THE EUROPEAN UNIVERSITY INSTITUTE

Department of Economics

CHANGES IN THE WORLD INCOME DISTRIBUTION: A NON-PARAMETRIC APPROACH TO CHALLENGE THE NEO-CLASSICAL CONVERGENCE ARGUMENT.

Alain DESDOIGTS

Thesis submitted for assessment with a view of obtaining
the Degree of Doctor of the European University Institute

Florence, June 1994

THE EUROPEAN UNIVERSITY INSTITUTE

Department of Economics

CHANGES IN THE WORLD INCOME DISTRIBUTION: A NON-PARAMETRIC APPROACH TO CHALLENGE THE NEO-CLASSICAL CONVERGENCE ARGUMENT.

Alain DESDOIGTS

The thesis committee consists of:

Prof. Daniel Cohen, CEPREMAP, Paris
Prof. Wolfgang Härdle, Humboldt University Berlin
Prof. Alan Kirman, EUI, Supervisor
Prof. Danny Quah, LSE
Prof. Robert Waldmann, EUI

Contents.

1/ Introduction.

The emergence of new technologies in industrial countries has increased the importance of economies of scale in production so that competitive markets will no longer be able to ensure Pareto efficient outcomes. Reasons for the occurence of large scale production may be purely technological or may derive from the high fixed costs of research and development (R&D).

These features are departures from the Neo-classical growth theory, which is based on competitive markets, identical rational individuals, diminishing returns to capital and labour and constant returns to scale. Within this framework, because marginal products are positive and exhibit diminishing returns to each input, the physical capital grows more slowly when per capita income exceeds its equilibrium level and more rapidly when per capita income is below equilibrium. As a consequence, the dynamics are such that an economy will reach a state of steady growth where capital will grow apace with the labour force. In this setting, Solow (1956) assumes that sustained growth is ensured by an unexplained technological progress. Secondly, because of the assumption of diminishing returns to each input, per capita capital growth is only transitory and the determinants of the propensity to save do determine only per capita level of income and consumption, leading to a situation in which any public intervention affects income levels, but is unsuccessful regarding the rates of growth in the long-run.

An economy's growth experience being a summary of all activities of an entire society, is a huge conceptual framework which must be explained before its process and the interactions between economic and non-economic, i.e. political and institutional, factors can be identified. Endogenous growth theory proposed an alternative approach whereby steady-states are issued by an endogenous explanation of the source of technological change. This led, on the one hand, to a revival of interest in growth modelisation which emphasizes the micro-foundations of the growth process by identifying engines of growth and its mechanics, as well as the motivating forces that affect the accumulation of reproducible factors. On the other hand, macroeconomists returned to growth theory and turned to development because Endogenous growth theory shows that distortions and policy interventions affect both the level of income and the steady-state growth rate.

With the existence of scale economies which in accordance to the Marshall's argument, will be external to the firm but internal to the industry, the forces motivating and generating continuous technological change become endogenous and obey some explicit rules. Romer (1986) allows for a convex technology involving external economies through the accumulation of knowledge reached by the community over the years, so that per capita output will now exhibit increasing returns to scale in reproducible factors within a decentralized market. Within this framework, technological change will result from the decisions of the households becoming thus, endogenous. Then, assuming that the stock of knowledge is a public good, the presence of externalities lead to sub optimal equilibria. The resource allocation will not be Pareto optimal because firms may, for instance, act as "free-riders" and subsidies encouraging the accumulation of the reproducible factor which acts as the engine of growth, will lead to the social optimum. In other words, tax policies that encourage investment can raise the social return and thus the rate of growth. Therefore, the presence of externalities strengthens the role of the state in the optimal dynamics of some relevant economic variables such as education and R&D.

Romer assumes, as in the Neo-classical framework, that knowledge is disembodied; that is, it consists of books, publications which allow the aggregate capital stock to be used as a proxy for the aggregate stock of knowledge, thus allowing to concentrate on a model with a unique state variable. However, knowledge, i.e. technological progress, might be issued from the accumulation of human capital. It is said to be embodied in individuals and allowed to grow without any limit as if there were perfect inter-generational transfers (Lucas; 1988). There can be no doubt that, apart from the physical capital accumulation, improvements in technology associated with human capital accumulation contribute greatly to growth. The influences responsible for increasing returns are strengthened by the discovery of non-natural resources and the uses for them, and the growth of scientific knowledge in general.

Economies of scale introduced in these frameworks are still purely technological. The next step is the allowing for economies of scale because of the presence of high fixed costs in the production of new knowledge, so that marginal costs are smaller than the average costs. In a competitive framework, firms would suffer losses and therefore be constrained to behave as monopolists when setting the price of their new technics. Technological change is endogenous but arises, here, when self-interested people have the opportunity to benefit. In other words, the incentive for innovations holds in the monopoly rent as in a Schumpeterian economy. This kind of framework (See, among others, Romer, 1990; Grossman and Helpman, 1989 a-b) relies on a combination of such factors and some external economies because accumulation of knowledge although being partially excludable, is also assumed to be a non-rival good, and can therefore improve the R&D sector efficiency, leading social returns of production of knowledge to be greater than private returns. Again, the decentralized equilibrium will be sub optimal because of maximization of entrepreneurial profits which involve elements of imperfect competition. At this stage, Endogenous growth is important as it allows any institution to intervene and support the allocation of resources and to coordinate both domestic and international policies. This has permitted the growth theory to make it up with development and international trade theories.

Basically, by assigning no other source of growth than an exogenous technological progress, the Neo-classical framework which assumes a perfect international capital mobility, failed to account for permanent international differences across countries. This has stimulated work on Endogenous growth. Indeed, because of the assumed increasing marginal productivity of knowledge, both the growth rates and the per capita income level depend on the economy's initial physical and/or human capital endowments, i.e. history. Therefore, a change in the initial capital stock will now generate a permanent change in per capita income and consumption, leading a whole economy to exhibit "hysterisis" phenomena. First, under the condition of increasing returns to scale, the "primitive" accumulation or the early stage of development which allows for greater productivity, can explain the faster growth and cumulative inequalities between relatively prosperous and relatively poor countries. Secondly, externalities due to the public character of knowledge will also give rise to permanent disparities across countries because they tend to cluster around geographical centres like cities, countries and political entities (Lucas; 1988). Thus, the scale of these externalities will not extend to all forms of economic activity. They will be such that a threshold level of development will be necessary if the less developed countries are to catch up on innovations from leading countries.

The arguments which underly the following discussion are twofold. On the one hand, the empirical literature that uses inter-country income differences in order to analyze the implications of inequality for individuals to enjoy material welfare, relies either on trends in Gross Domestic Product (GDP) and/or statistical indicators like Lorenz curves, Gini coefficients and relative variance around the mean levels of relative productivity. The extent and size of changes in inequality during the post-Second World War period, have been examined by Summers, Kravis, and Heston (1984), among others, who find that the world as a whole experienced a "very slight rise" in inequality over this period.

On the other hand, the Neo-classical growth theory argues that backward countries will be able to catch up a leader whose productivity improvements are expected to slow down as the highest productivity level comes closer to its steady-state value. One central argument to this key economic issue of convergence across a world-wide set of countries, is that learning and imitating an existing technology might be easier than inventing and testing a new one. Thus, international transfers of knowledge are given a central role in the economic issue of convergence, as initially poorer countries are expected to be able to make use of current best practice procedures already in use in the more productive economies. This argument is challenged by "new" models belonging to the so-called Endogenous growth theory, which allow for permanent differences in growth rates leading therefore to the existence of multiple equilibria. What is more, under some initial conditions, it is impossible to realize the potential to catch up, leading some countries in a development trap.

Empirical studies focusing on the convergence hypothesis grew apace with those "new" frameworks which consist in identifying original engines of growth and mechanisms of transmission of the latter in the economy which must be the outcome of internal forces thus challenging the introduction of an exogenous growth rate of technological progress which acts as an "ignored" issue as in the original Solow paper (1956). Basically, for the catch-up hypothesis to be accepted, productivity growth rates across countries must be inversely related to their initial levels. In other words, within a cross-section coutry analysis, a negative coefficient on the initial level of real income provides evidence of convergence and indicates some type of diminishing returns for the production function expressed in terms either of per capita or per worker output. However, from a large body of empirical papers, it emerged that the results are very sensitive to both the sample and the period selected. De Long (1988) suggests that evidence for convergence for the richest countries may be due to sample selection bias. Other studies suggest that evidence for the industrial countries is relatively robust to sample choice, but does not extend to the poorer countries, at least in terms of a simple negative correlation between initial income and subsequent growth of per capita income.

A clear insight about what is meant by convergence within the Neo-classical framework is provided by Romer, Mankiw, and Weil (1990). Basically, poor economies are indeed expected to grow faster than rich ones if they are homogeneous with respect to their steady-state value and the exogenous rate of technological progress. In other words, only conditional convergence can be inferred from the Neo-classical growth model. Convergence occurs only once a set of control variables like investment, school-enrolment, political stability etc..., meant to control for microeconomic heterogeneity, is added to the empirical model. On the one hand, there exists strong evidence of convergence once the determinants of each own country's steady-state has been controlled for (Romer, Mankiw, and Weil (1990); Barro (1991)). On the other hand, this is not sufficient to reject the existence of stable multiple equilibria because data might more consistently support a theory of different aggregate production functions, leading thus to local rather than global convergence. This point has been made by Durlauf and Johnson (1992) whose findings exhibit greater and more significant partial associations between growth rates and initial incomes among homogeneous groups of countries in terms of both initial income and literacy rate, than the one estimated for the whole sample.

It remains as the consistency of the estimators is crucially dependent on the model specification and because of the heterogeneity of the major empirical studies, that among economists, no consensus exists about the issue of global versus local convergence, i.e. the existence of multiple versus a unique and global equilibrium. Indeed, a major drawback to these empirical approaches, is that the functional form of the model is specified directly or indirectly via the specification of the production function leading to non "robust" estimation as illustrated in the Levine and Renelt sensitivity analysis (1992). Therefore, in order to avoid misspecification of functional form which usually leads to inconsistent estimators, a more flexible methodology is needed in order to get new insights about the information present in both the structure of incomes and growth rates distributions across a world-wide set of economies. Indeed, it is rather difficult to discern consistent relationships among the existing empirical literature as it uses a wide range of "right-hand-side" variables, use different countries, measure variables differently, and employ different data sets (Levine and Renelt; 1991). In order to highlight these problems, we propose an extensive use of non-parametric smoothing techniques which provide comprehensive graphical descriptions of the data, and which give direct information about the problem at hand; that is, with no a priori information, we want first to discuss the extent and direction of changes in global inequality over time and then to discuss the catch-up hypothesis for a large set of countries, observations of which are issued by the United Nations International Comparison Programme (ICP).

Parametric analysis uses the information available in the data only from the postulated model and thus, may be biased in our situation thus requiring methods which are robust for misspecification of functional form. There is a large variety of non-parametric methods on which has been written a considerable amount of literature. We will concentrate here on using kernel density estimation. Indeed, because of their flexibility, kernel density estimates, which are the most commonly used non-parametric density estimates, apart from the histogram, in both univariate and multivariate situations, are conceptually straightforward, and easily computed. Secondly, the asymptotic theory involved is well-developed. They are frequently used as a suitable mean of displaying the features of a data set.

They allow to estimate descriptive features such as the number and locations of modes present in the structures of incomes and growth rates distributions as well as tail behaviours and skewnesses... Indeed, some stylized facts which are not immediately apparent may arise when using non-parametric techniques, which are well-known for detecting structures deviating from traditional parametric forms. Note that Quah (1993 b) established that coefficients of arbitrary signs in such cross-country regressions are consistent with an unchanging cross-section distribution of incomes and therefore turned as well to hint at the usefulness of non-parametric estimation in discussing convergence and cross-sectional dynamics.

As already mentioned, tests of per capita income convergence in the parametric empirical literature consist in regressing the average growth rates over a finite period and across countries on the initial per capita income conditioning on hypothesized explanatory variables like investment, school-enrolment, government expenditures...The latter allow for control of differences in the permanent growth component, which might differ across a set of heterogeneous countries.

The "new" growth theory stresses country characteristics such as education levels or political stability, as dominant determinants of growth. Thus, the aim of these cross-country empirical studies also consists in examining empirical linkages between long-run growth rates and hypothesized economic, political and institutional variables. More than 50 variables have been found to have a significant partial association with growth in at least one regression. As the "robustness" of these variables, except the share of income devoted to investment, has been challenged (Levine and Renelt; 1992), meaning that researchers cannot agree on the confidence they should place in these kinds of results, we propose comparing these empirical findings with some new information which arises when using non-parametric techniques which are well-known for diagnostic checking of an estimated parametric model. Basically, traditional parametric methods nest on strong assumptions so that results may not be robust to specifications like linearity in the regression curve where Y and X are assumed to be normally distributed.

Misspecification of the hypothesized functional form may have serious consequences for both the econometric result as well as for recommanding policy changes that are expected to increase the chances of the lagging countries to perform better. Therefore, the approximation of the conditional mean of average growth rates given any explanatory variable, requires the use of more flexible functional forms escaping from a finite set of parameters, thus, doing justice to the amount of information which is available. Again, and because *it lets the data speak for themselves*, non-parametric techniques like kernel estimation, which is understandable at an intuitive level and a suitable tool for detecting special features like monotonicity or unimodality, provide new insights of how growth rates might depend on any of these variables and allow to transmit clear and straightforward readable information. Non-parametric techniques as they do specify no functional form at all, are robust for misspecification and allow consistent estimation of a regression model where the only specification that is involved concerns the choice of the dependent variable and the independent variables. Note that because of its computational efficiency, we will use the so-called WARPing (Weighted Averaging of Rounded Points) framework which also benefits from all the statistical properties of traditional kernel methods (Scott, 1995b; Härdle 1990b).

Finally, within general equilibrium processes, the elaboration of growth models with technological progress needs some restrictive assumptions about the consumer's preferences, i.e. uniform and proportional expansion of demand. Otherwise, they will exhibit neither quantitative nor qualitative solutions. However, deviations from homotheticity may alter production patterns and growth rates (Lucas; 1988). Secondly, steady states where everything grows in exact proportion are not realities in situations where the proportions and the growth rate are themselves the outcome of internal economic forces (Kaldor; 1985). Therefore, we propose first to challenge the homotheticity assumption and to examine empirical linkages between some expenditure patterns and growth, which are the last component of income and which have not received much attention from researchers. We will focus our attention on the last variable of interest introduced in almost all empirical studies on the basis of the "new" growth theory; that is, human capital accumulation proxied by both aggregated and disaggregated households' education expenditure patterns.

The discussion will proceed as follows: Section 2 provides a brief empirical literature review. In Section 3, both the data and the methodology used in order to discuss the economic issue of convergence, are presented. The direction and the extent of income inequality is discussed in section 4. Convergence and the amount of mobility within the system are finally analyzed in Sections 5 and 6.

The empirical linkages between some traditional components of income, i.e. investment and government expenditures, at both an aggregated and disaggregated level, and long-run growth are examined in Section 7. In Section 8, there follows a brief description of world-wide expenditure patterns and how some of them may be related to growth. Finally, Section 9 summarizes and concludes on the findings of this empirical study.

2/ "Catching-up": A Brief Empirical Literature Review.

In the empirical history, the theoretical puzzle of convergence is better known as the phenomenon of opportunity to catch-up. Even though this is yet to be fully explained, empirical evidence suggesting convergence in terms of labour productivity and examining the growth performance across both homogeneous and heterogeneous groups of countries has emerged since the mid-eighties.

The central idea of the catch-up hypothesis is concerned with the level of technology embodied in a country's capital stock. Indeed, the expectation that labour productivity will increase more quickly in backward countries is mainly because learning and imitating an existing technology should be easier than inventing and testing a new one. Secondly, countries with lower levels of industrialisation may have greater returns relative to the most advanced economies, in training labour, then, reallocating it between agriculture and industry. Thirdly, they might have greater opportunities to exploit the possibilities of advanced scale-dependent technologies.

- *Average growth rates versus initial levels of productivity: a negative association is interpreted as convergence.*

As a consequence, productivity growth rates across countries are expected to be inversely related to their initial levels. In other words, a negative coefficient on the initial level of real income provides evidence of convergence and indicates some type of diminishing returns for the production function expressed in terms either of per capita or per worker output. Abramowitz (1986) and Baumol (1986), using Maddison's 1870-1979 data[1], support the hypothesis of convergence in productivity levels for a set of sixteen presently industrialized countries[2]. Indeed, both studies do exhibit a long-run tendency towards convergence of levels of per capita product or, alternatively, of per worker product. Baumol is able to explain 88% of growth only considering the initial Gross Domestic Product per worker hour, and simply plotting the resulting trends clearly shows that convergence occurred over the last century. Abramowitz also looked at the association between initial levels and average growth rates of labour productivity. Starting with a sub period from 1870 to 1890, he then computed the above rank correlation for nine subsequent key years from 1870 to 1979[3]. He finds an inverse correlation which confirms the potential to catch up for technologically backward countries. What is more, the estimated coefficient is higher as the length of the sub period under study increases, suggesting that the higher a country's productivity level in 1870, the more slowly that level grows in the following century.

---

[1] Maddison's estimates of productivity levels are mainly extrapolations of base levels established by Kravis, Heston, and Summers (1978a) in their International Comparisons of Real Product and Purchasing Power. The up to data version of this data table is presented in the following section.

[2] The countries are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Italy, Japan, Netherlands, Norway, Sweden, Switzerland, United Kingdom, United States.

[3] The years are 1870, 1890, 1913, 1929, 1938, 1950, 1960, 1973 and 1979.

Abramowitz also computed the averages of the productivity levels of the various countries relative to that of the United States as well as the relative variance around the mean level for the above nine key years. If the hypothesis of convergence is accepted, the means should tend to one while the variances are expected to tend to zero as the length of the period increases. While the variance among the relative productivity levels of the 15 follower countries does indeed continuously decline from 0.5 in 1870 to 0.15 in 1979, despite being almost entirely a post-Second World War phenomenon, the calculated averages of GDP per worker hour relative to the US on the other hand, exhibit zero indication for convergence[4]. Nevertheless, the author claims that the evidence for a reduction among countries in productivity differentials remains fairly strong as convergence is a powerful and continuous element in the growth performance displayed by the countries included in the sample.

However, if the potential to catch up is indeed expected to be realized, countries are constrained by their own "social capability." This allows Abramowitz to justify the existence of a leader standing out from the followers who, initially, were latecomers lacking experience in large scale production and commerce, and definitely needing to improve both their general and technical education. Note that institutional commitments might also act as a constraint on convergence; some will forge ahead while others will fall behind. However, once countries have reached a threshold level of development in terms of "social capability," the process of catching-up will be driven by interactions between followers and leaders via flows of capital, final goods, applied knowledge...

---

[4] Abramowitz (1986; pp. 393) argues that: "this is because the productivity increases reflect more than gaps in technology and in reproducible capital intensity, with respect to which catch-up is presumably possible," like resource endowments and unequal abilities to persue paths of progress.

Baumol proposes a fairly similar explanation to this sharing process, based on the international public goods property of productivity policy. He suggests that spillovers from leader economies to followers have played a crucial role in affecting productivity growth. Indeed, the benefits of innovation and investment flowing from leaders to followers may lead to feedbacks from the latter which will try to exploit the possibilities of advanced scale-dependent technologies by import substitution and expansion of exports, thus leading the economies into the so-called "Schumpeterian race." Finally, convergence is expected within this sharing process because followers have more to learn from the most advanced economies meaning that benefit flows might be unbalanced in favour of the followers, at least among countries which exhibit similar patterns of consumption and production. Indeed, Baumol added poor, intermediate, and centrally planned economies to its original sample of presently industrialized countries. Then, observing this new cloud of observations, he argued that the less developed countries will not benefit to the same extent by the sharing process involved in the "Schumpeterian race." These countries might be unable to assimilate modern technologies and their productivity level will not necessarily converge to those of industrial nations before they reach a threshold level of development. Therefore, Baumol concluded that there might exist more than one "convergence club" and while the convergence phenomenon extends to both intermediate and centrally planned economies as suggested by post-Second World War data, the poorer less developed countries show no such trend.

- *Sample selection bias.*

If we have stressed the fact that the above sample consisted of presently industrialized countries, it is because to observe convergence versus divergence may depend crucially on the sample selection criteria[5]. Bradford De Long (1988; pp. 1139) comments on the ex-post sample of countries which belongs to what Baumol calls the "convergence club" and which suffers from selection bias: "only a regression run on an ex-ante sample, a sample not of nations that have converged but of nations that seemed likely to converge, can tell us whether growth since 1870 exhibits convergence." Therefore, he built up what he called the "once rich twenty-two" sample including countries possessing at this time a social capability for rapid industrialisation and removing Japan but including nations richer than Finland (the second poorest country after Japan in Maddison's sample). This led him to add countries like Argentina, Chile, East-Germany, Ireland, New Zealand, Portugal and Spain to the original sample, all of which provided at the beginning of the period some strong evidence of potential to belong to the presumed "convergence club." However, the author was struck by the failure of Chile, Argentina and Portugal to converge. After trying to explain for convergence first in terms of ex-ante political stability as Abramowitz suggested, and then in terms of religious establishment as the "Protestant ethic" argument suggests[6], the author concludes that evidence in favour of convergence is no greater than that of forces making for divergence.

Bradford De Long also observed that although the standard deviations of Log GDP continuously declined over the last century for Maddison's sixteen, it slightly increased for the once rich twenty-two. Therefore, the evidence for the richest countries may be due to sample selection bias. Other studies suggest that the evidence for the industrial countries is relatively robust to sample choice, but does not extend to the poorer

---

[5] Abramowitz notes (pp. 394): "[my sample] is a biased sample in that its members consist of countries all of whom have successfully entered into the process of modern economic growth."

[6] The ex-ante measure of democracy is defined as inclusion in the electorate of more than half the adult male population. The religious establishment is characterized in the regression model by a dummy variable which is one for Protestant, one-half for mixed, and zero for Catholic nations.

countries, at least in terms of a simple correlation between starting values and subsequent growth.

In reply to this insightful comment, Baumol and Wolff (1988) estimated coefficients of variation, i.e. standard deviations divided by the mean for a set of 19 European countries[7], divided into sub groups where in the first one figure out the top eight countries in terms of GDP per capita in 1870, in the second one, the top 9, and so on, until the top 14. Only the top eight sub group supports the convergence hypothesis -i.e. decreasing coefficients of variation over time. Note that convergence mainly occurred from 1860 to 1913. On the other hand, the remaining groups display growing divergence until 1880 while this process began to erode toward the end of the period, so that globally, the catch-up hypothesis should be rejected as heterogeneity increases within the sample.

In a second step, Baumol and Wolff analyze post-Second World War growth performance for a set of 72 countries issued by the Summers and Heston data Table. They built up and plotted, respectively on the X and Y-axis, an ex-ante stratification of countries, calculating moving averages of 1950 per capita income for subsets of 10 countries and their corresponding average growth rates over the period. Their results suggest first the existence of a development trap for the initially poorest countries, while near the median, the potential to catch up can be realized. Finally, they re-ranked the Summers-Heston countries on a 1950 basis and calculated the coefficients of variation over time, corresponding to sub groups ranking from the top 10 to the top 60 countries. Then, plotting some of these time series, they showed that convergence occurred for the top

---

[7] The authors used Bairoch's estimates of GNP percapita (1976) for 19 European countries where data are available from 1830 to 1913.

10 up to 14 countries. On the other hand, if homogeneisation initially occurred for larger groups of countries, it was definitely lost during the second half of the period under study - i.e. 1965-80.

It emerges from these papers that no definitive conclusion can be drawn either in favour of or against convergence. Indeed, the results are very sensitive to both the sample and the period which are selected. Convergence can be observed within a group of countries which, ex-post, has reached very similar patterns of productivity levels. However, what about convergence between countries which belonged initially to a homogeneous group in terms of "social capability" and countries which are "out of the race?" Endogenous growth theory allowing for persistent growth rates differences and for the existence of a development trap and multiple equilibria, appears, in the long-run, to better fit the reality.

Thus, in order to accept the catch-up hypothesis, one must rule out the sample selection bias and test the parameter stability over different periods to control for cyclical differences. Dowrick and Nguyen (1989) extensively tested the convergence hypothesis among 24 OECD countries in the post-Second World War period. First, measures of dispersion (coefficient of variations and standard deviation of Log output) support the null hypothesis of convergence as they both decline continuously from 1950 to 1985. Still, while the process of convergence appears to be weak since 1973, nonparametric tests show that the poorer half of the sample has grown faster than the richer half not only over the entire period but also over three sub periods (1950-60, 1960-73, 1973-85). Secondly, a regression analysis of growth rates of aggregate output[8] given the initial real GDP per

---

[8] The authors refer the growth rate of the dependent variable to aggregate output rather than to output percapita. Indeed, they argue that "a crucial distinction exists between, on the one hand, the

worker allows for more than 50% of growth differences among OECD countries (even excluding Japan from the original sample). As in previous papers, the negative sign of the coefficient supports the idea that countries with initially lower incomes tended to grow faster over the period. Finally, as controlling for capital and labour input growth leads to an increase in the magnitude and statistical significance of the coefficient, OECD post war convergence does not seem to have been due to either higher rates of investment or to more rapid rises in labour participation in the poorer countries. Thus, these results strongly support the total factor productivity catch-up hypothesis. To extend the original sample to a world wide set of economies, it can be shown that TFP catch-up has been a dominant and stable feature of the world as a whole with the exception of the poorer of the non OECD economies, which have been, on the other hand, characterized by rather low investment ratios relative to rapidly expanding populations[9].

Although post-war data support convergence among the industrialized countries, there is no doubt that very different growth performances have been experienced by the less developed countries as well as by some of the countries which ex-ante exhibited some potential to catch up, e.g. South-American economies. Note as well that if one crucial issue remains to explain how come countries either achieved convergence or felt behind, this might not be assessed without controlling for productivity catch-up, on the one hand, and on the other hand, it appears that the magnitude and the significance of either total factor

---

convergence (or divergence) of percapita income or labour productivity and on the other hand, any tendency for catching-up in levels of total factor productivity (TFP). Of course, TFP catch-up implies a tendency for income levels to converge, but such a tendency may be masked or exaggerated if factor intensity growth varies systematically with income." (Dowrick and NGuyen; 1989, pp.1010).

[9] Note that no sample selection bias arises as "the estimate of the rate of TFP catch-up which has been derived from the OECD sample (excluding Japan) is very similar to the estimated rate for the ex-ante selection of countries which were relatively rich in 1950". Dowrick and NGuyen (1989; pp. 1022)

productivity or per capita productivity catch-up, increases as one adds controlling variables, e.g. investment share, employment, in the model.

- *Conditional convergence.*

The key economic issue of convergence as well as economic growth in a cross-section of countries has, since De Long's criticism, been extensively examined. Along the same lines as Abramowitz and Baumol, Barro (1991) calculated the average growth rates for a set of 98 countries in the period 1960-85. He, then tested whether these average growth rates across economies are inversely related to the absolute level of per capita GDP. No significant relationship - either negative as one would have expected according to the catch-up hypothesis, or positive which would have rejected the hypothesis- was discovered. However, by adding school-enrolment observed at the beginning of the period, as a proxy to human capital in his regression model, he finds respectively negative and positive significant partial associations between growth rates and 1960 GDP per capita and school-enrolment at both the secondary and primary levels.

In this "modified sense," he thus finds that the data do support the convergence hypothesis underlying the Neo-classical growth model; that is, poor countries are expected to catch up with richer countries if they exhibit high human capital levels relative to their level of per capita GDP. Then, as suggested by some arguments of traditional growth models, he also tested the influence of fertility and investment on both per capita GDP and the stock of human capital. He finds that for a given value of per capita GDP, low (high) levels of human capital were associated with high (low) net fertility and low (high) ratio of private investment, while per capita GDP is not significantly related to fertility. He also finds that growth is inversely related to the share of government consumption in GDP and negatively related to measures of political instability like the number of revolutions and coups per years, and the number of political assassinations. Finally, growth rates are inversely correlated with a proxy for market distortion[10].

At this stage of the discussion, and comparing "Barro regressions" results with previous studies presented earlier, we have to note first that technological transfer is largely embodied in human capital, and maybe to a larger extent than in physical capital, due for instance, to developments in the technology of information and communication as suggested by Dowrick and Nguyen (1989). Secondly, political stability reveals itself to be quite important as a determinant of growth, thus challenging Bradford De Long's result (1988). Finally, as Barro (1990) argues, the negative association between the government consumption share of GDP and growth performance might find an explanation in the absence of the direct influence of the former on private productivity leading only to lower saving and growth through the distortion effects from taxation or government expenditures programmes.

It should be remembered that the Neo-classical growth model considers only equilibrium growth paths within a closed economy, where a full utilisation of labour and capital are established through sufficiently flexible factor prices. Thus, it is a justification for regarding a trend of a single country as a steady-state equilibrium and cannot directly be extended to comparisons among countries. Indeed, one needs to make some specific assumptions about preferences and the underlying technology -i.e. the production function. The latter must be identical across countries if one wants to argue that the Neo-classical framework predicts that initially poorer economies will tend to grow faster in per capita

---

[10] Market distortion is measured by the average deviation from unity of the Summer-Heston purchasing power parity for investment in 1960.

terms. A clear insight about what is meant by convergence within the Neo-classical growth model is available in Barro and Sala-i-Martin (1992). Basically, they argue that the Neo-classical framework implies conditional convergence; that is, poor economies are indeed expected to grow faster than rich ones but only if they are homogeneous with respect to their steady-state value and the exogenous rate of technological progress. If this is likely to be true for the US states[11], can it be applied to a wide range of countries? Still, note first that this may justify the absence of a significant relationship in Barro results (1991), on the one hand, and the negative correlation displayed by the OECD sample in the Dowrick and Nguyen (1989) study, on the other hand. Finally, as the negative association between growth rates and initial Log income per capita becomes greater and more significant for an OECD sample, and/or shows up within a world wide cross-section analysis when one adds controlling variables like school-enrolment, political stability, investment, etc., they argue that convergence is only conditional, suggesting that including additional variables within the empirical model allows some of the cross-section heterogeneity to be hold constant in both the steady-state values and the rate of technological progress.

A human capital-augmented version of the Solow model -i.e. where the saving rate is exogenous- constructed by Romer, Mankiw and Weil (1990), also confirms the convergence hypothesis. Such a model implies that controlling for population growth and saving rates of human and physical capital, disparate economies are converging over time to a similar level of per capita output. Indeed, they have tested for both unconditional convergence -i.e. without conditioning on hypothesized explanatory variables like school-enrolment, private investment, population growth, etc.- and for conditional convergence. Note, however, that the control variables are not, as in previous papers, ad-hoc additions to

---

[11] Indeed, Barro and Sala-i-Martin (1992) find a clear and significant evidence for the catching-up phenomenon to occur for the US states over various periods from 1840 to 1988, at least if diminishing returns to capital set in very slowly.

the equations but are directly issued from their human capital-augmented version of the Solow model. Thus, without conditioning and after having divided the world into non-oil, intermediate and OECD countries, they confirm the results found by Dowrick and NGuyen (1989) and find as Barro (1991), no significant relationship between the logarithm of initial per capita income and the log. difference GDP per working age person, between 1960 and 1985 for both the other samples. Then, conditioning by the rates of investment or alternatively of saving, population growth and human capital, a significant negative relationship arose in all three samples showing therefore a "strong evidence for convergence."

Again, we lay stress on the fact that Solow (1956) in arguing that differences across countries may come from different initial levels of technology, was not claiming to explain either inequalities in income distribution or the decrease of relative advantage with continued trade in the economy as a whole. Indeed, he only tried to explain for growth rates among industrialized countries and to account for the main features of the US economic growth performance. The Romer & Al. pro-convergence study cited above stresses that the Solow model predicts only that per capita income in a given country converges to that country's steady-state value. This argument justifies empirical evidence for convergence only once one has controlled for the determinants of the steady-state. In other words, one can only infer conditional convergence from the Neo-classical growth model.

- *The null hypothesis: is the equilibrium unique?*

However, even if a set of control variables, meant to control for microeconomic heterogeneity, does indeed support the convergence hypothesis, this does not automatically lead to the rejection of the existence of stable multiple equilibria in long-run per capita output as argued by "new" growth models. In other words, if differences in growth rates are apparently compatible with the view that each country has access to an identical concave production technology, data might more consistently support a theory of different aggregate production functions, leading to heterogeneity in production technologies and thus to local rather than global convergence. This point has been analyzed by Durlauf and Johnson (1992) who provide evidence of the existence of groups of countries exhibiting local convergence, suggesting thus that there may be several "convergence clubs[12]." In order to support the multiple equilibria perspective -i.e. convergence once one has isolated economies which are associated with the same equilibrium- they exogenously split the Summers-Heston data into sub groups based upon two control variables: per capita output and the adult literacy rate at the beginning of the period. They then reran the Romer & Al. human capital augmented version of the Solow model for these sub sets of countries. Their null hypothesis supporting a multiple regime perspective has to be accepted because equality of coefficients across the different groups is rejected. Secondly, the coefficients corresponding to the partial association between growth rates and initial Log. income per capita and exhibited by the sub samples, are greater and more significant than that estimated for the whole sample, suggesting faster convergence rate among the subgroups than is displayed by the single regime.

In a second step, they allow the data to endogenously identify economies with similar "laws of motion." A technique described in Breiman & Al. (1984) provides a general nonparametric way of identifying multiple data regimes using more than one control variable and without imposing exogenously the number of sample splits. Again, evidence for local rather than global convergence arises. The world we are facing is a

---

[12] Durlauf and Johnson (1992; pp. 2-3) argue: "one difficulty with the body of cross-section studies is that they often do not make clear the nature of the null and alternative models associated with a particular statistical test. For example, does a negative β in a regression restricted to a group of advanced industrialized economies such as the OECD represent evidence supportive of the Solow model as opposed to the Romer-Lucas class of models? The answer is no."

world divided into four sub groups and while Romer & Al. could explain 46% of overall growth variation, they find for the poorest economies they explain 57%, for the poorest and highest intermediates, respectively, 52% and 57%, and for the highest income economies, 82% of the total growth variation, such that segregating countries into groups by initial conditions improves overall model fit. One of the most striking features is how much coefficients differ across the identified sub groups, while within these new data regimes, the evidence for convergence concerns only the poorest and the highest intermediate income countries. Indeed, the estimated coefficient on the initial Log. output per capita becomes respectively negative and positive, and both insignificant for the lowest intermediate and for the highest income countries.

Results found by Durlauf and Johnson would appear to support Endogenous growth theory and the existence of multiple equilibria because they provide, maybe insighted by Levine and Renelt (1991), some reasons for choosing what countries to include in a particular sample and because they discuss about how the results change when they use different selection criteria. However, convergence versus divergence within each sub group appears to depend dramatically on the "split selection criteria." As the consistency of the estimators is crucially dependent on the model specification, no consensus exists among economists about the issue of global versus local convergence. Indeed, a drawback to these empirical approaches, is that the functional form of the model is specified directly or indirectly via the specification of one or several production functions such that the "robustness" of the estimation varies according to the model specification. Therefore, in order to avoid misspecification of functional form which usually leads to inconsistent estimators, we need a more flexible methodology. We propose here to use a methodology which has a lot to say in a world-wide per capita income distribution context: non parametric analysis which avoids the settings of restrictions -i.e. a finite set of parameters- on the unknown distributions under study and to exhibit "convergence clubs" escaping from the "split selection criteria."

- *Dynamics and evolving distributions.*

Kirman and Tomasini (1969) pioneered the use of discrete stochastic processes, where per capita income series are assumed to be random variables generated over time in order to discuss the amount of mobility within the system. The system under study consists in 44 countries over a period of 16 years from 1951 to 1966. The first scenario they propose is a ranking of each country in states denoted by i from 1 to 44 for each year over the entire period. They first assume that "if any country is in rank i at some time the probability that it will move to rank j is the same no matter what country it is." Secondly, they divide the 16 year period into two sub periods of eight years and estimate the probability for a country to be in any of the 44 possible ranks given its rank at the beginning of the period. If the process is stationary -i.e. a country's probability of changing rank will be the same whenever one examines the transition probabilities- then both distributions might be identical and the process would be in a state of equilibrium. On the other hand, if each country tends to its own steady state -i.e. to its permanent rank- one should observe less shifting of ranks in the second sub period. However, a problem arises if one wants to measure the transition probability. Indeed, the calculated probabilities are not independent, as any change in the rank of one country must be offset by a similar change for another country.

A second scenario consists in introducing a Markov process to represent the transition probabilities. It is illustrated by the probability for a country of a change of ranks in period T+1 given the rank occupied in time T. The results show that mobility might arise within blocks of ranks without any possibility of moving outside them; that is, not all ranks are accessible to any country and one might expect local rather than global convergence as found by Durlauf and Johnson (1992). The final scenario is then straightforward and consists in setting up classes of countries by size of per capita GDP.

Note that this procedure has the advantage that the probabilities are independent of each other. Then, varying both the number of states and the bounds of these states in order for the results to be independent of these latter, they calculate the probability for a country to move from one class to another over both sub periods allowing for upward as well as downward movements. They found an almost zero probability for countries to move downward as "they managed to achieve a positive growth over the total sub periods." However, there was empirical evidence that the wealthier countries outdistanced their poorer counterparts implying a greater upward movement of the richer countries.

In the empirical literature examined above, the time-averaged growth rate is invariably used as a proxy to the equilibrium growth path which characterizes an economy. This time-trend is implicitly assumed to be stable over time so that deviations from this trend are regarded as cyclical disequilibrium paths of a temporary nature. However, as suggested by Quah (1993 c), if this implicit assumption is rejected[13], cross-country regressions like the ones discussed above are likely to be misleading in trying to exhibit stable relationships between the proposed explanatory variables and the time-averaged growth rate. Quah (1993 b) also shows that coefficients of arbitrary signs in such regressions are consistent with an unchanging cross-section distribution of incomes. This result contradicts the "catching-up" hypothesis which states that each country eventually becomes as rich as the others -i.e. the cross-section dispersion diminishes over time.

---

[13] Quah (1993 c) evaluates the trend growth rates stability. Using the Summers-Heston data table, and dividing the period from 1960-85 into two sub periods of equal size, he then plotted the average growth rates as well as the standard deviation of Log. income per capita experienced by each country during the two sub periods. If the implicit assumption of stability is accepted, data should concentrate around the 45°-line. Instead, he observed that changes in trend were significant, as 78% of the observations were below the 45°-line, and that 72% of the economies available in the sample experienced sharp increases in their income variability, suggesting therefore instability in underlying long-run growth patterns.

This led him to turn to an alternative methodology in order to analyze the laws of motion experienced by different economies. Especially, he hinted (See Quah; 1993 b-c) at the usefulness of nonparametric estimation in order to discuss convergence and cross-sectional income dynamics for a large set of countries. His goal is to focus directly on the evolving long-run distribution of incomes across countries. First, mapping one distribution at a time T into another at a time T+τ, he wants to assess the speed of convergence of the evolving distributions and their respective intra mobility properties. If some regularities arise from this mapping -e.g. a distribution tending towards a mass point, or alternatively divergence manifesting through multimodality- one can indeed expect the long-run properties of the distribution of incomes across countries to be displayed. Therefore, using a nonparametric estimate, and arbitrarily discretizing the set of possible values of income relative to the world average into six grids (0, 1/4, 1/2, 1, 2, ∞), he first estimated point in time per capita income densities of the cross country distribution of these normalized series for different sub periods over the entire sample period (1962-85). As the cross section distribution did not appear to be collapsing, he estimated the probability that an economy starting at time T in one of these five intervals remains or alternatively transits to another state. Probabilities have been calculated over different time horizons ranging from one year to 23 years. The results suggest that the world today is becoming a world partitioned into either very poor or very rich -i.e. "a two-camp world"- where to escape from the development trap is very unlikely. Indeed, a striking feature is the immobility over a 23 years time horizon in the transition probability for both the extreme states; that is, the poorest have 75% probability of being stuck in the development trap, while the richest with all probability (at least 95%) remain rich. On the other hand, the middle income class is likely to vanish over a long-run horizon with an equal probability of rising or falling. Thus, the relative gap between rich and poor is expected to widen.

