CeLEG Working Paper Series

CONVERGENCE TOOLS AND MIXTURE ANALYSIS

Michele Battisti *Christopher F. Parmeter*

Working Paper No. 07
April 2010

Center for Labor and Economic Growth
Department of Economics and Business
LUISS Guido Carli
Viale Romania 32, 00197, Rome – Italy
http://www.luiss.edu/celeg

CONVERGENCE TOOLS AND MIXTURE ANALYSIS

MICHELE BATTISTI AND CHRISTOPHER F. PARMETER

Abstract. In this paper we employ traditional β, σ, γ-convergence tools to assess growth behaviour of groups of countries in connection with mixture densities. In the presence of mixtures, the standard methods to assess convergence are likely to be misleading. With respect to some recent papers dealing with issues of clustering through mixture densities of GDP, we assess what happens when the densities are multivariate, so that they contains traditional growth determinants. Our results show that, taken as global measures, convergence is an uncommon feature, whereas focusing on the individual components of the mixture we find pockets of convergence. Additionally, we focus attention on the classic determinants of growth and how these variables behave over time and within groups. Here the findings suggest that the behavior of countries within components is driven by similarities in the behavior of some of the growth determinants and that the well known result of a changing shape of income distribution from 1960 to 2000 is strongly related to changes in relationships among these variables began during 1980s.

1. Introduction

The divergence in cross-country per capita incomes is by now a stylized fact. However, a renewed theoretical interest has emerged attempting to explain the mechanisms in place that could have generated the large disparities that have led to the 'emerging' polarization of income. Zeira (2007) develops a model of skill-biased technical change where skilled workers replace unskilled workers. Due to the fact that the decision to adopt technology is endogenous, this skill replacement does not occur uniformly across countries. This fact can generate disparities in wages for both kinds of workers, as well as total factor productivity, leading to further disparities in income per capita.

Juxtaposed with this work, Galor & Mountford (2006) have argued that the rapid expansion of international trade over the last 50 years have generated demographic transitions which have influenced the shape of both the distribution of the world population and per capita income. Their analysis suggests that international trade worked in an asymmetric fashion across countries with differing incomes; rich countries used the gains from trade to invest in education whereas

Date: April 30, 2010.
Key words and phrases. Convergence, Nonparametric, Mixture densities.
Michele Battisti, Department of Studies on Law, Politics and Society, University of Palermo, Piazza Bologni 8, 90134 Palermo, Italy. Christopher F. Parmeter, Department of Agricultural and Applied Economics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0401. The research on this project has benefited from the comments of Oded Galor, Daniel Henderson, Paul Johnson, Mario Lavezzi, and Thanasis Stengos.

poor countries directed these gains towards population growth. These two differing channels then impacted growth in differing manners, which led to the polarization now standard in an analysis of cross country growth.

Here we seek to study the joint distribution of cross country income per capita, human capital, population growth and total factor productivity to discern how these variables co-evolved over time. We begin by using standard convergence measures to detect patterns which imply a narrowing of the individual distributions. We then use univariate mixture models to detect clustering of countries, based on per capita income, and then look for channels of convergence for the standard Solow growth determinants. After this analysis we gravitate towards a more generalized framework by employing multivariate mixture models to detect patterns in the joint distribution of income per capita *and* the standard determinants.

Our results are striking. First, focusing on basic convergence of either income per capita or any of the determinants we do not detect the presence of a collapsing distribution. When we move to univariate mixture models we see the common feature that the world has been gravitating towards two poles, a rich and a poor and within each of these poles it appears that the standard growth determinants are collapsing. Extending our results to bivariate and trivariate analysis we see distinct groupings emerge along various avenues. Thus, no unified theme emerges. We find clustering along income and total factor productivity, clustering along income per capita and human capital and clustering along income per capita and physical capital accumulation.

Here we build on the recent contribution of Pittau, Zelli & Johnson (2010) who stress the importance of searching for components in the mixture density to more aptly characterize clubs that may be converging to a common steady state. Our contribution tries to extend this approach by looking to multidimensional clustering based on mixtures, so that we look for multivariate densities given not only by GDP, but also by growth determinants such as human and physical capital and total factor productivity. In this regard we compare mixture of univariate densities (GDP) with mixture of multivariate densities over the period 1960-2000. By looking at the number of clusters, transition of countries among clusters, changing distances of clusters over time we may get insights about the determinants of polarization and the reason why several peaks emerged. We use mixture densities to determine the presence of clubs in the global distribution on per capita income.

