamics of regional employment growth rates across regions (dynamics of the regional shocks). There is a positive and very high correlation between aggregate unemployment and intra-distribution mobility of the regional variable. Regarding the behaviour of the aggregate together with the quantile location (shape of the cross-section distribution) the contemporaneous correlation is negative for the lower quantiles and positive for the higher. In other words, lower quantiles moving up and higher moving down correspond to a lower aggregate unemployment rate.

3. If we have conditioned out the aggregate and country-specific effects correctly, we can conclude that there are some interesting relations between the aggregate and disaggregate fluctuations in the labour market. The intra-distribution mobility in the cross section distribution of employment growth rates (indices *ev* and *hm*) helps to predict aggregate unemployment.

Aggregate unemployment causes the maximum and the dynamics on the shape of the regional shocks distribution (40th and 60th percentiles and the maximum) helps to predict aggregate unemployment rates.

ENDNOTES

1 Idiosyncratic or disaggregate is understood as specific to a sector, region, skill or similar categories.

2 Either unemployment was previously below its natural rate or this rate has raised.

3 In line with the traditional approach in Business Cycle literature.

4 Bean & Pissarides (1990).

5 NAIRU is the unemployment rate compatible with price stability.

6 Jackman, Layard & Savoy (1990).

7 Ignores vacancies but it can be a good proxy to mismatch.

8 See for example Bentolila and Dolado (1990).

9 Formally: *Turbulence index* = $\sum|\Delta(N_i/N)|$, where N_i is employment in category *i* and *N* is total employment.

10 See Layard, Nickel and Jackman (1990).

11 They define predictable versus unpredictable component of the employment growth rates.

12 Danny Quah (1994a,1996).

13 Most of the literature looks at the correlation coefficient.

14 It exhibits a similar order of magnitude in both, cross-section and time series dimensions.

15 See Stokey and Lucas (1989).

16 Notice that characterising the disaggregates by using the standard deviation most likely we will loose a big deal of the information contained in (M_t, q_t).

17 Notice that the Markov process (M_t, q_t) is not necessarily stationary.

18 They have been used in Shorrocks(1986) and Genewe, Marshall and Zarking (1986) among others.

19 The calculations and graphics have been executed using Danny Quah's Time Series Random- Fields shell *tSrF*

20 Europe in this paper refers to the countries in the European Union.

21 Decressin and Fatas (1994) study in detail this issue for a similar data base and for the same regional partition. They compare mobility in Europe with that in the US to see whether it may compensate for the absence of flexibility of the exchange rate as a policy instrument. Buiter (1995) argues that the sort of labour mobility that can be a substitute of the flexibility in the exchange rate must be a temporal one, and notices that temporal migration does not happen even in the US and the monetary union is viable there.

22 For a more detailed description of the regions see Decressin, J and A. Fatas (1993).

23 I.e. whether past values of one variable help to predict values of the other variable.

24 Other possibilities of conditioning out the aggregate shocks would be including in the regression variables such as oil prices, etc. (conditioning on the causes of the common shocks) .

25 Estimating $g_{it} = \beta_i \; \Delta \; log \; N_{et} + \gamma \; \Delta \; log \; N_{ct} + \varepsilon it$ by pooled *OLS*, (under Swamy assumption) will yield unbiased and consistent estimator.

TABLE 1a : Contemporaneous Correlation: Unemployment and Regional Relative Employment Rates (quantiles)

	U	0.2	0.4	0.6	0.8	1.0
U	-	-0.771	-0.750	-0.305	+0.642	+0.794
q0 0.2	-	-	+0.801	+0.136	-0.745	-0.713
q1 0.4	-	-	-	+0.356	-0.67	-0.779
q2 0.6	-	-	-	-	+C.305	-0.279
q3 0.8	-	-	-	-	-	+0.645
q4 1.0	-	-	-	-	-	-

TABLE 1b : Contemporaneous Correlation: Unemployment and Regional Relative Employment Rates(mobility indexes)

U	-	*hm*	*ev*	*e2*	ar
U	-	+0.743	+0.819	+0.798	+0.795
hm	-	-	+0.973	+0.968	+09572
ev	-	-	-	+0.991	+0.976
e2	-	-	-	-	+0.975
ar	-	-	-	-	-

TABLE 2a : Granger Causality. Exclusion Restriction Marginal Significance Level *
Regional Relative Employment Rate (quantiles)

Quantile	System Lag Length 2		3	
q0:0.2	0.315,	0.652	0.222,	0.592
q1:0.4	0.030,	0.643	0.004,	0.679
q2:0.6	0.325,	0.794	0.370,	0.777
q3:0.8	0.107,	0.946	0.023,	0.659
q4:1.0	0.158,	0.005	0.093,	0.010

* For each lag length the first column is the Marginal Significance Level for excluding the corresponding quantile in the VAR for unemployment, the second one is for excluding unemployment from the VAR for the quantile.