In another paper, Quah (1993a) normalizes per capita income (X_i) for each country relative to the US which can be seen for the period under study as the leader one should catch-up, if the convergence hypothesis is true. Again, using nonparametric methods, he estimates the joint density $f(X_{i,T}, X_{i,T+\tau})$, where $\tau=1, 5, 10, 15$ years, rescaling this latter to obtain a conditional probability for each $X_{i,T}$. The corresponding graphs show a persistence for economies to remain where they are relative to the US, confirming the result presented above -i.e. "the rich tend to get richer, and the poor poorer." However, it should be noted that some intradistribution mobility do arise as the probability transition horizon is extended. This procedure was then redone, allowing for some conditioning information like the investment share of GDP and secondary school enrolment. The cross country distributions showed the same structure, i.e. immobility, especially for the lowest income classes as previously.

Stochastic processes might seem attractive for the study and discussion of the following null hypothesis: whether either the absolute or relative gap between the wealthier countries and their poorer counterparts is widening and/or to assess whether poor and rich countries will tend to converge in terms of levels of income per capita. This empirical review tried to illustrate that the evidence for convergence is not "robust" to sample choice and/or to split selection criteria for a world-wide set of economies. The absence, among economists, of consensus is largely due to the heterogeneity of these empirical studies which, as surveyed by Levine and Renelt (1991), use a wide range of "right-hand-side" variables, use different countries, measure variables differently, employ different data sets, such that it is very difficult to discern consistent relationships. Therefore, the robustness of these results has been carefully examined through a sensitivity analysis in an extremely useful paper by Levine and Renelt (1992) who found that only growth and the share of investment can be considered as a "robust" association while the consistency of the partial association between average growth rates and initial incomes is crucially dependent on the model specification such that slight alterations in the list of explanatory variables can over turn the results found in the above empirical literature. In other words, traditional parametric methods nest on strong assumptions such that results may not be robust to specifications.

This led us to turn to a more flexible methodology which specifies no functional form at all and which is robust for misspecification. The fast computational facilities now available, will lead us to extensively use nonparametric analysis and in particular nonparametric density estimation procedures. Indeed, nonparametric estimation is the benchmark when one wants to make comparisons with parametric forms and provides some clues about stylized facts which are not immediately apparent. Therefore, the next section will introduce and present both the data and the statistical tools which will allow us to discuss the key economic issue of convergence examined in the present section.

3/ Nonparametric Density Estimation.

Kernel density estimation is a method for estimating the structure of unknown distributions, and is the most commonly used non-parametric density estimate apart from the histogram because it is understandable at an intuitive level, conceptually straightforward, easily computed, and possesses mathematical properties which allow to escape from the major drawbacks of the histogram. Indeed, the asymptotic theory involved is well-developed. First, non-parametric techniques provide comprehensive graphical descriptions of the data and give direct information about the problem at hand. Secondly, because of recent computational facilities, this field is no longer confined to theoretical statisticians, but is also used by applied researchers in economics.

As discussed above, parametric analysis uses the information available in the data only from the postulated model and thus may be biased in our situation. Therefore, we propose to use non-parametric density smoothing which appears to be the right benchmark as it *lets the data speak for themselves*, when one wants to provide additional information about stylized facts which are not immediately apparent. Thus, we expect to obtain better understanding of the structure of the data by estimating the unknown density $f(x)$ from our observations $\{X\}_{i=1}^{118}$. However, we want to emphasize that non-parametric analysis cannot replace the parametric analysis. Indeed, one must be aware that while kernel smoothing is an important tool in many ways, it does not solve all problems. What we claim is that a great deal can be learned from a data table like P.W.T5 and a computing environment like XploRe 3.1 (1993)[1], about the key economic issue of convergence. As Härdle and Scott (1992; pp. 98) put it: "non-parametric techniques for density smoothing are known for their flexibility and their ability to detect structures deviating from a postulated parametric model." In this section, we present kernel density estimates as well as their statistical properties. We then describe the WARPing (Weighted Averaging of Rounded Points) method which benefits asymptotically from the statistical efficiency of kernel methods in both low and high dimensions, while avoiding their computational costs (Scott, 1995b; Härdle, 1990b).

3.1/ The Data.

The Penn World Table (Mark 5) displays a set of national accounts economic time series covering a large number of countries. All of the country estimates are based on the international comparisons of national outputs produced by the United Nations International Comparison Project (ICP). It contains for 118 countries[2] the following annual estimates of per capita income covering a period from 1960 to 1985: real gross domestic product both in current international prices and expressed relative to 1985 international prices. Secondly, both per-equivalent adult and per-worker real gross domestic products expressed relative to 1985 international prices are also available. The concept of constant prices in which the quantities of both all years and countries are valued at the international prices of the base year, i.e. 1985, are useful because inflation effects

---

[1] XploRe 3.1 is an interactive, open statistical computing environment which allows for application of both univariate and multivariate applications as well as interactive graphical representation.

[2] A complete list of countries as well as their respective quality rating is available in Appendix 1. Note that a grade of "D" received by a country from Heston & Summers identifies countries whose real income figures are based on very little primary data.

keep current-prices time-series comparable over time[3]. Following Heston, Kravis and Summers' paper (1984; pp. 240-241), it might be "of some interest to know how countries' current standards of living compare. This is probably better captured by real private consumption per capita" expressed in current international prices. Finally, government consumption, investment and population are also available.

The problem of international comparability is partly that of finding a common numeraire to apply to different countries. International comparisons of GDP have generally been based on their nominal values, which are obtained by converting the data available in the national accounts to a common currency, e.g. the US Dollar. However, it is now well-known that exchange rates defined as the prices of domestic currency in terms of foreign currency, which reflect the relative prices of internationally traded goods, tend to understate the real income of less developed countries by a factor which is systematically and inversely related to the country's per capita GDP (Summers, Kravis and Heston; 1980). Indeed, the prices of tradables relative to nontradables tend to be more expensive in less developed countries and tend to decrease as incomes grow. Recall here that the purchasing power parity doctrine postulates that the equilibrium exchange rate at which the currencies of two countries will trade, will be determined by the relative price levels of the countries. Then, because the exchange rate is based on the relative prices of internationally traded goods, this understatement and the distortion become larger as the country's income is low relative to the United States from which the currency is used in order to convert the national accounts into US Dollar. Thus, one of the most attractive feature of this database appears to be that expenditure entries are denominated in a common set of prices in a

---

[3] Summers & Heston (1991; pp. 344) note that: "the real percapita GDP expressed in constant prices suffers from a fixed base problem: after a while, relative prices change, and the base year weights become less and less appropriate... A way of mitigating the declining appropriateness of the base year weights for comparison years distant from 1985, is to bring changing relative prices into the analysis explicitly through a chain index... The merit of such a real expenditure's measure is the fact that its growth rate for any period is based upon international prices most closely allied with the period." It is defined in terms of a chain of ratios of consecutive years' percapita GDP. Each element of the chain is the ratio of percapita $GDP_{j,\, t+1}$ to percapita $GDP_{j,\, t}$ where both the numerator and denominator are valued in the t-th year's international prices.

common currency so that international comparisons can be made not only over time but also "inter spatially" -i.e. between countries. More precisely, this data table is such that the price parities rather than exchange rates, denominated in the country's national currency expressed relative to the US Dollar, are used to convert the countries' national currency expenditures to a common currency unit, the international Dollar[4].

What are the consequences of converting national accounts into international Dollar with PPPs rather than into US Dollar through nominal exchange rates? We have already noted that even though the purchasing power parity does not hold in its absolute version because some deviations may arise in the short run (some time may be needed for exchange rates to adjust so as to equalize price levels), Kravis and Lipsey (1983), as well as I.C.P. benchmark studies find that not only do nominal exchange rates differ significantly from corresponding PPPs', thus justifying the rejection of the nominal exchange rate for the conversion, but they do so in a systematic way: the national price level of a country defined as the purchasing power of the domestic currency (the number of Francs required to buy in France the same goods and services a Dollar buys in the US) divided by the exchange rate (the number of Francs required to buy a Dollar on the foreign exchange market), is a rising function of the level of its income or stage of development. This leads conversion to PPP Dollars to favour the lower incomes more than it does the higher income, as the real per capita measures as presented in P.W.T5 are the data issued from the national accounts over the price levels defined as the PPP over the exchange rate. Therefore, the conversion based on purchasing power parity -i.e. real exchange rate- adopted in the I.C.P. appears to have a cost although to a definitely slighter extent than when the conversion is based on nominal exchange rate (Kravis and Lipsey; 1990).

---

[4] "In this international dollar currency relative prices of individual goods are set at the (weighted) average of relative prices for the same goods in all countries, and the level of prices is normalized so that the GDP of the United States is the same in international Dollars as in American dollars." (Summers & Heston, 1991; pp. 334). "The price parity for a detailed category, is of the form $(p_{ij}/p_{i,US})$. Thus, the ratio of a country's expenditure in its own currency to the category price parity referred to above, $(p_{ij}q_{ij})/(p_{ij}/p_{i,US})$ is equal to the quantity valued at the US category price. Such US priced quantities for any category are directly comparable across countries." (Summers & Heston; 1991, pp. 341)

Whether the series expressed in international prices are more relevant than the countries' national accounts expenditures is also of interest. In other words, could empirical differences between growth rates based on national and international prices bias some of the results? Among other empirical regularities, Summers and Heston (1991; pp. 360-361) find that "growth rates based on international prices can differ significantly from those based on national prices; but when they do, it is nearly always the case that relative prices within the countries have changed substantially over the growth rate period." Finally, As Summers, Kravis, and Heston largely contributed to the build up of this data table and as they used these data in their study on "Changes in World Inequality", the reader would do better refer to their description of the sources of the data (K.H.S.; pp. 238-241) which underlie our study. Indeed, the authors are very careful to document the procedures they use in constructing indexes, and they highlight any problem about which they think users of the data should be aware.

3.2/ An Alternative to Traditional Parametric Models for Exploring Univariate and Multivariate Data without making Specific Distributional Assumptions.

3.2.1/ Kernel Density Estimation.

We are therefore given observations $\{x_i\}_{i=1}^{118}$ of a random variable from an unknown distribution with density $f(x)$ where $X_1,\ldots,X_{118}$ are assumed to be independent and identically distributed observations. The problem then is to estimate this unknown density from our sample in order to provide a way of understanding and representing the behaviour of the random variable. Non-parametric estimation allows us to avoid any pre-specified functional form of the density, e.g. log normal, Gamma, Beta, which are the usual forms used to represent income distributions, and therefore *to let the* *data speak for themselves*[5]. It especially allows us to focus on valuable indication of features such as skewness and multimodality in the data.

Let us return to some primary notions in the theory of statistics. First, note that our sample space Ω is finite. Indeed, it contains a finite number of elements or points, say $n = N(\Omega)$, where $n(\Omega)$ is the size of Ω; that is 118. If Ω is the totality of outcomes, the function or random variable denoted by X makes some real numbers correspond to each outcome of our experiment. This distribution of values can be described by using *density functions*. Thus, there exists a finite or denumerable set of real numbers, say $\{x_i\}_{i=1}^{118}$, so that X takes on value only in that set, and X is therefore assimilated to a discrete variable. A naive estimate, usually called a discrete frequency function or discrete density function, of our discrete random variable X at a point x, may be:

$$f(x) = P[X = x_i] \qquad \text{if } x = x_i, i = 1, \ldots, 118$$

$$f(x) = 0 \qquad \text{if } x \neq x_i$$

Then, define $f(\cdot)$ as a function with domain Ω the real line and counter domain the interval $[0,1]$, and using the indicator function,

$$f(x) = \sum_{i=1}^{118} P[X = x_i] I_{x_i}(x),$$

where $I_{x_i}(x) = 1$ if $x = x_i$ and $I_{x_i}(x) = 0$ if $x \neq x_i$.

Because the number of outcomes (N) is finite, it is obvious that the probability of each outcome is n^{-1} and defined this way, $f(x)$ is an equally likely probability function over Ω= [a,b], where a and b are taken to encompass the data -i.e. [Min(X), Max(X)].

---

[5] All the following analysis will closely follow the methodology and the tools developed in Silverman (1986) and Härdle (1990b) which both provide a very readable account of the area. Note as well that there exist apart from kernel methods, other non-parametric density estimation techniques which are surveyed in Silverman (1986; Chapter 2).

The earliest nonparametric estimate of a univariate density $f(x)$ was the histogram. The histogram has been used to provide a visual clue to the general shape of $f(x)$ and to estimate it without specifying any formal parametric structure. As a first step in our data analysis, one might draw a histogram from the sample $\{x_i\}$. Note that, even though it might appear obvious, the shape of the histogram will directly depend on the way we divide the real line. Thus, the first step in estimating $f(x)$ consists in partitioning Ω into a grid (or mesh), thereby controlling for the amount of smoothing inherent in the procedure. Let us choose to partition the sample space into m (j=1,..., m) non overlapping bins ($B_j = [b_{i,j}, b_{i,j+1}]$) where $a = b_{i,1} < b_{i,2} < \ldots < b_{i,m+1} = b$, and count the relative frequency of observations falling into each of the bins, we can write as above:

$$f(x) = P(x \in B_j) \qquad \text{if } x \in B_j, j = 1, \ldots, m$$

$$f(x) = 0 \qquad \text{otherwise}$$

Then, as the probability for observations of $X = \{x_i\}_{i=1}^{118}$ to fall into an interval B_j, is the relative frequency of observations falling into this bin, we can rewrite the above formulation replacing probability by proportion and given our sample by:

$$P(x \in B_j) = \frac{1}{n(b_{i,j+1} - b_{i,j})} \#\{(x_1, x_2, \ldots, x_{118}) \in B_j\}$$

We can now characterize the histogram of a one dimensional density function by the following expression:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{118} n_j I_{B_j}(x)$$

where: n_j is the number of sample values falling into B_j and $\sum_{j=1}^{m} n_j = n$,

$h = b_{i,j+1} - b_{i,j}$, is usually called the window width or smoothing parameter,

and I_{B_j} is the indicator function of the jth bin.

However, the histogram as a density estimate presents some undesirable features. For instance, the shape of any histogram is very sensitive to the choice of its origin (See, for instance, Silverman: 1986, pp. 10-11). Secondly, averaging the observations falling into a fixed bin is likely to be noisy according to the subjectively chosen bin width. Therefore, Rosenblatt (1956) proposed putting smooth kernel weights in each of the observations leading to average over scaled kernel functions centred in the points x_i.

The univariate kernel density estimate of f will have the form:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{(x - x_i)}{h}\right)$$

where the choice of the kernel function and the smoothing parameter, also called a bandwidth, will determine the performance of $\hat{f}_h(x)$ as an estimate of $f(x)$. The usual class of kernels (K) consists of functions which satisfy the conditions described above in order to be probability density functions and able to be used in the previous equation; that is:

$K(x) \geq 0$, and $\int K(x)dx = 1$ which ensures that $\int \hat{f}_h(x)dx = 1$.

where, K will usually be a continuous, bounded, and symmetric function, e.g. normal density, so that observations close to the center of the bin will be given greater weights. The choice of the kernel determines the shape of the bumps while the window width determines their width.

Finally, we have to note that all the definitions and arguments presented in this sub section as well as the methods of univariate density estimation can be extended to several random variables and the generalisation to d dimensions such that $X_1, X_2, \ldots, X_d$ are joint discrete random variables, is straightforward. Indeed, rather than areas, one will deal with volumes in R^d using d smoothing parameters. That is, the multivariate density estimate of f will have the form:

$$f_h(x) = (1/nh^d) \sum_{i=1}^{n} K((x-x_i)/h), \qquad x \in R^d$$

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

3.2.2/ Statistical Properties of the Kernel Estimates.

Having a fixed finite sample, the main problem to be solved when one wants to use nonparametric estimation, is the length of the "window width" which will lead to both an unbiased and stable density estimate, i.e. consistent. Again, the smoothing parameter which regulates the degree of smoothness is called the bandwidth (h) and must be optimal when choosing a compromise between the problem of under smoothing (for the reduction of the bias) and over smoothing (in order to reduce the variance).

Indeed, one will recommend using a particular statistical procedure only if the estimate possesses desirable properties. One of them consists in unbiasedness. Recall first that an estimate is unbiased if the mean of its distribution equals the function being estimated. That is,

$$E\left[\hat{f}_h(x)\right] = f(x).$$

Then, Härdle (1990b; Chapter 2, pp. 55) gives a proof that univariate nonparametric density estimates are asymptotically unbiased for $f(x)$ and without any conditions on $f(x)$, if $h \to 0$.

$$E\left[\hat{f}_h(x)\right]=\frac{1}{n}\sum_{i=1}^{n}E\left[K_h(x-x_i)\right]$$

where $K_h(u)=h^{-1}K\left(\frac{u}{h}\right)$ is the rescaled function with bandwidth h.

$$E\left[\hat{f}_h(x)\right]=E\left[K_h(x-X)\right]$$

$$E\left[\hat{f}_h(x)\right]=\int K_h(x-u)f(u)du$$

Make now the change of variable, $u=x+sh$:

$$E\left[\hat{f}_h(x)\right]=\int K_h(s)f(x+sh)ds$$

If we let $h\rightarrow 0$, we see that

$E\left[\hat{f}_h(x)\right]\rightarrow f(x)\int K_h(s)ds=f(x)$ when $h\rightarrow 0$ as $\int K_h(s)ds=1$,

with $Bias\left[\hat{f}_h(x)\right]=E\left[\hat{f}_h(x)\right]-f(x)\rightarrow 0$ when $h\rightarrow 0$.

The estimate is thus asymptotically unbiased when the bandwidth h converges to zero. Therefore, the bias increases as the bandwidth h increases, and to choose h close to 0 will indeed lead to an unbiased estimate, but also to a very noisy representation of the data.

If we increase h, we average over more data or in our case over more kernel functions, and thus, the amount of smoothing in the estimate increases, leading to a more stable density estimate. The stability of the estimate which is measured by the variance is the second interesting point. Still, note that the bias in the estimation of $f(x)$ does not depend directly on the sample size, meaning that using a large sample does not necessarily lead to reducing the bias.

Indeed, a more important property that an estimate must fulfil is consistency; that is,

$$\hat{f}_h(x)\xrightarrow{p}f(x)$$

Literally, does the estimate converge in probability to the true distribution of our observations? Thus, we have to find a criteria with which to assess whether our estimate is a "good, a bad or an ugly" one. A measure of goodness in order to discuss the closeness of the estimate $\hat{f}_h(x)$ to the true density f is called the Mean Squared Error of the estimate. It is defined as follows and by standard properties of mean and variance leads to:

$$MSE(x) = E_f\left[\left\{\hat{f}_h(x) - f(x)\right\}^2\right]$$

$$MSE(x) = E_f\left[\left\{\left(\hat{f}_h(x) - E\left[\hat{f}_h(x)\right]\right) - \left(f(x) - E\left[\hat{f}_h(x)\right]\right)\right\}^2\right]$$

$$MSE(x) = E_f\left[\left(\hat{f}_h(x) - E\left[\hat{f}_h(x)\right]\right)^2\right] - 2E_f\left[\left(\hat{f}_h(x) - E\left[\hat{f}_h(x)\right]\right)\right]\left(f(x) - E\left[\hat{f}_h(x)\right]\right)$$

$$+\left(f(x) - E_f\left[\hat{f}_h(x)\right]\right)^2$$

$$MSE(x) = Var\left(\hat{f}_h(x)\right) + \left\{Bias\left(\hat{f}_h(x)\right)\right\}^2$$

We must note that the mean squared error is the sum of two non negative quantities. It shows how the consistency, variance and bias of an estimate are related[6]. We already know that the bias will tend to zero as the smoothing parameter moves closer to zero. On the other hand, Härdle (1990b; chapter 2, pp. 57-58) shows that the variance, being nearly proportional to (1/nh), is negatively related to the smoothing parameter. In other words, if one wants a stable density estimate in order to minimize the variance, he has to ensure that enough observations fall into the bin or to choose h large. This is contradictory to the unbiasedness property. This is the trade-off problem introduced above, concerning the choice of the bandwidth. A small (large) h will penalize the stability (bias) of the density estimate.

---

[6] Note that another measure of accuracy relates to how well and closely the entire curve $\hat{f}_h(x)$ estimates $f(x)$. One such a measure of goodness of fit is the Integrated Squared Error defined as:

$$ISE\left(\hat{f}_h(x)\right) = \int\left[\hat{f}_h(x) - f(x)\right]^2 dx.$$

One better understands now why we emphasized the finiteness of our sample. Indeed, the Mean Squared Error of kernel density estimation will converge to zero if $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$. The choice of the smoothing parameter as well as the choice of the kernel one uses when estimating how some observations are distributed remains of crucial importance in density estimation.

3.2.3/ WARPing Approximation of the Kernel Density Estimate using the Quartic Kernel

Given our random sample $X_1, X_2, \ldots, X_{118}$, we would like to estimate the unknown density $f(x)$ which gave rise to this random sample. However, if kernel estimators are conceptually straightforward and easily computed, this is not necessarily unexpensively.

Here, we estimate a kernel density using the WARPing method described in Härdle (1990b) and proposed originally by Scott (1985b). The latter proposed benefitting from the statistical efficiency of the kernel methods in both low and high dimensions, while avoiding their computational cost. His motivations are that traditional estimates are not well-suited to modern data analysis, which requires interactive computing for large data sets with usually more than one random variable.

We have already mentioned that an undesirable feature of the histogram estimate is its dependence on the arbitrarily chosen origin, meaning that estimates with different mesh origins may have different subjective features and number of modes. Indeed, holding fixed the bandwidth (h), any density estimate using a particular origin might be considered as an eventual and possible one though giving different impressions on the location of the peaks in the density. Thus, if one estimates the histogram for different bins locations and with equal bin widths, and in a following step averages these histograms estimates, he will make this latter independent of any origin. This method is known as ***Weighted Averaging of Rounded Points*** and kernel can be approximated by WARPing.

The estimation of densities by WARPing consists of the three usual steps in order to estimate probability functions:

-1/ As in the histogram case described above, one will "bin" the data. That is, he will divide the real line into bins:

$$B_j = [(j-1)h, jh] \quad \text{with } j = 1,\dots,m.$$

Then, the observations contained in each bin centered at the origin $x_0 = 0$, will be counted. Thus, the information about the data will be reduced from the set of observations $\{X_i\}_{i=1}^{118}$ to the set of non empty bins B_j and where n_j is the number of observations in bin B_j, and its corresponding index function I_{B_j}, that one can assimilate as defined above, to a Bernoulli variable.

Now, we want to avoid the histogram's undesirable property, i.e. its dependence on the choice of the origin. If one shifts each bin B_j, by an amount of, say $\frac{lh}{M}$ to the right with $l = 0, 1, \dots, M-1$, we will have M histograms with M different origins and,

$$\hat{f}_{h,l}(x) = \frac{1}{nh}\sum_{j=1}^{m} n_{j,l} I_{B_{jl}}(x) \quad \text{with } B_{j,l} = \left[\left(j-1+\frac{l}{M}\right)h, \left(j+\frac{l}{M}\right)h\right]$$

Averaging now over these M histograms leads to:

$$\hat{f}_h(x) = \frac{1}{Mnh} \sum_{l=0}^{M-1} \sum_{j=1}^{m} n_{j,l} I_{B_{jl}}(x)$$

-2/ Note that data enter only in the form of frequencies in the bins $B_{j,l}$. Therefore, one must create weights which will have all three properties of a probability function and defined as:

$$W_M(l) \geq 0$$

$$W_M(l) = W_M(-l)$$

$$\sum_{l=1-M}^{M-1} W_M(l) = M \text{ which ensures that } \int \hat{f}_h(x)dx = 1$$

This specification of weighting function for Averaged Shifted Histograms, i.e. $W_M(l) = 1 - \frac{|l|}{M}$, allows to approximate a bigger class of density estimates. Here, we have chosen to use the Quartic kernel[7] which can be approximated by WARPing if M, the number of steps is large enough, so that the Averaged Shifted Histogram weighting function corresponding to the Quartic kernel, is (Scott: 1985b, pp.1033).

$$W_M(l) = \frac{15M^4}{16M' - l}\left[1 - \left[\frac{l}{M}\right]^2\right]^2 \qquad l = 1-M,\ldots,0,\ldots,M-1.$$

-3/ One will finally weight the observations falling into each non empty bin calculated in the first step with the weightings which have been evaluated in 2/. Thus, after having discretized the data into small bins (starting at zero) and applied convolution with a

---

[7] Scott (1985b) notes that the choice of the kernel is not critical. Indeed, Marron and Nolan (1989) introduced the concept of canonical kernels. Basically, two kernel density estimates with different kernels K_1, K_2, which give the same amount of smoothness to the data if the bandwidths h_1, h_2 differ through a multiplying factor. A table for the so-called "canonical bandwidth transformation" is available in Härdle (1990b; Chapter 2, pp.76).

discretized (Quartic) kernel in the domain, the WARPing estimate is defined as (See Härdle; 1990b, Chapter 1, pp 29-30):

$$\hat{f}_h(x) = \frac{1}{nh} \sum_j I_{B_j^*(x)} \sum_{l=1-M}^{M-1} W_M(l) n_{b_{(j+l)}}$$

where $I_{B_j^*(x)}$ is the indicator function for bin $B_j^* = [j\delta, (j+1)\delta]$ and $\delta = \frac{h}{M}$ is the window width so that each original bin B_j with bandwidth h, has been divided into M subbins of bandwidth δ, n_l is the number of observations in bin l and $W_M(\cdot)$ is the discretized Quartic kernel assigning a positive weight to each observation in the sample when estimating $\hat{f}_h(x)$. This procedure is known as Weighted Averaging of Rounded Points, where the rounded points are the number of observations falling into the smaller bins with length δ.

This method has the statistical efficiency of kernel methods while being computationally comparable to histogram methods as it performs smoothing operations on the bin counts rather than the raw data as in traditional kernel density estimation[8]. Note finally, that these non-parametric density estimates have the same statistical properties as traditional kernel density estimates as $M \to \infty$, in both univariate and multivariate cases. Kernel density estimates by WARPing for given bandwidths and displayed in the following discussions, can be calculated in XploRe via the macro *denest* for the univariate case and via the macro *denest2* for the bivariate case. They are presented in Appendix 2.

---

[8] For a formal comparison of computational costs between WARPing and traditional kernel density estimation, see, for instance, Härdle and Scott (1992).

The remaining problem concerns the choice of the bandwidth, which must be appropriate and optimal although we do not yet know the density $f(x)$. Indeed, practical application of kernel density estimation is crucially dependent on the choice of the smoothing parameter, which may not be issued by subjective choices. Therefore, a number of methods that automatically choose a value for the associated smoothing parameter have been developed. Finally, note that evaluation of density estimation is especially important for optimization of the smoothing parameter. This central issue is presented in the next section in which we estimate income densities for our sample of countries available from P.W.T5.

4/ Univariate Kernel Density Estimates of the World Income Distribution.

The estimation of the density of income distributions using kernel rather than ad-hoc specific functional forms, e.g. lognormal, is one of the current applications in econometric literature. For example, Hildenbrand and Hildenbrand (1986), and Deaton (1989) use kernel density estimation to estimate the income distribution respectively in the UK and Thailand, in order to analyze, for the former study, income effects for an arbitrary demand function and for the latter, rice patterns of demand. The resulting densities are very different from their corresponding lognormal fits and exhibit at least two modes in the a priori unknown distribution of incomes. Using like Hildenbrand and Hildenbrand, the Family Expenditure Survey, Park and Marron (1990) compare data-driven bandwidths methods for selecting the optimal amount of smoothing inherent in a kernel density estimator.

A common task before starting with kernel density estimation, consists in asking why is it necessary to use this kind of sophisticated technic rather than a simple histogram? In order to construct an histogram, we need to choose both an origin and a binwidth. On the one hand, the choice of the latter will control the amount of smoothing inherent in the procedure. On the other hand, the choice of the origin remains crucial to detect the location

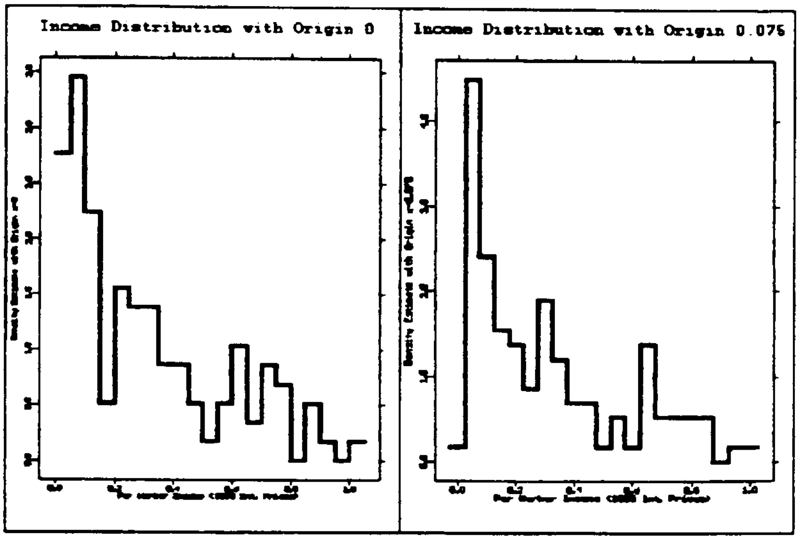


Figure 4.1.

Histograms of per worker income distribution for the year 1985,
with binwidth h=0.05 and origins x_0= 0 (Left-hand box) and 0.075 (Right-hand box).
(Observations have been normalized relative to their maximum)

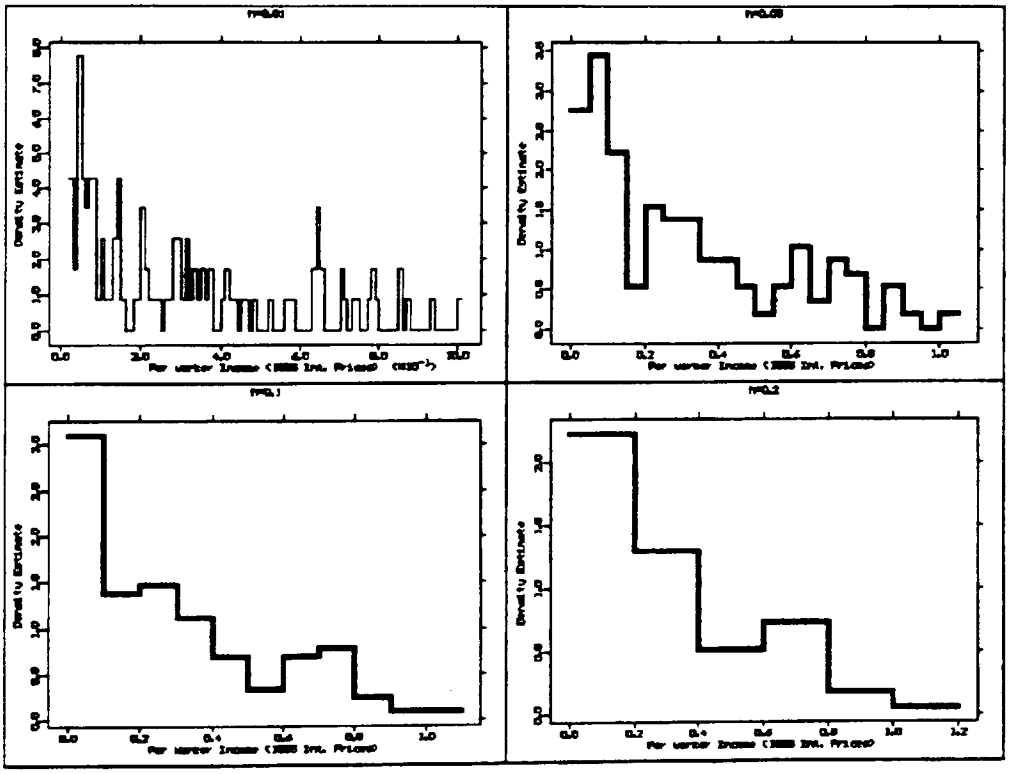


Figure 4.2.

Histograms of per worker income distribution for the year 1985
with origin x_0= 0 and binwidths h= 0.01, 0.05, 0.1, 0.2.
(Observations have been normalized relative to their maximum)

of the bumps present in the distribution. This is illustrated in Figures 4.1-2. Indeed, we observe, for a given bin width, how much the shape of the income distribution of the per worker income measure for the year 1985, differs dramatically according to the arbitrarily chosen origin and independently of any choice of degree of smoothing (See Figure 4.1). Secondly, as the dissection of the real line into bins must be carried out a priori, there exists a wide range of choices not only of an origin but also of coordinate direction of the grid (See Figure 4.2). Finally, note for the remaining of the discussion, that these problems are exacerbated when extending the graphical presentation to bivariate data.

In the previous section, we stressed that the quality of the information is crucially dependent on the accuracy of the density estimate. On the other hand, the latter is dramatically dependent on the choice of the smoothing parameter which, in order to be appropriate, should minimize some criterion function such as the MSE or the ISE and not be issued by subjective choices as in many effective data analysis. One technique widely studied and used by both theoretical statisticians and applied researchers to select the optimal bandwidth associated with the unknown distribution of any data set, is the least-squares cross-validation for kernel density estimation.

4.1/ Least-Squares Cross-Validation for Kernel Density Estimation.

Here, we are concerned with the choice of the window width h which determines the degree of smoothness imposed on $f(x)$. In analyzing the desirable properties that nonparametric estimates should possess, we showed that both unbiasedness and consistency were dependent on the smoothing parameter. Therefore, the optimal h will explicitly depend on the unknown distribution with density $f(x)$. An over small h will result in a very noisy representation of the data, reflecting therefore too much of the sample variability. On the other hand, to choose an over large h will lead to the smoothing away of important features of the underlying density. A traditional and automatic method for choosing the optimal bandwidth is cross-validation (CV). We propose here to use the traditional and commonly used least-squares cross-validation which has been proposed independently by both Rudemo (1982) and Bowman (1984).

Among others, an alternative measure of accuracy which is usually preferred to MSE criteria in order to assess how closely $\hat{f}_h(x)$ estimates $f(x)$, is the Integrated Squared Error, $ISE(h) = \int \left(\hat{f}_h(x) - f(x)\right)^2 dx$. It measures the distance between $\hat{f}_h(x)$ and $f(x)$. We know that $\hat{f}_h(x)$ will be a consistent estimate of $f(x)$ if $ISE(h)$ is minimum.

$$ISE(h) = \int \left(\hat{f}_h(x) - f(x)\right)^2 dx$$

$$ISE(h) = \int \left(\hat{f}_h(x)\right)^2 dx - 2\int \hat{f}_h(x) f(x) dx + \int \left(f(x)\right)^2 dx$$

Note that as $\int \left(f(x)\right)^2 dx$ does not depend on h and that $\int \left(\hat{f}_h(x)\right)^2 dx$ is directly computable from the data, e.g. by WARPing[1], we are therefore looking for a bandwidth (h) which will minimize the following quantity:

$$ISE(h) - \int \left(f(x)\right)^2 dx = \int \left(\hat{f}_h(x)\right)^2 dx - 2\int \hat{f}_h(x) f(x) dx$$

---

[1] See Appendix 3 for the realization of the cross-validation with the XploRe macro *dencvl*.

We are now left to estimate the last term, $\int \hat{f}_h(x)f(x)dx$. Since $\hat{f}_h(x)$ is a random variable itself as a function of our random variable X, $E\left[\hat{f}_h(x)\right]$ is defined and equal to $\int \hat{f}_h(x)f(x)dx$.

$$E_f\left[\hat{f}_h(x)\right] = \frac{1}{n}\sum_{i=1}^{n} f_{h,i}(x_i)$$

Recall that if $\{x_i\}_{i=1}^{118}$ denote the available observations of our random variable X with density f, the kernel estimate $\hat{f}_h(x)$ of $f(x)$ with bandwidth h is defined as:

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\bullet - X_i}{h}\right)$$

Then, the algorithm for computing $E\left[\hat{f}_h(x)\right]$ consists in removing a single value, e.g. x_i, from the sample. Secondly, the density estimate at that x_i will be computed from the remaining n-1 sample values. That is:

$$\hat{f}_{h,i}(x) = \frac{1}{(n-1)h}\sum_{i \neq j} K\left(\frac{(x - X_j)}{h}\right)$$

Finally, one will choose h to optimize our given criteria involving all values of $\hat{f}_{h,i}$ with i=1,2,...,n. A data-driven bandwidth h which will asymptotically minimize $ISE(h) - \int (f(x))^2 dx$ is given by:

$$\hat{h} = Arg \min CV(h)$$

with $CV(h)$, the least-squares cross-validation function, such that:

$$CV(h) = \int \left(\hat{f}_h(x)\right)^2 dx - 2n^{-1}\sum_{i=1}^{n} \hat{f}_{h,i}(x_i)$$

Basically, there are two reasons for the popularity of the least-squares cross-validation. First, it is at first glance a simple fully-automatic method. Secondly, Stone (1984) shows that if f is bounded and very mild conditions hold on the kernel K, then

$Arg \min CV(h)$ is asymptotically optimal, as it converges to the bandwidth which minimizes $ISE(h)$, therefore ensuring that it is a very useful estimate.

However, the performance of this method has often been challenged by both theoretical investigations and simulation studies[2]. On the one hand, the poor performance often recorded by least-squares cross-validation, is due to the great variability in $Arg \min CV(h)$, which more often than not tends to undersmooth the data. On the other hand, there is no consensus among theoretical statisticians about the method which should be selected when estimating smoothers issued by different data sets. Indeed, results from different simulations differ greatly according to the data sets under study. For instance, Park and Turlach (1992) made simulations of normal mixtures densities exhibiting unimodal, bimodal, and trimodal densities. They conclude that least-squares cross-validation indeed shows more structure and sometimes more than is present in the underlying density, although, on the other hand, it is relatively accurate in locating true modes. As the problem is that of knowing what precisely we want from our smoothers, it must be recalled that here we are interested in the economic issue of convergence. In other words, we want to provide some indications of the shape of the underlying distribution of income as well as the number and locations of the modes which will allow us to discuss the existence of a unique versus multiple equilibria. This is the reason for the use of the least-squares cross-validation method in the following discussion.

---

[2] See among others, Hall and Marron (1987a), Park and Marron (1990) and Sheather (1992)...

4.2/ Changes in the World Income Distribution.

The extent and size of changes in inequality over the post-Second World War period have been examined, among others, by Summers, Kravis, and Heston (1984) using the table issued by the International Comparison Programme. They first divided a set of 127 countries grouped by stage of development: the poorer developing countries, the middle developing countries, industrialized countries, and the centrally planned economies. Then, depicting the corresponding trends of Log. percapita income, they found that the industrialized countries grew faster than those in the middle, while the low end was plagued by slow growth. Secondly, a few numbers illustrate the difference in the growth performance experienced by the sub groups. For instance, Gini coefficients calculated using different measures of real gross domestic product, fall sharply from 0.302 in 1950 to 0.129 in 1980 for their set of industrialized countries, while the middle income sub group exhibits a small drop from 0.269 to 0.258. However, both the low income countries and the world as a whole experienced a tiny rise respectively from 0.103 to 0.112 and from 0.493 to 0.498. This led them to argue that the world as a whole experienced a "very slight rise" in inequality over the 1950-80 period. However, it should also be noted that, considering current standards of living, there is a clear monotone trend towards increasing inequality. We will discuss this issue with non-parametric kernel density estimates of the world income distribution. Indeed, these technics because of their flexibility, will give some new insights about the information present in the structure of incomes across a world wide set of economies.

The arguments underlying the following discussion are twofold. Indeed, in the Neo-classical framework, the mechanism which ensures the existence of an optimal decentralized equilibrium -i.e. marginal products are positive and exhibit diminishing returns to each input- inhibits at the same time a permanent and continuous process of growth if one rules out technological progress which acts as an "ignored" issue in the original Solow (1956) paper. Whatever the propensity to save is exogenous or derived from the households' behaviour, the dynamic solution of the Neo-classical framework[3], which guides the evolution of per effective worker capital, is so that at the equilibrium[4], all savings are used in order for the investment flow to get a capital growth at the same rate as labour in order just to fit out the new workers with new capital. Therefore, any interventionist policy which would consist in increasing the propensity to save will only lead to a transitory increase in the growth rate and therefore in the per capita income level. Within this framework, two economies characterized by similar preferences -i.e. an identical propensity to save- and different initial conditions are expected to converge and to reach identical per capita capital and output levels. We give here a very readable insight that this expectation is nothing but true.

In Figure 4.3, we have plotted the estimated per capita income[5] densities for a sample of 117 countries available in the group of countries[6] presented in P.W.T5 and for the years 1960 and 1985.

On the one hand, the estimated density for the year 1960 turns out to be unimodal. On the other hand, a two-mode shape of the kernel density estimate by

---

[3] In the original Solow paper: $dk/dt= sf(k) -gk$, where k is the per capita capital stock, s the fixed marginal propensity to save, and the labour force grows at a constant relative rate (g) independently of any economic variable in the system.