These results suggest that, in the spirit of Brock & Durlauf (2001), no unified theory exists to explain the great divergence over the last 50 years. However, it appears that both the skill-biased theory of Zeira (2007) and the trade timed theory of Galor & Mountford (2006) are both viable explanations for subsets of countries and their associated growth patterns. The use of multivariate clustering analysis is important because it allows one to study jointly the behavior of income per capita as well as its potential sources and moves beyond the conditional mean approach which has been argued to be limited in discerning shape and structure of the underlying density. Moreover, by focusing on the joint evolution of these variable we can more readily understand how each variable is influence by another, as opposed to looking for patterns in a set of univariate results.

The remainder of the paper is structured as follows. Section 2 reviews the literature on the evolution of the distribution of per capita income and the associated clubs that have emerged. Section 3 discusses our three empirical measures of convergence more formally and presents generic results without respect for potential clustering. Section 4 discusses both the univariate and multivariate econometric clustering methods we will employ in this study. Our results are presented in Section 5 while Section 6 highlights our findings and offers suggestions for future work.

2. Empirical Literature on Income Divergence

One of the most debated and heavily studied issues in growth empirics regards the presence of income convergence among countries (e.g., Durlauf, Johnson & Temple 2005). Here researchers are interested in the characterization of the distribution of income (or aspects thereof) and its behavior over time. Depending on the specific feature of the distribution that is studied different classes of convergence are identified. In general no consensus exists on the 'best' measure to claim convergence of cross-country incomes although measures that capture shape and location changes are informative.

More recently, attention has focused less on convergence of the countries of the world and more on pockets of convergence. This stems from the influential work of Durlauf & Johnson (1995) who suggested that parameter heterogeneity was an important factor when thinking about cross-country convergence. To assess convergence, either globally or in pockets, the empirical literature primarily relies on the notion of β-convergence (Barro 1991), which focuses attention on either the mean (absolute convergence) or conditional mean (conditional convergence) of the distribution of cross-country income. Since other aspects of the distribution are important to detecting the presence of convergence, both σ and γ-convergence (Barro & Sala-i-Martin 1991, Boyle & McCarthy 1997, respectively) were introduced as complements to β-convergence measures. These measures account for a decreasing spread and intra-distributional churning, respectively, both of which can help to illuminate the potential for convergence in general.

To more aptly characterize the behavior of the entire distribution Quah (1993*a*, 1993*b*, 1996*a*, 1996*b*) drew attention to a more comprehensive approach to studying distribution dynamics which can capture more of the underlying features of the long run distribution, such as twin peakedness. None of the aforementioned measures capture this important underlying issue.[1] Thus, it is possible, in the presence of clubs that β, σ or γ-convergence is occurring within the clubs but is not representative of the entire distribution. Thus, while certain groups of countries may display convergent type behavior, the same measure(s) is meaningless when looking over the entire distribution. Consequently, given this difficulty with blindly applying convergence measures in the presence of additional features of the distribution of income, it is useful to identify and distinguish the 'clubs' within a distribution prior to focusing on convergence. This step will allow the researcher to delineate between changes across members of clubs and changes across the clubs themselves (i.e. when there is a transition of a country in a different cluster over time).

Empirical work related to clustering of income/growth patterns include Durlauf & Johnson's (1995) regression trees, Desdoigts's (1999) projection pursuits and Canova's (2004) predictive density approach. The first method is a nonlinear extension of the idea of principal components analysis that allow to reduce multidimensional data to smaller dimensions by projecting them on one or two dimensions, while the second one looks to breaks in income densities after an income ordering of the observations (EU regions). The focus of Desdoigts (1999) is on a classification based on many economic countries characteristics as in Durlauf & Johnson (1995) while in the great part of empirical literature included Canova (2004) the interest is mainly on income distribution. This

[1] Classic theoretical contributions on this issue can be found in Azariadis & Drazen (1990), Galor (1996) and, in the context of unified growth theory, by Galor (2007).

series of research looks at homogeneous clustering in a static sense, so that the determination of clusters is fixed over time.