TABLE 2b : Granger Causality. Exclusion Restriction. Marginal Significance Level *
Regional Relative Employment Rate (mobility indexes)

Mobility Indexes	System Lag Length 2	3
hm	0.323, 0.162	0.048, 0.400
ev	0.113, 0.488	0.009, 0.587
e2	0.205, 0.102	0.170, 0.037
ar	0.560, 0.517	0.677, 0.689

* For each lag length the first column is the Marginal Significance. Level for excluding the corresponding mobility index in the VAR for Unemployment, the second one is that for excluding unemployment from the VAR for the mobility index.

TABLE 3a : Contemporaneous correlation: Unemployment and Regional Employment Growth Rates after conditioning out the Europe-wide and country-specific effects. (quantiles)

	U	q0 0.2	q1 0.4	q2 0.6	q3 0.8	q4 1.0
U	1	- 0.7601	- 0.693	- 0.124	+0.769	+0.807
q0 0.2	-	-	+0.719	+0.114	- 0.725	-0.683
q1 0.4	-	-	-	+0.490	- 0.616	- 0.683
q2 0.6	-	-	-	-	+0.180	- 0.123
q3 0.8	-	-				- 0.816
q4 1.0	-	-	-	-	-	-

TABLE 3b: Contemporaneous correlation: Unemployment and Regional Employment Growth Rates after conditioning out the Europe-wide and country-specific effects. (mobility indexes)

	U	ar	*e2*	*ev*	*hm*
U	1	0.769	0.784	0.837	0.8014
ar	-	-	0.940	0.948	0.967
e2	-	-	-	0.983	0.982
ev	-	-	-	-	0.993
hm	-	-	-	-	-

TABLE 4a : Granger Causality. Exclusion Restriction. Marginal Significance Level * Unemployment and Regional Employment Growth Rates after conditioning out the Europe-wide and country-specific effects (quantiles).

Quantile	System Lag Length 2	3	4
q0 0.2	0.208, 0.387	0.304, 0.308	0.277, 0.416
ql 0.4	0.015, 0.843	0.003, 0.847	0.001 0.780
q2 0.6	0.216, 0.932	0.001, 0.539	0.009 0.002
q3 0.8	0.445, 0.332	0.183, 0.383	0.152, 0.119
q4 1.0	0.095, 0.004	0.060, 0.131	0.304, 0.003

* For each lag length the first column is the Marginal Significance Level for excluding the corresponding quantile in the VAR for Unemployment, the second entry is that for excluding unemployment from the VAR for the quantile.

TABLE 4b : Granger Causality. Exclusion Restriction. Marginal Significance Level * Unemployment and Regional Employment Growth Rates after conditioning out the Europe-wide and country-specific effects (mobility indices)

Mobility Indexes	System Lag Length 2	3	4
hm	0.125, 0.608	0.124, 0.411	0.106, 0.127
ev	0.033, 0.778	0.000, 0.394	0.001, 0.322
e2	0.062, 0.144	0.000, 0.174	0.001, 0.152
ar	0.377, 0.478	0.374, 0.764	0.261, 0.113

* For each lag length the first column is the Marginal Significance. Level for excluding the corresponding mobility index in the VAR for Unemployment, the second one is that for excluding unemployment from the VAR for the mobility index.

TABLE 5a : DIRECTION OF THE CAUSALITY
Unemployment and Regional Employment Rates

QUANTILES

	2 lags	*3 lags*
q0	X	X
q1	→	→
q2	X	X
q3	X	→
q4	←	←→

MOBILITY INDICES

	2 lags	*3 lags*
hm	X	→
ev	X	→
e2	X	←→
ar	X	X

→ : from disaggregate to aggregate
← : from aggregate to disaggregate
X : no causality, (msl >0.1)

TABLE 5b: DIRECTION OF THE CAUSALITY
Unemployment and Regional Employment Rates
(conditioning out Europe-wide and country-specific effects)

QUANTILES

	2 lags	*3 lags*	*4 lags*
q0	X	X	X
q1	→	→	→
q2	X	→	←→
q3	X	X	X
q4	←→	←→	←

MOBILITY INDICES

	2 lags	*3 lags*	*4 lags*
hm	X	X	X
ev	→	→	→
e2	→	→	→
ar	X	X	X

→ : from disaggregate to aggregate
← : from aggregate to disaggregate
X : no causality (msl >0.1)

DATA APPENDIX : Variables and sources and specific samples:

National employment. and National Labour Force
Source: *OECD Labour Force Survey.*
Time Sample: 1960-1990
Cross-Section Sample: 11 countries: France, Germany, Italy, Spain, UK, Belgium, Denmark, Greece, Ireland, Netherlands, Portugal.