[4] That is when $dk/dt=0$.

[5] The available measures of income are nominal measures while the income disposable for consumption or savings would have been a better proxy in order to discuss about the extent of inequality. However, our goal in this section is only to provide some rough insights about the evolution of the income distribution during the 1960-85 period. We are not claiming to discuss about the relevance of the measures of income we are using as a proxy to measure economic welfare.

[6] We removed Kuwait which acts as an outlier when income measures are expressed relative to 1985 prices.

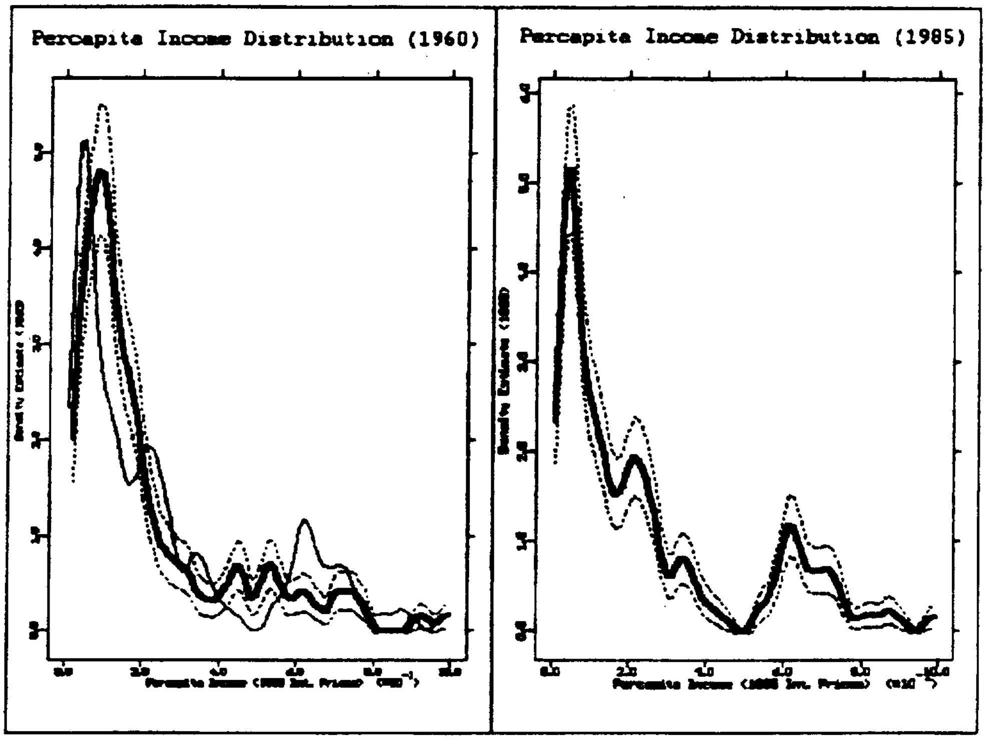


Figure 4.3.

Kernel density estimates by WARPing for the per capita income distributions (thick line); with bandwidth h= 0.05 and for the years 1960 and 1985, and 90% confidence limits (dashed lines). (1985 International Prices)

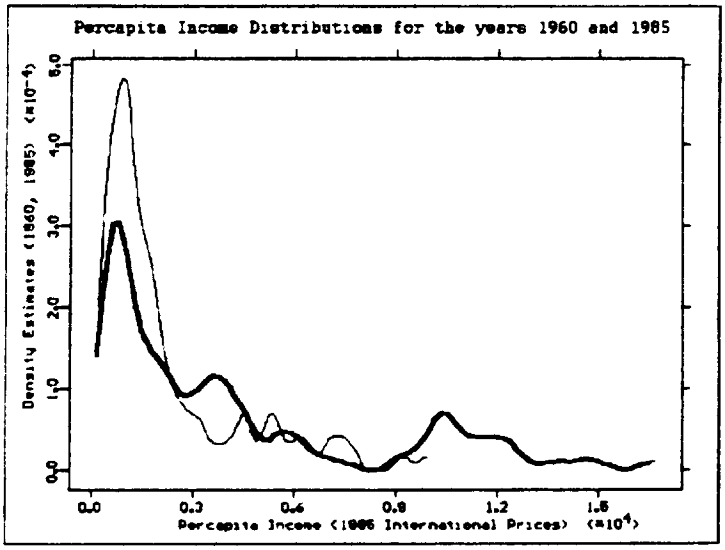


Figure 4.4.

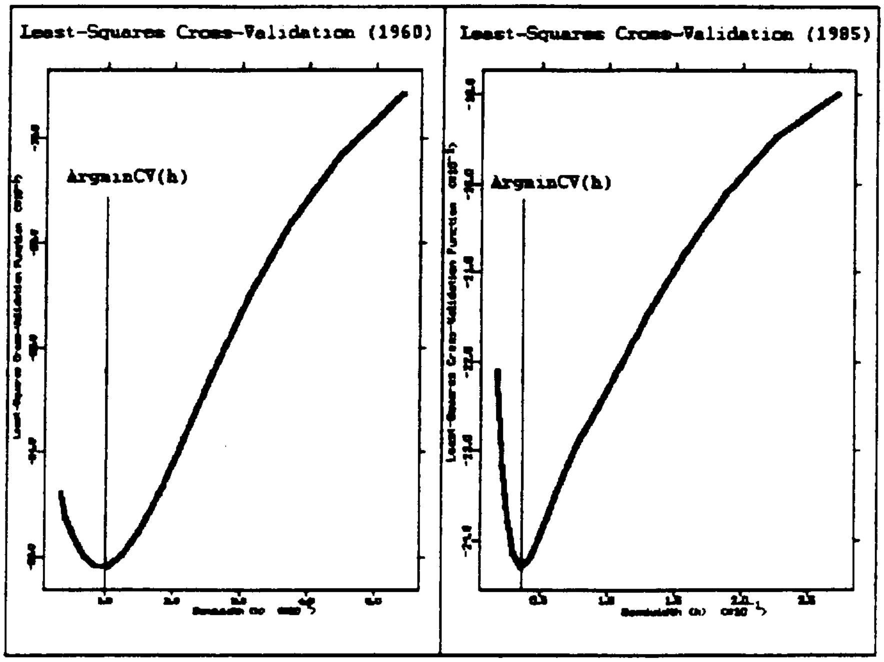
Kernel density estimates by WARPing for the per capita income distributions; with bandwidths h= 500, 850 respectively for the years 1960 (solid line) and 1985 (thick line). (1985 International Prices)

WARPing arises when looking at the year 1985. Indeed, a second mode arose followed by a wave which confirms the results found by Summers, Kravis and Heston (1984), i.e. that the middle-oil countries tend to catch-up with the industrialized countries. Finally, while some intermediate economies tended to converge, there is no doubt that some, though not all, of the poorest have been out of the "race". In order to be definitely aware of this increasing inequality and avoiding any measure of dispersion, we can plot these two figures using an identical scale (See Figure 4.4).

We note that variations of the smoothing parameter h do not change these features. In Figure 4.5, we plot the least-squares cross-validation functions (CV(h)) for both 1960 and 1985 and give some of their respective points estimates. Note that in order to compute the cross-validation functions, both random variables for the years 1960 and 1985 have been normalized so that observations are encompassed into a [0,1] interval. Indeed, note that the bandwidths which minimize the least-squares cross-validation functions estimated for both 1960 and 1985 do not smooth the density estimates to the same extent. On the one hand, the smoothing parameter which leads to a consistent estimate of the per capita income density estimate in 1960 appeared to oversmooth the data, while the opposite -i.e. undersmoothing- is true when considering the density estimate for 1985.

Therefore, the optimal smoothing parameter, once the data have been normalized, will be the one which minimizes the sum of the cross-validation functions corresponding to each random variable for both the 1960 and 1985 years. More formally, recall that:

$$CV_j(h) = ISE_j(h) + Cste_j, \text{ where } j = 1,2.$$



Least-Squares Cross-Validation (1960)

Least-Squares Cross-Validation (1985)

Figure 4.5.

Least-Squares Cross-Validation functions using a Quartic kernel, for the 1960 (Left-hand box) and 1985 years (Right-hand box). (Per capita income has been normalized relative to Max(x))

Bandwidth	CV(h) Function (1960)	Bandwidth	CV(h)Function (1985)
0.0350552	-2.47603	0.0176513	-2.21035
0.0420662	-2.52545	0.0211816	-2.31809
0.0504795	**-2.55027**	0.0254179	-2.37814
0.0605754	-2.57414	0.0305015	-2.41398
0.0726904	-2.59929	*0.0366018*	*-2.42918*
0.0872285	-2.61616	0.0439222	-2.41807
0.104674	*-2.61885*	**0.0527066**	**-2.38438**
0.125609	-2.5983	0.063248	-2.34204
0.150731	-2.55251	0.0758976	-2.29798
0.180877	-2.47934	0.0910771	-2.25608
0.217053	-2.3761	0.109293	-2.20623
0.260463	-2.24572	0.131151	-2.14485
0.312556	-2.10331	0.157381	-2.07971
0.375067	-1.9609	0.188857	-2.01164
0.45008	-1.82904	0.226629	-1.94825
0.540096	-1.71439	0.271955	-1.90027

Then, the optimal bandwidth is:

$$Arg\min\left[\sum_{j=1}^{2} CV_j(h)\right]$$

Finally, as we are interested in inequality -i.e. the absolute extent and size of changes in the income distribution- the univariate density estimates plotted at the bottom of the pages are computed using the appropriate rescaled optimal bandwidth issued by the cross-validation.

Note as well that the modality structure remains inside the 90% confidence bands computed for the density estimates of our running data sets[7]. Basically, this means that the true underlying density lies within these bands with a 90% probability. What is more, and this remains true in the remaining discussion, the modes which arise in 1985 no longer fit within these confidence bands. Thus, both their existence and locations provide some strong evidence in favour of the dramatic shifts in the income distribution within the period under study.

Alternatively, we have considered whether the two mode shape of the density estimate by WARPing either disappears or is strengthened when other measures of income are used. Indeed, the returns may be those of one person or three or more men, women, or children. In the Neo-classical framework, the "intensive" production function is based on the labour force and nothing is said about the way this should be proxied, i.e. population, workers, adults, etc... We note that a three modes shape arises for the year 1985 when a

---

[7] Formulas and computer algorithms for confidence bands can be found in Bickell and Rosenblatt (1973) as well as in Härdle (1990b).

per worker income measure is used and that this inequality does rise to a slighter extent than it does for the per capita income measure[8] (See Figures 4.6-4.7).

The scales of our density estimates do not reveal a "very slight rise" as Summers, Kravis, and Heston found in inequality for the 1950-80 period. On the contrary, even though some countries caught up in terms of per capita income, the absolute gap between the poorest and the richest has been considerably increasing over the whole period, thus corroborating the Endogenous growth theory, which argues in favour of cumulative inequalities between relatively prosperous and relatively poor countries, leading the latter to a "development trap."

Next, how do countries' current standards of living compare? This measure is quite relevant, as high economic welfare is identified with relative abundance of economic goods. This is shown in Figures 4.8-4.9. The world's current standards of living distributions show an even stronger trend towards increasing inequality than the one displayed in Figure 4.2. However, no measure of satisfaction, health, happiness and/or content is available in P.W.T5 and we will restrict our discussion by considering that income remains of interest as a measure of welfare.

However, this does not allow us to definitely reject the Neo-classical growth theory. Indeed, if one introduces technological progress which acts as the "ignored" issue, i.e. the residual, in the Solow model or different preferences, both per capita income and consumption differences might increase either in a transitory manner or tend towards infinity as time goes on. Consequently, two economies with different initial conditions might observe an increasing gap between their respective per capita income and/or

---

[8] Note that no significant change has been observed when using a per-equivalent adult gross

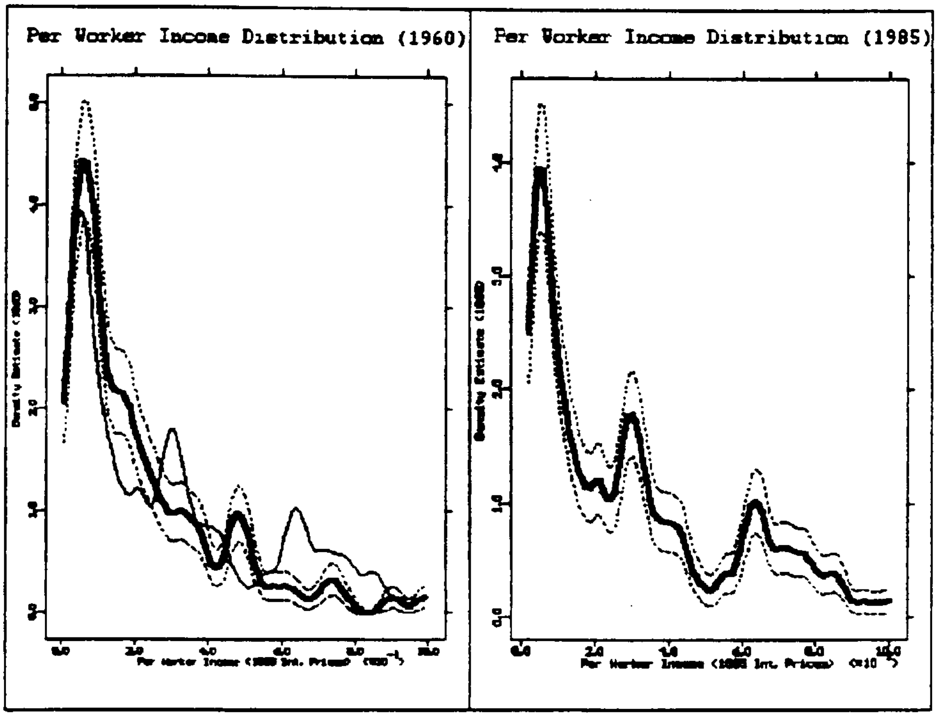


Figure 4.6.

Kernel density estimates by WARPing for the per worker income distribution (thick line); with bandwidth h= 0.06 and for the years 1960 and 1985, and 90% confidence limits (dashed lines). (1985 International Prices)

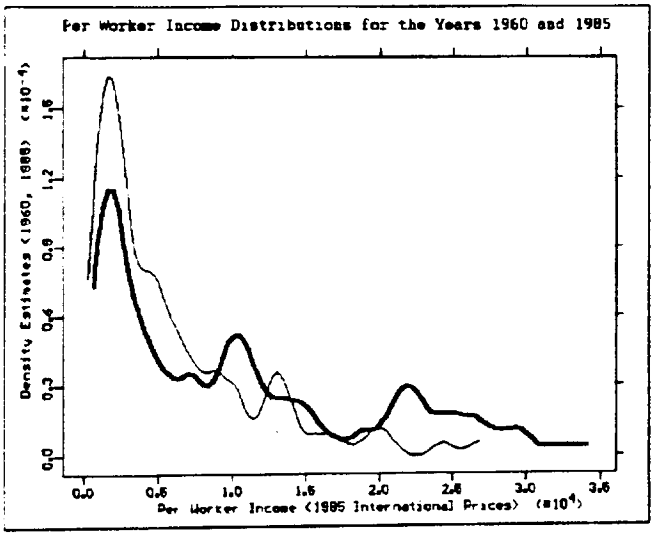


Figure 4.7.

Kernel density Estimates by WARPing for the per worker income distributions; with bandwidths h= 1620, 2060 respectively for the years 1960 (solid line) and 1985 (thick line). (1985 International Prices)

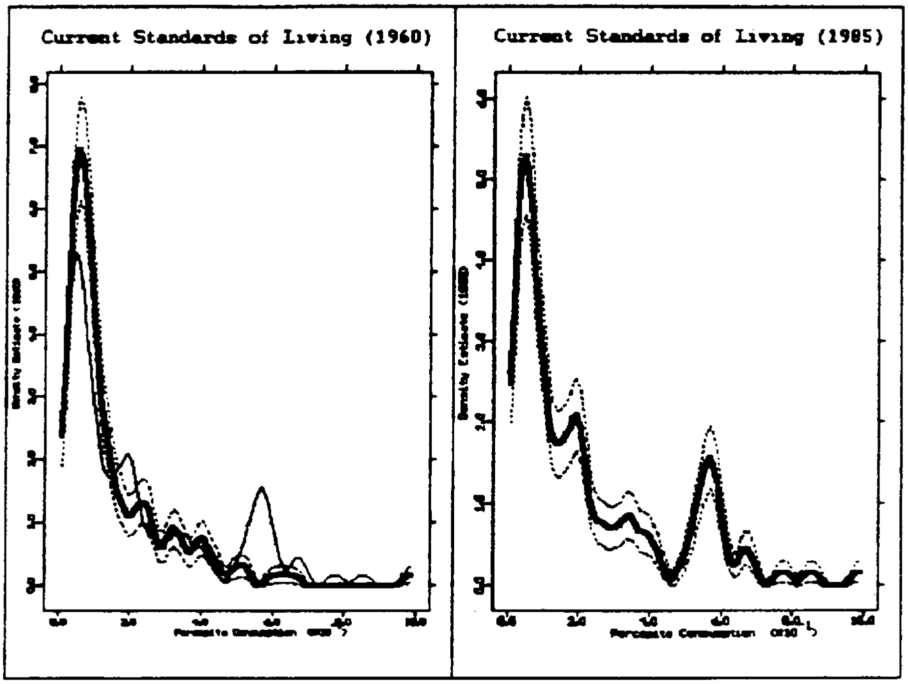


Figure 4.8.

Kernel density estimates by WARPing
for the world current standards of living distribution (thick line);
bandwidth h= 0.05 and for the years 1960, 1985, and 90% confidence limits (dashed line).
(Current International Prices)

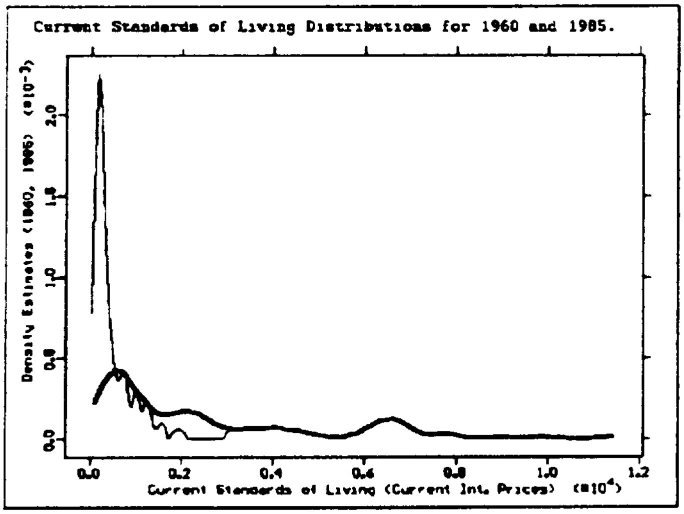


Figure 4.9.

Kernel density estimates by WARPing
for the world current standards of living distributions;
bandwidths h= 155, 695, respectively for the years 1960 (solid line), and 1985 (thick line).
(Current International Prices)

consumption level. However, as technological progress is assumed to be a non rival and non-excludable good such that poorer countries are expected to be able to make use of current best practise procedures already in use in the more productive economies, and because they provide greater returns to investment, they should therefore exhibit larger growth rates if there exists a perfect capital mobility across countries.

Because of technological progress due, for instance, to the accumulation of technological knowledge, the leader may be expected to exhibit significant growth rates - i.e. greater than the growth rate corresponding to the steady-state in the Solow model and such that $dk/dt \neq 0$- though not as large as in the poorer countries because of the assumed diminishing returns to each input and identical preferences and technologies. For instance, applying a 5% (10%) growth rate to the leader country (to the followers) with respectively a 100$ (10$) initial level of income, the absolute gap will increase from 90$ to 94$, while the relative gap will reduce from (1-(10/100))= 0.9 to (1-(11/105))= 0.895, so that what is really meant by convergence within a Neo-classical framework where technological progress is taken into account, is convergence in relative economic welfare. In other words, one expects convergence in the incomes issued by a logarithmic transformation. Indeed, we observed that some bumps and peaks did arise during the period under study so that taking logarithms will lead to something closer to symmetry, and, thus, will rule out the strong skewness present in the original income distributions. This is considered in the following section.

5/ Local versus Global Convergence.

We have just observed that per capita incomes when looking at the absolute gap among countries are far from converging. Therefore, we are led to ask how growth rates behave and whether we can draw some conclusions about the existence of different "convergence clubs." The absolute gap between countries has risen considerably over the post Second World War period: what about the relative gap -i.e. per capita and/or per worker incomes expressed in logarithms? Does it tend to disappear or not? Again, these estimates are expected to provide some indications of the shape of the underlying distribution as well as the number and locations of the modes within the system. We already noted that the choice of a logarithmic transformation applied to income measures, finds a justification in the strong skewness present in the income distribution. Note that the strong skewness present in the previous income distributions, is definitely due to the fact that African countries consist of more than one third of the observations present in the sample. Therefore, the choice of the logarithmic transformation justifies itself in order to improve the homogeneity of our sample and to get visual impressions which are somehow closer to symmetry.

We know that in order to get a consistent estimate of the true density underlying our observations, we must choose a bandwidth which will minimize a criterion function. Here again, we use the least-squares cross-validation method whose bandwidth tends to the one which minimizes the Integrated Squared Error. The cross-validation functions which are displayed in Figures 5.1(a-b) concern respectively the per capita and per worker income measures for 1985 and have been estimated using a Gaussian kernel where the optimal bandwidth is given above the figures. As already noted, the choice of the kernel does not have a great influence on the resulting density estimate if the bandwidth is optimal. Therefore, densities displayed in the following discussion are kernel density estimates using a Quartic kernel as in the previous section, after having applied the appropriate canonical kernel transformation to the optimal bandwidth issued from the least-squares cross-validation using a Gaussian kernel.

In Figures 5.2-5.3, univariate kernel density estimates issued by the logarithmic transformation of our incomes measurements are displayed. Whatever the origins are, these figures show that the recent literature focusing on the existence of multiple equilibria and arguing for divergence across countries fits the reality better than the Neo-classical framework does. Moreover, these figures show some strong evidence for local convergence rather than global convergence. However, the figures where density estimates for both 1960 and 1985 are plotted together with an identical scale might lead the reader to think that, instead of local convergence which indeed is somewhat clearer in 1985 than at the beginning of the period, we are facing a world where the middle-class countries are disappearing. Indeed, the density estimates in Figures 5.2-3, show that the mode of distribution in 1960 mainly consists of middle income countries. On the other hand this mode tends to vanish in 1985 while two extreme modes arise. If this is the case, Quah (1993a-b) was right when arguing that we are moving towards a "two-camp world" with a middle income class vanishing and where the poor countries are becoming poorer relative

Figure 5.1.a.

Least-Squares Cross-Validation function using a Gaussian Kernel,
for the per capita income measure and for the year 1985,
(Log. income measures have been normalized relative to their maximum)

ArgminCV(h)= *0.016*

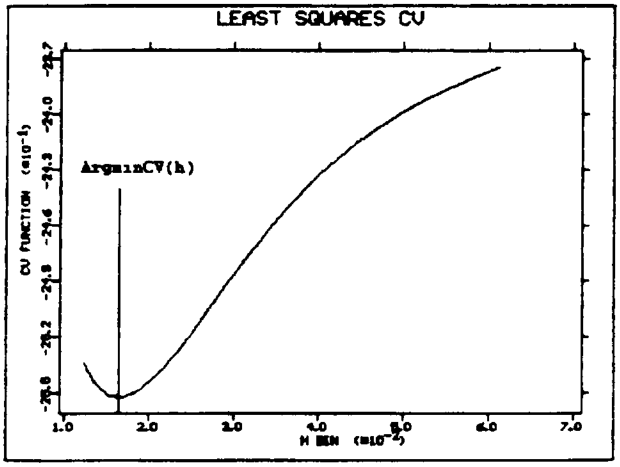
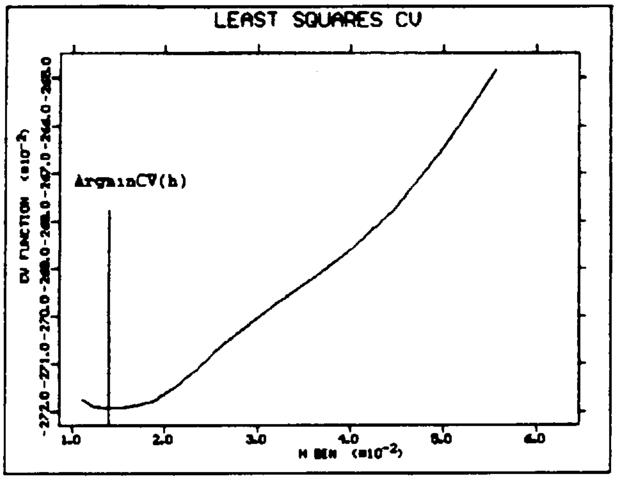


Figure 5.1.b.

Least-Squares Cross-Validation function using a Gaussian kernel,
for the per worker income measure and for the year 1985,
(Log. income measures have been normalized relative to their maximum)

ArgminCV(h)= *0.013*



to the richer countries. It should be remembered that Quah obtained this result by normalizing the same data relative to the world average per capita income and using "fractile Markov chains," while we have here chosen to study the relative behaviours using a logarithmic transformation of the data, which is another way of discussing the economic issue of convergence versus divergence, but, and this is not of little importance regarding previous studies, without imposing class groupings. Still, whether we are moving to a world with only very poor and/or very rich countries or whether the present existence of several "convergence clubs" is only transitory (this is the same question) is debatable. Indeed, looking at the density in 1960, whether it would have been possible to guess this feature of divergence because the two extreme modes already gave some hints to their future existence, is still an open question as nobody has found a reliable way to predict the behaviour in both the short and long-runs of the world-wide income distribution.

Next, it should be noted that some striking features about convergence arise when looking at the density estimates when a per worker income measure is used. Examining Figures 5.4-5.5, we note first the existence of three "convergence clubs" for the poorest countries meaning that some of the poorer countries might be close to catch up in relation to the richer countries. Indeed, the cross-section distribution of per worker incomes appears clearly to collapse. Secondly, the two modes on the right of the density estimate in 1985 seem to encompass more observations. This multimodality confirms the results found in the previous section and we are thus back to the "Barro regressions" results; that is, fertility in poor countries may act against convergence. It is probable that it does, but if this is the only reason for not observing convergence, then it should be kept in mind that the Solow model claims that when the steady-state is reached, then capital is expected to grow apace with the labour force and that Solow did not say that population is the proxy which must be used when estimating the labour force. In the Neo-classical model, convergence is indeed expected in terms of marginal productivity and Figures 5.4-5.5 give

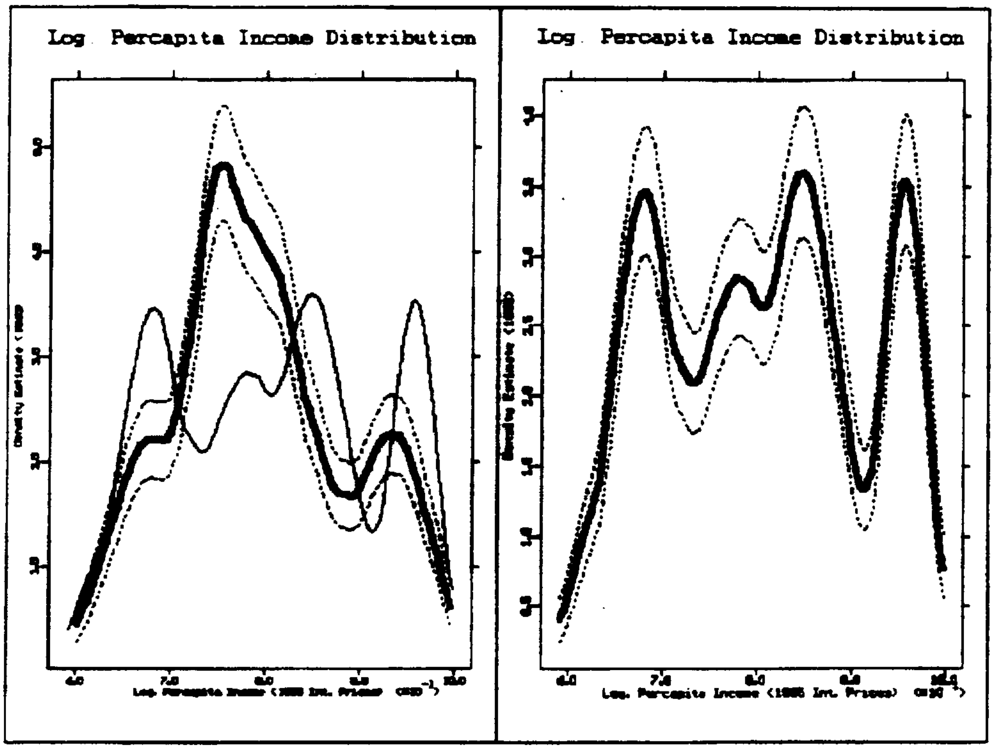


Figure 5.2

Kernel density estimates by WARPing for the logarithmic per capita income distribution (thick line), with bandwidth h= *0.042* and for the years 1960 and 1985, and 90% confidence limits (dashed lines). (1985 International Prices)

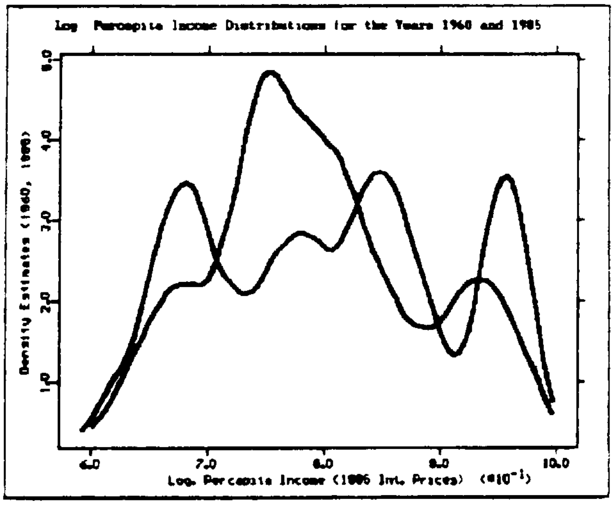


Figure 5.3

Kernel density estimates by WARPing for the logarithmic per capita income distribution, with bandwidth h= *0.042* and for the years 1960 and 1985.
(1985 International Prices)

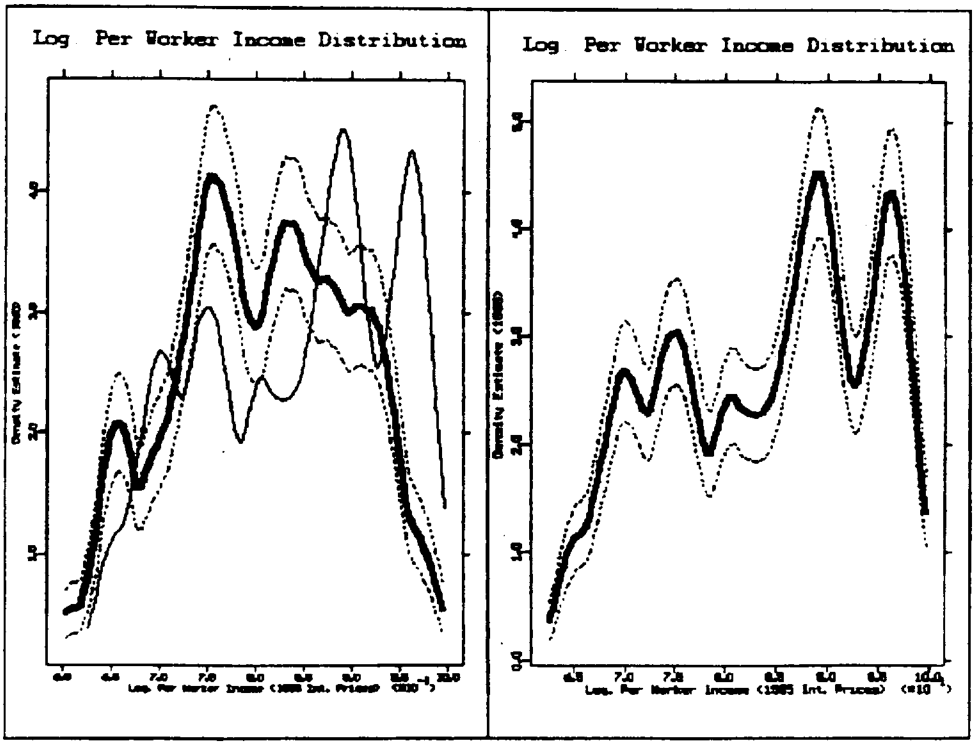


Figure 5.4

Kernel density estimates by WARPing for the logarithmic per worker income distribution (thick line), with bandwidth h= *0.034* and for the years 1960 and 1985, and 90% confidence limits (dashed lines). (1985 International Prices)

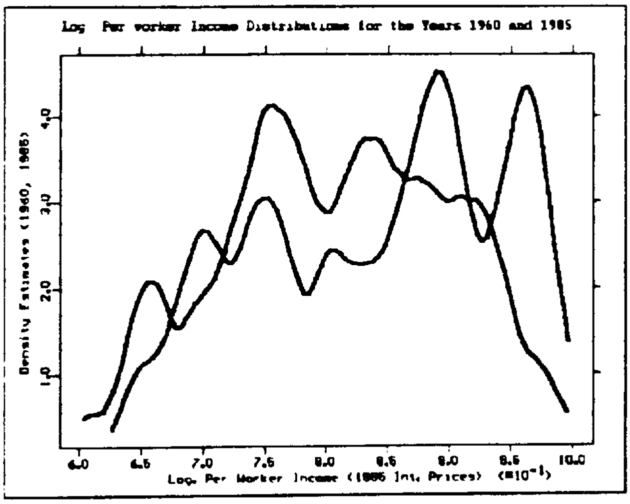


Figure 5.5

Kernel density estimates by WARPing for the logarithmic per worker income distribution, with bandwidth h= *0.034* and for the years 1960 and 1985. (1985 International Prices) some evidence that this might be happening, if one considers that GDP per worker is a better measure of productivity change. However, the problem with these estimates, is that they are points in time density estimates and do not give clear insights about the amount and the extent of mobility within the system.

Finally, looking at the confidence bands, note to what extent density estimates for the years 1960 and 1985 exhibit different structures where the concentration of mass dramatically shifted within a 25 year period. They are such that the reader could place a "90% confidence level" in the strong evidence in favour of local rather than global convergence.

The right-hand boxes in Figures 5.2 and 5.4 show some evidence, at least at the point in time estimations for the year 1985, of local rather than global convergence. The next step becomes straightforward and consists in identifying the countries falling into the bumps displayed by these density estimates. These groups of countries for the year 1985 and for both per capita and per worker[1] income measures are shown in Tables 5.1-2. These tables illustrate as Fischer (1991) did, that experiences across countries are very similar within regions. This latter, among other explanations, suggests that this might be due to the common influence of particular industrialized country partners, e.g. Japan in Asia, the US for Latin America. Finally, note that the choice of the grids in order to class the countries into groupings does not depend on *ex-ante* control variables like the initial level of output and/or the literacy rate at the beginning of the period. We first *let the data speak for themselves* and then we gathered together the countries into classes exhibited by the density estimates.

---

[1] The sample is reduced to 116 countries when the per worker income is used as these observations are missing for Guinea Bissau in the data table we have been given.

Now, let denote by ξ_q, the qth quantile of both per capita and per worker random variables such that ξ_q is the smallest number satisfying $P[X \leq x_i] \geq \xi_q$, where the x_i are the local minima. From the latter, one can by "tatonnement" compute the quantiles which fit those local minima. The results are given in the following table.

Per capita Income		$\xi_{0.31}$	$\xi_{0.49}$	$\xi_{0.80}$
Per worker Income	$\xi_{0.16}$	$\xi_{0.33}$	$\xi_{0.44}$	$\xi_{0.76}$

The inter quartile range ($\xi_{0.31}$- ξ_0) will identify the 31% poorest countries for the per capita income measure in all years of the period under study and the 20% richest countries will belong to the inter quartile range (ξ_1- $\xi_{0.80}$). Note first that the two groups regrouping the richer countries encompass more countries when the per worker income measure is used, and that the 30% poorest countries can now be divided into two groups. Next, we computed the income values corresponding to those quantiles for the year 1960. This procedure gives the first hints to the amount of mobility in the distribution of countries over the period. Thus, both upward and downward movements are available in Tables 5.3-4. The country classifications for both measures of income and for the year 1960 are available in the Appendices 4.a-b.

Whether the countries belonging to the first groups for both income measures have been out of the race because they fell under colonial rules is an open question. Note, however, that countries like Ethiopia, Afghanistan and Nepal which have never been colonized did not grow at an "impressive" rate. The period under study begins at exactly the same time as the beginning of independence and although the African continent's colonial background definitely hindered their development, there can be no doubt that their own political organisation -e.g. in Burma, Tanzania, Uganda and Zambia- since 1960 has also contributed either to stuck or to their being led, as the downward movements suggest, into a development trap (See, for instance, Reynolds; 1983). Reynolds also notes that differences in growth rates do not necessarily hold in factor endowments. Indeed, Zaire, a relatively resource-rich country remains very poor, while countries with poor natural resources like Hong-Kong, Japan, Korea, Singapore and Taiwan could make any agnostic believe that "miracles" do exist. A better explanation might be export-led growth and the fact that technological spillovers confine to the industrial sector (See, for instance, Dowrick and Gemmel; 1991). Many other comments could illustrate the composition of these "convergence clubs", but we leave this discussion to researchers with a deeper knowledge of the situation and its historical background. Still, note that the identified upward and downward movements in Tables 5.3-4, are consistent with other empirical studies. For instance, Easterly & Al. (1993) focused on the low persistence of growth rate differences across countries. Displaying the scatter plot of each country growth rate for the per worker income measure and for 115 countries issued from PWT5 over two periods (1960-73 and 1974-88), they conclude that persistence of growth rate differences across countries is low with a rank correlation of 0.21. Thus, we can associate our results to a basic fact and such that instability of relative incomes concerns countries which have experienced large and persistent differences in their growth experience, like on the one hand, the famous Asian Gang of Four and Botswana and, on the other hand, Chad, Guyana, Madagascar, etc... At least, countries which would require a case study approach -i.e. a pragmatic empirical work- in order to make more specific conclusions about the performance of recommanded policy changes, are made available avoiding any subjective choices.

However, there are two kinds of problems with this kind of procedure. First, keeping the inter quartile ranges fixed make the transitions dependent on each other. In other words, if one country moved from Group 1 to Group 3, this change must be offset by either a similar change for another country or by two changes for two different countries, e.g. one country will move from Group 3 to Group 2 and another from Group 2 to Group 1. Therefore, even though the range of arrangements is quite large, this dependency acts against a reliable estimation of the amount of mobility in the system. Secondly, the amount of mobility exhibited in Tables 5.3 and 5.4 is crucially dependent on the choice of the smoothing parameter as it determines the shape of the density estimates when averaging the kernel functions associated with the observations falling into the bins.

The mobility in the system remains a crucial topic for the study of the economic issue of convergence. Until now, we have estimated point in time density estimates which converge to the true distributions of our observations. Looking at Figures 5-6 and 5.7, one becomes aware that the mobility within the system cannot be ignored and that the above procedure definitely underestimates it. The amount and the extent of mobility are considered in the next section.