Focusing on income distribution evolution, Paap & Van Dijk (1998) began what is now a burgeoning interest within the growth empirics literature on the use of mixture models. Mixture models are usually employed to analyze properties of unknown clusters of observations (McLachlan & Peel 2000) and have been employed in a wide array of settings ranging from biology and medicine to marketing. A recent notable contribution of Pittau et al. (2010) assesses the characteristics of GDP per capita distribution over 1960-2000 by comparing mixture models and nonparametric density methods. The mixture approach is in stark contrast to studying the distribution of income/growth from the studies of Bianchi (1997) and Henderson, Parmeter & Russell (2008) who focus exclusively on modality of the entire distribution, which as noted by Pittau et al. (2010) is neither necessary nor sufficient to capture the underlying components of the mixture density.

Another recent bulk of literature allows covariates to influence the conditional mean, known as the mixture regression approach. Following this line of research some recent works are Alfo, Trovato & Waldmann (2008) Paap, Franses & Van Djik (2005), Battisti & Di Vaio (2008) and Owen, Videras & Davis (2009). The first two use a panel dimension while the latter ones employ cross section estimation. The common finding, aside from geographical and time period differences, is that growth processes are heterogeneous for groups of countries. Another common feature of this strand of literature is the determination of a given number of clubs for the whole period under examination that is usually found to be equal to two clusters. We advocate and we will show that, in the light of findings of Quah (1996*b*) and Henderson et al. (2008) this is an assumption that is not confirmed by the data, while the polarization that appeared after Second World War is better described by a changing number of clubs, ranging from one to three.[2]

Beyond the search for components in a mixture density, we apply traditional measures of convergence (β, σ and γ) to each of our estimated clubs to determine the insights these methods provide ignoring the general criticisms levied against the use of any summary statistic measure to

[2]It is important to stress that this may be a transitory or a permanent phenomenon. For instance the world income distribution built by Bourguignon & Morrisson (2002) over 1820-2000 does not confirm that multimodality is a recent feature of income distribution.

characterize growth. This is an interesting investigation as membership in a club (the identification of finite mixtures) does not imply that any concept of convergence necessarily applies. This approach also differs from that of Feyrer (2008) and Johnson (2005) who looked at the distribution dynamics of each of the Solow variables using the dynamic approach of Quah, which are treated as independent of the evolution of growth in this setup. Instead, we use mixture analysis to determine which countries, with high probability, are part of a basin of attraction and then investigate how the Solow variables for these countries behave.[3]

3. Convergence Measures and Initial Results

The traditional analysis of absolute β-convergence stems from a growth regression (Barro & Sala-i-Martin 1991) which in compact can be written

$$\frac{Y_T - Y_t}{T - t} = g_{T-t} = \alpha + \left(1 - e^{\beta t}\right) Y_t + \varepsilon, \tag{1}$$

where g_{T-t} is equal to the average growth rate of income per capita, β is the convergence rate, Y_t and Y_T are respectively the logarithm of initial and final income per capita (so that $T \geq t$) and ε is an error term with the usual properties of mean zero and variance σ^2. What is typically estimated is the term $1 - e^{\beta t}$, which, in the presence of convergence is negative. Conditional β convergence is similar in nature to absolute β-convergence except that additional determinants of economic growth, such as human capital and trade, are included in the regression. The presence of σ-convergence is instead determined via the inequality

$$\sigma_T^2(Y) - \sigma_t^2(Y) < 0. \tag{2}$$

When the cross-country dispersion is diminishing over time there is a tendency for the distribution to collapse. However, finding σ-convergence is not indicative of this behavior as it is more likely that while the variation is diminishing over time, it is approaching a nonzero limit. Finally, to account for churning within the distribution that cannot be captured by either β or σ measures we

[3]Our approach here is closely related to the work of Henderson & Zelenyuk (2007) who look for similar behavior in the efficiency of countries. Their classification was based on development status of the countries in their analysis rather than through econometric mixture techniques.

assess γ-convergence as[4]

$$\frac{\sigma^2\left[\rho(Y_{it})+\rho(Y_{iT})\right]}{\sigma^2(2\rho(Y_{it}))}<1, \tag{3}$$

where ρ is the absolute rank of per capita income across countries. The logic of the rank is that if the variance is the same there is no change in numerator and denominator so that the index is equal to one, otherwise one should experience movements in the ranking of distribution over time. This can be thought of as a measure of leapfrogging, so that if there is a decreasing variance together with a index far from 1, we have evidence of country mobility and reduced dispersion.