European Employment (Nt) and European Labour Force (Lt): calculated by adding country variables.
Unemployment rate. Defined as ut = (Lt - Nt) / Lt

Regional Employment.
Source: *OECD, Regional Employment and Unemployment 1960-87*
Time-sample: 1960-1990 (max), annual data.
Cross-section sample: 51 regions

France: 8 regions, Germany: 8 regions, Italy:11 regions, Spain:7 regions(Bentolila and Dolado), UK: 11 regions, Belgium :1 region, Denmark: 1 region, Greece :1 region, Ireland :1 region, Nether.: 1 region, Portugal: 1 region.

REFERENCES

Abraham, Katherine G., (1991), "Mismatch and Labour Mobility: Some Final Remarks" in F.P. Schioppa, Ed.(1991) *Mismatch and Labour Mobility*, Cambridge: Cambridge University Press, 453-481.

Abraham, Katherine G. and Katz, L.F., (1986), " Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?", *Journal of Political Economy*, 94, 507-22

Barro, Robert and Sala i Martin, X., (1991), "Convergence Across States and Regions", *Brookings Papers on Economic Activity*, (1):107-182, April.

Bean, Charlie. R., R. Layard and Nickell, S., (1986), "The Rise in Unemployment : A Multi Country Study"

Bentolila, Samuel and Dolado, J.J., (1991), " Mismatch and Internal Migration in Spain, 1962-86, " in F.P. Schioppa, Ed.(1991) *Mismatch and Labour Mobility*, Cambridge: Cambridge University Press, 453-481.

Buiter, W., (1995) " Macroeconomic Policy During a Transition to Monetary Union".*mimeo*

Dolado, Juan J. and Lütkepohl, H., (1994), "Making Wald Tests Work for Cointegrated VAR Systems" CEMFI W.P. no. 9424

Decressin, Jorg and Fatas, A., (1993)," Regional Labor Market Dynamics in Europe and Implications for EMU" CEPR no. 1085.

Entorf, H., (1993), " Do Aggregate Measures of Mismatch Measure Mismatch? A Time Series Analysis of Existing Concepts", INSEE-CREST Discussion Paper No. 9318

Evans, G. W., (1193), " Sectoral Imbalance and Unemployment in the United Kingdom: 1963-84. *Oxford Economic Papers* 45 , 440-456

Genewe, J., Marshall, R.C. and Zarking, G.A.,(1986), " Mobility indices in Continuous Time Markov Chains", *Econometrica* 54(6):1407-1423, Nov.

Kazamaki. E., (1994) "Sectoral Employment Shifts and Unemployment" *Labour* 8(3) 377-394.

Lilien, D. M., (1982), " Sectoral Shifts and Cyclical Unemployment" *Journal of Political Economy* 90(4) 777-793)

Lippi, Marco and Reichlin, L., (1994)," Diffusion of Technical Change and the Decomposition of Output into Trend and Cycle" *Review of Economics Studies* , January.

Neeling J., (1987) "Sectoral Shifts and Canadian Unemployment". *The Review of Economics and Statistics* 69: 718-723.

Quah, D., (1992),"The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds" *Econometrica* 60(1), 107-118, January

Quah, D., (1993)," Exploiting Cross section Variation for Unit Root Inference in Dynamic Data" Discussion Paper, EM/93/270

Quah, D., (1994a)," One Business Cycle and one trend from (many) many Disaggregates" *European Economic Review,*

Quah, D., (1994b)," Empirics for Economic Growth and Convergence" Working Paper, Economics Dept., LSE, September.

Quah, D., (1996)," Aggregate and Regional Disaggregate Income Fluctuations" *Empirical Economics* . Forthcoming.

Quah, D. and Sargent, T., (1993), " A Dynamic Index Model for Large Cross Sections" in *Business Cycles Indicators and Forecasting*, ed. James Stock and Mark Watson, vol 28 (Chicago IL: University of Chicago Press and NBRE) chapter 7 pp 285-306.

Shorrocks, A. F., (1978) "The measurement of Mobility" *Econometrica* 46(5), 1013-1024, September.