TABLE 5.1

Country Classification for the year 1985
(Per capita Income expressed relative to 1985 prices)

Group 1	*Group 2*	*Group 3*	*Group 4*
Benin	Botswana	Algeria	Hong-Kong
Burundi	Cameroon	Congo	Israel
Central Af.	Cape Verde Is.	Gabon	Japan
Chad	Egypt	Mauritius	Saudi Arabia
Ethiopia	Ivory coast	So. Africa	Singapore
Gambia	Morocco	Tunisia	Austria
Ghana	Swaziland	Iran	Belgium
Guinea	Zimbabwe	Iraq	Denmark
Guinea Bissau	China, P.R.	Jordan	Finland
Kenya	Pakistan	Korea, Rep. of	France
Lesotho	Philippines	Malaysia	Germany
Liberia	Sri Lanka	Syria	Iceland
Madagascar	Thailand	Taiwan	Italy
Malawi	Dominican, Rep.	Cyprus	Luxembourg
Mali	El Salvador	Greece	Netherlands
Mauritania	Guatemala	Ireland	Norway
Mozambique	Nicaragua	Malta	Sweden
Niger	Bolivia	Portugal	Switzerland
Nigeria	Guyana	Spain	Great Britain
Rwanda	Paraguay	Turkey	Canada
Senegal	Papua New Guinea	Yugoslavia	Trinidad & Tob.
Sierra Leone		Barbados	U.S.A.
Somalia		Costa Rica	Australia
Sudan		Jamaica	New Zealand
Tanzania		Mexico	
Togo		Panama	
Uganda		Argentina	
Zaire		Brazil	
Zambia		Chile	
Afghanistan		Colombia	
Bangladesh		Ecuador	
Burma		Peru	
India		Surinam	
Nepal		Uruguay	
Haiti		Venezuela	
Honduras		Fiji	

TABLE 5.2

Country Classification for the year 1985
(Per worker Income expressed relative to 1985 prices)

Group 1	*Group 2*	*Group 3*	*Group 4*	*Group 5*
Burundi	Benin	Botswana	Algeria	Hong-Kong
Central Af.	Ghana	Cameroon	Congo	Israel
Chad	Ivory Coast	Cape Verde. Is.	Egypt	Japan
Ethiopia	Kenya	Swaziland	Gabon	Saudi Arabia
Gambia	Lesotho	Pakistan	Mauritius	Singapore
Guinea	Liberia	Philippines	Morocco	Syria
Madagascar	Mauritania	Sri Lanka	So. Africa	Austria
Malawi	Nigeria	Thailand	Tunisia	Belgium
Mali	Senegal	El Salvador	Iran	Denmark
Mozambique	Sierra Leone	Honduras	Iraq	Finland
Niger	Somalia	Jamaica	Jordan	France
Rwanda	Sudan	Bolivia	Korea. Rep.	Germany
Tanzania	Zambia	Guyana	Malaysia	Iceland
Togo	Zimbabwe		Taiwan	Italy
Uganda	Afghanistan		Cyprus	Luxembourg
Zaire	Bangladesh		Greece	Netherlands
Burma	China. P.R		Ireland	Norway
India	Haiti		Malta	Spain
Nepal	Papua New Guinea		Portugal	Sweden
			Turkey	Switzerland
			Yugoslavia	Great Britain
			Barbados	Canada
			Costa Rica	Mexico
			Dominican Rep.	Trinidad & Tob.
			Guatemala	U.S.A.
			Nicaragua	Venezuela
			Panama	Australia
			Argentina	New Zealand
			Brazil	
			Chile	
			Colombia	
			Ecuador	
			Paraguay	
			Peru	
			Surinam	
			Uruguay	
			Fiji	

TABLE 5.3

Upward and Downward movements from one group to another from 1960 to 1985 for the per capita income expressed relative to 1985 prices.

UPward Movements

Group 1 to Group 2
Botswana
Cameroon
Cape Verde Is,
Egypt
Morocco
Zimbabwe
China, P.R.
Pakistan

Group 1 to Group 3
Korea, Rep. of

Group 2 to Group 3
Congo
Jordan
Taiwan

Group 3 to Group 4
Hong-Kong
Japan
Singapore

DOWNward Movements

Group 2 to Group 1
Benin
Ghana
Liberia
Madagascar
Mozambique
Nigeria
Senegal
Sudan
Zambia

Group 3 to Group 2
Sri Lanka
Guatemala
Nicaragua
Guyana

Group 4 to Group 3
Barbados
Uruguay
Venezuela

TABLE 5.4

Upward and Downward movements from one group to another from 1960 to 1985
for the per worker income expressed relative to 1985 prices.

UPward Movements

Group 1 to Group 2
Lesotho
China, P.R.

Group 1 to Group 3
Botswana

Group 2 to Group 3
Cameroon
Swaziland
Thailand

Group 2 to group 4
Congo
Egypt
Gabon

Group 3 to Group 4
Morocco
Korea, Rep. of
Taiwan

Group 4 to Group 5
Hong-Kong
Japan
Singapore
Syria
Spain

DOWNward Movements

Group 2 to Group 1

Chad
Madagascar

Group 3 to Group 1
Mozambique

Group 3 to Group 2
Ghana
Mauritania
Nigeria
Senegal
Sudan

Group 4 to Group 2
Zambia

Group 4 to Group 3
Sri Lanka
El Salvador
Jamaica
Bolivia
Guyana

Group 5 to Group 4
Iraq
Barbados
Argentina
Chile
Uruguay

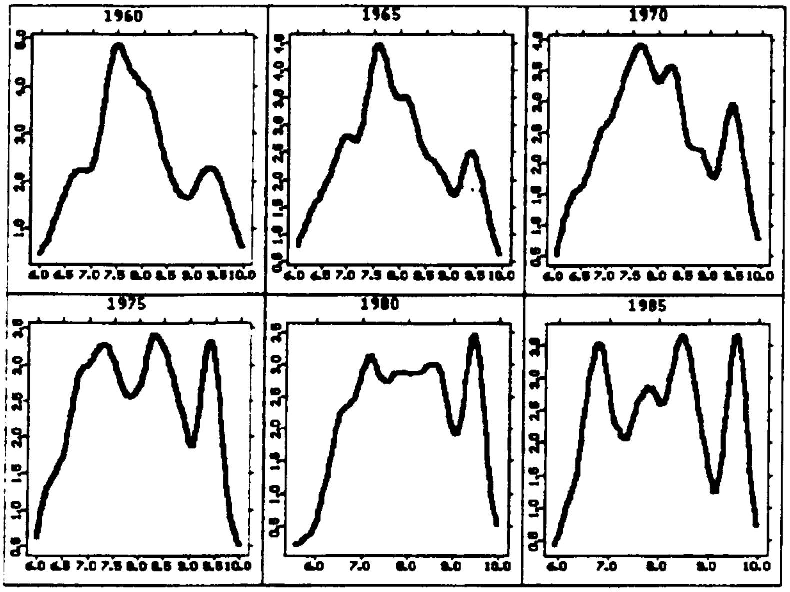


Figure 5.6.

Kernel density estimates every fifth year by WARPing for the logarithmic per capita income distribution, with bandwidths h= 0.042.
(1985 International Prices)

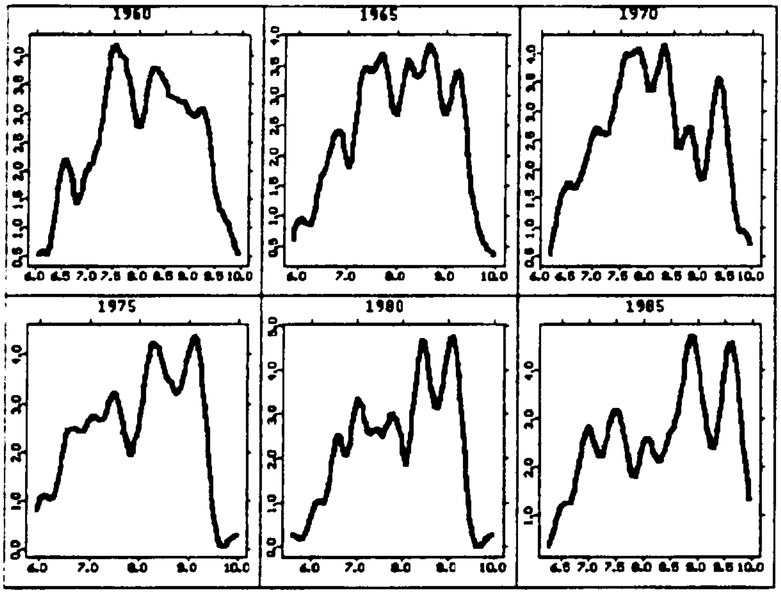


Figure 5.7.

Kernel density estimates every fifth year by WARPing for the logarithmic per worker income distribution, with bandwidths h= 0.034.
(1985 International Prices)

Before leaving the issue about the existence of multiple versus a unique and global equilibrium, we would like to take advantage of non-parametric techniques. For bivariate distributions in general, the mean of Y in the conditional density of Y given X=x will be some function of x, say m(x), and the equation Y=m(x) when plotted in the xy plane gives the regression curve for Y. That is,

$$E[Y|X=x]=m(x)=\frac{\int yf(x,y)dy}{f(x)}$$

Then, if we have n independent observations $\{(X_i,Y_i)\}_{i=1}^{118}$ the regression relationship is estimated in the following way:

$$Y_i=\hat{m}_h(X_i)+\varepsilon_i$$

where, here, $\hat{m}_h(x)=\frac{\int y\hat{f}_h(x,y)dy}{\int \hat{f}_h(x,y)dy}$, is the Nadaraya-Watson kernel estimator, and ε is a random variable denoting the variation of Y around $\hat{m}_h(x)$.

Note that the approximation of m(x) commonly called "smoothing," is quite similar to that in density estimation and supposes no prior knowledge on m (See, for instance, Härdle; 1990a). In density estimation, we considered the points in a small neighbourhood around x and weighted the frequencies of observations falling into a window with bandwidth h centred in x. Here, the procedure consists in weighting the response variables near a point x and is known as *local averaging*. Still, note that non-parametric techniques are not used here either to avoid traditional parametric model misspecification nor they are seen as a benchmark in order to compare with any parametric forms. They are used as a suitable tool in order to illustrate a scenario of local rather than global convergence. Note that a formal presentation of the techniques used in the present discussion, is available in Section 7.

Here, we need non-parametric smoothing, because a scenario which would exhibit local rather than global convergence, can be written in the following way. Let X_i (Y_i), be the logarithm of either per capita or per worker income for the year 1960 (1985) and run the following regression smoothings without imposing any functional form at all,

$$\begin{aligned} Y_i &= \hat{m}_h(X_i) + \varepsilon_i \\ X_i &= \hat{m}_h(Y_i) + \eta_i \end{aligned}$$

Then, the scenario of *local convergence* is such that $\mathrm{Var}(\eta_i)$ is expected to be "much greater" than $\mathrm{Var}(\varepsilon_i)$[2].

Before presenting the results, we need to choose the amount of smoothing inherent to our scenario: that is, we must choose, in analogy with the density estimation problem, a bandwidth which will regulate the size of the neighbourhood around x and which will minimize some criterion function. At an intuitive level, Deaton (1989) illustrates the choice concerning the degree of smoothness in the estimation procedure through a traditional task which consists in smoothing a time-series of daily stock returns, by calculation of a moving average. Indeed, in order to remove some of the noisiness present in the observations, one can plot instead of the day return itself, an average of the returns for the d previous days, the day itself, and the d succeeding days. Then, one understands at an intuitive level, that the bigger is d, the smoother will be the resulting plot. The same idea applies to the choice of the bandwidth which will determine how near observations have to be in order to contribute to the average at each point. Note that the performance of the estimated regression smoothers is controlled through the cross-validation method described in Section 7 such that the bandwidth used in the models estimation is the one which minimizes the sum of the average of the sum of squared prediction errors of both smoothings, after a normalization of the observations relative to their maximum. The resulting optimal bandwidths for both the percapita and per worker income measures are respectively equal to 0.165 and 0.19.

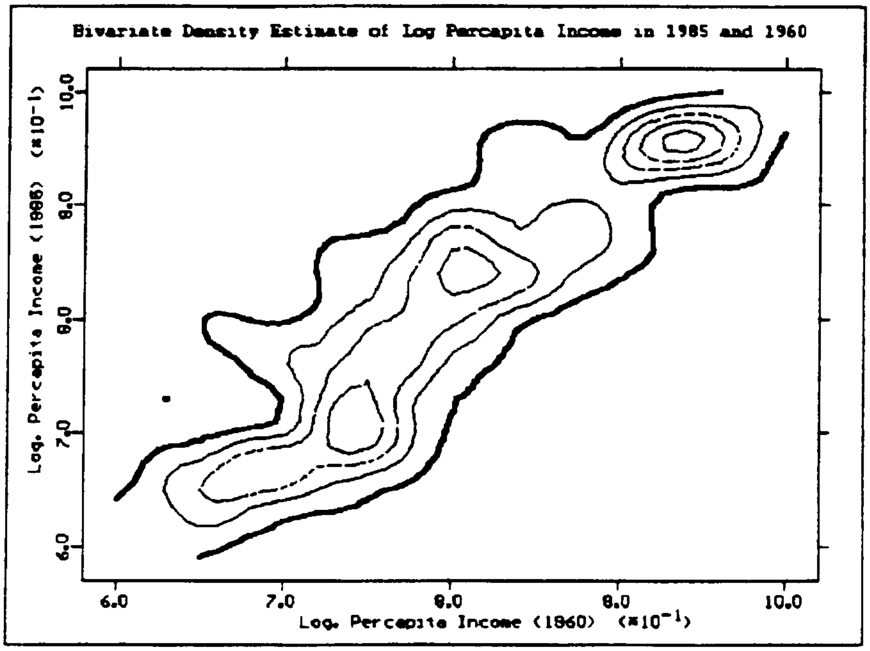
---

[2] We thank Robert Waldmann for having suggested this scenario.

Unfortunately and as already mentioned, although both the per capita and per worker income density estimators for 1985 do show evidence in favour of local convergence, they are point in time density estimates and do not give any clear insights about the cross-sectional dynamics. The above scenario suggesting local rather than global convergence was expected to support the above evidence of local convergence. However, results do not corroborate our expectations. Indeed, we find for the per capita income measure that $Var(\eta_i)= 509.10^{-5}$ while $Var(\varepsilon_i)= 535.10^{-5}$. That is, both variances are very close from one to each other and the inequality runs in the opposite way of the expected one. On the other hand, if the scenario of local convergence applies to the per worker income measure with $Var(\eta_i)= 66.10^{-4}$ while $Var(\varepsilon_i)= 64.10^{-4}$, the latter is only ≅3% smaller.

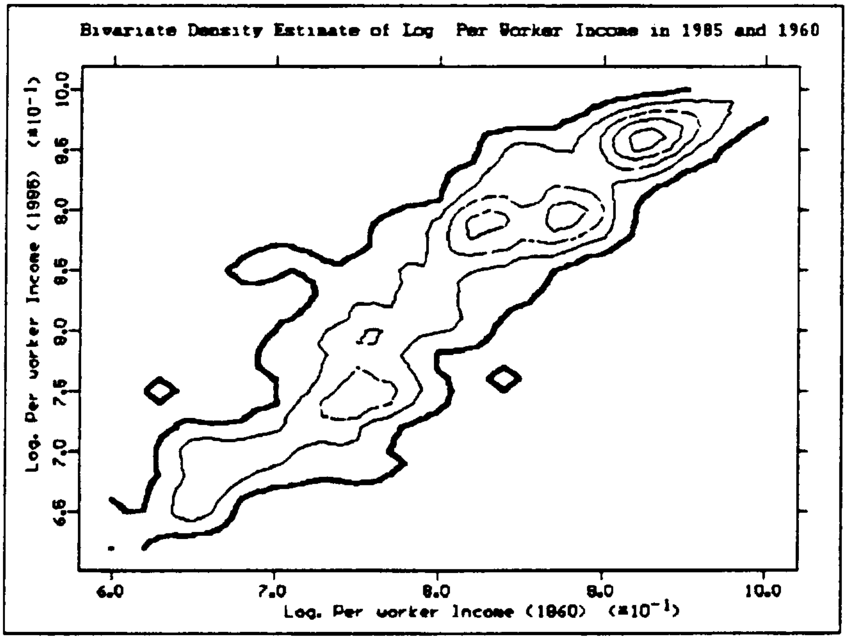
This leads us to estimate the joint density in order to look at the bivariate distribution of our random variables. To this end, we plot in Figures 5.8-9, the equal probability contours of bivariate kernel density estimates of log. incomes in 1985 and 1960, for both per capita and per worker income measures. These contours give a clear readable impression of relative heights and thus of the concentration of mass. Note first that the contours exhibit some strong evidence of local rather than global convergence especially for both the highest and lowest classes of income. Thus, if our scenario failed to support the local convergence, this is due to the large and persistent mobility experienced by countries which initially belonged to the middle income class, some catching-up and others falling behind. Again and as already suggested by Quah (1993b), this shows the limits of univariate density estimates of incomes, which are point in time density estimators and which cannot therefore give any clear insights about long-run regularities in cross-sectional dynamics.

Figure 5.8



Equal probability contours (0.1, 0.3, 0.5, 0.7, 0.9) of bivariate kernel density estimate of Log. per capita income in 1985 and 1960, with bandwiths h= *0.042*, in both X and Y-directions.

Figure 5.9



Equal probability contours (0.1, 0.3, 0.5, 0.7, 0.9) of bivariate kernel density estimate of Log. per worker income in 1985 and 1960, with bandwiths h= *0.034*, in both X and Y-directions.

6/ Original Scenarios to Illustrate the Economic Issue of Convergence.

We showed in the previous section that the world we were facing in 1985 was one which exhibited local rather than global convergence. We also noted, as illustrated by Quah (1993c), that this local convergence might only be transitory if the world tends towards a "two-camp world", divided into very rich and very poor countries, with the middle income countries either catching up or falling behind. We also stressed that those point in time density estimates do not give good insight in terms of mobility within the system. In other words, the snapshots exhibited in the two previous sections appear to be limited when one wants to discuss the transition probabilities -i.e. the probabilities for countries to converge or to diverge. Recall here the Neo-classical convergence argument: if some countries have already reached their steady-state, then backward countries are expected to grow faster than the "leader" as the framework consists in decentralized markets with diminishing returns to each input. Note the two key expressions in this argument: first, the existence of backward countries relative to a leader and, secondly, that the former are expected to grow faster than the latter. This argument can be expressed in terms of a posteriori probabilities which are expected to reflect some long-run regularities in cross-sectional dynamics along the lines pioneered by Quah (1993a).

First, one normalizes the per capita income of each country in the sample relative to the US, which is commonly presented as the leader for the post-Second World War period. Thus, a new random variable (X) is created, such that a country i characterized by an observation $x_i < 1$ will be a country which is, according to the Neo-classical argument, expected to "catch-up" the leader and/or to grow faster than this latter. The "catch-up" or phenomenon of convergence is such that:

S_1: $P[\, x_i(t) < 1 \mid x_i(t) < x_i(t+n)]$ and $P[\, x_i(t) \geq 1 \mid x_i(t) > x_i(t+n)]$

While, the alternative scenario -i.e. where the relative gap is constant or increasing- can be written.

S_2: $P[\, x_i(t) < 1 \mid x_i(t) \geq x_i(t+n)]$ and $P[\, x_i(t) \geq 1 \mid x_i(t) \leq x_i(t+n)]$

The results will be summarized and presented in the following way:

	$P[\, x_i(t) < 1]$	$P[\, x_i(t) \geq 1]$
$P[x_i(t) > x_i(t+n)]$	A (S_2)	D (S_1)
$P[x_i(t) = x_i(t+n)]$	B (S_2)	E (S_2)
$P[x_i(t) < x_i(t+n)]$	C (S_1)	F (S_2)

Therefore, our experiment involves two random variables of interest and consists in dividing our sample of countries into countries initially poorer or richer than the US and at the same time, to look at the value they take relative to the leader over a n-years horizon performing a pooled cross-section, time-series investigation such that we eliminate cyclical components, thus, doing justice to the notion of business cycles. On the one hand, the outcome A will represent the proportion of occurences in which countries are initially poorer that the US and that they felt behind after n years by an amount which will be automatically determined by the discretization into grids of the coordinates, also occurred.

On the other hand, outcome C considers the proportion of occurences in which countries are initially poorer than the US and that they caught up after n years also occurred. Finally, in case B, countries are selected such that they must be initially poorer than the US and such that the relative gap remained constant over a n-years horizon.

As noted in the methodology, kernel density estimates can easily be drawn for multivariate densities. Here, as we are left with two joint discrete random variables of interest, we will first estimate by WARPing the joint density of ($x_i(t)$, $x_i(t+n)$) and then derive the corresponding a posteriori five and ten years transition probabilities -i.e. n=5,10[1]. The marginal smoothing parameters are such that the set of possible income values is discretized into grids at 0, 5, 10,..., 120% of the US income, so that the probability for a country to convergence will be to jump from an initial income class, say 20% of US income, to a greater income class -i.e. >20%. At this stage, the reader may think that upward movements from, say a 20% to a 25% grid may not be really significant in order to definitely accept the convergence hypothesis. However, one must first keep in mind that the periods under study only cover either five or ten-years horizons so that the amount of mobility and the extent of mobility -i.e. the rate of convergence- have to be distinguished. In other words, the opportunity to catch-up is one thing, the velocity at which it proceeds, is another one. Secondly, a country belonging to the 20% grid relative to the US in the year 1960, will exhibit a per capita income of ≅1995 international Dollars. If this latter jumps to the 25% grid in 1965, it will exhibit a ≅2915 international Dollar per capita income which is equivalent on average and for the period, to a ≅7.6% growth rate

---

[1] This methodology is directly issued by Quah (1993a). He estimated these joint densities and their corresponding graphs showed some strong evidence for observations to concentrate on the 45°-degree line, so that the relative gap between countries is very likely to remain constant. However, as the time separating two income measures is increasing, he finds that intra-distribution mobility occurs and that mobility appears to be more pronounced for both high income countries and for countries which initially exhibited some "potential to catch-up." The point here is that the probability transitions do not appear in his study. Instead, in order to produce directly understandable and readable graphs, he conditioned the joint densities given $P[x_i]$. Note that as we want to estimate the probabilities to catch-up, we cannot rule out the stylized fact already developed in the previous sections that countries are not uniformly distributed along the real line.

per year. This example gives insight into the extent of mobility which is exhibited in the following transition probabilities. The five and ten years a posteriori transition probabilities for both the per capita and per worker income measures, are presented respectively in Tables 6.1 and 6.2, while the corresponding graphs for the per capita income measure are available in Appendix 5. As least-squares cross-validation carry over the multivariate case with no essential modification, note that the underlying a posteriori probabilities to the bivariate kernel densitiy estimates, are issued by an estimation which used optimal bandwidths in each component. Secondly, as the Stone theorem (1984) applies equally to the multivariate case, both bandwidths tend therefore to the ones which minimize the Integrated Squared Error.

Observations concentrate on the 45°-degree line as illustrated by the number of occurences of case B, and neither bivariate density estimates for five and ten years transition exhibit large upward or downward movements. However, the amount of mobility is rather significant. Maybe more interestingly, both probabilities for the relative gap either to increase or to remain constant, are decreasing as n, the length of the period, increases. This might be interpreted as an argument in favour of the convergence hypothesis; that is, more countries experience catch-up over longer horizons. However, there still remains a great probability to diverge on the one hand, while the relative gap remaining constant is a very likely occurence. As noted above, the present study covers 25 years so that estimating the 15 and 20 years transition probabilities may lead the results to be greatly dependent on what happened from the mid-70s' (1979 oil shock, inverted Phillips relationship between inflation and activity in terms of employment, raw materials terms of trade deterioration, etc) to the mid-80s' which are known as the lost decade for most developing countries, especially in Africa and Latin America which developing economic policy focused on structural adjustment -i.e. a combination of macroeconomic stabilization measures to restore domestic and external equilibrium. In other words, cycles

TABLE 6.1

5-Years Transition Probabilities.

Per capita Income	$P[\, x_i(t) < 1]$	$P[\, x_i(t) \geq 1]$
$P[x_i(t) > x_i(t+5)]$	0.292 (S_2)	0.005 (S_1)
$P[x_i(t) = x_i(t+5)]$	0.305 (S_2)	0.001 (S_2)
$P[x_i(t) < x_i(t+5)]$	0.396 (S_1)	0.001 (S_2)

Per Worker Income	$P[\, x_i(t) < 1]$	$P[\, x_i(t) \geq 1]$
$P[x_i(t) > x_i(t+5)]$	0.111 (S_2)	0.000 (S_1)
$P[x_i(t) = x_i(t+5)]$	0.528 (S_2)	0.000 (S_2)
$P[x_i(t) < x_i(t+5)]$	0.361 (S_1)	0.000 (S_2)

TABLE 6.2

10-Years Transition Probabilities.

Per capita Income	$P[\, x_i(t) < 1]$	$P[\, x_i(t) \geq 1]$
$P[x_i(t) > x_i(t+10)]$	0.236 (S_2)	0.005 (S_1)
$P[x_i(t) = x_i(t+10)]$	0.277 (S_2)	0.001 (S_2)
$P[x_i(t) < x_i(t+10)]$	0.481 (S_1)	0.000 (S_2)

Per Worker Income	$P[\, x_i(t) < 1]$	$P[\, x_i(t) \geq 1]$
$P[x_i(t) > x_i(t+10)]$	0.104 (S_2)	0.000 (S_1)
$P[x_i(t) = x_i(t+10)]$	0.378 (S_2)	0.000 (S_2)
$P[x_i(t) < x_i(t+10)]$	0.518 (S_1)	0.000 (S_2)

and macroeconomic policy changes definitely influence the probabilities meaning that only five and ten years transition probabilities, which allow both expansion and recession phases to be covered, are relevant for our scenarios. Note as well that the same bivariate density estimates as applied to the per worker income measure by Quah, exhibit a smaller amount of mobility where the probability to diverge almost disappears (≅10%) for both five and ten years transition probabilities. This confirms the difference observed in the shapes of the logarithmic per capita and per worker income measures plotted in the previous section. Again, note that the probability to converge is larger as n increases.

Therefore, the immobility on the 45°-line, exhibited in the graphs available in Appendix 5, does not give a good insight into the amount of mobility within the system. Indeed, the matrices of a posteriori transition probabilities are such that not only rich and middle income countries have experienced either upward or downward movements. In other words, mobility is not confined to middle and high income classes; low income classes may catch up and middle and high income classes may fall behind. Unfortunately, this feature of the world does not appear either in the graphs or in the probabilities issued by the above scenarios of convergence versus divergence. Both the development trap and the amount of immobility found in the previous literature find an explanation in what one might call the "unequalizing spiral." Once one normalizes the data relative to the leader, a country starting at the 5% (respectively at the 60%) grid in 1960 will exhibit a ≅500 (≅5990) international Dollar per capita income. In order to catch up, in our density estimates, i.e. to move at least to the 10% (65%) grid, it should reach in 1965 a per capita income of ≅1150 (≅7580) international Dollar which is equivalent on average to a ≅17% (≅4.8%) growth rate per year. This example illustrates why countries at the lowest income classes, even though they exhibit greater growth rates relative to the leader, get stuck in the development trap and why the amount of immobility appears to be greater at these stages of development even though the real line is divided into very small grids. However, as argued before, one must distinguish the amount from the extent of mobility. Therefore, we are left to measure these latter in another way.

If a backward country relative to the leader is expected to catch-up, then the former must obviously exhibit greater growth rates. The point here is to normalize growth rates relative to the US growth rate. We create a new random variable (Y) where a backward country i is expected to exhibit $y_i>1$ if the convergence hypothesis is to be accepted. More formally, we propose two new scenarios which will involve the two discrete random variables as defined above, and such that:

S_3: $P[x_i < 1 | y_i > 1]$ and $P[x_i \geq 1 | y_i \leq 1]$

The alternative scenario which might be seen as the probability for the absence of opportunity to catch up will be defined this way:

S_4: $P[x_i < 1 | y_i \leq 1]$ and $P[x_i \geq 1 | y_i > 1]$

The results will be presented as previously, where the a posteriori probabilities given in boxes C and D will correspond to the convergence hypothesis, while the alternative hypothesis will be shown in the remaining boxes. Note that boxes B and E correspond to the case where the relative gap remains constant while in cases A and F, it is increasing. Case C considers the proportion of occurences of being a backward country and to have experienced, at the same time, faster growth relative to the US.

	$P[x_i < 1]$	$P[x_i \geq 1]$
$P[y_i < 1]$	A (S_4)	D (S_3)
$P[y_i = 1]$	B (S_4)	E (S_4)
$P[y_i > 1]$	C (S_3)	F (S_4)

The corresponding transition probabilities are shown in Tables 6.3 and 6.4. As one would have expected, these new scenarios lead to a great amount of mobility. The probability for the relative gap to remain constant almost disappears even over a five-years horizon. Note that, again, the probability to converge within our system increases as n, the length of the period, increases and that convergence is more likely when using a per worker income measure. Furthermore, the underlying probability issued from the estimation of bivariate density of average growth rates and per worker income relative to the US, reaches ≅70% over a ten-years horizon. At first sight, this appears to act strongly in favour of the catch-up hypothesis. However, and as already discussed, the world we were facing in 1985, exhibited local rather than global convergence, and these scenarios are not appropriate to a discussion of the existence of multiple equilibria.

The corresponding three-dimensional perspective plots of bivariate kernel density estimates which led to these transition probabilities, as well as equal probability contours of these latter are available in Figures 6.1-4[2], with grids every 5% in the X-direction and 0.5% in the Y-direction. They highlight why previous studies using parametric models did not lead to a consensus about the key economic issue of convergence for a world-wide set of countries and why the sample selection criteria is so crucial in estimating the partial association between average growth rates and initial levels of productivity. Note that the height of the joint density estimates represent the fraction of countries at the different levels of income and their associated average growth rates relative to the US over both five and ten years horizons. Then, for any class of per capita income, the most likely occurrence is to experience growth very close to the leader. Indeed, growth rates appear to be "normally" distributed around ≅1, although they are either slightly positively or slightly negatively

---

[2] The transition probabilities have been computed every fifth year, such that Figures 6.3-5 correspond to the joint density estimates $f(x_i, y_i)$ where x_i is the initial percapita income of country i relative to the US one at time t and y_i is the annual average growth rate normalized to the US over the five and ten following years. The procedure of averaging consisted in geometrical mean of indices base 100 in the previous year.

TABLE 6.3

5-Years Transition Probabilities.

Per capita Income	$P[x_i < 1]$	$P[x_i \geq 1]$
$P[y_i < 1]$	0.447 (S_4)	0.007 (S_3)
$P[y_i = 1]$	0.070 (S_4)	0.002 (S_4)
$P[y_i > 1]$	0.472 (S_3)	0.002 (S_4)

Per Worker Income	$P[x_i < 1]$	$P[x_i \geq 1]$
$P[y_i < 1]$	0.338 (S_4)	0.007 (S_3)
$P[y_i = 1]$	0.064 (S_4)	0.001 (S_4)
$P[y_i > 1]$	0.582 (S_3)	0.008 (S_4)

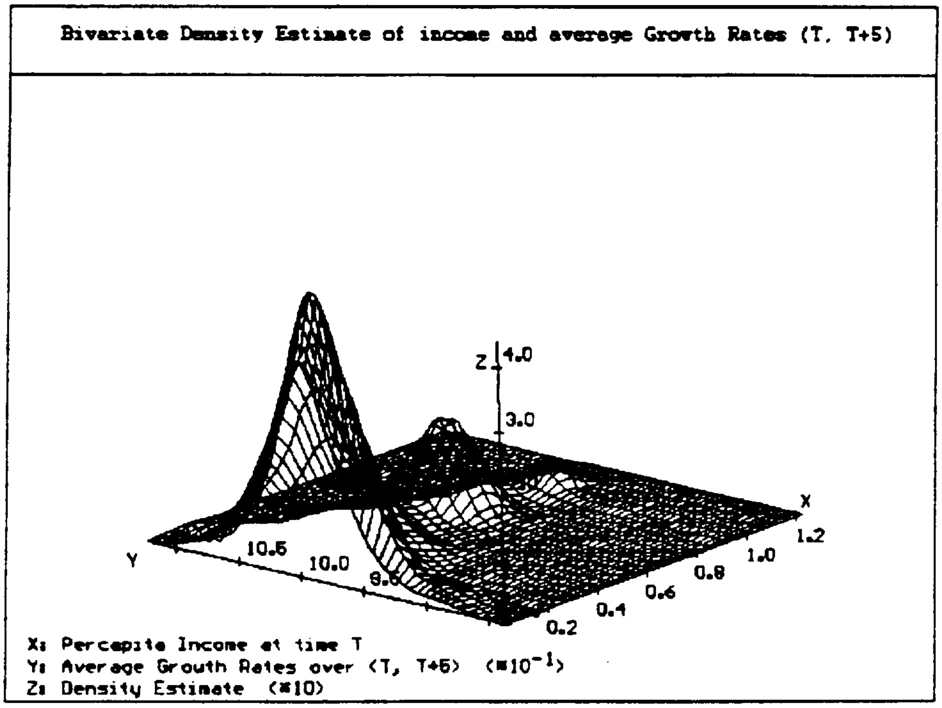
TABLE 6.4.

10-Years Transition Probabilities.

Per capita Income	$P[x_i< 1]$	$P[x_i \geq 1]$
$P[y_i < 1]$	0.405 (S_4)	0.007 (S_3)
$P[y_i = 1]$	0.045 (S_4)	0.001 (S_4)
$P[y_i > 1]$	0.540 (S_3)	0.002 (S_4)

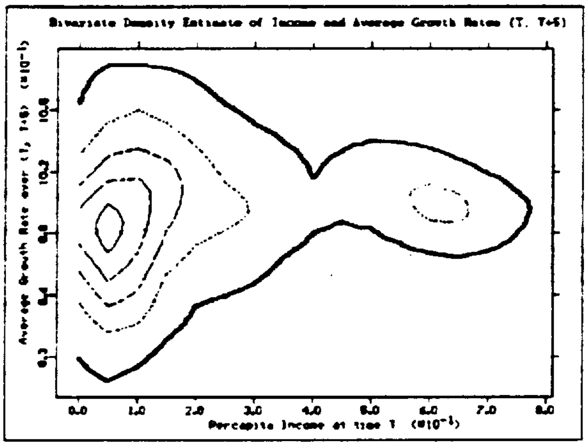
Per Worker Income	$P[x_i< 1]$	$P[x_i \geq 1]$
$P[y_i < 1]$	0.258 (S_4)	0.006 (S_3)
$P[y_i = 1]$	0.034 (S_4)	0.001 (S_4)
$P[y_i > 1]$	0.693 (S_3)	0.008 (S_4)

Figure 6.1.



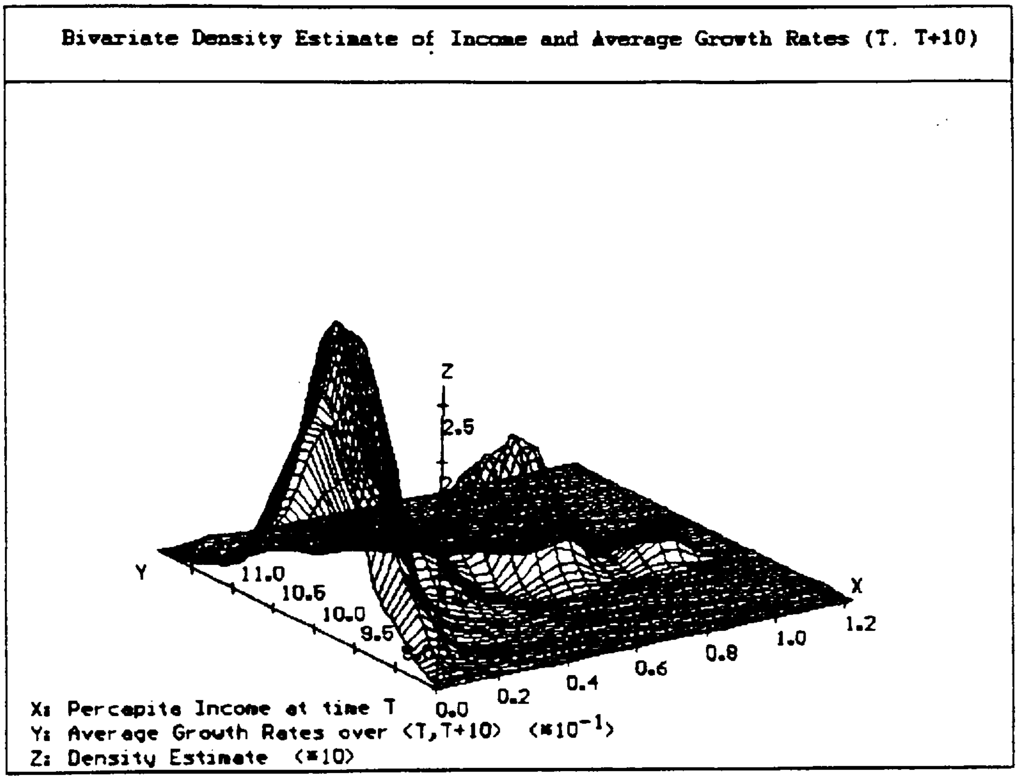
5-Years Transition Probabilities.
Three-dimensional perspective plot of bivariate kernel density estimate of average growth rates (T, T+5) and per capita income (T), with bandwidths $h_X = 0.1$ and $h_Y = 0.025$.

Figure 6.2.



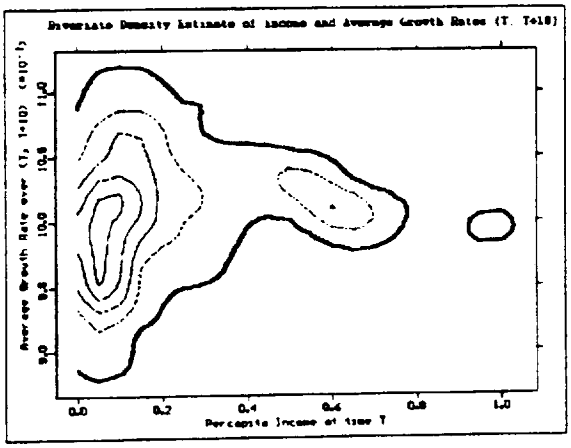
5-Years Transition Probabilities.
Equal probability contours Plot (0.1, 0.3, 0.5, 0.7, 0.9) of bivariate kernel density estimate of average growth rates (T, T+5) and per capita income (T), with bandwidths $h_X = 0.1$ and $h_Y = 0.025$.

Figure 6.3.



10-Years Transition Probabilities.
Three-dimensional perspective plot of bivariate kernel density estimate of average growth rates (T, T+10) and per capita income (T), with bandwidths h_X= *0.1* and h_Y=*0.025*.

Figure 6.4.



10-Years Transition Probabilities.
Equal probability contours plot (0.1, 0.3, 0.5, 0.7, 0.9) of bivariate kernel density estimate of average growth rates (T, T+10) and per capita income (T), with bandwidths h_X= *0.1* and h_Y=*0.025*.

skewed according to the per capita income class. Therefore, two new characteristics about the opportunity to catch-up arise from these joint density estimates.

First, it is clear now that it is not only the privilege of countries which have already reached a "threshold level of development", at least when this is proxied by the per capita level of productivity. Secondly, the possibility of falling behind is not confined to the lowest income class. Still, note from Figures 6.2 and 6.4 that the extent to which countries will either catch up or fall behind, decreases as they belong to higher income classes. Note as well, that the extent of mobility becomes rather explicit when looking at the scales on the Y-direction. The extent of mobility, i.e. the rate at which convergence and divergence proceed, clearly appears to be greater as n, the length of the period, increases. Still, note that these latter are not large enough in order for the poorest countries, to escape from the "unequalizing spiral" illustrated in the previous scenarios. Again, this explains how come the amount of immobility appears to be so great when one wants to assess for the long-run regularities present in the cross-sectional dynamics of incomes. Finally, note that equal probability contour plots for the bivariate kernel density estimate of average growth rates and per worker income are available in Appendix 6.

This is even more obvious when looking at Figure 6.5, where average growth rates relative to the US, over both five and ten years horizons, are displayed. Note that some measures of location and dispersions are also available in Table 6.5. Again, there is a strong evidence for the growth rates to concentrate around ≅1 with a very small dispersion around the mean, though twice larger for the 10-years horizon compared to the 5-years horizon. Still, note that average growth rates over a ten-year horizon and for the per worker income measure, are distributed around ≅3% of growth when expressed relative to the US. This justifies the stronger evidence for convergence observed in the third scenario which

Figure 6.5

Kernel density estimates by WARPing of average growth rates over 5 (left box) and 10 (right box) years horizons (solid line). respectively with bandwidths h= *0.012* and *0.015*, and 90% condidence limits (dashed lines).