In order to check for our three measures of convergence we use a sample of 74 countries that appear yearly in the Penn World Table 6.3 from 1960 to 2000.[5] This small group of countries, for whom we may assess the role of variables such as human and physical capital accumulation, is much more homogeneous so joint measures of polarization may be more difficult to detect. Investment rates and workforce growth are taken from Penn World Table 6.3 as well, while human capital is constructed as years of schooling, adjusted to account for returns to schooling (Hall & Jones 1999), as in Feyrer (2008) and Henderson et al. (2008), given in (Psacharopoulos 1994).

In Table 1 we look for evidence of β, σ and γ-convergence across decades by using respectively the PWT logarithm of output for worker (RGDPWOK).[6] The data show that there is no evidence of absolute β-convergence, while presenting both σ-divergence and high rank correlation with the latter implying few changes in rankings. Divergence seems to be the norm. By including the traditional variables capturing physical and human capital investment and the growth rate of population we obtain the standard result that conditional β-convergence is present and around 1% for the period 1970-1990.

Table 1 about here

Table 2 highlights the same convergence measures for the MRW variables, where the change in each variable has been standardized to avoid scale problems. We see that our human capital proxy is strongly divergent in the first two decades and differences seem to persist over time whereas

[4]See Boyle & McCarthy (1997, 1999).

[5]The list of countries is provided in Table 7.

[6]Results are unchanged for RGDPCH and RGDPL. For the variance measure, here we follow the standard deviation of income per worker following Dalgaard & Vastrup (2001).

both the physical capital and TFP growth experienced stronger changes in the last thirty years and a more persistent behaviour. By looking at the traditional accumulation variables a different pattern emerges among human and physical capital, while differences in TFP per worker and per capita depend on working age population participation. Globally considered these results highlight a marked discontinuity that appeared during the eighties and some evidence of a minor change in the last decade.

Table 2 about here

Now the question we assess is that, given the increasing multimodality behaviour highlighted, for instance, by Henderson et al. (2008), are these measures meaningful? The logic of the question is given in the homogeneity assumption of a common growth pattern that lies behind equation (1), then if this homogeneity holds only for subsamples of countries should we should expect to have different results by looking across these subsets? Secondly, if there is multimodality linked to the presence of different clubs then we have to take into account the distinction among position changes inside the same cluster and outside it and γ-convergence is unable to distinguish this effect.

4. Econometric Mixture Analysis

4.1. **Univariate Mixture Modelling.** In order to check for convergence within groups we may simply choose *a priori* groups for example OECD vs Non-OECD, or we may endogenize group membership. By a statistical point of view we may suppose that if the distribution is not unimodal we could have more different subdistribution hidden in a Gaussian density. It implies that this could be a strong constraint because the mean could be not the real peak of the subdistributions but a value that for instance lies in the right tail of the poorer group and in the left tail of the richest one. A more general approach encompasses the standard one because if one finds that for instance a solution with two distributions gives parameters that are very close, it means that these are overlapping so one may revert to the standard one-segment approach. A way to avoid preliminary assumptions on the membership is through mixture densities, where the density function for observation i is given by

$$f_i(Y_{t,i}|\theta) = \sum_{s=1}^{k} \psi_s f_s(Y_{t,i}|\theta_s). \tag{4}$$

Here ψ_s is the probability to "stay in group s" for country i. By using a normal density and assuming a structure with S groups, we sum the conditional probabilities to obtain the unconditional density as

$$\ln \Theta = \ln \sum_{s=1}^{S} \psi_s (2\pi\sigma_s^2)^{-\frac{1}{2}} \exp\left\{ \frac{-\left(Y - \mu_{Y,s}\right)^2}{2\sigma_{Y_i,s}^2} \right\}, \tag{5}$$

where $(\mu_s, \sigma_s, \psi_s)$ are the specific segment parameters for the distribution of Y (given that the probabilities sum to one, we have that one of the ψ_i is a linear combination of the others, i.e., for $k = S$, we have that $\psi_S = 1 - \psi_1 - \psi_2 - \cdots - \psi_{S-1}$).

Two issues are related to the set of probabilities and to the number of segments: firstly, given that the probabilities are unknown a reliable solution with this problem is to treat the probabilities as missing data and maximize the solution through the EM (expectation-maximization) algorithm (Dempster, Laird & Rubin 1977). More specifically we do not have priors on the initial membership probabilities[7] so we need to select a starting set of values to operationalize this procedure.[8] We select as our solution the mixture probabilities which deliver the the highest value of the logarithm of the likelihood function. An additional complicating factor is the number of segments, which must be chosen *a priori* to operationalize the procedure as well. The finite mixture approach does not *ex poste* identify the number of components, so we have to choose *a priori* number of segments and then test for the number of segments that is most informative. The common approach to testing for the correct number of segments relies upon maximizing a likelihood criteria such as the Bayesian Information Criterion (BIC); see Fraley & Raftery (2002). We use the framework of (5) to reestimate the usual convergence measures as in Table 1, but we assess these measures inside clusters over time.