(Per capita income expressed relative to 1985 International Prices)

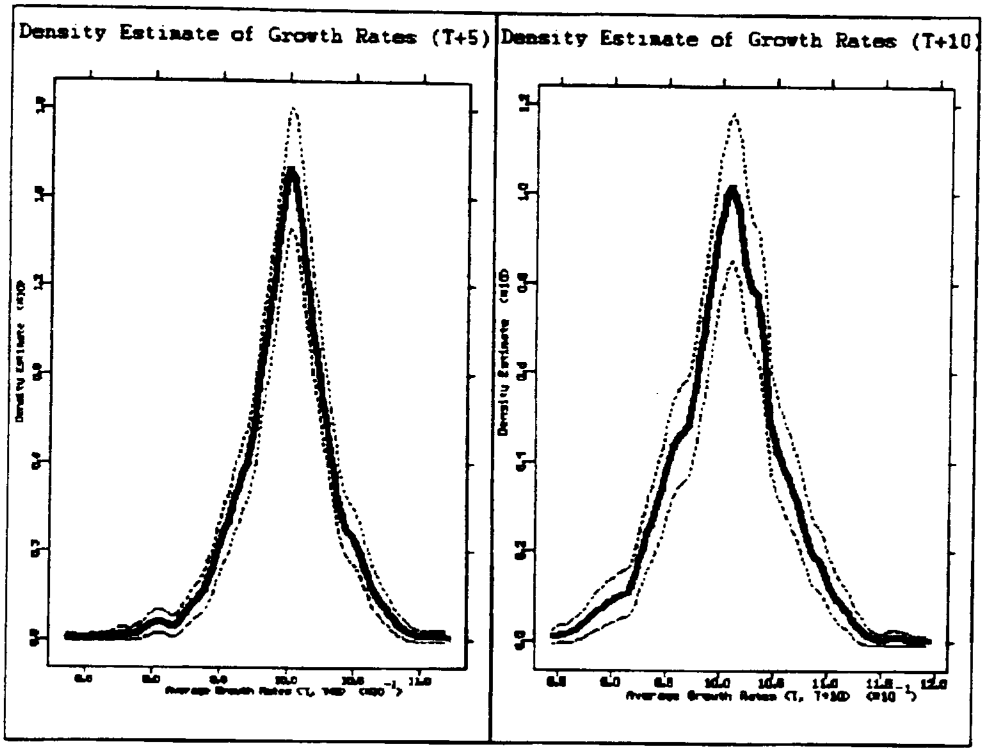


TABLE 6.5

	Average Growth Rates (T, T+5) (Per capita Income)	Average Growth Rates (T, T+10) (Per capita Income)
Mode	1.0018	1.0095
Mean	1.0021	1.0092
Median	1.0030	1.0011
Variance	0.0011	0.0024
Skewness	-0.0260	0.1674

	Average Growth Rates (T, T+5) (Per worker Income)	Average Growth Rates (T, T+10) (Per worker Income)
Mode	1.0086	1.0288
Mean	1.0107	1.0279
Median	1.0116	1.0286
Variance	0.0012	0.0024
Skewness	-0.0278	-0.0122

would have exhibited a slighter evidence in favour of convergence if observations had been normalized, for instance, relative to the world average.

On the one hand, the concentration of mass around ≅1 strongly prevents convergence and shows how much traditional cross-country growth regressions, when assuming a specific parametric form for the joint density of growth rates and initial levels of productivity, tend to over-summarize the amount of information present in the data, thus missing some important features in the underlying structures of incomes and growth rates. On the other hand, note that only the poorest countries belong to both the right and left tails of our density estimates. Countries belonging to the right tails confirm the convergence hypothesis, while countries in the left tails confirm the Endogenous growth theory which allows for permanent differences in growth rates, and in a way that under some initial conditions, there is no possibility to realize the potential to catch-up, leading some countries into a development trap and thus corroborating the argument in favour of multiple equilibria. This also illustrates that economies are integrated in interdependent processes, although they are independent political entities. Government policies affect the equilibrium of other countries through mechanisms which have to be determined within a global economy. Indeed, a global perspective is needed in a world facing a structural crisis in international trade, debt repayments and unemployment. As Kaldor (1985) argues, a global and balanced growth might be preferred to nationalist interests that lead to outcome inequalities.

A limit to these results is that they are either univariate or bivariate estimates of growth rates versus levels of productivity. However, what is expected from the Neo-classical framework is conditional convergence, i.e. differences in the permanent growth component are controlled by conditioning variables, like investment, government expenditures, school-enrolment... This point is analyzed in the following section. Note that as the empirical distribution of the observations itself will not necessarily tell us a lot on the relationship between our variables of interest, we will introduce kernel non-parametric regression curve estimation, structure and behaviour of which, e.g. monotonicity or unimodality, is of particular interest when one wants to infer on the influence of any of these variables on average growth rates, and to compare with the previous empirical literature which used parametric forms in order to identify growth determinants.

7/ The Association between some Income Components and Growth at both an Aggregated and Disaggregated Level.

Recall that tests of per capita income convergence in the parametric empirical literature consist in regressing the average growth rates over a finite period and across countries on the initial per capita income conditioning on hypothesized explanatory variables such as investment, school-enrolment, government spending, etc. The underlying idea of conditional convergence is that each country is expected to converge towards its own steady-state value which will be identical across a set of homogenous economies. These differences in the permanent growth component are controlled by these conditioning variables.

On the other hand, these empirical studies also have the goal of examining empirical linkages between long-run growth rates and hypothesized economic, political and institutional variables. The huge amount of literature already existing on this issue led Levine and Renelt (1992; pp. 942) to argue: "given that a range of 50 variables have been found to be significantly correlated with growth in at least one regression, readers may be uncertain as to the confidence they should place in any one study." Among the regressors are variables relating to trade and trade policy; political and social stability and rights; human capital; and macroeconomic policy and outcomes. Therefore, they proposed examining whether the conclusions from these existing studies are robust or fragile to small changes in the conditioning set of information. Their regression model is such that apart the explanatory variable of interest, they always include a set of variables which commonly enter most of the past empirical studies -i.e. investment share, initial level of real GDP in 1960, initial secondary school-enrolment rate, average annual rate of population growth- and introduce all possible combinations of the remaining controlling variables but such that the set of explanatory variables never goes up to eight. They then identify the highest and lowest values of the coefficient of the variable of interest for all the hypothesized models and check whether they cannot be rejected at the 5% level. A partial correlation is said to be "robust" if these coefficients remain significant and of the same sign as expected from the theoretical framework. On the other hand, if one of these two extreme coefficients is not significant or changes sign, the reader might feel less confident of the partial association between growth and the variable of interest. They refer to this statistical inference as "fragile." Basically, almost all results appeared to be fragile except a positive, robust correlation between growth and the share of investment and between the investment share and the ratio of international trade to GDP. Therefore, the conclusions provided in almost all traditional empirical cross-country growth regressions are very sensitive to slight alterations in the right-hand-side variables tempering therefore the confidence one should place in the conclusions of existing studies.

Traditional parametric methods nest on strong assumptions, so that, as illustrated by Levine and Renelt (1992), results may not be "robust" to specifications of functional form, and/or serial correlation assumptions, measurement error processes, distributional assumptions... Indeed, misspecification of the hypothesized functional form may have serious consequences for both the econometric result as well as for recommanding policy changes that are usually expected to increase the chances of the lagging countries to perform better. Therefore, the approximation of the conditional mean of average growth rates given any explanatory variable, requires the use of more flexible functional forms avoiding a finite set of parameters, thus doing justice to the amount of information which is available.

Again, nonparametric smoothing techniques are the right benchmark to obtain new insights into the general relationship between growth rates and investment, and/or the government expenditures shares and the question of whether these relations are linear. Non-parametric techniques, as they do specify no functional form at all, are therefore robust for misspecification and allow consistent estimation of a regression model where the only specification which is involved, concerns the choice of the dependent variable and the independent variables. Kernel estimation[1], which is understandable at a intuitive level, is a suitable tool for estimating the average growth rate given a certain percentage of income devoted to investment. We will focus our attention on special features like monotonicity or unimodality.

7.1/ Univariate Kernel Regression Smoothing.

Recall that if (X,Y) is a two-dimensional random variable, then E[Y| X=x] is called the regression curve of Y on x. As already mentioned, for bivariate distributions in general, the mean of Y in the conditional density of Y given X=x will be some function of x, say m(x) and the equation Y=m(x) when plotted in the xy plane gives the regression

---

[1] Most of literature on nonparametric regression function estimation deals with the kernel method and its variants which are motivated from local averaging and because the asymptotic theory involved is well-developed. The reader interested in alternative smoothing techniques may, for instance, refer to Härdle (1990a).

curve for Y. It is a curve which gives the location of the mean of Y for various values of X in the conditional density of y given X=x; that is,

$$E[Y|X = x] = \frac{\int y \cdot f(x,y)dy}{f(x)}$$

Where $f(x,y)$ denotes the joint density of (X, Y) and $f(x)$ the marginal density of X.

Then, if we have n independent observations $\{(X_i, Y_i)\}_{i=1}^{n}$ the regression relationship is commonly modelled as

$$Y_i = m(X_i) + \varepsilon_i$$

where ε is a random variable denoting the variation of Y around m(x).

Note that if the form of the joint density is known, then m(x) may be determined analytically. However, when the former is unknown, m(x) is unknown as well. Usually, the form of m(x) is assumed to be linear where the joint density of Y and X is normal. These assumptions lead to the well-known problem of robustness of the models under study, which, when it arises, invalidates any inference and thus the estimation itself. Instead, non-parametric models are obtained without making any assumption about the functional form of m(x), and thus are more flexible than parametric ones.

The approximation of m(x), commonly called "smoothing", is quite similar to that in density estimation and supposes no prior knowledge of m. Indeed, for the latter, we considered the points in a small neighbourhood around x and weighted the frequencies of observations falling into a window with bandwidth h centred in x. Here, the procedure consists in weighting the response variables near a point x. Härdle (1990a; pp. 16) notes that "a quite natural choice is the mean of the response variables near a point x. This local average should be constructed in such a way that it is defined only from observations in a small neighbourhood around x, since Y-observations from points far away from x will have, in general, very different mean values. This *local averaging procedure* can be viewed as the basic idea of smoothing."

In order to weight the observations Y_i depending on the distance of X_i to x, the following nonparametric regression smoother is used:

$$\hat{m}_h(x) = n^{-1}\sum_{i=1}^{n} W_{h,i}(x)Y_i$$

where $\{W_{h,i}(x)\}_{i=1}^{n}$ denotes a sequence of weights which depends on the smoothing parameter h and the sample $\{X_i\}_{i=1}^{n}$ of the explanatory variable.

Recall now that a kernel is a continuous, bounded and symmetric real function K which integrates to one, the weight sequence for kernel smoothers is defined by

$$\{W_{h,i}(x)\}_{i=1}^{n} = \frac{K_h(x - X_i)}{\hat{f}_h(x)}$$

where $\hat{f}_h(x) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i)$, is the kernel density estimate,

and where $K_h(u) = h^{-1}K\left(\frac{u}{h}\right)$ is the rescaled kernel function with bandwidth h.

This leads to the popular Nadaraya-Watson estimate[2]:

$$\hat{m}_h(x) = n^{-1}\frac{\sum_{i=1}^{n} K_h(x - X_i)Y_i}{n^{-1}\sum_{i=1}^{n} K_h(x - X_i)}$$

---

[2] This form of the estimate was proposed by Nadaraya (1964) and Watson (1964).

In particular, if the kernel is chosen to be a unimodal density function with zero mode, for instance, let the kernel be the density of the standard normal distribution, then the closer x to X_i, the more weight is put on Y_i. In the following empirical examination, the kernel regression estimate will be computed via the WARPing technique and using a Quartic kernel as for the density estimates. The approximation of the Nadaraya-Watson estimate by WARPing is described in Härdle (1990b; Chapter 5, pp. 137-138) and its realisation within XploRe is available in Appendix 7. First, the real line is discretized into bins and the absolute frequency of data points (X_i) as well as the corresponding sum of Y_i's falling into each non empty bin, are computed. Secondly, weights are created to appropriately weight, in a third step, the computed sums (nominator) and frequencies (denominator). Finally, it remains to combine both nominator and denominator.

As for density estimation, the choice of the bandwidth is crucial to control for the consistency of the estimates which will lead to an unbiased and stable "smoothing." The problem here is to determine how close the estimated curve will be to the true underlying curve? Indeed, as the smoothers will average over observations with different mean values, depending on the chosen "window-width," which regulates the size of the neighbourhood around x, the bias will increase with h, while interesting features hold in the data might disappear with over large bandwidths. Indeed, note that $\hat{m}_h(X_i) \rightarrow Y_i$ as $h \rightarrow 0$,while at any arbitrary point x, $\hat{m}_h(x) = n^{-1}\sum_{i=1}^{n} Y_i = \overline{Y}$ as $h \rightarrow \infty$, so that the bandwidth determines the degree of smoothness.

As for density estimation, $\hat{m}_h(x)$ will be a consistent estimate of the regression curve m(x), if $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$. Again, we face the trade-off problem concerning the choice of the bandwidth h. As our sample is finite, a small (large) h will penalize the stability (bias) of the regression estimate.

We propose here to use the cross-validation method in order to control for the performance of the regression smoothers. A convenient measure of accuracy of $\hat{m}_h(x)$ is the Average Squared Error. Let us define the cross-validation function as the average of the sum of the squared prediction errors,

$$CV(h) = n^{-1}\sum_{i=1}^{n}((Y_i - \hat{m}_{h,i}(X_i))^2 w(X_i)$$

where $\hat{m}_{h,i}(X_i) = (n-1)^{-1}\dfrac{\sum_{i\neq j} K_h(x - X_j)Y_j}{(n-1)^{-1}\sum_{i\neq j} K_h(x - X_j)}$, the leave out estimate computed the same way as the Nadaraya-Watson estimate, e.g. via the WARPing technique which allows fast computation of the cross-validation bandwidth (See macro *regcvl* from XploRe in Appendix 8), but without the i-th observation. The nonnegative weight function $w(\cdot)$ is a convenient tool with which to drop observations at the boundary of X. Then, the bandwidth ($h = Arg \min CV(h)$) which will minimize the prediction error, will be asymptotically optimal leading to a consistent estimate of m(x). Finally, as for density estimates, asymptotic confidence bands for m(x) will be computed according to the Bickel and Rosenblatt (1973) theorem. We refer the reader to Härdle (1990a, Chapter 4, pp.116-117) for the construction of these approximate confidence bands.

7.2/ Investment and Growth.

Investment share of GDP is consistent with both the Neo-classical and a wide variety of "new" growth models. Levine and Renelt (1991) found that of the 41 empirical studies they surveyed, 33 included this variable in their modelling. As already noted, the partial association between growth and the investment share has been found "robust" in terms of the Levine and Renelt (1992) null hypothesis. Furthermore, this finding is consistent with a wide range of previous studies.

We will not review all the existing literature here, but will restrict our attention to two of the empirical studies which have already been presented in Section 2; that is, the Romer & Al. (1990) pro-convergence study and the local versus global convergence study by Durlauf and Johnson (1992).

The estimation of the textbook Solow model by Romer, Mankiw and Weil (1990), leads the investment share's coefficient to be of the predicted sign and highly significant for two of the three samples -i.e. for non-oil countries excluding countries whose data received a grade of "D" from Summers and Heston or whose populations in 1960 were less than one million (75), and OECD countries (22), but not for non-oil producers (98 countries). Furthermore, the overall fit of the Solow model appears to be quite satisfactory as the adjusted R^2 is 0.59 for the intermediate sample. Note, however, that their human capital augmented version of the Solow model leads to a drop in the size of the physical capital investment coefficients such "that the elasticity of income with respect to the stock of physical capital is not substantially different from capital's share in income (Romer & Al.; 1990, pp. 27)." This led the authors to conclude that capital receives approximately its social return, suggesting thus the absence of substantial externalities. Furthermore, despite the absence of these latter, they find that the accumulation of physical capital has a greater impact on income per capita than the Solow model implies.

Durlauf and Johnson (1992), having split the Romer & Al. 98 countries sample into high initial income/high initial literacy and low initial income/low initial literacy, found an investment share coefficient which is over twice as large for the former compared to the latter, both of them being highly significant. Hence, as illustrated by these two empirical studies, there is a consensus among researchers that growth is significantly and positively related to the share of income devoted to investment, but the size of the coefficient -i.e. the social returns to capital accumulation- associated with this partial association is highly dependent on the sample selection criteria and on the set of controlling variables included in the model, so that we want to provide some informations, which will necessarily be robust to misspecification, about the dynamics of the relation between output and investment.

Among other variables, the share of Gross Domestic Product devoted to private and public domestic investment is available in the Penn World Table (Mark 5) for the list of countries available in Appendix 1 and for the period from 1960 to 1985. It is either expressed relative to 1985 priced output or in current priced output. In the present section, we will use the former -i.e. 1985 international Prices. Indeed, during the period under study, inflation increased and has been much greater in Sub-Saharian Africa and particularly in Latin America than in Asia (Fischer; 1991). As the latter finds a predominantly negative relationship between inflation and growth, we prefer here, because of the use of pooled time-series cross-section regressions, to take advantage of the concept of constant prices because inflation effects keep current-prices to be comparable over time.

We will present in this section four kinds of results drawn on the evidence from major studies and examine the partial association between growth and investment. First, we model the regression relationship between per capita income playing the role of the explanatory variable (X_i) and the share of income devoted to private and public domestic investment. Secondly, we will determine the average growth rates relative to the world average for given shares of gross product devoted to investment giving particular attention either to monotonicity or unimodality. Thirdly, both the equal probability contour of the bivariate kernel density estimate of investment share and average growth rates relative to the world average as well as the joint density itself will be computed. Finally, as Quah (1993 b) found a significant amount of instability in countries' long-run growth patterns during the period under study -i.e. 1960-85[3]- the data from these four modellings are such that the entire period has been divided into five subperiods of five years each. An arithmetical mean over the subperiods has been calculated for the explanatory variable and growth rates are geometrical means. The cross-section data for each sub period have been pooled so that both differences in income and changes over the entire period for each country are taken into account[4]. Indeed, simple cross-sectional regressions ignore information that might be available in the time-series of data within each country.

Kernel estimators are well-known for performing badly in the tails. Therefore, in order for the mean response of the explained variable versus the explanatory variable not to be dramatically influenced by very few observations which may be located far from the main body of the data, we will apply a logarithmic transformation to the data of the explanatory variable so that the smoother as well as the optimal bandwidth issued by the cross-validation will be computed from this transformation. Finally, note that the regression smoothings graphed herein are smooth estimates having performed the appropriate inverse exponential rescaling.

Nonparametric regression smoothing is, because of its flexibility, a suitable tool for the detection of structures deviating from a postulated parametric model. For example, Levine and Renelt (1992) in their "sensitivity" analysis of the investment share versus the initial percapita income, show a "fragile" relationship between these two variables. Indeed, the coefficient was close to zero, varying from 0.008 to -0.002 according to the different

---

[3] See as well Easterly & Al. (1993) who find that growth rates are highly unstable over time. Indeed, they find a correlation across decades of countries growth rates of income per capita, which runs from 0.1 to 0.3, while on the other hand, country characteristics appear to be persistent.

[4] Fischer (1991) dividing the world into regions and the 1960-88 period into three subperiods, finds that even if each region experienced a sharp increase in the rate of investment between the first two periods, this did not lead to any increase in the growth rate. Still, note with Levine and Renelt (1991) that five year averages might not necessarily eliminate all cyclical components because business cycle frequencies may be different across countries and across variables within the same country.

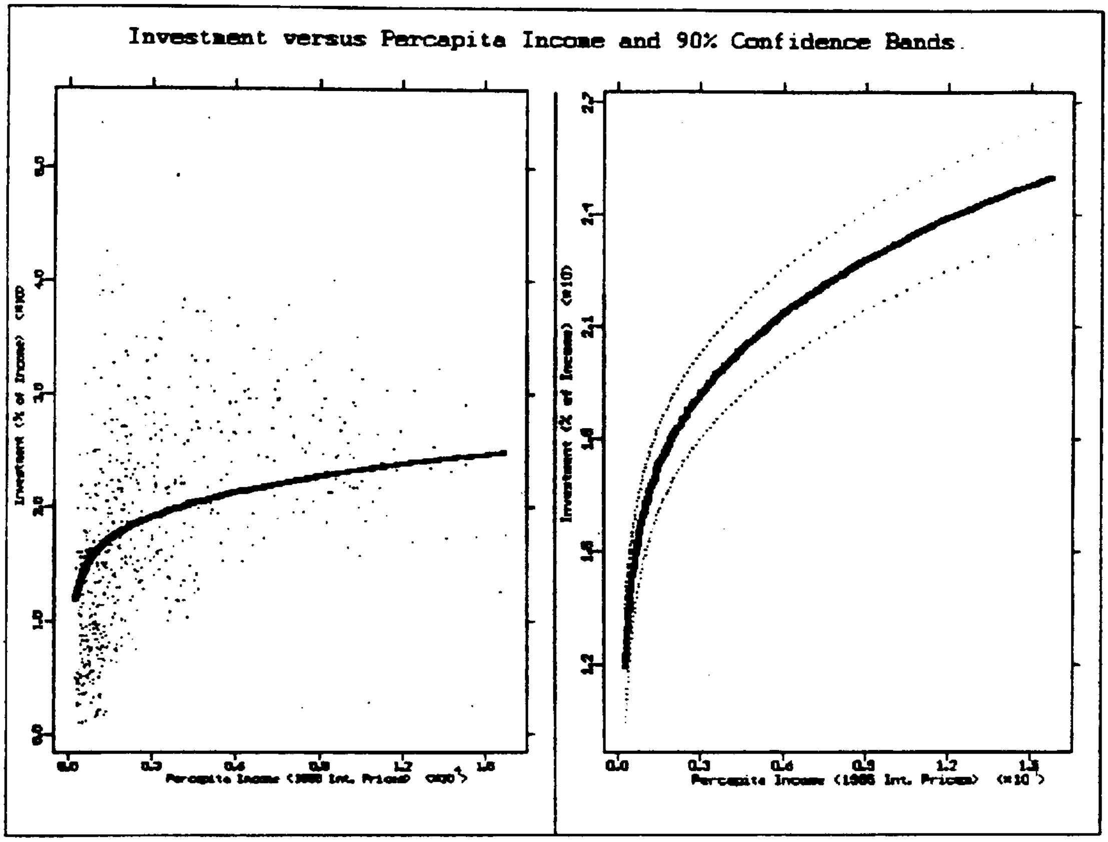
sets of control variables included in the modelling. Instead, and as shown in Figure 7.2.1.a, the relationship appears to be logarithmic so that the slope is rather high for the lowest income classes. Note from Figure 7.2.1.b, that the bandwidth which led to this regression curve minimizes the Averaged Squared Error and that the true regression relationship lies, with a 90% probability, within the confidence bands plotted in the right-hand box of Figure 7.2.1.a.

As shown in Figure 7.2.1.c[5], the mean response of growth rates relative to the world average given the share of income devoted to private and public domestic investment varies within a very small range, i.e. [-1.5%,+0.5%]. On the other hand, the curve is steeper when starting at a low investment share, thus, confirming Durlauf and Johnson's (1992) findings. Indeed, increasing the share of income devoted to investment from +ε% to 15% leads the average growth rate to rise from -1.5% of growth to the world average growth rate itself. Devoting more income to investment does not lead to a very "sensitive" response of the mean growth rate and the partial association between average growth rates relative to the world average and investment share is approximately linear and flattens out among the highest shares of income devoted to investment. This logarithmic structure of the mean response of growth rates relative to the world average versus the share of income devoted to private and public domestic investment confirms the "robust" positive partial association usually found by previous parametric studies. However, if we do not claim to be challenging the existing consensus on this empirical linkage, we are struck by the rather small extent to which the mean growth rate responds to increases in the share of income devoted to investment. Finally, note that the above smoother reflects only intercountry averages and can not be applied to any single country.

---

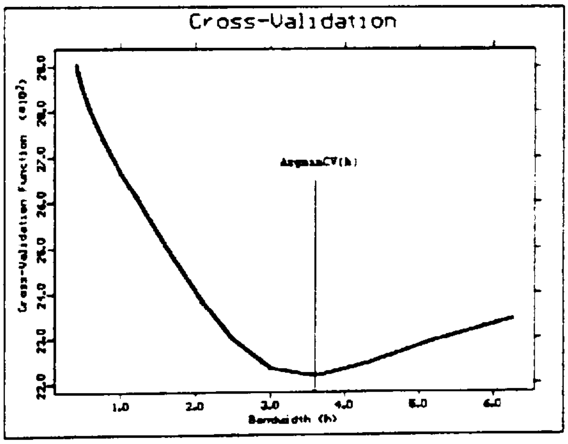
[5] Note that the observations which escape from the main body of the data on the X-axis and which would have led to an inverted U-shape of the smoother without the logarithmic transformation refer to Gabon which devoted ≅70% of its income to investment during the first three years of the 1975-80 sub-period and to Irak over 1980-85, which started to fight against Iran in 1980 until 1988.

Figure 7.2.1.a



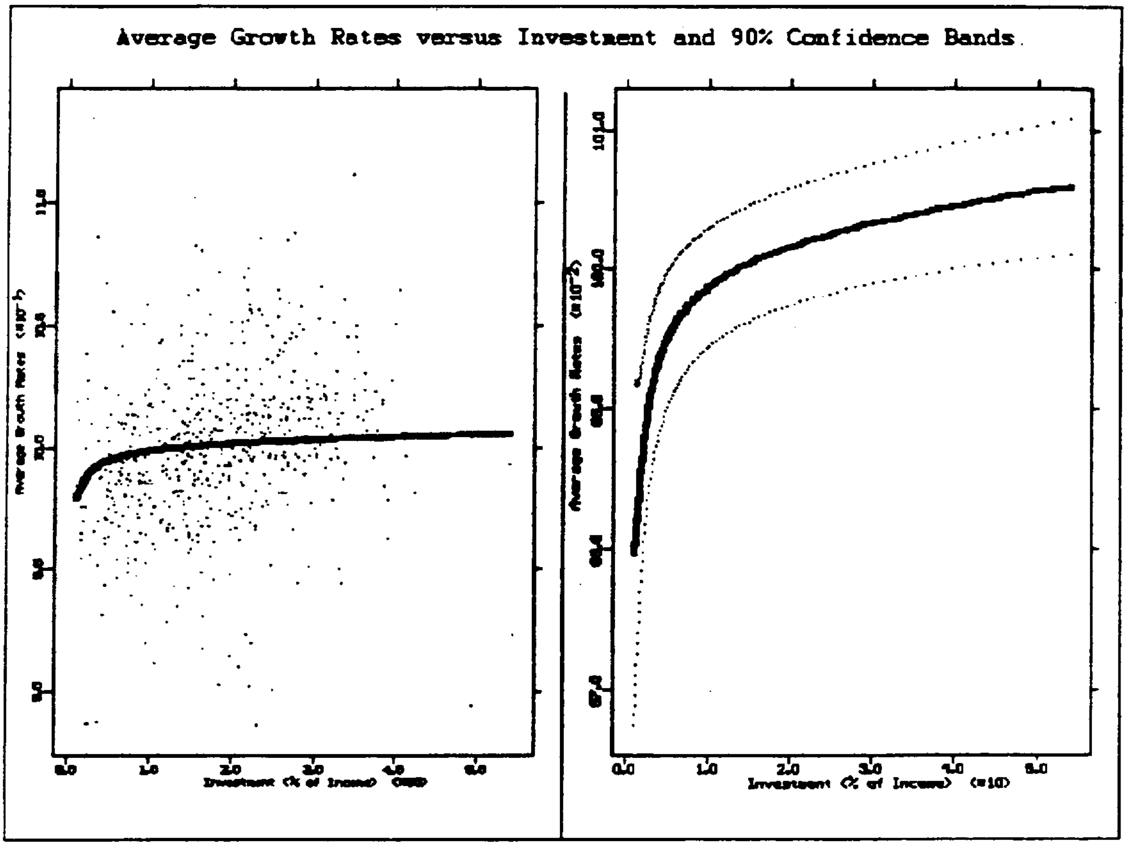
Plot of **Private and Public Domestic Investment** share (% of Income) versus per capita income observations (points), and true regression curve m(X) (thick line), with bandwidth h= *3.6* and 90% confidence bands (dashed lines).

Figure 7.2.1.b



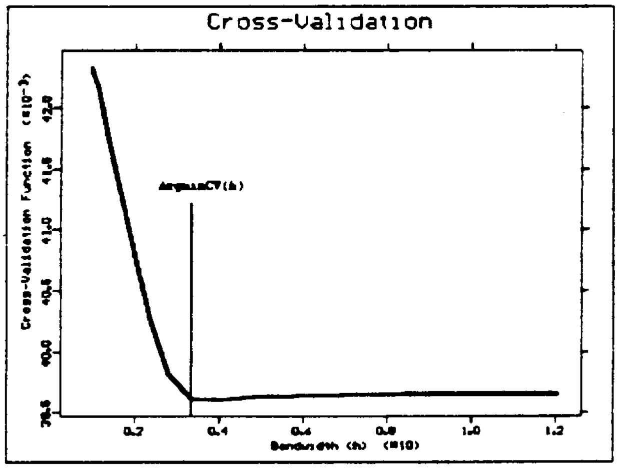
Cross-Validation function using a Quartic kernel for the original Nadaraya-Watson estimate of **Private and Public Domestic Investment** share versus per capita income. ArgminCV(h)= *3.6* (issued from the logarithmic transformation of the X variable).

Figure 7.2.1.c



Plot of average growth rates versus **Private and Public Domestic Investment** share (points)
and true regression curve m(X) (thick line),
with bandwidth h= *3.2* and 90% confidence bands (dashed lines).

Figure 7.2.1.d



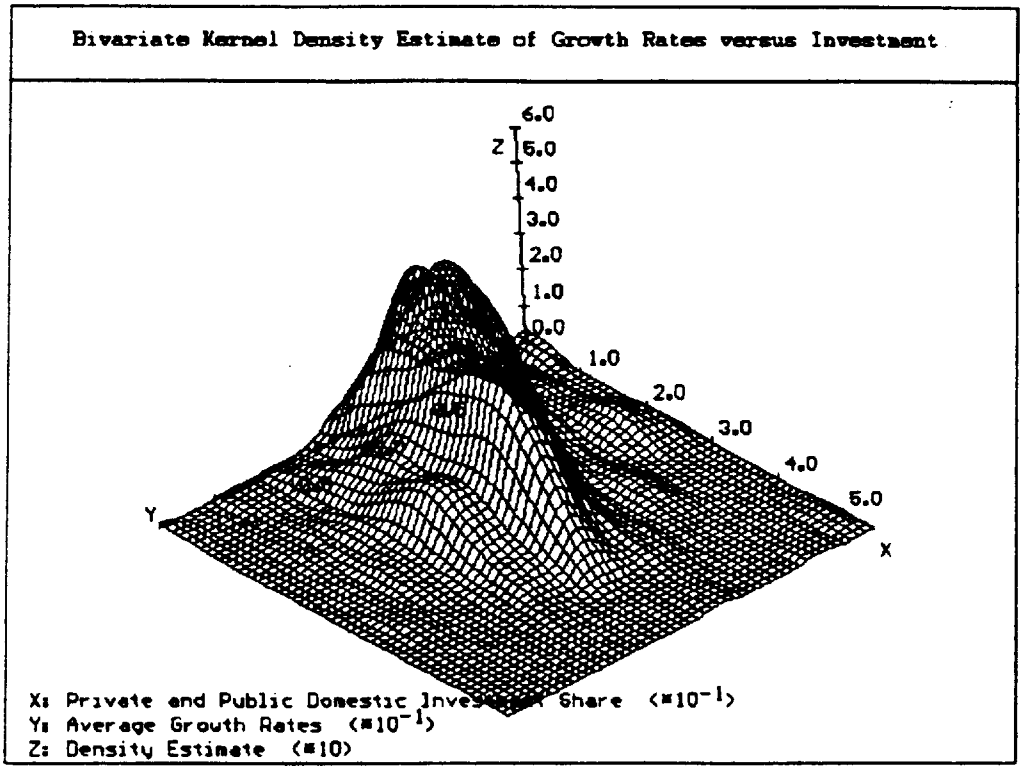
Cross-Validation using a Quartic kernel for the original Nadaraya-Watson estimate
of the average growth rates versus **Private and Public Domestic Investment** share,
ArgminCV(h) = *3.2* (issued from the logarithmic transformation of the X variable)

Therefore, in order to provide additional information about the distribution of growth rates relative to the world average and the share of income devoted to private and public domestic investment, we propose to graph three-dimensional perspective plots of the bivariate kernel density estimate of these two variables as well as equal probability contours of this latter (See Figures 7.2.1.e-f). These non-parametric techniques are used to enhance our understanding of patterns in the data without using regressions. Such graphs can be thought of as smoothed histograms, but this time in three dimensions. The joint estimate is such that the grids for X and Y are equal, respectively, to 1% in the X-direction and 0.5% in the Y-direction and such that bandwidths in each component are asymptotically optimal. Therefore, observations are counted and weighted, not in an interval band but in a two-dimensional elliptical band, so that point estimates are available for each half-percentage of growth and each percentage of income devoted to private and public domestic investment. The height of the joint density estimate represents the fraction of countries at the levels of investment share and growth rates relative to the world average represented by the co-ordinates along the base. Thus, a point estimate gives the underlying a posteriori probability to grow at a definite rate when devoting a definite share of income to investment.

Figure 7.2.1.e gives a clear impression of relative heights and thus of the concentration of mass. Note though not directly readable from the graph, that in accordance with the corresponding regression smoothing available in Figure 7.2.1.c, growth rates are "normally" distributed around -2% of growth relative to the world average from 1% to 7% of income devoted to private and public domestic investment, around -1% from 8% to 12%, around 0 from 13% to 21% and around +1% from 22% to50%.

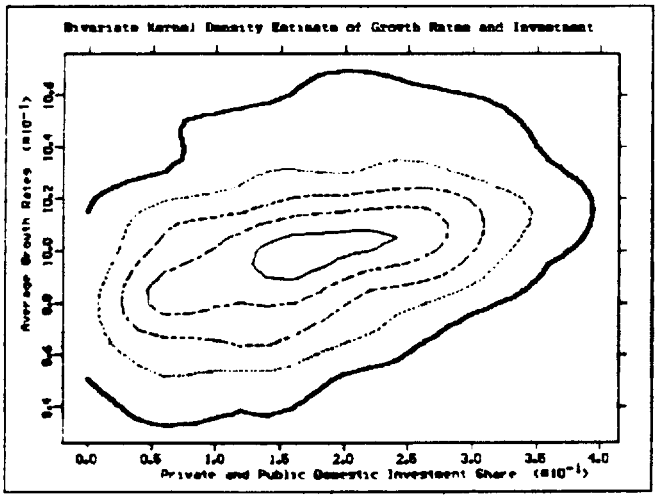
Note that the joint density surfaces associated with the bivariate density contours fill out the skeletal information in the regression function in Figure 7.2.1.c.

Figure 7.2.1.e



Three-dimensional perspective plot of bivariate kernel density estimate of **Private and Public Domestic Investment** share and average growth rate relative to the world average. with bandwidths h_X=*0.075* and h_Y=*0.02*.

Figure 7.2.1.f



Equal probability contours (0.1, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density estimate (f(X, Y)) with bandwidths h_X=*0.075* and h_Y=*0.02*.
X-direction = **Private and Public Domestic Investment** share
Y-direction = average growth rate relative to the world average.

The contour plot, whose points linked by a contour have the same density, shows the dramatic diversity of growth experiences whatever the share of income devoted to investment is; for example, the mean response of growth rates relative to the world average and devoting ≅15% of its income to investment, is about 0, while the range of growth rates experienced for this percentage is enormous [-6%, +6%], showing thus the wide heterogeneity in growth patterns for a given private and public domestic investment share.

At first sight, these results may appear to be more confusing than helpful. However, as Bradford De Long and Summers (1991; pp. 452) argue: "the fact that equipment's share in total investment varies so widely and the centrality in historical discussions of growth suggest the importance of disaggregating investment in considering its relation to economic growth. If machinery and structures contribute differently to growth, then analyses of the relationship between capital accumulation and growth are likely to be misleading." This led these authors to focus on equipment investment, i.e. technology embodied in machinery. This appears, thus corroborating traditional thought, to be a prime determinant of rates of productivity growth, having higher social returns than any other forms of investment. Using data issued by the Penn World Table (Phase IV; 1980), they found a "clear, <u>strong</u> and robust" statistical relationship, showing that countries which devoted the highest shares of their income to equipment also experienced the highest growth rates over the period from 1960 to 1985[6]. They argue that previous studies have been carried out at an inappropriate level of aggregation. Indeed, countries with a low share of investment relative to income are allowed to experience faster growth relative to the world average, meaning that arguments developed along the lines discussed in Bradford De Long and Summers' paper remain crucial and determinant in explaining the wide heterogeneity in the structure of growth rates given the share of income devoted to

---

[6] See also Bradford De Long, "Productivity and Machinery Investment: A Long Run Look 1870-1980." NBER Working Paper #3903 (November 1991).

investment. In other words, "the key to growth is not so much the accumulation as the effective use of resources, combined with an outward-oriented economy [which] conforms to market forces" (Bradford De Long and Summers; 1991, pp. 484-485). However, a combination of other factors such as economies of scale, education, and terms of trade doubt less remains crucial as determinants of growth.

As we dispose of disaggregated data corresponding to the basic headings used in Phase IV of the International Comparison Program for 1980, we are able to investigate and to compare, using the above methodology, the relationship reported in the De Long and Summers' study discussed above[7].

First, as the authors found that there was little information in the producer's transportation component of durables, which they suggested is due to "differences in the need for transportation caused by differences in urbanization and population density," (De Long and Summers; 1991, pp. 449), they focused on electrical and non electrical machinery and its association with productivity growth, i.e. using a per worker income measure. Note that, in fact, the mean responses of growth rates relative to the world average given the shares of income devoted to construction are approximately linear over the whole distribution and are like a horizontal line displayed at the world average growth rate level (See Figure 7.2.2.a). On the other hand, the mean responses of growth rates versus both non electrical and electrical machinery show steep curves which do not seem to flatten out among countries which devote constantly higher shares of their income to both these disaggregated components of domestic capital formation (See Figures 7.2.2.c-d).

---

[7] Unfortunately, the statistical division diskette underlying Phase IV of the International Comparison Programme for 1980, provides per capita quantities valued at international prices for a set of 60 countries and only for the year 1980. These countries are listed in Appendix 9 with their respective three letter World Bank codes. Domestic capital formation is composed of producer durables and construction. The former is then disaggregated into machinery, electrical machinery and transportation. The latter, on the other hand, is divided into residential, non-residential and non-building construction. The basic headings are available in Appendix 10.

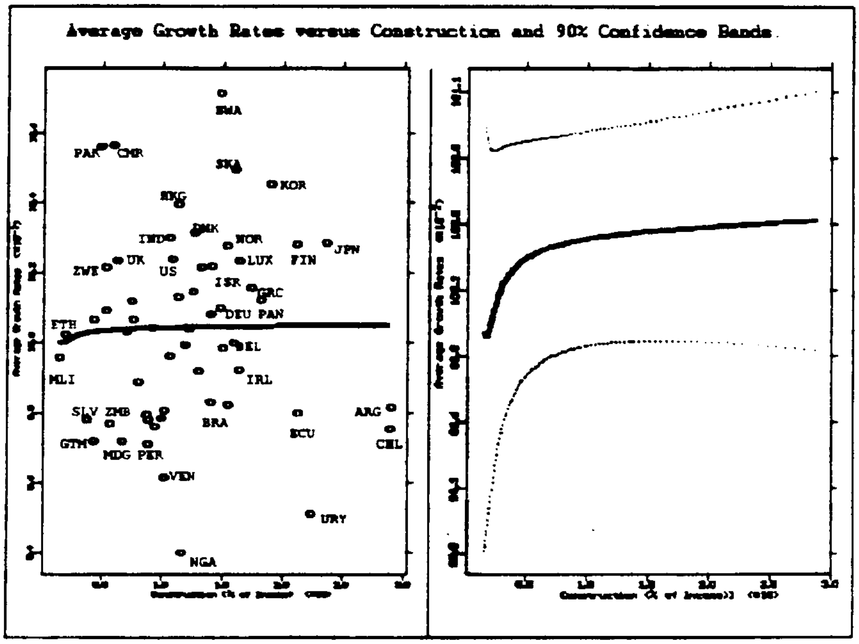
Indeed, looking at the figures in the right-hand boxes, where the associations are graphed using a larger scale, the average response of growth rates is such that a country which is able to increase by 2% the share of its income devoted to non electrical machinery, is expected to experience a rather sharp rise of 1% in its growth rate. A zoom of Figure 7.2.2.d. shows an even stronger association: each extra percent of income devoted to electrical machinery is expected to lead to an extra percent of growth.

Recall, here, that international transfers of knowledge and/or technology are given a central role in the economic issue of convergence as initially poorer countries are expected to be able to make use of current best practise procedures already in use in the more productive economies. Then, identification of countries as displayed in Figures 7.2.2.a-d, illustrates that learning and imitating an existing technology in order for the poorest to catch up with the richest, is far to be realized for a large number of countries. This might suggest first, that for technological transfers to occur, they must be largely embodied in human capital as suggested by Lucas (1988), maybe in a larger extent than in physical capital, and that technical progress in one country may not necessarily spillover to actually affect techniques used in other countries[8]. Adverse conditions might indeed lead a country to never be able to catch up and therefore to specialize in traditional products so that its deficit in the R&D activity will never be overcome, leading to "hysterisis" phenomena within the whole economy as suggested by the Endogenous growth theory and illustrated by most of African countries. Secondly, as Pasinetti (1981) argues: any investigation into technological progress must require hypothesis on the evolution of consumers' preferences as income increases. This issue is discussed in Section 8.

---

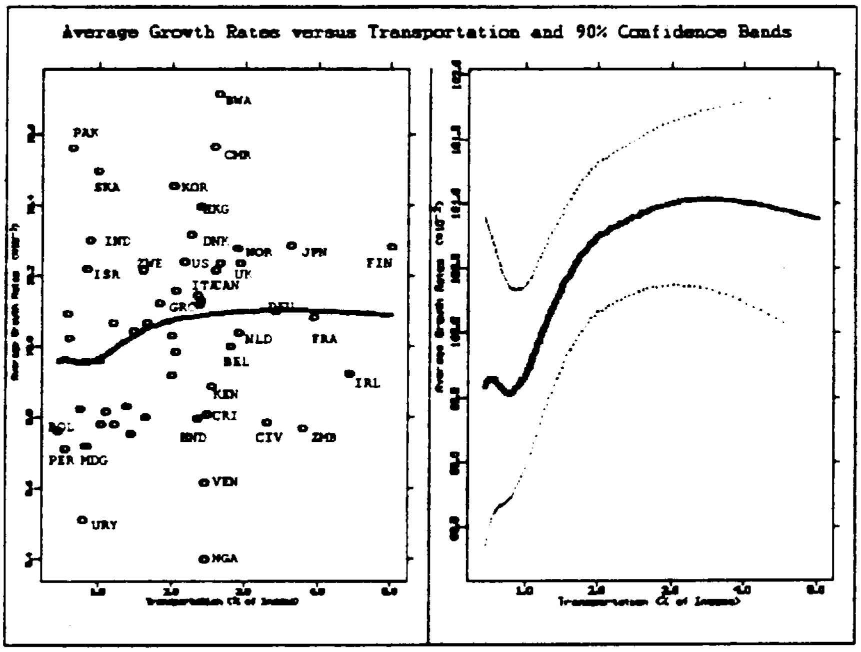
[8] A good illustration consists in the car sector in South Korea which has largely benefited from technical assistance from Japan and the US. Furthermore, the former, which provides the more suitable products to South Korean technological needs, proposes as well a training to the use of their equipments. Secondly, policies in terms of technology adopted by South Korea and Taiwan, do not only consist in learning and imitating existing technologies. Indeed, they focus on improving the population education at both low and high skills levels, in order to be able to improve the imported technology with an ultimate goal which will consist in innovating themselves (See, for instance, Lorot & Schwob; 1987).

Figure 7.2.2.a



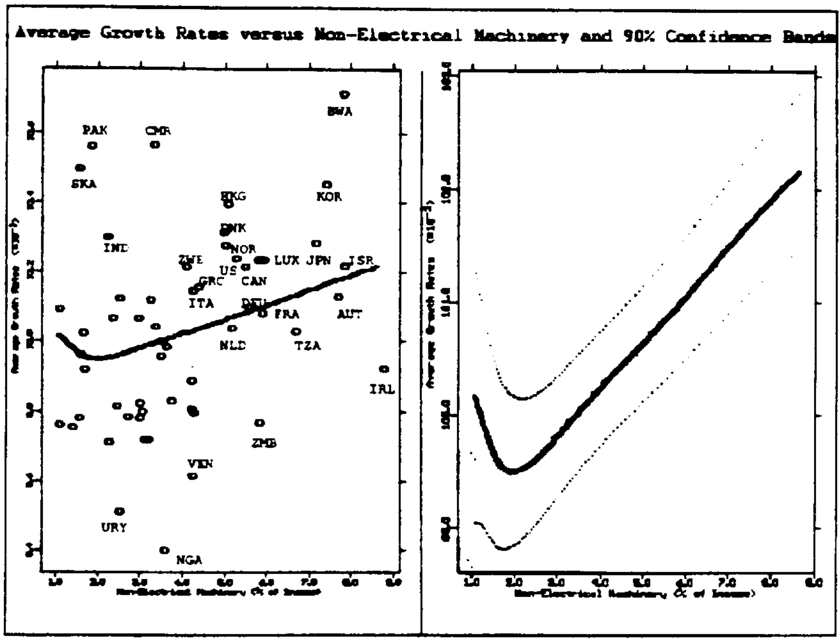
Nadaraya-Watson estimate of the average growth rate (Y) versus **Construction** share (X) (solid line) and 90% confidence limits (dashed lines) with bandwidth h= 2.3.

Figure 7.2.2.b



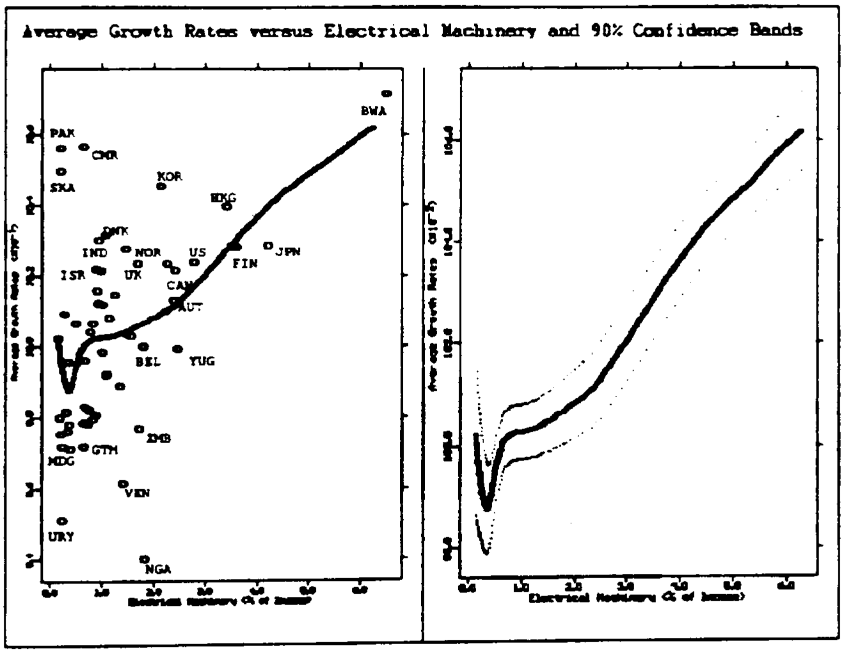
Nadaraya-Watson estimate of the average growth rate (Y) versus **Transportation** share (X) (solid line) and 90% confidence limits (dashed lines) with bandwidth h= 1.

Figure 7.2.2.c



Nadaraya-Watson estimate of the average growth rate (Y)
versus **Non-Electrical Machinery** share (X) (solid line) and 90% confidence limits (dashed lines)
with bandwidth h= 1.2.

Figure 7.2.2.d



Nadaraya-Watson estimate of the average growth rate (Y)
versus **Electrical Machinery** share (X) (solid line) and 90% confidence limits (dashed lines)
with bandwidth h= 1.

However, while this finding corroborates the results obtained by De Long and Summers which did not consider disaggregated components of machinery, we also note that the concentration of mass differ dramatically between both non electrical and electrical components, with the result that the strong association between growth and the latter seems to be dominated by a very small number of observations located in the right tail of the distribution, e.g. Botswana.

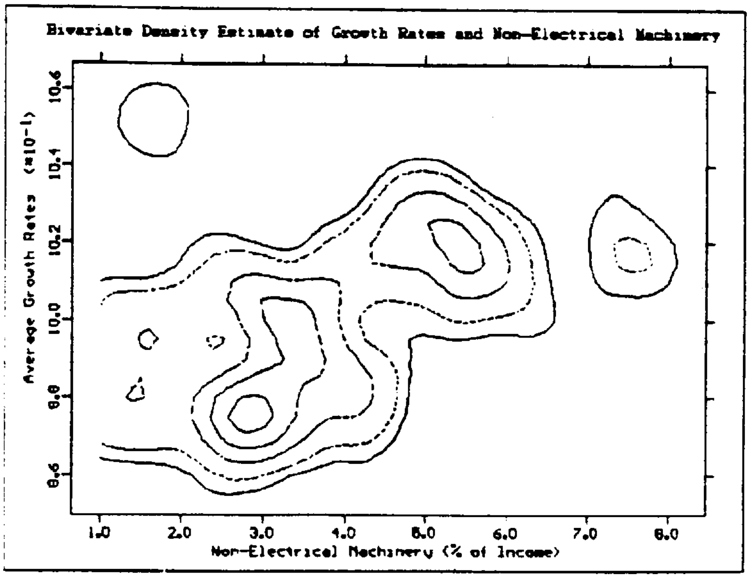
The joint distributions of these four disaggregated components of domestic capital formation are available in Figures 7.2.3.a-d. These give a visual impression of where growth rates relative to the world average tend to concentrate for a given share of income devoted to each of the components of private and public domestic investment. Note first that growth rate patterns are more homogeneous in countries with higher equipment investment, so that the probability of experiencing slower growth relative to the world average is almost zero when devoting more than 5% (2%) of its income to non electrical (electrical) machinery. Secondly, note the surprising bimodal distribution which arises from the bivariate density estimate of growth rates and transportation (See Figure 7.2.3.c), which may indicate that the producers' transportation component of durables carries some information which does not arise when using a traditional linear fit.

Thus, accumulation of physical capital can be associated with economic stagnation and/or recession. The now well-known case of Sub-Saharian African countries which investment has been oriented more than often towards construction and infrastructures indirectly productive[9], corroborates this finding. Differences in growth experiences are more a matter of effective use and allocation of resources compared to the accumulation of physical capital per se.

---

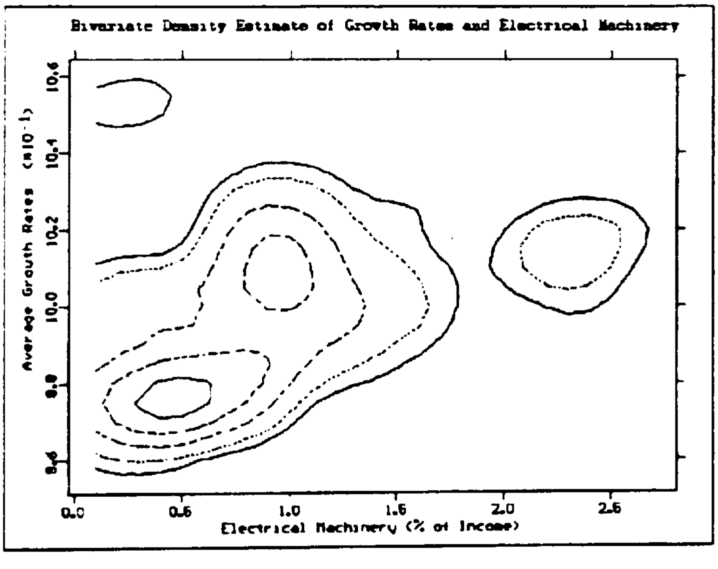
[9] See for instance Hugon (1993).

Figure 7.2.3.a



Equal probability contours[1] (0.2, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density estimate (f(X,Y)) with bandwidths h_X=*1* and h_Y=*0.02*.
X= **Non Electrical Machinery** share,
Y=average growth rates relative to the world average.

Figure 7.2.3.b



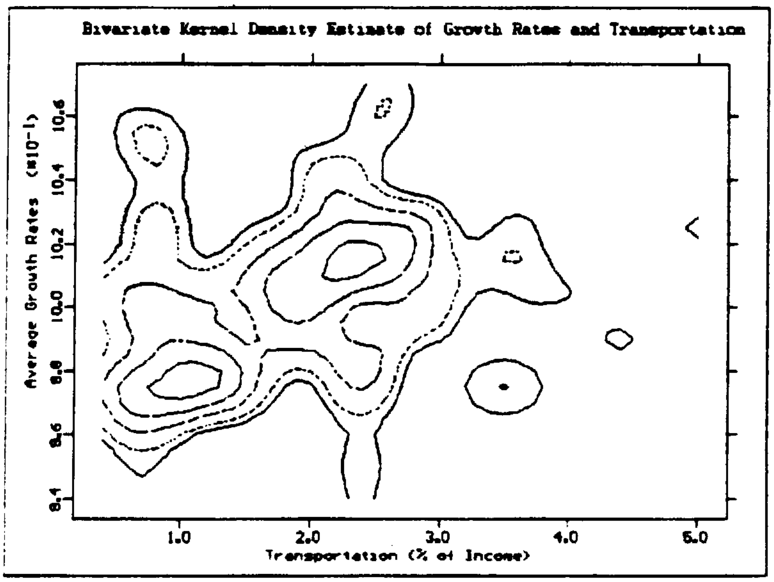
Equal probability contours[2](0.2, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density estimate (f(X,Y)) with bandwidths h_X=*0.5* and h_Y=*0.02*.
X= **Electrical Machinery** share,
Y=average growth rates relative to the world average.

---

[1] The marginal smoothing parameters have been set to 0.1 in the X-direction and to 0.005 in the Y-direction.

[2] The marginal smoothing parameters have been set to 0.1 in the X-direction and to 0.005 in the Y-direction.

Figure 7.2.3.c

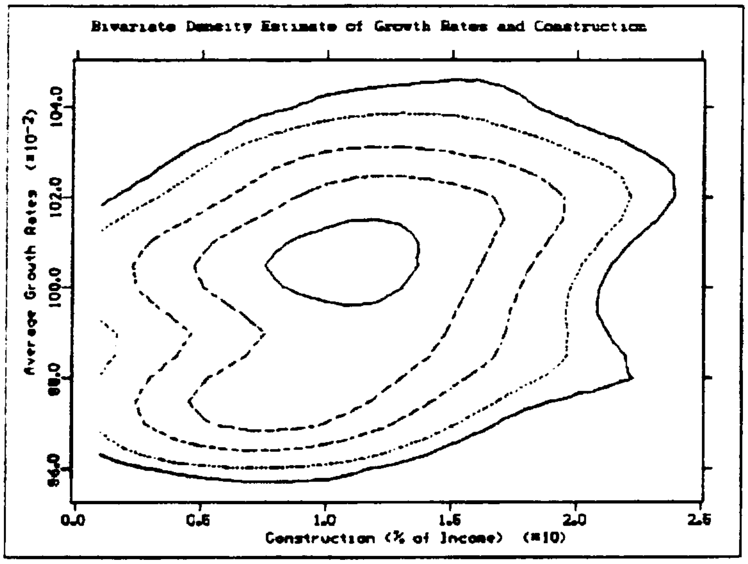


Equal probability contours[1] (0.2, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density estimate (f(X,Y)) with bandwidths h_X=0.5 and h_Y=*0.02*.

X= **Transportation** share.

Y=average growth rates relative to the world average.

Figure 7.2.3.d



Equal probability contours[2](0.1, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density estimate (f(X,Y)) with bandwidths h_X=*10* and h_Y=*0.02*.

X= **Construction** share,

Y=average growth rates relative to the world average.

---

[1] The marginal smoothing parameters have been set to 0.1 in the X-direction and to 0.005 in the Y-direction.

[2] The marginal smoothing parameters have been set to 1 in the X-direction and to 0.005 in the Y-direction.

7.3/ Government Expenditures and Growth.

Among other macroeconomic-policy indicators, both aggregated and disaggregated measures of fiscal policy have frequently been introduced in the conditioning information set since Barro's work (1991). As already discussed in Section 2, Barro finds a negative partial association between the overall size of the government in the economy and long-run growth rates, and these latter are insignificantly related to the share of public investment. Barro (1990) justifies this finding by arguing that government consumption has no direct effect on private productivity, and only leads to lower saving and growth through the distortion effects of taxation or government expenditure programmes. Within this framework, growth increases with taxation and expenditures at low levels and then decreases when the distortionary effects of taxation exceed the beneficial effects of public goods. Therefore, government expenditures and growth are expected to be positively correlated when government expenditures are below the optimal amount, negatively related when they are above, such that the association between growth and government expenditures is expected, within this framework, to exhibit an inverted U-shape.

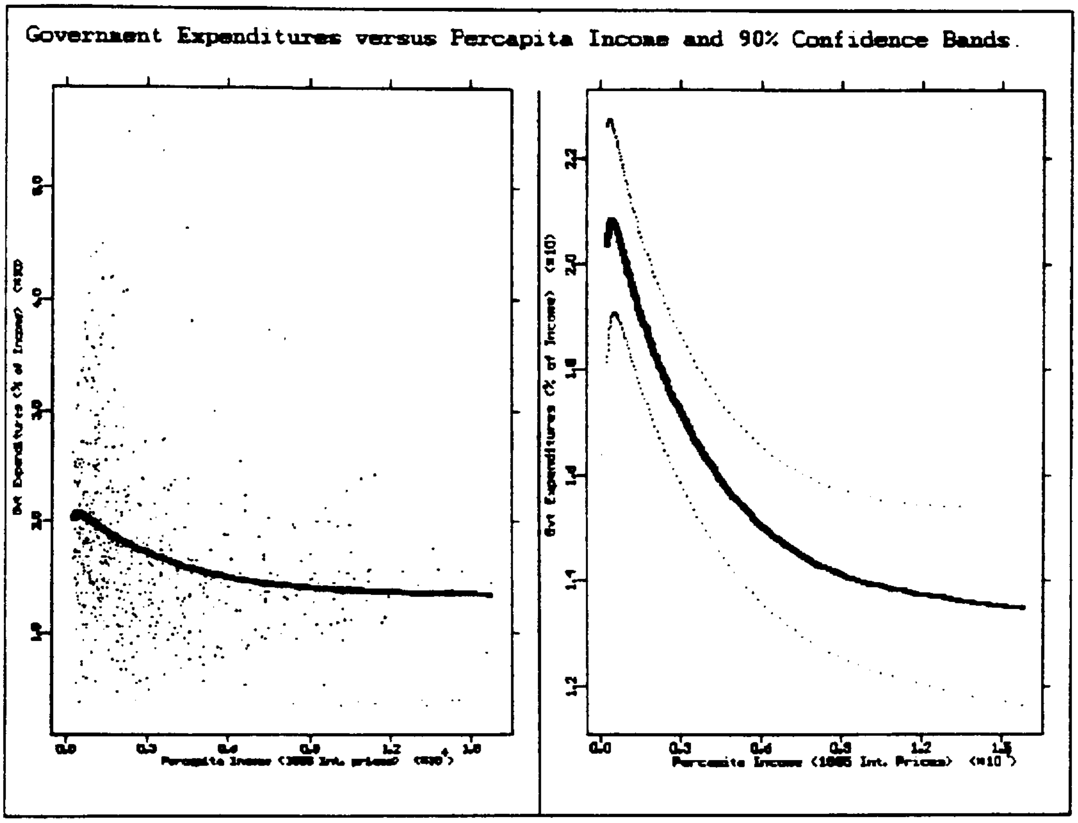
On the other hand, the sensitivity analysis of cross-country growth regressions, examined by Levine and Renelt (1992), shows that the role of the government in economic activity, proxied by the ratio of government consumption expenditures to GDP, is "fragile." Even though the coefficient on government expenditures share is always negative whatever the set of conditioning variables, it is significant only for very particular conditioning sets. This leads the authors to conclude, according to their null hypothesis, that the association between growth rates and the role of the government in economic activity at an aggregate level, is not "robust."

The association between the share of government expenditures and income together with the empirical linkage between the former and growth rates relative to the world average are presented in Figures 7.3.a-f, with their corresponding cross-validation functions as well as 90% confidence bands. Note first that the mean response of the government expenditure to GDP ratio decreases as the level of development increases. The average overall role of government in the economy is therefore greater in poorer countries. Paraguay, for instance, is a country which devoted ≅50% of its income to government expenditures during the period under examination and which has been characterized by "political stability," as Stroessner has been invariably reelected from 1954 to 1989. On the other hand, Paraguay has been characterized by growth rates greater than the world average growth rate, so that it is an open question whether this rapid growth is due to "political stability," as suggested by "Barro regressions," or, in the case of South Korea, a politically unstable country, government of which devotes 3.8% (1985) of its GDP to research and development in the electronic industry[10], to the provision of growth-promoting R&D and taxes designed to close the gap between private and social costs. Still, note that if the mean response of growth rates given the share of government expenditures to GDP decreases as the latter increases, the extent to which average growth rates respond to increases in the role of fiscal policies in a country's economic development is rather small, as shown in Figure 7.3.c. Finally, the three-dimensional perspective plot of the bivariate kernel density estimate of growth rates relative to the world average and the government consumption share of income provides additional information to that issued by the regression smoothing. Indeed, growth rates appear to be normally distributed either slightly positively or slightly negatively skewed around the mean level according to the share of income devoted to government expenditures.

---

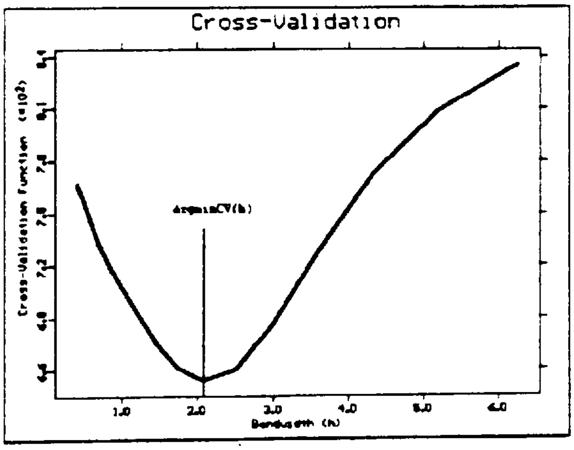
[10] Lorot & Schwob (1987).

Figure 7.3.a



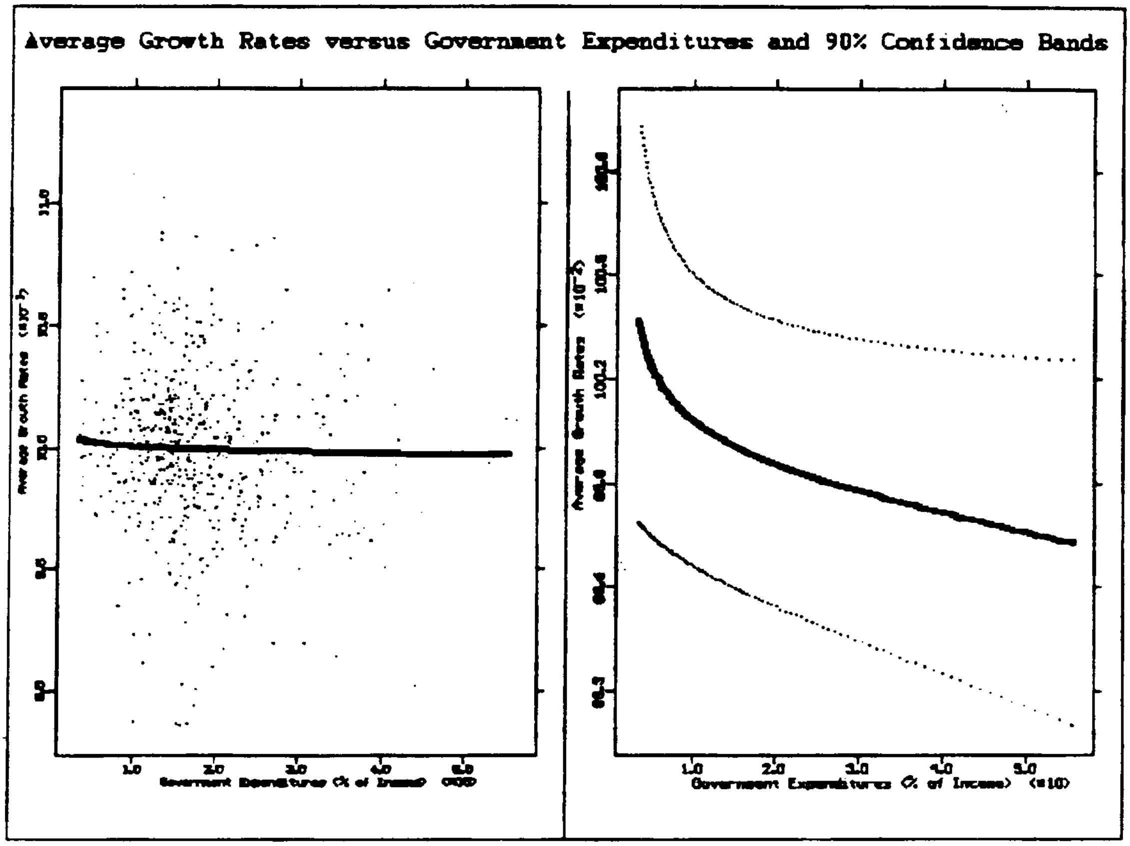
Plot of **Government Expenditures** Share (% of Income) versus per capita income observations (points) and true regression curve m(X) (thick line), with bandwidth h= *2.1* and 90% confidence bands (dashed lines).

Figure 7.3.b



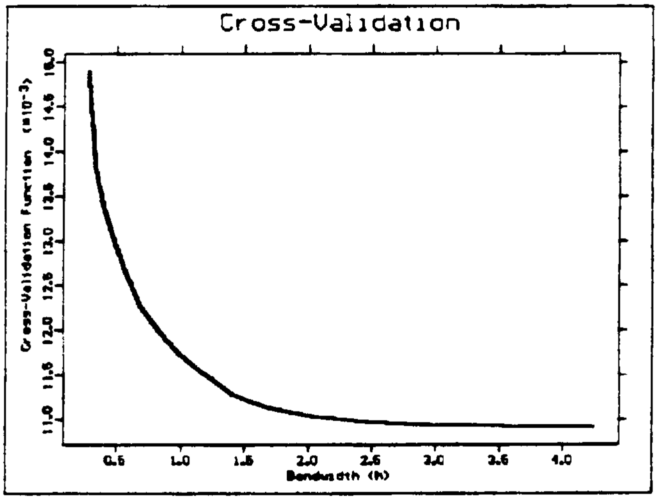
Cross-Validation function using a Quartic kernel for the original Nadaraya-Watson estimate of **Government Expenditures** share versus per capita income.
ArgminCV(h)= *2.1*.

Figure 7.3.c



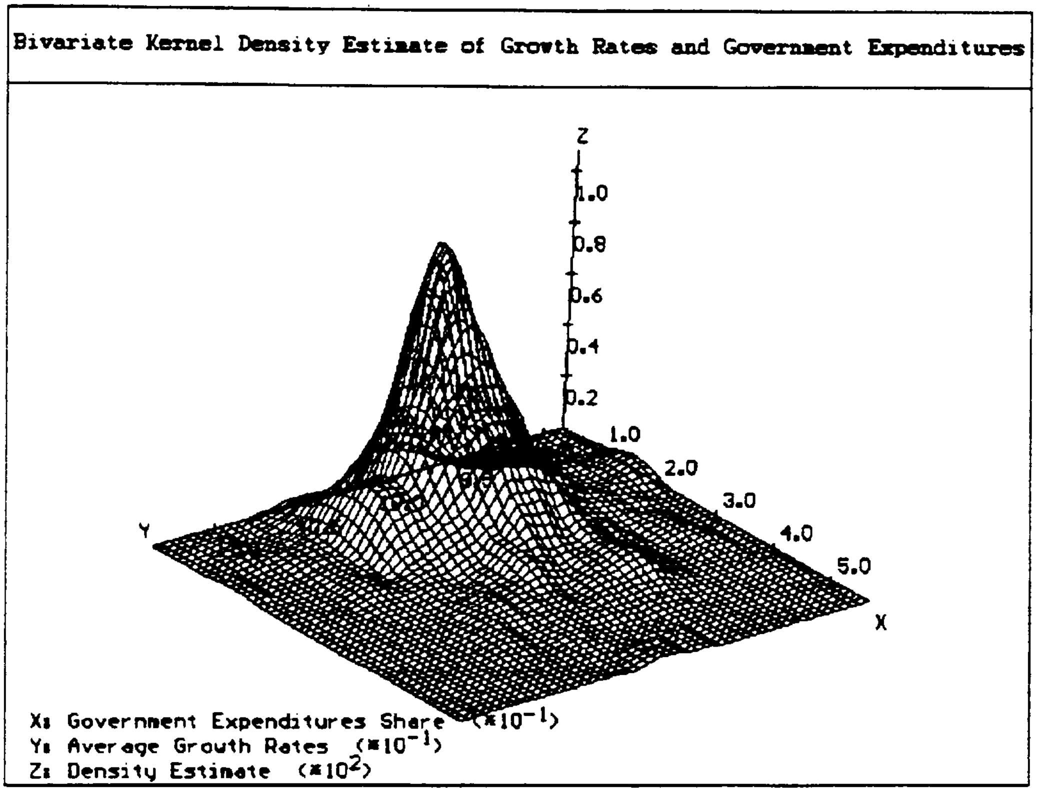
Plot of average growth rate versus **Government Expenditures** share (points)
and true regression curve m(X) (thick line), with bandwidth h= 2.6 and 90% confidence bands (dashed lines).

Figure 7.3.d



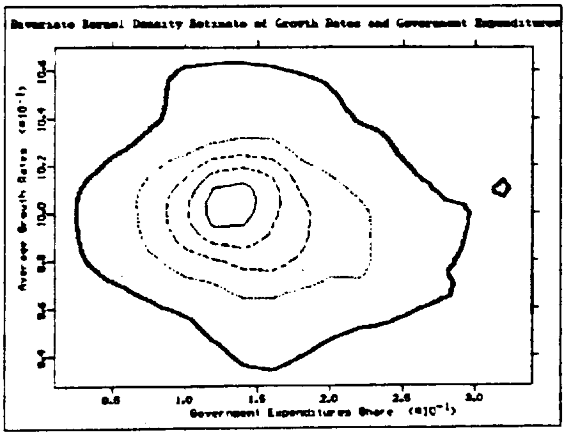
Cross-Validation Function using a Quartic kernel for the original Nadaraya-Watson Estimate
of the average growth rate versus **Government Expenditures**.

Figure 7.3.e



Three-dimensional perspective plot of bivariate kernel density estimate
of **Government Expenditures** share and average growth rates relative to the world average.
with bandwidths h_X=*0.06* and h_Y=*0.02*.

Figure 7.3.f



Equal probability contours (0.1, 0.3, 0.5, 0.7, 0.9)
of the bivariate kernel density estimate (f(X, Y)) with bandwidths h_X=*0.06* and h_Y=*0.02*.
X-direction= **Government Expenditures** share
Y-direction= average growth rates relative to the world average.

The above aggregate measure of fiscal policy with average per capita growth rates in this cross-section study might not be the right measure to proxy the size of the government in economic activity. As Barro (1991) argues: the effect of government expenditures on economic growth may depend on the allocation of these funds. This led him to remove government spending on education and defence which are more "like public investment than public consumption; in particular, these expenditures are likely to affect private-sector productivity or property rights, which matter for private investment."[11] However, Levine and Renelt's results for this new variable, are insignificant when altering the conditioning information set. Note as well that none of the other fiscal variables such as the ratio of government capital formation, government education expenditures and government defence expenditures to GDP, has been found to be "robustly" correlated to growth rates.

Therefore, as for investment at an aggregated level, the role of the government in economic development may be studied in terms of efficiency of resource allocation as suggested by Endogenous growth theory and the experiences of Singapore, South Korea and Taiwan. The size of the government in the economic activity proxied with the share of income devoted to public consumption appears to be rather unreliable as a claim for "laissez-faire" policies.

Thus, our findings corroborate both Bradford De Long and Summers' (1991) and Levine and Renelt's (1992) findings, and lead to the conclusion that differences in growth trends are a matter of effective use and allocation of resources rather than the accumulation of physical capital per se. As for investment, empirical linkages between fiscal policy and growth would require a detailed analysis of the composition of government expenditures and the structure of the tax system.

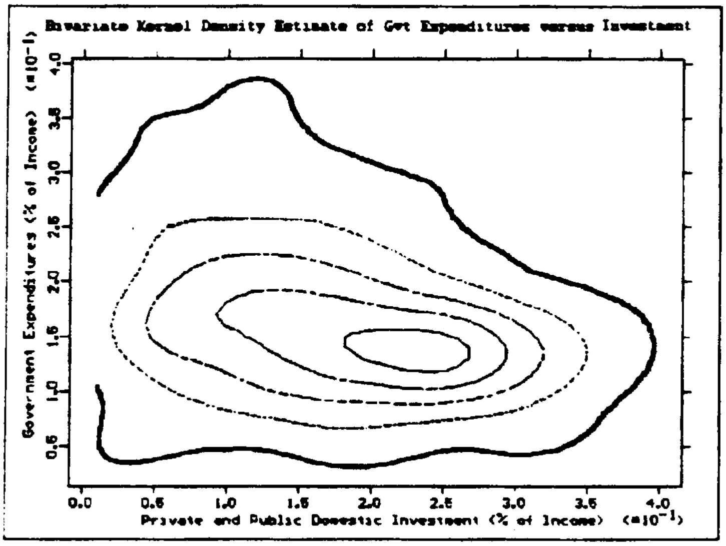
---

[11] Barro (1991; pp. 430).

Finally, we note from Figure 7.3.g that the joint distribution of government consumption share of GDP and the private and public domestic investment share is another good illustration of what Levine and Renelt call a "fragile" relationship. Indeed, the investment behaviour appears to be dramatically heterogeneous for a given share of income devoted to government consumption at least where 70% of the observations appears to concentrate. This contradicts traditional thought that governments which run large budget deficits are more than often governments out of control and that budget deficits necessarily crowd out investment. Note as well that Fischer (1991) found that data suggested only weakly that countries which grow faster do better on the account of the balance of payments. Fischer then, justifies the weakness of the association suggesting that this latter derives in part from variations in the tightness of constraints on borrowing. Again, from Figure 7.3.h which shows equal probability contours of the bivariate kernel density estimate of average growth rates versus the net foreign balance, one could have guessed that this relationship is nothing but "strong." Indeed, there exists a dramatic range of growth experiences whatever the country is facing a surplus or an external deficit. What a country exports must be just as important as whether it exports. In order to explain why a large trade imbalance might not be as harmful as claimed by international agencies, one can refer to the preferential access to imported intermediate inputs needed for producing exports and adopted by the Korean and Taiwanese governments over the period. Therefore, and as already highlighted by Levine and Renelt (1991), one must be extremely careful in drawing causal inferences which are not interpreted within a theoretical framework.

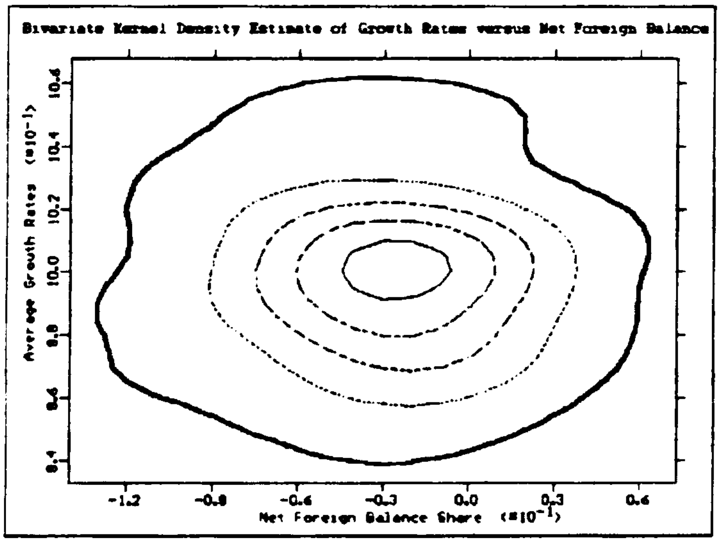
Again, non-parametric techniques provide information which is directly readable from the graphs and fill out the skeletal information available from more traditional techniques. Indeed, it is interesting to see how some economic issues can be illuminated by flexible displays of bivariate relationships. However, it becomes much more difficult to use these well, or to display the results when more than two variables are involved. As already

Figure 7.3.g



Equal probability contours[1] (0.1, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density estimate (f(X,Y)) with bandwidths $h_X = 0.075$ and $h_Y = 0.06$.
X= **Private and Public Domestic Investment** share,
Y= **Government Expenditures** share.

Figure 7.3.h



Equal probability contours[2](0.1, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density estimate (f(X,Y)) with bandwidths $h_X = 0.07$ and $h_Y = 0.02$.
X= **Net Foreign Balance** share,
Y=average growth rates relative to the world average.

---

[1] The marginal smoothing parameters have been set to 0.01 in the X-direction and to 0.01 in the Y-direction.

[2] The marginal smoothing parameters have been set to 0.01 in the X-direction and to 0.005 in the Y-direction.

discussed and surveyed by Levine and Renelt (1991), over 50 variables have been found to be significantly correlated with growth in at least one regression, and one cannot imagine bivariate displays of all possible combinations. The last component of income which has not yet received particular attention from researchers studying empirical linkages between macroeconomic variables and long-run growth, is private consumption. Therefore, the next section will focus on the patterns of demand in a world-wide cross-section demand analysis.

8/ World-Wide Expenditure Patterns: Education and Growth.

Pasinetti (1981; pp. 68-69) argues: "...any investigation into technical progress must necessarily imply hypothesis...on the evolution of consumer's preferences as income increases... Increases in productivity and increases in income are two facets of the same phenomenon. Since the first implies the second, and the composition of the second determines the relevance of the first, the one cannot be considered if the other is ignored."

However, the elaboration of growth models with technological progress needs some restrictive assumptions about the consumer's preferences, i.e. uniform and proportional expansion of demand. In the absence of this economic property, growth models which are general equilibrium processes within decentralized markets will exhibit, neither a quantitative nor a qualitative description of the solution.

Our aim here is to focus on the pattern of education expenditures to increases in income, on a worldwide cross-section demand analysis. First, within a single country, empirical evidence finds that per capita demand for each commodity does not expand proportionally, with the result that utility functions are not homothetic. Secondly, "with homothetic utility, the composition of the world demand will remain as fixed as income grows. We know that income elasticities for important classes of goods differ significantly from unity...This force will create comparative advantages...altering production patterns and growth rates as it does" (Lucas; 1988, pp. 34).

Models of Endogenous growth allow a distinction to be made between the effects that trade can have on the income levels and on the growth rates, providing many valuable lessons and a suitable framework for analyzing long-run growth. For a long time, the traditional literature has misunderstood how international trade affected growth and what the dynamic effects of international trade were on growth. Basically, in the Neo-classical theory, gains from trade have a level effect on income and consumption analogous to a one-time shot upward in production possibilities. This does not induce sustained and permanent increases in growth rates as proponents of outward-oriented policies have argued in the traditional theory of international trade. Instead, Endogenous growth theory has immediate implications for the effects of increasing trade in goods and flows of ideas, and for the benefits of economic integration on the growth rates as well as on the welfare levels.

However, both trade and Endogenous theories as well as empirical analysis have generally assumed that preferences are identical and homothetic across countries. Thus, global integration through international trade in goods and assets may improve welfare even in a country stuck in a development trap characterized by low domestic rates of growth. For instance, Helpman and Grossman (1990) argue that "a country that specializes its production activities in sectors that offer little prospect for technological progress may nonetheless benefit *as much as* others from the advances that are made in the more progressive sectors of the world economy". This argument comes directly from the Utilitarism theory, as the global and/or joint welfare has been maximized. As a consequence, inequalities become acceptable and legitimate as the poorest will also benefit from this welfare maximization. This must be interpreted as rational altruism. In other words, in order for the richer countries to be able to be interested in the left-out characterized by low productive capacities, one must encourage the more productive countries. However, these benefits rely on a worldwide income elasticity equal to unity for any class of goods. They find a justification in the so-called "orthodox" economy which analyzes economic behaviours independently of any social structures, postulating universal both motivations (Utilitarism) and substantial rationality. If one measure remains in order to account for inequality, we can imagine that this is consumption. However, we are aware that the problem is neither that of getting as big a "cake" as possible nor that of knowing how to divide it. Indeed, this cake is not homogeneous and individuals may not all have the same needs. In other words, there is not only one cake and equality and/or inequality may only be coherent with a number of cakes; that is, with several dimensions.

Hence, our first aim is to discuss the relevance of the assumption of homotheticity for a worldwide class of countries. Secondly, we will try to show how preferences can act as a constraint to growth. In other words, we would like to address the economic implications of nonhomothetic preferences on growth.

8.1/ Non-Homothetic Preferences.

To our knowledge, only a few papers (See, for instance, Hunter and Markusen, 1988; Theil and Finke, 1985; Finke, Lu and Theil, 1984) have been written on empirical evidence for nonhomothetic preferences across countries. First, Hunter and Markusen used a Linear Expenditure System derived from a simple Cobb-Douglas utility function for which the origin has been displaced; they found significant deviation from homotheticity in both economic and statistical terms. Secondly, Theil & Al. using the Working model, tabulated income elasticities, and rejected the homotheticity assumption across countries.

Here, following Deaton (1989), we will estimate bivariate kernel density estimates of some basic headings and per capita income, without assuming any pre-specified functional form of the stylized fact and with a minimum of unecessary assumption. In other words, the only specification which is involved concerns the choice of the dependent and independent variable: expenditure shares on goods will depend on per capita income[1].

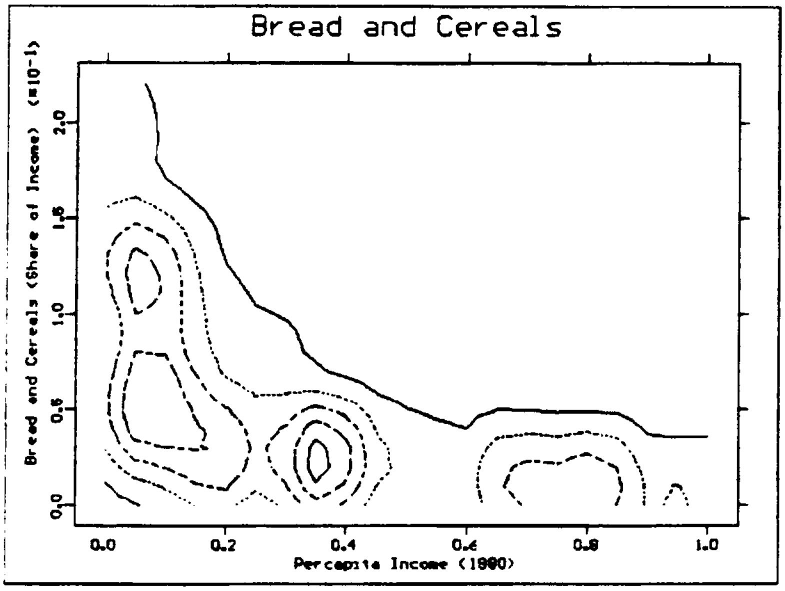
Phase IV of the International Comparison Programme for 1980 provides per capita quantities of basic headings valued at international prices for 60 countries at world average prices. Thus, we have for each country i the per capita income (x_i) and the expenditure y_{ic} on commodity c, where the list of headings contains 125 specific private consumption goods and services (See Appendix 11). The vector of quantities for each of the summary and detailed categories, expressed in international Dollars, reflects the physical quantity structure of each country in a comparable way, so that detailed international comparisons can be made of the composition of real final expenditures on GDP.

If the world works the way economists' models suggest, countries that have similar incomes (respectively similar price structures) should have similar quantity structures. In other words, they should absorb the different categories of goods in the same proportions. Instead, the quantity structures of our set of 60 countries appear to be nothing but similar, exhibiting a dramatic heterogeneity in consumption patterns, especially for the lowest income classes. Consumption patterns for four classes of goods are highlighted by flexible displays of contours plots which are a directly readable way of seeing the role of real per capita income in determining whether quantity compositions are similar or not (See Figures 8.1.a-d).

---

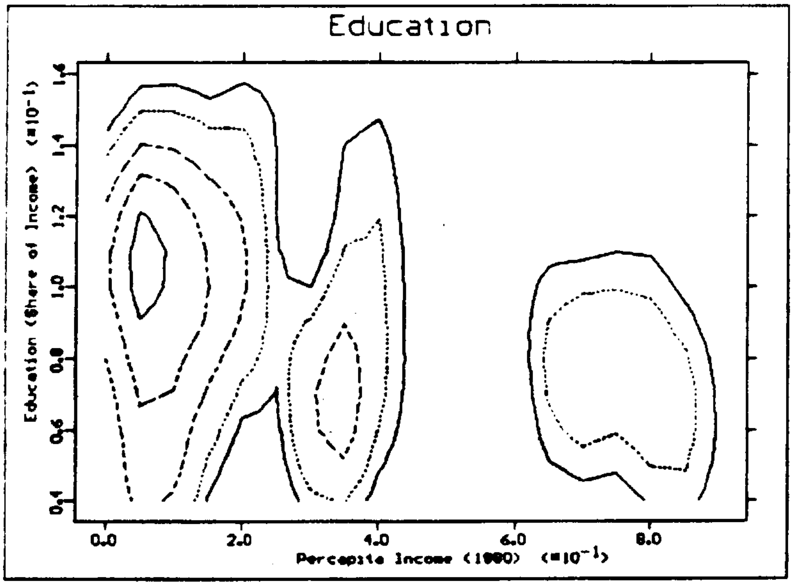
[1] See also Hildenbrand and Hildenbrand (1986) for an aggregate version of the law of demand using non-parametric estimates of densities and Engel curves from the British survey data.

Figure 8.1.a



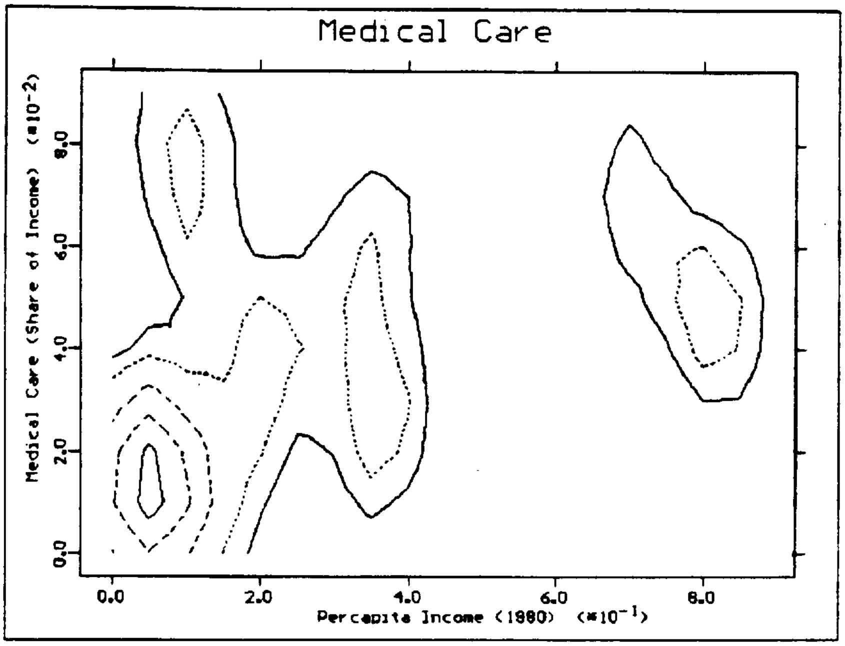
Equal probability contours (0.1, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density (f(X,Y)) estimate with bandwidths h_X=0.1 and h_Y=0.05.
Y= **Bread and Cereals** (Share of income),
X= Per capita Income (1980).

Figure 8.1.b



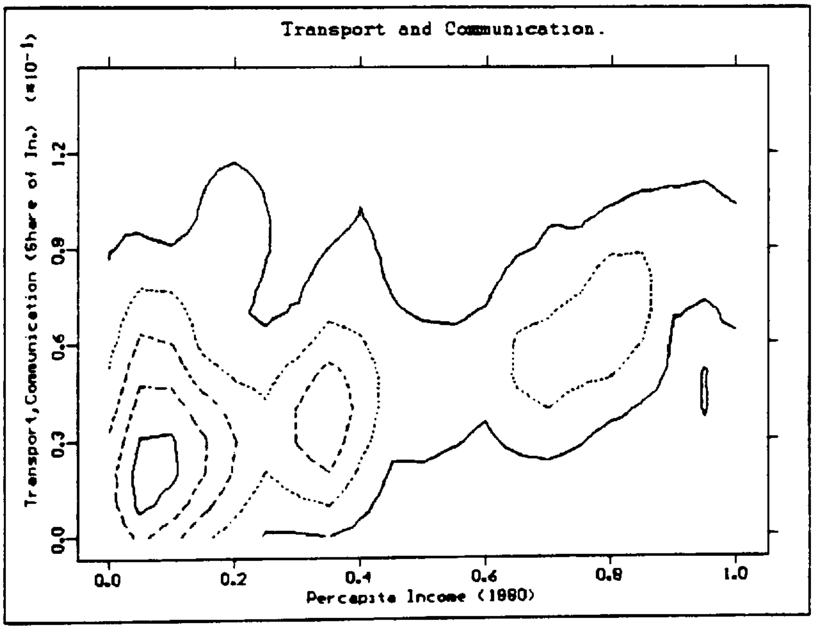
Equal probability contours (0.2, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density (f(X,Y)) estimate with bandwidths h_X=0.1 and h_Y=0.04.
Y= **Education** (Share of Income),
X= Per capita Income (1980).

Figure 8.1.c



Equal probability contours (0.2, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density (f(X,Y)) estimate with bandwidths h_X=0.1 and h_Y=0.03.
Y= **Medical Care** (Share of income),
X= Per capita Income (1980).

Figure 8.1.d



Equal probability contours (0.1, 0.3, 0.5, 0.7, 0.9) of the bivariate kernel density (f(X,Y)) estimate with bandwidths h_X=0.1 and h_Y=0.03.
Y= **Transport and Communication** (Share of Income),
X= Per capita Income (1980).

Shares of income devoted to both bread and cereals as well as to education confirm Engel's law, declining as living standards rise. Again, the height of the joint distribution represents the fraction of countries at the levels of standards of living and classes of goods by the coordinates along the base. Note, however, that although such contour plots give a clear impression of the concentration of mass, they might be dominated by information about the tails of the distribution where there may be very few observations. On the other hand, they encompass all the information, which is definitely over summarized when using cross-tabulation and linear regressions.

Note, for instance, the greater diversity in bread and cereals budgeting patterns as countries become poorer (See Figure 8.1.a). There is, in fact, a dramatic range of behaviour for the lowest income classes while almost all observations for the highest income classes concentrate between $+\varepsilon$ % and 5% of their total gross domestic product. A more striking feature is that this is similar for the consumption patterns of education (See Figure 8.1.b). Indeed, these appear to be more homogeneous among the richest countries and the average budget spent on education obviously decreases as standards of living rise.

Note as well the increasing diversity in medical care expenditure patterns as income increases; the poorest countries devote a very small share of their income to therapeutical products and health services (See Figure 8.1.c). Therefore, even though we present here only four classes of goods, the bivariate density contours exhibit a dramatic diversity in expenditure patterns across countries. Specific factors reflecting non-typical individual behaviours, like an urban/rural differentiation and/or ethnic determinants might partly explain this heterogeneity in consumption patterns. Thus, the traditional assumption introduced in economists' models, which is that similar countries in terms of income, should have similar quantity structures, has to be rejected. Steady states where everything grows in exact proportions do not exist in reality; the proportions, or what the growth rate is itself the outcome of internal forces (Kaldor; 1985). In order to understand them, households' behaviours might be interpreted taking into account social structures where the relevance of the *Homo oeconomicus* applied indifferently to both occidental and "primitive" societies, must be challenged maybe starting from the traditional anthropology as opposed to the microeconomic theory (Hugon; 1993). For instance, numerous African authors emphasize the incompatibility between the African representation and the occidental motivations -i.e. individualism, organisation, productivity, accumulation, which are specific to merchant and capitalist societies within a competitive economic organisation. Economic irrationality might find an explanation in non-typical economic behaviours which in Africa for instance, are nothing but a priority compared to symbolic activities within a community logic.

As an exhaustive analysis of all hypothesized relationships between expenditure patterns at a disaggregated level and growth would require much space and might be limited in terms of additional information, we propose focusing on a specific basic heading which has received a great deal of attention in studies examining empirical linkages between growth and intangible variables of interest; that is, education.

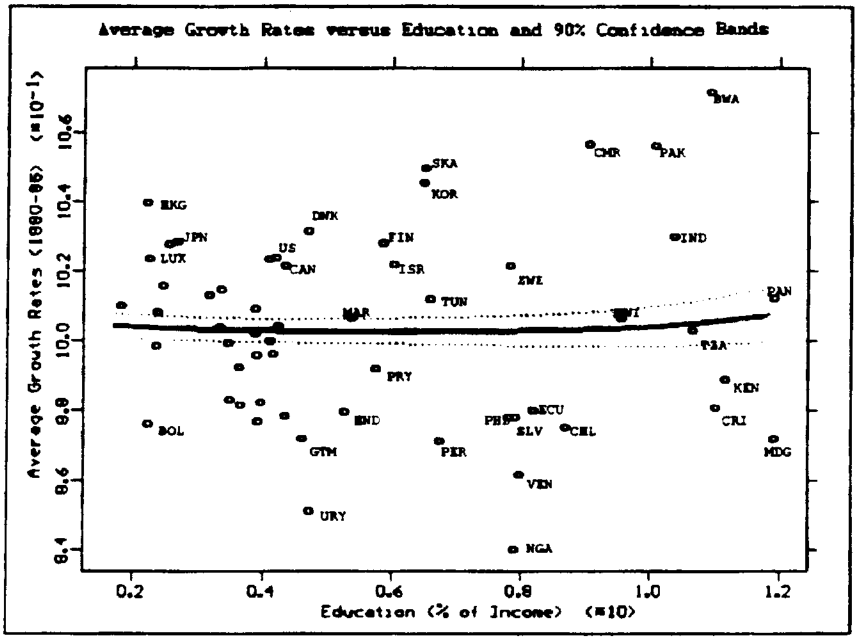
8.2/ Education Expenditure Patterns and Growth.

The shift of emphasis from disembodied technical innovation to human capital accumulation as the key source of growth and development is one of the major themes of the new economic development literature issued by a number of models of Endogenous economic growth (See, for instance, Lucas, 1988 and Romer, 1989 and 1990). Indeed, within the Neo-classical framework, physical accumulation of capital cannot lead to sustained and permanent growth. Therefore, researchers focused on different sources of growth such as "learning by doing," by using, by selling. They also highlighted the role of technical and scientific knowledge in R&D as well as accumulation of human abilities and skills through education expenditures. Within these frameworks, technological progress results from the households' behaviour. They decide to invest in the R&D sector and education. Then, because human capital is assumed to be a non rival good, policies -i.e. subsidies- encouraging both education and R&D will be Pareto improving and will raise the social returns and thus the rate of growth. An assortment of theoretical papers motivated the introduction of human capital proxies in cross-country growth regressions. Especially, they led Barro (1991) to examine the relationship between growth and the stock of human capital proxied by school-enrolment observed at the beginning of the period under study. As noted in section 2, he found a positive and significant partial association between growth rates and 1960 school-enrolment at both the primary and secondary levels, which has been found, with some qualifications, to be "robust" to slight alterations in the list of independent variables (Levine and Renelt; 1992). Secondly, the Romer & Al. (1990) human capital augmented Solow model explains about 80% of the international variation in income per capita, suggesting thus the role for the level of human capital variables to explain the rate of growth of output.

In our study, human capital accumulation is proxied neither by school-enrolment nor by adult literacy rate. Instead, we focus on social returns from households' education expenditures at both an aggregated and disaggregated level. In the United Nations International Comparison Project (Phase IV), the education expenditure category is divided into education fee, compensation for education, and commodities for education.

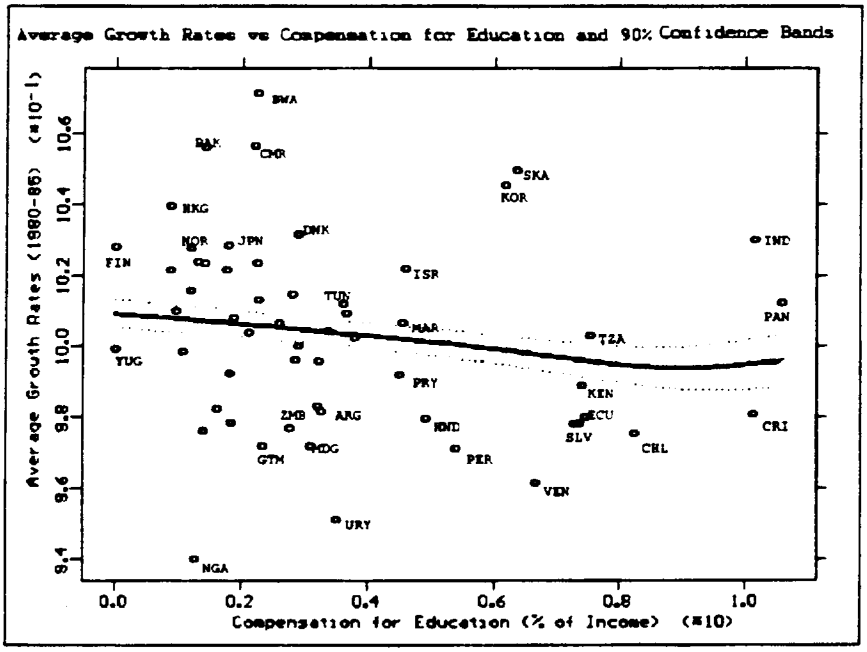
Figures 8.2.1.a-b display poor information in both aggregated education expenditures and one of its component, compensation for education. Furthermore, the mean response of average growth rates decreases as the share of income devoted to compensation for education increases. Note as well and as already insighted, that poorest countries devote largest shares of their income to education and compensation for education than richest countries do. On the other hand, both education fee and commodities for education may be crucial variables for understanding subsequent growth (See Figures 8.2.2-3). Indeed, a zoom of these figures show that a country which increases the share of its income devoted to commodities for education (education fee) from 1% (4%) to 2% (8%) can be expected to experience an increase of two percentage points per year relative to the world average. However, note that poorest countries appear to forsake the last component of education expenditures, devoting the greatest share to compensation for education while the richest countries' patterns concentrate on expenditures devoted to commodities for education. Finally, there is a great similarity between expenditure patterns in industrialized countries, and how they are associated with subsequent growth, in commodities for education and investment in both electrical and non-electrical machinery. This suggests the need for statistics concerning the accumulation of human capital, which would specify the ways, knowledge, are acquired and transmitted both inter-spatially and inter-temporally. Human capital proxies should allow to cover different types of qualifications and to distinguish among different sectors of activities.

Figure 8.2.1.a



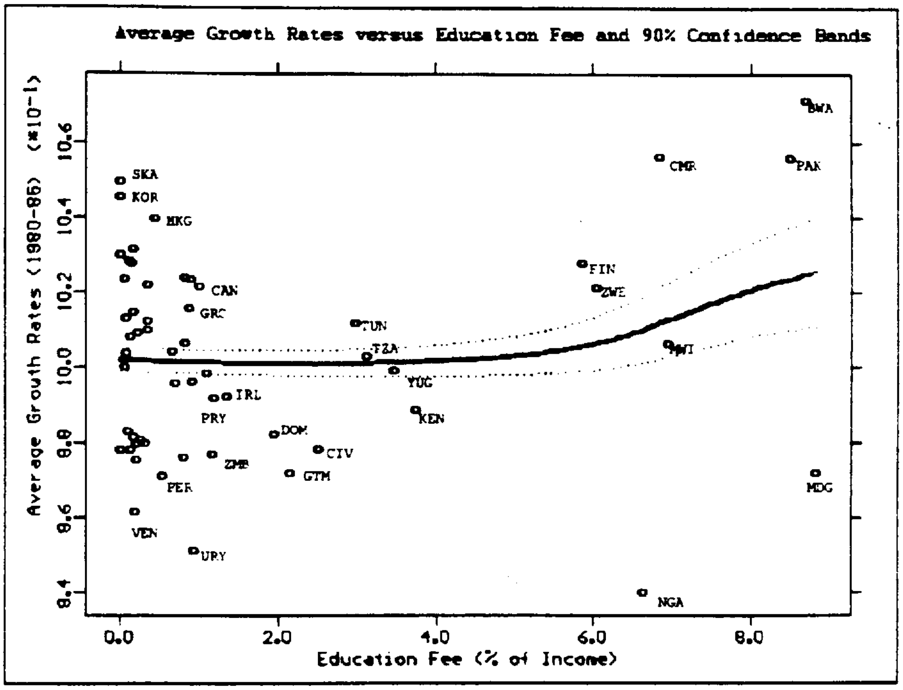
Nadaraya-Watson estimate of the average growth rate (Y)
versus **Education** share (X) (solid line), and 90% confidence limits (dashed lines)
with bandwidth h= 7.5.

Figure 8.2.1.b



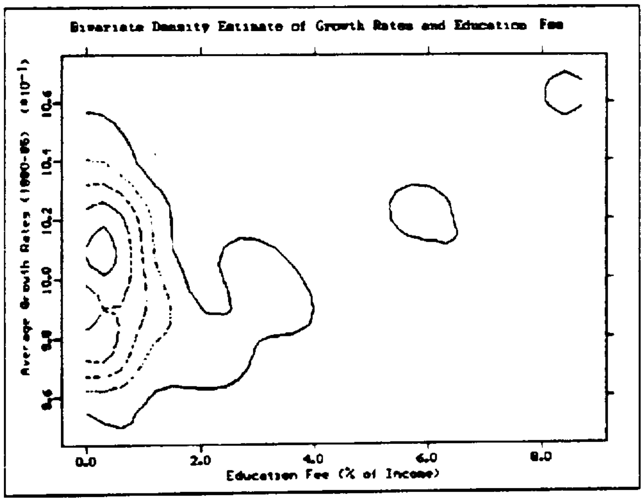
Nadaraya-Watson estimate of the average growth rate (Y)
versus **Compensation for Education** share (X) (solid line), and 90% confidence limits (dashed lines)
with bandwidth h= 7.

Figure 8.2.2.a



Nadaraya-Watson estimate of the average growth rate (Y) versus **Education Fee** share (X) (solid line), and 90% confidence limits (dashed lines) with bandwidth h= 6.5.

Figure 8.2.2.b

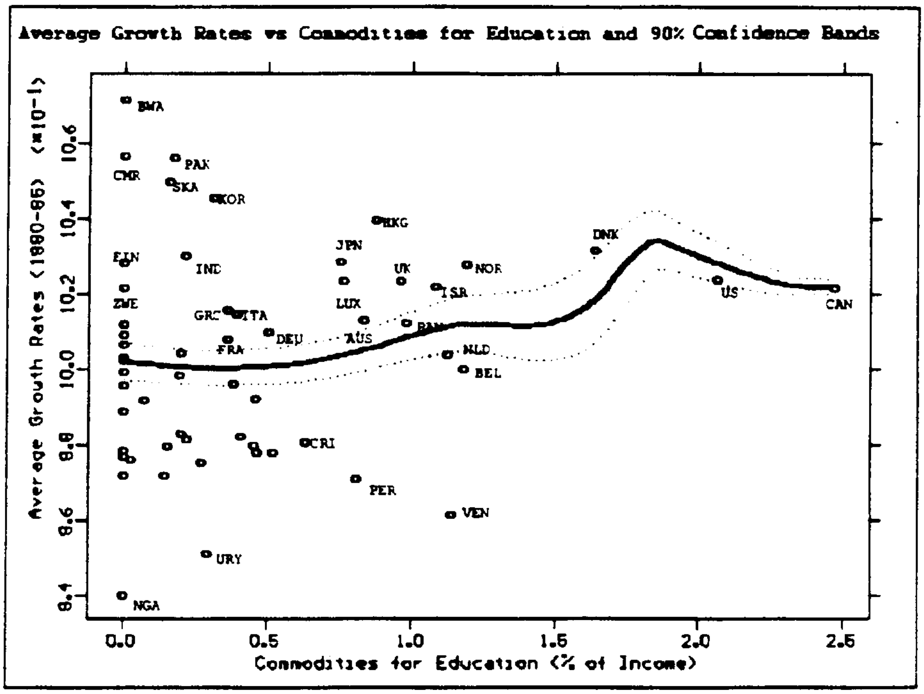


Equal probability contours of the bivariate kernel density[1] estimate (0.1, 0.3, 0.5, 0.7, 0.9) with bandwidths h_X=1.2 and h_Y=0.02.
Y= Average Growth Rates (1980-85),
X= **Education Fee** share (1980).

---

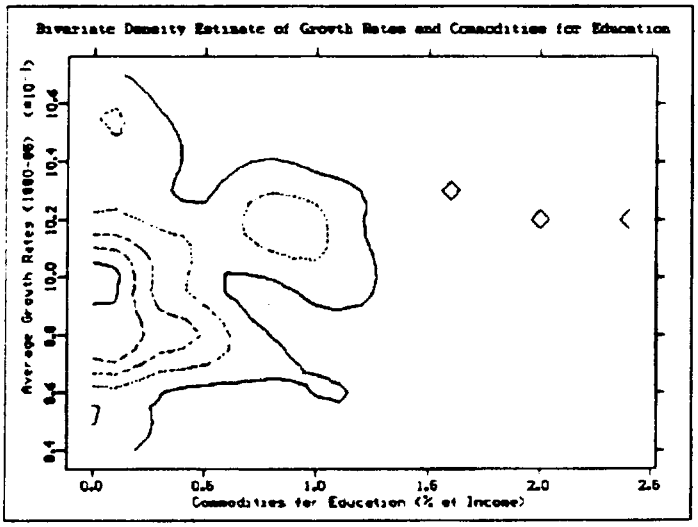
[1] The marginal smoothing parameters have been set to 0.3 in the X-direction and 0.005 in the Y-direction.

Figure 8.2.3.a



Nadaraya-Watson estimate of the average growth rate (Y) versus **Commodities for Education** share (X) (solid line), and 90% confidence limits (dashed lines) with bandwidth h=1.

Figure 8.2.3.b



Equal probability contours of the bivariate kernel density[1] estimate (0.1, 0.3, 0.5, 0.7, 0.9) with bandwidths h_X=0.3 and h_Y=0.02.
Y= Average Growth Rates (1980-85),
X= **Commodities for Education** share (1980).

---

[1] The marginal smoothing parameters have been set to 0.1 in the X-direction and 0.005 in the Y-direction.

9/ Summary and Conclusions.

The economic issue of convergence versus divergence among a world-wide cross-section of countries has been extensively analyzed since Romer (1986) proposed the first model of Endogenous growth which challenged the Neo-classical growth theory, allowing for the existence of a development trap; that is for the existence of multiple equilibria. Both theoretical and empirical literature did grow apace identifying original engines of growth and ever more sophisticated mechanisms of transmission of this in the economic activity. However, while some very interesting and highlighting arguments have been developed in the theoretical literature, there is as yet no consensus in order to guide empirical work; thus, any economic, institutional, and political variable might be considered when controlling for differences in steady state values in order to discuss the economic issue of convergence, and when looking at statistical inferences on the partial association between long-run growth and variables of primary interest. As noted by Levine and Renelt (1991), the absence of consensus is largely due to the heterogeneity of empirical studies which use a wide range of "right-hand-side" variables, use different countries, measure variables differently, employ different data sets, so that it is very difficult to discern consistent relationships.

If the parametric model is not true, the estimators may not be fully efficient. What is more, in parametric regression analysis, when the functional form is misspecified, these latter may not even be consistent. Indeed, a major drawback to these empirical approaches, is that the functional form of the model is specified directly or indirectly via the specification of the production function leading more than often to non "robust" estimations as illustrated by the Levine and Renelt (1992) sensitivity analysis. This justifies why we have once more challenged the Neo-classical convergence argument. To this end, and because previous research consists in parametric analysis, which uses the information in the data only from the postulated model and may be biased in our situation, we have chosen a non-parametric methodology which appears to be the right benchmark as it *lets the data speak for themselves*, when one wants to provide some additional information about stylized facts which are not immediately apparent and to make some comparisons with parametric forms developing procedures that can be used in the absence of a priori restrictions. Again, we stress that nonparametric analysis cannot replace the parametric one. However, with a data table like P.W.T5 and a computing environment like XploRe 3.1, much still remains to be said about the economic issue of convergence.

Non-parametric smoothing procedures can be employed as a suitable mean of estimating features of a data set. Therefore in a first step, we look at the extent to which income inequality rose during the period under study, i.e. 1960-85. If we accept that the current standards of living are an accurate measure of welfare, we find that inequality rose to a dramatic extent. Secondly, applying a logarithmic transformation to both per capita and per worker income measures in order to get something closer to symmetry, and estimating univariate kernel densities of these latter, we find some strong evidence of local rather than global convergence. Indeed, while the density estimate initially exhibits an unimodal structure, after 25 years it exhibit a fourth and fifth modes structure respectively for the per capita and per worker income measure. Note that, using this methodology, local convergence shows up without exogenously splitting the world into groupings, -i.e. escaping from any subjective choices. We then identify countries falling into the bumps exhibited by the density estimates in 1985 and countries falling into the same grids but in 1960. This procedure helps us to begin to understand the amount of mobility within the system. However, as might be expected, few countries moved from one group to another. Both upward and downward movements concern specific countries which would definitely require a case study approach -i.e. a pragmatic empirical work, in order to make more specific conclusions about the performance of recommanded policy changes. Still, note that some of them have already been extensively studied individually; for example, the "miracles" experienced by Japan, Hong-Kong, Singapore, South Korea and Taiwan, or conversely some African countries which have had problems since their independence. However, although the relative gap between the two extremes remains constant throughout the period, we illustrate as well that the distribution of income exhibits a fair amount of mobility.

The following section is dedicated to discussing both the amount and the extent of mobility within the system, and consists in original scenarios which allow to illustrate the economic issue of convergence. After having normalized the different income measures relative to US income, we estimate the joint density estimates (f(x(t), x(t+n)) in order to calculate the transition probabilities from one income class to another other five and ten years horizons. Indeed, according to the Neo-classical convergence argument a backward country is expected to catch up if a perfect capital mobility exists and if the framework consists in decentralized markets with diminishing returns to each input. We assimilate the opportunity to catch up to the following scenario:

$$S_1: P[\, x_i(t) < 1 \mid x_i(t) < x_i(t+n)] \text{ and } P[\, x_i(t) \geq 1 \mid x_i(t) > x_i(t+n)]$$

These joint densities have already been estimated by Quah (1993a). Observing that most of the observations, especially for the lowest income classes, were concentrated on the 45°-degree line, he concluded that mobility appears to confine to the middle and high income classes. However, the underlying a posteriori probabilities show a great amount of mobility that the graphs do not, and we, therefore, stressed that the amount and the extent of mobility must be distinguished. Indeed, the above scenario suggests that cross-sectional dynamics do not confine to middle and high income classes and the opportunity to catch up concerns as well the low income classes, while the possibility to fall behind is also likely to occur for both initially middle and high income countries. Secondly, for both measures of income, the probability to converge increases with n, the length of the period separating two observations. However the probabilities to diverge or for the relative gap to remain constant, is large enough to reject the scenario which acts in favour of convergence, at least for the period under study.

In order to highlight the amount of mobility within the system and especially to show that catching-up is not only a privilege of countries which have already reached a threshold level of development, and that falling behind does not concern only the lowest income classes, we propose looking at the scenario of convergence in using a new random variable (Y) which consists in the growth rates normalized relative to the US growth rate:

S_3: $P[x_i < 1 | y_i > 1]$ and $P[x_i \geq 1 | y_i \leq 1]$

Here, the probability that the relative gap will remain constant almost disappears, and the scenario of convergence is the most likely occurence especially for the per worker income measure. Furthermore, it is now clear that catching-up is not only the privilege of countries which have already reached a threshold level of development, at least when this is proxied by the level of productivity. Secondly, the possibility of falling behind is not confined to the lowest income class. However, the extent to which countries either catch up or fall behind decreases when they belong to higher income classes, and increases as the length of the period increases. Still, note that average growth rates relative to the leader are not large enough in order for the poorest countries to escape from what one might call the "unequalizing spiral" as illustrated by the previous scenario.

We estimated the growth rate distributions for both measures of income. There is a strong evidence that growth rates issued by the per capita income measure concentrate around 1 with a very small dispersion around the mean, thus justifying the small extent of mobility within the system. However, as the poorest countries belong to the tails of the distribution, this suggests that both convergence and permanent differences in growth rates characterize the world of today, meaning that the existence of multiple equilibria is more likely to occur than a unique and global equilibrium.

Finally, we suggest that this illustrates the integration of the economies in interdependent processes, even though nations appear to be independent political entities. Government policies may affect the equilibrium of other countries through mechanisms which have to be determined within a global economy. Indeed, a global perspective is needed in a world facing a structural crisis in international trade and debts repayments. As Kaldor (1985) suggests, a global and balanced growth might be preferred to nationalist interests that lead to outcome inequalities.

In traditional cross-country growth regressions, the "catch-up" hypothesis consists in regressing average growth rates on the initial income per capita conditioning on hypothesized explanatory variables such as investment, school-enrolment, government spending...These allow for differences in the permanent growth component to be controlled for. The goal of these empirical studies is also to examine empirical linkages between long-run growth rates and economic, political, and institutional variables. As Levine and Renelt (1992) have registered a range of 50 variables which are significantly correlated with growth in at least one regression, and as only one of them has been found to be "robustly" partially associated with growth, we proposed re-examining these relationships for some variables of primary interest, using non-parametric techniques.

Traditional parametric methods nest on strong assumptions so that, as illustrated by Levine and Renelt (1992), results may not be "robust" to the model specification. Misspecification of the hypothesized functional form may have serious consequences for both the econometric result as well as for recommanding policy changes that are usually expected to increase the chances of the lagging countries to perform better. Therefore, the approximation of the conditional mean of average growth rates given any explanatory variable, requires the use of more flexible functional forms avoiding a finite set of parameters. Non parametric smoothing techniques which do specify no functional form at all, are robust for misspecification and allow consistent estimation of a regression model.

On the one hand, we are struck by the rather small extent to which the mean growth rate responds to increases in the share of income devoted to investment. The mean response of growth rates relative to the world average given the share of income devoted to private and public domestic investment exhibits a logarithmic structure and indeed, varies within a very small range [-1.5%, +0.5%]; that is, the curve is steeper when starting at a low investment share, and flattens out among the highest shares of income devoted to investment. On the other hand, at an aggregated level, there exists a wide heterogeneity in growth patterns given the private and public domestic investment share. This dramatic diversity of growth experiences whatever the share of income devoted to investment is, clearly shows that neither countries which devote high shares of their income to investment will experience faster growth relative to the world average, nor will countries which devote a small part of their income to private and public investment necessarily fall behind.

The role of fiscal policies in a country's economic development, in accordance with the so-called "Barro regressions", is also examined. Growth rates appear to be normally distributed around the mean level -i.e. the world average growth rate- whatever the share of income devoted to government consumption is. This suggests, first, that provision of growth-promoting public goods and design taxes in order to close the gap between private and social costs as suggested by the "new" models of growth consists in a very relevant and useful argument. Secondly, the size of the government in the economic activity proxied with the share of income devoted to public consumption, appears to be rather unreliable in order to claim for "laissez-faire" policies.

These findings illustrate, corroborating Bradford de Long and Summers' (1991) as well as Levine and Renelt's (1992) results, that differences in growth trends are more a matter of effective use and allocation of resources rather than disembodied capital accumulation per se. Indeed, disaggregating domestic capital accumulation into construction, transportation, and both non electrical and electrical machinery, shows that growth rates patterns are quite homogeneous in countries with high equipment investment share, and such that the a posteriori probability to experience growth below the world average is almost zero when investment consists in machinery. Recall that international transfers of knowledge are given a central role in the economic issue of convergence. We also observed that learning and/or imitating an existing technology in order for the poorest to catch-up with the richest is far to be realized for countries belonging to the lowest income class. This suggests that for technological transfers to occur, they must be largely embodied in human capital as suggested by Lucas (1988), maybe in a larger extent than in physical capital, and that technical progress in one country may not necessarily spillover to actually affect techniques used in other countries.

Finally, as Pasinetti (1981) argues: any investigation into technological progress must require hypothesis on the evolution of consumer's preferences as income grows. This led us to turn to the last component of income which has not yet received particular attention from researchers studying empirical linkages between macroeconomic variables and long-run growth; that is, private consumption. However, the elaboration of growth models with technological progress needs some restrictive assumptions about the consumer's preferences, i.e. uniform and proportional expansion of demand. Lucas (1988; pp. 34) argues: "with homothetic utility, the composition of world demand will remain as fixed as income grows. We know that income elasticities for important classes of goods differ significantly from unity... This force will create comparative advantages...altering production patterns and growth rates as it does." Thus, we plot bivariate kernel density estimates of some basic headings and per capita income issued from the phase IV data table if the ICP Program. The traditional assumption introduced in economists' models which is that similar countries in terms of income should have similar quantity structures, has to be rejected. The quantity structures appear to be nothing but similar, exhibiting a dramatic heterogeneity in consumption patterns, especially for the lowest income classes. Steady-states where everything grows in exact proportions do not exist in reality; the proportions or the growth rate are themselves the outcome of internal forces (Kaldor; 1985). Or, following Hugon (1993), households' behaviours might be interpreted taking into account social structures where the relevance of the *Homo Oeconomicus* applied indifferently to both occidental and "primitive" societies must be challenged maybe starting from the traditional anthropology as opposed to the microeconomic theory.

The shift of emphasis from disembodied technical innovation to human capital accumulation as the key source of growth and development, is one of the major themes of the new economic development literature issued from a number of models of Endogenous growth. Thus, we focus as well on education expenditures at both an aggregated and disaggregated level. Our findings corroborate both theoretical arguments and empirical findings emphasizing the importance of human capital accumulation as an engine of growth. However, as for investment, differences in growth rates are a matter of effective allocation of resources rather than the accumulation of human capital per se, especially when a large share of its income is devoted to commodities for education. What is more, there is a great similarity between expenditure patterns in industrialized countries, and how they are associated with subsequent growth, in commodities for education and investment in both electrical and non-electrical machinery.

While we were at the institute for statistics and econometry at the Humboldt University in Berlin, spending more than ten hours a day "playing" with the computer, we came to a paper from Summers (1991). The author argues that too little time is spent understanding the broad patterns in the data compared to the time spent on sophisticated statistical techniques. Basically, he is struck by the negligible impact of formal econometric work on the development of economic science. In his conclusion, he goes even further suggesting that researchers, referees and editors show more interest in the demonstration of virtuosity and skills compared to the contribution which has being made.

As we do not claim here to have demonstrated much "technical virtuosity," we are led to ask whether the discussion undertaken into this empirical study, makes some contribution (let's say ε) to knowledge. We are aware that results reported previously will not consist in a crucial input to theory creation or the evolution of professional opinion. However, we hope to have illustrated that key economic issues can be investigated with non-parametric techniques which are far from over used in economic "science." Even though they consist in elaborate and rather sophisticated techniques, their usefulness finds a justification in the straightforward way data can be presented. Secondly, they place "little or no structure" on the data such that natural experiments within explicit probability models, can be examined through original scenarios. On the one hand, policy issues can be illuminated by flexible displays of bivariate relationships. On the other hand and as we restricted our discussion to simple two-variable situations, we want to lay stress on the fact that a natural following step in our investigations, will consist in applying these methods to the multivariate case. In high dimensions, these methods are less attractive because of the lack of simple but comprehensive graphs. However, as described in Härdle (1990a), there are a number of restricted structures that can be employed like the non-parametric additive models models or semiparametric models.

Of course, Summers will regret that it has required the "armor of a stochastic pseudo-world." However, even though this work cannot be considered as a "memorable" exercice, we hope to have brought new informations and insights about the economic issue of convergence and whether it fits the facts or not. We hope as well, if our interpretation of the evidence herein is correct, to have said something about the relevance of some of the arguments developed in the "new" growth literature.

References.

Abramovitz, M. (1986) "Catching-up, Forging Ahead, and Falling Behind." *Journal of Economic History, Vol. XLVI, N°2, 385-406.*

Barro, R.J. (1990) "Government Spending in a Simple Model of Endogenous Growth." *Journal of Political Economy, Vol. 98, N°5, 103-125.*

Barro, R.J. (1991) "Economic Growth in a Cross-Section of Countries." *The Quaterly Journal of Economics, Vol. 106, N°2, 407-443.*

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

Baumol, W.J. (1986) "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show." *American Economic Review, Vol. 76 N°5, 1072-1085.*

Baumol, W.J. and Wolff, E.N. (1988) "Productivity Growth, Convergence, and Welfare: Reply." *The American Economic Review, Vol. 78, N°5, 1155-1159.*

Bowman, A.W. (1984) "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates." *Biometrika, Vol. 71, 353-360.*

Bickel, P.J. and Rosenblatt, M. (1973) "On some global Measures of the Deviation of Density Function Estimates," *Annals of statistics, Vol. 1, 1071-1095.*

Breiman, L., Friedman, J.L., Olshen, R.A. and Stone, C.J. (1984) "Classification and Regression Trees." *Wadsworth, Belmont Ca.*

Deaton, A. (1989) "Rice Prices and Income Distribution in Thailand: A Non-Parametric Analysis", *The Economic Journal, Vol. 99, 1-37.*

De Long, Bradford, J. (1988) "Productivity Growth, Convergence, and Welfare: a Comment." *American Economic Review, Vol. 78, N°5, 1138-1154.*

De Long Bradford, J. (1991) "Productivity and Machinery Investment: A Long Run Look 1870-1980", *NBER Working Paper #3903.*

De Long Bradford, J. and Summers, L.H.(1991) "Equipment Investment and Economic Growth", *The Quaterly Journal of Economics, Vol. 106, N°2, 445-502.*

Dowrick, S. and Gemmel, N. (1991) "Industrialization, Catching Up and Economic Growth: a Comparative Study Across the World's Capitalist Economies." *The Economic Journal (March) 263-275.*

Dowrick, S. and Nguyen, D.T. (1989) "OECD Comparative Economic Growth 1950-85: Catch-Up and Convergence." *American Economic Review, Vol. 79, N°5, 1010-1030.*

Durlauf, S.N., and Johnson, P.A. (1992) "Local Versus Global Divergence Across National Economies." *LSE Financial Markets Group Discussion Paper Series, N° 131.*

Easterly, W., Kremer, M., Pritchett, L., and Summers, L.H. (1993) "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks." *NBER Working Papers, N°4474.*

Finke, R., Lu, W.H. and Theil, H. (1984) "A Cross-Country Tabulation of Own-Price Elasticities of Demand." *Economics Letter, Vol. 14, 137-142.*

Fischer, S. (1991) "Growth, Macroeconomics, and Development." *NBER Working Papers, N°3702.*

Grossman, G.M., and Helpman, E. (1989a) "Comparative Advantage and Long-Run Growth." *NBER Working Paper, N°2809.*

Grossman, G.M., and Helpman, E. (1989b) "Growth and Welfare in a Small Open Economy." *NBER Working Paper, N°2970.*

Grossman, G.M. and Helpman, E. (1990) "Hysteresis in Trade Patterns." *NBER Working Paper, N° 3526.*

Hall, P. and Marron, J.S. (1987a) "Extent to Which Least-Squares Cross-Validation Minimizes Integrated Square Error in Nonparametric Density Estimation." *Probability Theory and Related Fields, Vol. 74, 567-581.*

Härdle, W. (1990a) "Applied NonParametric Regression." *Econometrics Society Monographs, N°. 19, Cambridge University Press.*

Härdle, W. (1990b) "Smoothing Techniques: With Implementation in S." *Springer-Verlag, New-York Inc.*

Härdle, W and Scott, D.W. (1992) "Smoothing by Weighted Averaging of Rounded Points." *Computational Statistics, Vol. 7, 97-128.*

Hildenbrand, K. and Hildenbrand, W. (1986) "On the Mean Income Effect: A Data Analysis of the U.K. Family Expenditure Survey." *Contributions to Mathematical Economics, in Honor of Gérard Debreu, Ed. by W. Hildenbrand and A. Mas-Colell. Amsterdam: North Holland.*

Hugon, P. (1993) "L'économie de l'Afrique." *Editions La Découverte, Collection Repères.*

Hunter, L.C. and Markusen, J.R. (1988) "Per-Capita Income as a Determinant of Trade." *In R.C Feenstra Ed., Empirical Methods for International Trade. Cambridge. MA: MIT Press.*

Izenman, A.J. (1991) "Recent Developments in Nonparametric Density Estimation." *Journal of the American Statistical Association, Vol. 86, N° 413, 205-224.*

Kaldor, N. (1985) "Economics Without Equilibrium." *University College Cardiff Press.*

Kirman, A.P. and Tomasini, L.M. (1969) "A New Look at International Income Inequalities." *Economia Internazionale Vol. XXII, N°3, 3-27.*

Kravis, I., Heston, A. and Summers, R. (1978) "The United Nations International Comparison Project: Phase II: International Comparisons of Real Product and Purchasing Power." *Baltimore: John Hopkins University.*

Kravis. I.B. and Lipsey, R.E. (1983) "Toward an explanation of National Price Levels." *Princeton Studies in Internatiopnal Finance, No. 52.*

Kravis, I.B. and Lipsey, R.E. (1990) "The International Comparison Program: Current Status and Problems." *NBER Working Papers, N°3304.*

Levine, R. and Renelt, D. (1991) "Cross Country Studies of Growth and Policy: Some Methodological, Conceptual, and Statistical Problems." *World Bank Working Papers Series No. 608*

Levine, R. and Renelt, D. (1992) "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review, Vol.82, N°4, 942-963.*

Lorot, P & Schwob, T. (1987) "Singapour, Taïwan, Hong-Kong, Corée du Sud: les Nouveaux Conquérants?" *Ed. Hatier Paris.*

Lucas Jr, R.E. (1988) "On the Mechanics of Economic Development." *Journal of Monetary Economics, Vol. 22, 3-42.*

Maddison, A. (1982) "Phases of Capitalist Development." *New York: Oxford University Press.*

Mood, A.M., Graybill, F.A., and Boes, D.C. (1974) "Introduction to the Theory of Statistics." *McGraw-Hill International Editions, Statistics Series, 3rd Edition.*

Nadaraya, E.A. (1964) "On Estimating Regression." *Theory of Probability and its Applications, Vol. 10, 186-190.*

Pasinetti, L.L. (1981) "Structural Change and Economic Growth: a Theoretical Essay on the Dynamics of the Wealth of Nations." *Cambridge University Press.*

Park, B.U. and Marron, J.S. (1990) "Comparison of Data-Driven Bandwidths Selectors." *Journal of the American Statistical Association, Vol. 85, N°409, 66-72.*

Park, B.U. and Turlach, B.A. (1992) "Practical Performance of Several Data Driven Bandwidth Selectors." *Computational Statistics, Vol. 7, 251-270.*

Quah, D (1993a) "Dependence in Growth and Fluctuations Across Economies with Mobile Capital." *L.S.E. Working Paper (Preliminary Draft).*

Quah, D (1993b) "Galton's Fallacy and Tests of the Convergence Hypothesis." *L.S.E. Working Paper (June).*

Quah, D. (1993c) Empirical Cross-Section Dynamics in Economic Growth." *European Economic Review, Vol. 37, 426-434.*

Reynolds, L.G. (1983) "The Spread of Economic Growth to the Third World: 1850-1980." *Journal of Economic Literature, Vol XXI, 941-980.*

Romer, D., Mankiw, N.G. and Weil, D. (1990) "A Contribution to the Empirics of Economic Growth." *Quaterly Journal of Economics, Vol.429, 407-437.*

Romer, P.M. (1986) "Increasing Returns and Long-Run Growth." *Journal of Political Economy, Vol. 94, N°5, 1002-1037.*

Romer, P.M. (1989) "Human Capital and Growth: Theory and Evidence." *NBER Working Paper N° 3173.*

Romer, P.M. (1990) "Endogenous Technological Change." *Journal of Political Economy, Vol. 98, 71-102.*

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

Rosenblatt, M. (1971) "Curve Estimates," *Annals of Mathematical Statistics, Vol. 42, 1815-1842.*

Rudemo, M. (1982) "Empirical Choice of Histograms and Kernel Density Estimators." *Scandinavian Journal of Statistics, Vol. 9, 65-78.*

Silverman, B.W. (1986) "Density Estimation for Statistics and Data Analysis." *Chapman and Hall, London.*

Scott, D.W. (1985b) "Averaged Shifted Histograms: Effective Nonparametric Density Estimators in Several Dimensions." *Annals of Statistics, Vol. 13, 1024-1040.*

Sheather, S.J. (1992) "The Performance of Six Popular Banwidths Selection Methods on Some Real Data Sets." *Computational Statistics Vol. 7, 225-250.*

Smith, S.C. (1991) "Reconsidering the Real Sources of Export-Led Growth," *Industrial Policy in Developing Countries, Economic Policy Institute Working Paper.*

Solow, R.M. (1956) " A contribution to the Theory of Economic Growth." *Quaterly Journal of Economics, Vol. 70, 65-94.*

Stone, C.J. (1984) "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates," *Annals of Statistics, Vol. 12, 1285-1297.*

Summers, L.H. (1991) "The Scientific Illusion in Empirical Macroeconomics." *Scandinavian Journal of Economics, Vol. 93, N°2, 129-148.*

Summers, R., Kravis, I.B, and Heston, A. (1980) "International Comparisons of Real Product and its Composition: 1950-77." *Review of Income and Wealth, Series 26, N°1, 19-66.*

Summers, R., Kravis, I.B, and Heston, A. (1984) "Changes in the World Distribution." *Journal of Policy Modeling, Vol. 6, N°2, 237-269.*

Summers, R. and Heston, A. (1991) "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988." *The Quaterly Journal of Economics Vol. 106, N°2, 327-368.*

Theil, H. and Finke, R. (1985) "Income and Price Elasticities of Demand at Low Levels of Real Income," *Economics Letters, Vol. N° 14, 1-5.*

Watson, G.S. (1964) "Smooth Regression Analysis," *Sankhyà, Series A, Vol. 26, 359-372.*

XploRe 3.1.b (1993) "XploRe- a Computing Environment for Exploratory Regression and Data Analysing," *Institut für Statistik und Oekonometrie, Humboldt-Universität zu Berlin.*

APPENDIX 1

Africa		**Asia**		**Europe**		**Oceania**	
Algeria	D	Afghanistan	D	Austria	A^-	Australia	A^-
Benin	D^+	Bangladesh	C^-	Belgium	A	Fiji	D
Botswana	C	Burma	D	Cyprus	C	New Zealand	A^-
Burundi	D	China. P.R.	D	Denmark	A^-	New Guinea	D
Cameroon	C^-	Hong Kong	B^-	Finland	A^-		
Cape Verde	D	India	C	France	A		
Central Af.	D	Iran	C^-	Germany	A		
Chad	D	Iraq	D	Greece	A^-		
Congo	D^+	Israel	B	Iceland	B^+		
Egypt	D^+	Japan	A	Ireland	A^-		
Ethiopia	D^-	Jordan	D	Italy	A		
Gabon	D	Korea	B^-	Luxembourg	A^-		
Gambia	D	Kuwait	D	Malta	C		
Ghana	D	Malaysia	C	Netherlands	A		
Guinea	D	Nepal	D^+	Norway	A^-		
Guinea Bissau	D	Pakistan	C^-	Portugal	A^-		
Ivory Coast	C	Philippines	C	Spain	A^-		
Kenya	C	Saudi Arabia	D	Sweden	A^-		
Lesotho	D	Singapore	D	Switzerland	B^+		
Liberia	D	Sri Lanka	C^-	Turkey	C		
Madagascar	D^+	Syrian Arab	C^-	U.K.	A		
Malawi	D^+	Taiwan	D^-	Yugoslavia	B		
Mali	D^+	Thailand	C^-				
Mauritania	D						
Mauritius	D^+						
Morocco	C^-	**North America**		**South America**			
Mozambique	D						
Niger	D	Barbados	C	Argentina	C		
Nigeria	D^+	Canada	A^-	Bolivia	C		
Rwanda	D^+	Costa Rica	C	Brazil	C^-		
Senegal	C^-	Dominican Rep.	C	Chile	C		
Sierra Leone	D^+	El Salvador	C	Colombia	C		
Somalia	D	Guatemala	C	Ecuador	C		
So. Africa	C^-	Haiti	D	Guyana	D		
Sudan	D	Honduras	C	Paraguay	C		
Swaziland	D^+	Jamaica	C	Peru	C		
Tanzania	C^-	Mexico	C	Surinam	D		
Togo	D	Nicaragua	D	Uruguay	C^-		
Tunisia	C^-	Panama	C	Venezuela	C		
Uganda	D	Trin & Tob.	C				
Zaire	D	U.S.	A				
Zambia	D^+						
Zimbabwe	C^-						

APPENDIX 2

Appendix 2.a.

```
; *************************************************************************
; * DENEST macro **********************************************************
; *************************************************************************
;* Function: DENEST estimates a one dimensional density with bandwidth h*
; *
; *
; * Call:          y = DENEST (x h)
; *
; *->X             x  n x 1  matrix
; *                h         scalar
; * X->            y  m x 2  matrix
; *
; * (!)            the cols of X must be equal to 1.
; *                the bandwidth must be bigger than d.
; *
; *Example:        LIBRARY(smoother)   ; load the necessary library
; *                x = NORMAL(100)     ; generates a pseudo random variable
; *                with standard normal distribution.
; *                y = DENEST(x 1.1)
; *                gives the kernel density estimate using the quartic
; *                kernel for the generated data and using the
; *                bandwidth h=1.1
; *                the m x 2 matrix y contains in the first col the grid
; *                in the second col the density
; *
; * Comments: WARPing method, see W. Haerdle,
; *                "Smoothing Techniques with applications in S", Springer
; *                Verlag.
; *                The discretization binwidth d for the small bins is set
; *                to d=(max(x)-min(x))./100
; *
; * See also:      QUA         (macro), the quartic kernel
; *                SYMWEIGH    (macro), to create weights
; *
; *************************************************************************
; ** Wolfgang Haerdle, 910426 *********************************************
; ** Sigbert Klinke,   930219 *********************************************
; *************************************************************************
proc (fh)=denest(x h)
  error(cols(x)<>1 "DENEST: COLS(X) <> 1")
  d=(max(x)-min(x))./100
  error(h .<=d "DENEST: h smaller than d")
  (xb yb)=bindata(x d 0)  ; bin data starting from origin 0 in steps of d
  ;
  ; The output matrix xb contains the index of the non empty bins
  ; The output yb contains the absolute frequency of datapoints xi
  ;
  wy=symweigh(0 d/h h/d &qua)
  ;
  ; computes a weight sequence from a quartic kernel at a grid h/d points
  ; with starting point 0 and step d/h. It gives the discretized kernel
  ; weights of the quartic kernel
  ;
  wx=aseq(0 rows(wy))
  ;
  ; creates an additive sequence of the form xi = xi-1 +1
;
;   hd = ceil(h./d)                     ; these commands are needed
;   xb = xb|#(min(xb)-hd max(xb)+hd)    ; for the density estimate to
;   yb = yb|#(0 0)                      ; go to zero at the extremes
;
  (xc yc or)=conv(xb yb wx wy)        ; calculate density function
  fh=(xc*d)~(yc/(rows(x)*d))
endp
; *************************************************************************
```

Appendix 2.b.

```
;  *********************************************************************
;  * DENEST2 macro ******************************************************
;  *********************************************************************
;  * Function: DENEST2 estimates a two dimensional density with
;  *           bandwidth h and discretization binwidth d
;  *
;  * Call:          y = DENEST2 (x h d)
;  *
;  * ->X            x  n x 1  matrix
;  *                h         scalar
;  *                d  2 x 1  matrix or scalar
;  * X->            y  m x k  matrix
;  *
;  * (!)            the cols of X must be equal to 2.
;  *                the bandwidth h must be bigger than d.
;  *
;  * Example:      LIBRARY(smoother)
;  *                x = READ(nicfoo) ; read the data file nicfoo.dat
;  *                y = DENEST2(x 0.6 0.2)
;  *                SHOW(y d3d)
;  *                gives the kernel density estimate using the quartic
;  *                kernel for the netincome-food data using the bandwidth
;  *                h=0.6 and small bin 0.2
;  *                the m x k matrix y contains the density estimate so
;  *                that it can be shown by the dynamic 3D picture
;  *
;  * Comments: WARPing method, see W. Haerdle,
;  *                "Smoothing Techniques with applications in S", Springer
;  *                Verlag*
;  *
;  * See also:     QUA          (macro), the quartic kernel
;  *                SYMWEIGH     (macro), to create weights
;  *                GRID         (comnd), to create a grid
;  *
;  *********************************************************************
;  ** Wolfgang Haerdle, 910426 *****************************************
;  *********************************************************************
proc (y)=denest2(x h d)
   error(cols(x)<>2 "DENEST2: COLS(X) <> 2")
   if (rows(d).=1)                  ; modify small bin vector
      d=d|d
   endif
   (xb yb)=bindata(x d)
   m=trn(max(xb)-min(xb)+1)
      error(prod(m).>1000  "DENEST2: CAN'T CREATE MORE THAN 1000 DATAPOINTS")
   wx=#(0 0)
   n=h./d
   dh=d./h
   wy=symweigh(wx dh n &qua)
   wx=grid(wx 1 n)
   ;
   ; generates a grid with origin x and stepwidth 1.
   ;
      (xc yc or)=conv(xb yb wx wy)
      s=((xc+0.5)*diag(d))~(yc./(rows(x)*prod(d)))
   y=split(s m[1,2]) ; splits a matrix and glues the part together.
endp
;  *********************************************************************
```

APPENDIX 3.

```
;  **********************************************************************
;  * DENCVL macro *******************************************************
;  **********************************************************************
;  * Function: DENCVL computes the L-2 cross validation function for
;  *           density estimation in one dimension
;  *
;  * Call:         y=DENCVL(x)
;  *
;  * ->X           x  nx1 matrix
;  * X->           y  16x2 matrix
;  *
;  * (!)           the cols of X must be equal to 1.
;  *
;  * Example:  FUNC("examples\dencvl")
;  *               LIBRARY(smoother)
;  *               x = READ(buffa)
;  *               y = DENCVL(x)
;  *               gives the cross validation function using the quartic
;  *               kernel for the buffalo snowfall data
;  *               the m x 2 y matrix contains the CV function in the first
;  *               col the bandwidth h and in the second col the cv function
;  *
;  * Comments: WARPing method, see W. Haerdle (1990b; Chapter 4, formula 4.4.5),
;  *               "Smoothing Techniques with applications in S", Springer
;  *               Verlag
;  *               the discretization binwidth d for the small bins is set
;  *               to d=(max(x)-min(x))./100
;
;  * See also:     QUA         (macro), the quartic kernel
;  *               SYMWEIGH    (macro), to create weights
;  *
;  **********************************************************************
;  ** Wolfgang Haerdle, 910709 ******************************************
;  **********************************************************************
proc(cv)=dencvl(x)
   error(cols(x)<>1 "DENCVL: COLS(X) <> 1")
   d=(max(x)-min(x))./100
   n=rows(x)
   (xb yb)=bindata(x d)               ; bin data
   wx=0
   wy=1
   (xc nz or)=conv(xb yb wx wy)       ; nz contains # pts in bin
   m=3
   h=m*d
   cv=matrix(16 2)
   while (m<=33)
      h=h*1.2
      wy=symweigh(0 d/h m &qua)       ; create weights
      wx=aseq(0 rows(wy))
      (xc yc or)=conv(xb yb wx wy)    ; calculate density function
      fh=yc/(n*d)
      wm0=15*m^4/(16*m^4-1)
      i=(m-1)/2
      cv[i,1]=h
      cv[i,2]=d.*(fh'*fh)-2.*nz'*fh/(n-1)+2*wm0/((n-1)*h)
      m=m+2
   endo
endp
;  **********************************************************************
```

APPENDIX 4

Appendix 4.a

Country Classification for the year 1960
(Percapita Income expressed relative to 1985 prices)

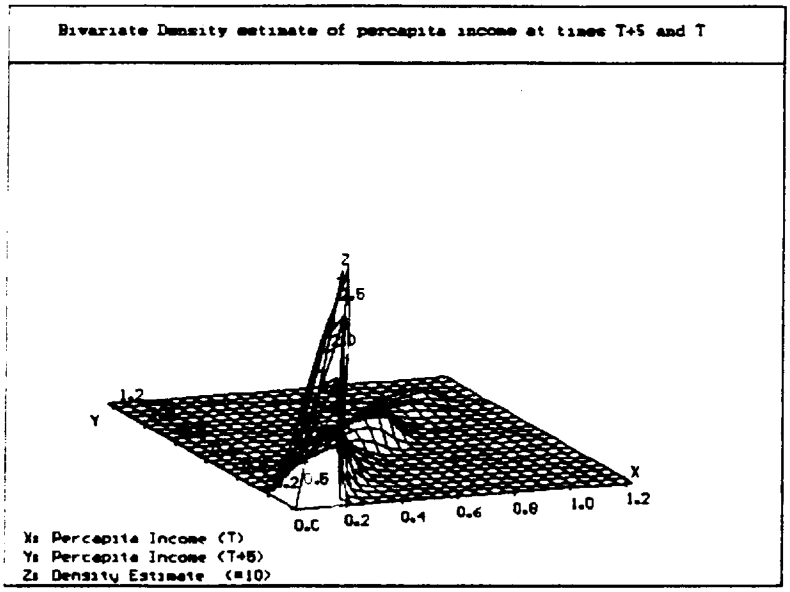
Group 1	*Group 2*	*Group 3*	*Group 4*
Botswana	Benin	Algeria	Israel
Burundi	Congo	Gabon	Saudi Arabia
Cameroon	Ghana	Mauritius	Austria
Cape Verde Is.	Ivory Coast	So. Africa	Belgium
Central Africa	Liberia	Tunisia	Denmark
Chad	Madagascar	Hong-Kong	Finland
Egypt	Mozambique	Iran	France
Ethiopia	Nigeria	Iraq	Germany
Gambia	Senegal	Japan	Iceland
Guinea	Sudan	Malaysia	Italy
Guinea Bissau	Swaziland	Singapore	Luxembourg
Kenya	Zambia	Sri Lanka	Netherlands
Lesotho	Jordan	Syria	Norway
Malawi	Philippines	Cyprus	Sweden
Mali	Taiwan	Greece	Switzerland
Mauritania	Thailand	Ireland	Great Britain
Morocco	Dominican Rep.	Malta	Barbados
Niger	El Salvador	Portugal	Canada
Rwanda	Bolivia	Spain	Trinidad & Tob.
Sierra Leone	Paraguay	Turkey	U.S.A.
Somalia	Papua New Guinea	Yugoslavia	Uruguay
Tanzania		Guatemala	Venezuela
Togo		Costa Rica	Australia
Uganda		Jamaica	New Zealand
Zaire		Mexico	
Zimbabwe		Nicaragua	
Afghanistan		Panama	
Bangladesh		Argentina	
Burma		Brazil	
China, P.R.		Chile	
India		Colombia	
Korea Rep. of		Ecuador	
Nepal		Guyana	
Pakistan		Peru	
Haiti		Surinam	
Honduras		Fiji	

Appendix 4.b

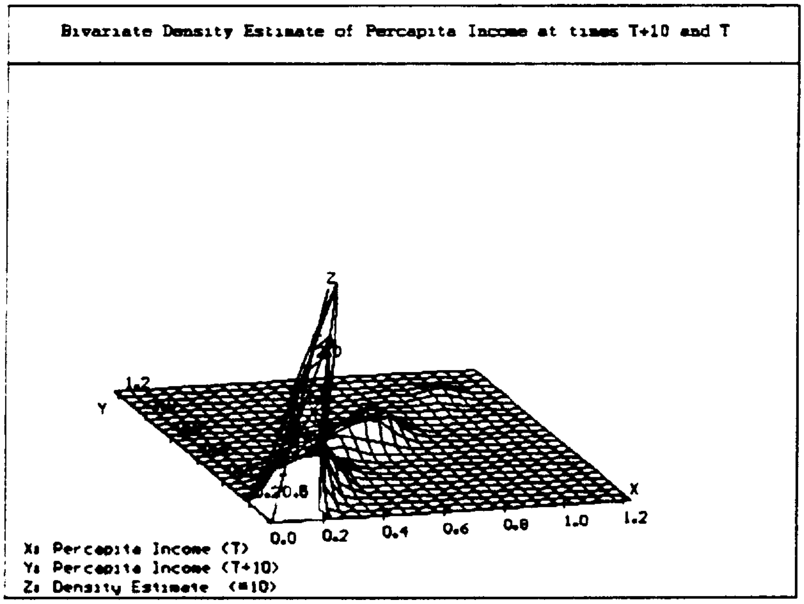
Country Classification for the year 1960
(Per worker Income expressed relative to 1985 prices)

Group 1	*Group 2*	*Group 3*	*Group 4*	*Group 5*
Botswana	Benin	Cape Verde	Algeria	Iraq
Burundi	Cameroon	Ghana	Mauritius	Israel
Central Af.	Chad	Mauritania	So. Africa	Saudi Arabia
Ethiopia	Congo	Morocco	Tunisia	Austria
Gambia	Egypt	Mozambique	Zambia	Belgium
Guinea	Gabon	Nigeria	Hong-Kong	Denmark
Lesotho	Ivory Coast	Senegal	Iran	Finland
Malawi	Kenya	Sudan	Japan	France
Mali	Liberia	Korea, Rep.	Jordan	Germany
Niger	Madagascar	Pakistan	Malaysia	Iceland
Rwanda	Sierra Leone	Philippines	Singapore	Italy
Tanzania	Somalia	Taiwan	Sri Lanka	Luxembourg
Togo	Swaziland	Honduras	Syria	Netherlands
Uganda	Zimbabwe		Cyprus	Norway
Zaire	Afghanistan		Greece	Sweden
Burma	Bangladesh		Ireland	Switzerland
China. P.R.	Thailand		Malta	Great Britain
India	Haiti		Portugal	Barbados
Nepal	Papua New Guinea		Spain	Canada
			Turkey	Mexico
			Yugoslavia	Trinidad & Tob.
			Costa Rica	U.S.A.
			Dominican. Rep	Argentina
			El Salvador	Chile
			Guatemala	Uruguay
			Jamaica	Venezuela
			Nicaragua	Australia
			Panama	New Zealand
			Bolivia	
			Brazil	
			Colombia	
			Ecuador	
			Guyana	
			Paraguay	
			Peru	
			Surinam	
			Fiji	

APPENDIX 5

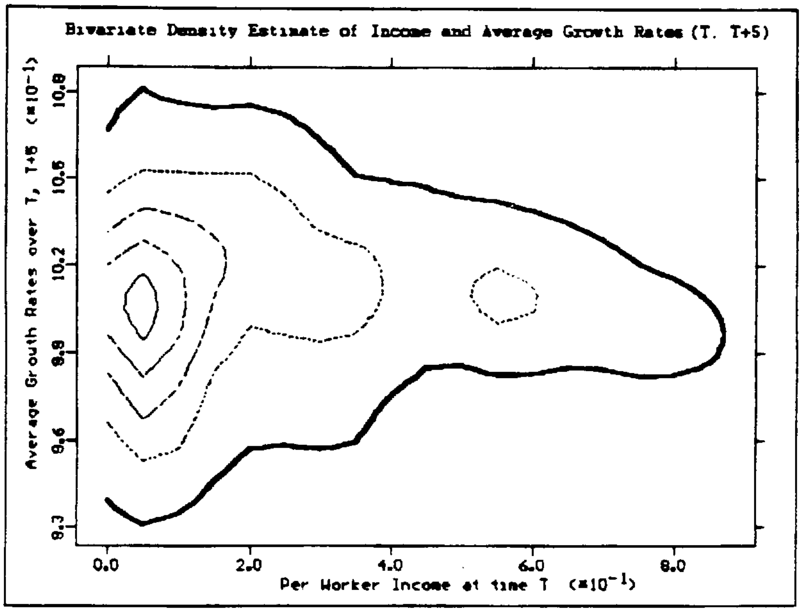


5-Years Transition Probabilities (Per capita Income)
Three-dimensional perspective plot of bivariate kernel density estimate
of per capita income (T+5) and per capita income (T), with bandwidths h= *0.1* in each component.

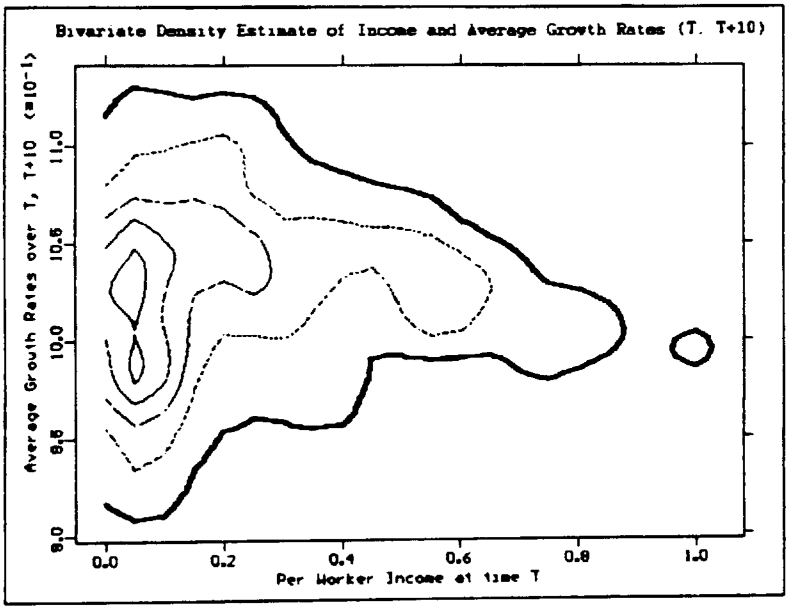


10-Years Transition Probabilities (Per capita Income)
Three-dimensional perspective plot of bivariate kernel density estimate
of per capita income (T+10) and per capita income (T), with bandwidths h= *0.1* in each component.

APPENDIX 6.



5-Years Transition Probabilities.
Equal probability contours plot (0.1, 0.3, 0.5, 0.7, 0.9) of bivariate kernel density estimate of average growth rates (T, T+5) and per worker income (T), with bandwidths h_X= *0.1* and h_Y=*0.025*.



10-Years Transition Probabilities.
Equal probability contours plot (0.1, 0.3, 0.5, 0.7, 0.9) of bivariate kernel density estimate of average growth rates (T, T+5) and per worker income (T), with bandwidths h_X= *0.1* and h_Y=*0.025*.

APPENDIX 7.

```
 ;  ***********************************************************************
 ;  * REGEST macro ********************************************************
 ;  ***********************************************************************
 ;  * Function: REGEST estimates a regression function with a quartic kernel and
 ;  *           bandwidth h
 ;  *
 ;  * Call:         y = REGEST (x h)
 ;  *
 ;  * ->X           x nx2 matrix
 ;  *               h     scalar
 ;  * X->           y mx2 matrix
 ;  *
 ;  * (!)           the cols of X must be equal to 2.
 ;  *               the bandwidth must be bigger than d.
 ;  *
 ;  * Example:  LIBRARY(smoother)
 ;  *               x = READ(geyser)
 ;  *               y = REGEST(x 0.4)
 ;  *               gives the kernel regression estimate using the quartic
 ;  *               kernel for the geyser data using the bandwidth h=0.4
 ;  *               the m x 2 matrix y contains in the first col the grid
 ;  *               in the second col the regression estimate
 ;  *
 ;  * Comments: WARPing method, see W. Haerdle,
 ;  *               "Smoothing Techniques with applications in S", Springer
 ;  *               Verlag
 ;  *               the discretization binwidth d for the small bins is set to
 ;  *               d=(max(x)-min(x))./100
 ;  *
 ;  * See also: QUA        (macro), the quartic kernel
 ;  *               SYMWEIGH   (macro), to creat weights
 ;  *
 ;  ***********************************************************************
 ;  ** Wolfgang Haerdle, 910426 *******************************************
 ;  ** Wolfgang Haerdle, 930502, paf at end removed ***********************
 ;  ***********************************************************************
proc(mh)=regest(x h)
   error(cols(x).<>2 "REGEST: COLS(X) <> 2")
   d=(max(x[,1])-min(x[,1]))./100
   error(h.<=d  "REGEST: h smaller than d")
   (xb yb)=bindata(x[,1] d 0 x[,2])    ; bin data in x and sum of y's
   wy=symweigh(0 d/h h/d &qua)         ; create weights for quartic kernel
   wx=aseq(0 rows(wy))
   (xc yc or)=conv(xb yb wx wy)        ; smooth x's and y's
   y=(d.*xc)~(yc[,2]./yc[,1])          ; combine nominator and denominator
;   mh=paf(y y[,2].<>NAN)              ; remove missings at the end
   mh =y
endp
 ;  ***********************************************************************
```

APPENDIX 8

```
;  *****************************************************************************
;  * REGCVL macro *************************************************************
;  *****************************************************************************
;  * Function: REGCVL estimates a regression function with a quartic kernel
;  *           and bandwidth h
;  *
;  * Call:         y = REGCVL (x)
;  *
;  * ->X           x nx2 matrix
;  * X->           y mx2 matrix
;  *
;  * (!)           the cols of x must be equal to 2.
;  *               the bandwidth must be bigger than d.
;  *
;  * Example:  FUNC("examples\regcvl")
;  *               LIBRARY(smoother)
;  *               x = READ(nicfoo)
;  *               cv = REGCVL(x)
;  *               gives the kernel regression  cross validation function
;  *               for the quartic kernel with the nicfoo data
;  *               the m x 2 matrix cv contains in the first col the grid
;  *               in the second col the cv function
;  *
;  * Comments: WARPing method, see W. Haerdle,
;  *               "Smoothing Techniques with applications in S", Springer
;  *               Verlag
;  *               the discretization binwidth d for the small bins is set
;  *               to d=max(x)-min(x))./100
;  *
;  * See also:    QUA        (macro), the quartic kernel
;  *               SYMWEIGH   (macro), to create weights
;  *
;  *****************************************************************************
;  ** Wolfgang Haerdle, 911018 ***********************************************
;  *****************************************************************************
proc (cv)=regcvl(x)
   library(smoother)
      error(cols(x).<>2 "REGCVL: COLS(X) <> 2")
      d=(max(x[,1])-min(x[,1]))./100
   (xb yb)=bindata (x[,1] d 0 x[,2])    ; bin data in x and sum of y's
   yo=yb[,2]./yb[,1]
   n=rows(x)
   ncv=10
   dp0=d*(15/16)
   i=1
   h=3*d
   cv=matrix(ncv 2)
;
     while (i<=ncv)
          wy=symweigh(0 d/h h/d &qua)     ; create weights for quartic kernel
         wx=aseq(0 rows(wy))
         (xc yc or)=conv(xb yb wx wy)    ; smooth x's and y's
         yc=paf(yc or.<>0)               ; use only indices with observations
         mh=(yc[,2]./yc[,1])             ; Nadaraya-Watson smoother
         t=(yo-mh)
         dp=dp0/h
         tr=1- (dp./(yc[,1]))   ; trace of smoother matrix
         res=t./tr              ; compute residuals
         cv[i,1]=h
         cv[i,2]=res'*res       ; sum of squared prediction errors
         i=i+1
         h=h*1.2
     endo
;
endp
;  *****************************************************************************
```

APPENDIX 9

Europe	**Africa**	**Asia**	**South-America**	**North-America**
Belgium (BEL)	Botswana (BWA)	Israel (ISR)	Argentina (ARG)	U.S.A
Denmark (DNK)	Cameroon (CMR)	Hong-Kong (HKG)	Bolivia (BOL)	Canada (CAN)
France (FRA)	Ethiopia (ETH)	India (IND)	Brazil (BRA)	
Germany (DEU)	Ivory Coast (CIV)	Indonesia (IDN)	Chile (CHL)	
Greece (GRC)	Kenya (KEN)	Japan (JPN)	Colombia (COL)	
Ireland (IRL)	Madagascar (MDG)	Korea R.P. (KOR)	Costa Rica (CRI)	
Italy (ITA)	Malawi (MWI)	Pakistan (PAK)	Dominican Rep. (DOM)	
Luxembourg (LUX)	Mali (MLI)	Philippines (PHL)	Ecuador (ECU)	
Netherland (NLD)	Morocco (MAR)	Sri Lanka (SRA)	El Salvador (SLV)	
U.K.	Nigeria (NGA)		Guatemala (GTM)	
Austria (AUT)	Senegal (SEN)		Honduras (HND)	
Finland (FIN)	Tanzania (TZA)		Panama (PAN)	
Hungary (HUN)	Tunisia (TUN)		Paraguay (PRY)	
Norway (NOR)	Zambia (ZMB)		Peru (PER)	
Poland (POL)	Zimbabwe (ZWE)		Uruguay (URY)	
Portugal (PRT)			Venezuela (VEN)	
Spain (ESP)				
Yugoslavia (YUG)				

APPENDIX 10

DOMESTIC CAPITAL FORMATION.

Producer Durables

- *Machinery:*
  - Products of processing.
  - Tool, Finished metal
  - Agricultural machinery
  - Machine tools for metal
  - Mining equipment etc...
  - Machinery for various specialities
  - Other machinery equipment
  - Office, Data processing
  - Precision, Optical instruments
- *Electrical Machines*:
- *Transportation:*
  - Motor vehicles, Engines
  - Other Transport equipment

Construction:

- *Residential:*
  - Family Dwellings
- *Nonresidential*:
  - Agricultural buildings
  - Industrial buildings
  - Buildings for market services
  - Buildings for nonmarket services
- *Non-Building*:
  - Transport routes
  - Other civil engineering

Change in Stock.

APPENDIX 11

Food, Beverage, Tobacco

Food

Bread and Cereals: Rice. Flour. other Cereals. Bread. Biscuits. Cakes. etc... Noodle, Macaroni. etc... Other Cereal products.

Meat: Beef & Veal. Pork. Lamb Goat Mutton, Poultry, Delicatessen. Meat Preparation. Other Meats.

Fish: Fish fresh/frozen. Fish dried/Smoked. Other Seafoods. Fish processed. Tinn

Milk. Cheese. Eggs: Milk Fresh. Milk Preserved. other Milk Products. Cheese. Eggs.

Oils and Fats: Butter, Margarine. Edible Oils. Other Fats.

Fruits and Vegetables: fresh Fruits. Dried Fruits. Nuts, Fruits Frozen. Juice. Fresh vegetables. Dried Vegetables. Frozen Preserved Vegetables. Potatoes. Tubers. Poc Poteto.

Other Food: Sugar, Coffe, Tea. Cocoa. Jam Syrup Honey. Chocolate. Confectionary, Ice Cream. Condiments. Spices.

Beverages

Non-Alcoholic beverages: Mineral Water. Soft Drinks.

Alcoholic Beverages: Other Alcoholic Beverages. Wine, Beer.

Tobacco: Cigarettes, Cigars, Cigarillos, Other Tobacco.

Clothing and Footwear

Clothing: Men's Garments, Women's Garments, Children's Clothing, Clothing Accessories, Repairs to Clothing.

Footwear: Footwear Mens. Womens, children, Infant. Repairs to Footwear.

Gross Rent, Fuel, Power

Gross Rents: Rents of Apartments, Rents of Houses, Repairs Maintenance of house.

Fuel and Power: Electricity, Town Natural Gas, Liquid Fuels, other Fuels.

House Furnishings, Operation

Furniture: Furniture, Fixtures, Floor Coverings.

Household Textiles.

Appliances: Refrigerators, Freeze, Cooking, Washing, Heating, Sewing, knitting Machine, other Household Appliances.

Other Households Goods, Services: Glassware, Tableware, Cutlery, Kitchen, Cleaning Maintenance Prod., Laundry, Dry Cleaning, Domestic Services and others.

Medical Care

Pharmace, Therapeutical Production: Pharmaceutical Products, other Medical Products, other Therapeutic Goods.

Health Services: Services of general Practice, Services of Dentists, Services of other Outside, Medical Analysis, Medical Personnel, Other than Medical Personnel, Current Consumption of Government.

Transport and Communication

Equipment: Motor Cars, Other Personal Transport.

Operation costs: Tires, Tubes, Accessories, Repair Charges, Motor Fuels, Oil, Grease.

Purchased Transport: Parking, Toll, etc... Local Transport, Rail, Bus Transport, Air, Sea, other.

Communication: Postal communication, Telephone, Telegraph.

Recreation and Education

Equipment for Recreation: Radio Sets, T.V. Sets, Record, Cassette Recorder, Camera, Phonograph Equipment, Sports goods, Accessories.

Recreational Services: Cinema, Theater, Concert, other Recreational Services.

Books, Periodicals, Newspaper: Books, Brochures, Magazines, Newspaper.

Education: Education Fee, Compensation for Education, Commodities for Education.

Miscellaneous Goods, Services: Barber, Beauty shops, Toilet Articles, Jewellery, Watch, other Personal Goods, Writings, Drawing Supplies, Restaurants, Catering Services.