4.2. **Multivariate Mixture Modelling.** A further generalization of this idea of clustering is to focus instead on the joint distribution. This provides several interesting insights which classical mixture regression models may blur. First, given that no 'regression' error is present issues of endogeneity are not present in this framework. Moreover, issues of correct specification of the conditional mean (Durlauf et al. 2005, Maasoumi, Racine & Stengos 2007) do not occur. Second,

[7]If membership probabilities depended upon another set of variables we could model this through either a logit or a probit model. This is the so called concomitant variable approach, (see Grün & Leisch 2007).

[8]We repeat this process a large number of times (100 for our examples) to avoid locating in a local maxima.

as opposed to the work of Feyrer (2008) and Henderson et al. (2008), instead of searching for identical patterns in the factors of cross country output (human capital or total factor productivity say) with growth itself, we can now connect clustering directly by searching for groups of countries based on all of the variables simultaneously.

To describe multivariate cluster analysis, consider multivariate data $\mathbf{X}$, which in our example could be both GDP per capita and human capital. Now the density of each of the n p-variate observations, $\mathbf{X}_i$, is (Symons 1981):

$$f_i(\mathbf{X}_i|(\psi_k, \mu_k, \Sigma_k)) = \sum_{k=1}^{s} \psi_k N_p(\mathbf{X}_i|\mu_k, \Sigma_k), \tag{6}$$

where $N_p(\cdot)$ is the p-variate normal density given as

$$N_p(\mathbf{z}|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{-(1/2)(\mathbf{z}-\mu_k)'\Sigma_k^{-1}(\mathbf{z}-\mu_k)}.$$

The components or clusters in both these models are ellipsoidal, centered about their means.

The covariances Σ_k determine additional geometric features. Each covariance matrix is parameterized by an eigenvalue decomposition of the form

$$\Sigma_K = \lambda_K D_K A_K D_K^T, \tag{7}$$

where D_k is the orthogonal matrix of eigenvectors, A_k is a diagonal matrix whose elements are proportional to the eigenvalues of Σ_k, and λ_k is a scalar (Fraley & Raftery 2007). D_k gives the orientation the k_{th} cluster, while λ_k highlights the volume occupied by cluster and A_k is the shape. We may specify the clusters to have identical or different variance-covariance matrices, but an additional perspective which we may restrict are the volume and orientation of the clusters. A different orientation among the variables across clusters, as we see below in the results, may indicate that for different levels of the variables the relationships between them are either decreasing or increasing. In a growth context it may indicate a raw test for production function homogeneity. For instance, a bivariate mixture density could have a positive orientation for one cluster and a negative orientation in the other (see Banfield & Raftery 1993).

If we write the A_k=diag$\alpha_{1k}, \alpha_{2k}, ...\alpha_{pk}$ in decreasing order, with p equal to the number of dimensions, then the number of α_{ik} greater than 1 explain the number of relevant dimension for the cluster (i.e. with $\alpha_{2k} > 0$ the k_th-cluster is concentrated about a two-dimensional plane in p-space, Banfield & Raftery (1993)). More parsimonious specifications are available which allow equal covariances or equal volume, by setting each A_k or D_k equal across segments.

Two issues in finding clusters in a multivariate setting are the choice of criteria and the limitations of the computational solution (see Fraley & Raftery 2002). On the first issue, BIC had been found to be an adequate criteria; for example in Fraley & Raftery (1998)[9]. The second limitation has been discussed recently by Fraley & Raftery (2002) who highlight as a deficiency of multivariate clustering with high-dimensional data, the number of parameters per component in a mixture of multivariate normal densities grows as the square of the dimension of the data, so when the dimension of data is high relative to the total number of observations the variance-covariance matrices of the mixtures may be singular. This means that for clustering with a large number of preliminary variables, as in Desdoigts (1999), a dimension reduction strategy may be appropriate prior to cluster analysis.

5. Results

5.1. **Univariate Clustering Results.** Focusing on univariate clustering in the distributions of both GDP per worker and the traditional determinants encompassing the Solow model, our results in Table 1 reveal high levels of clustering across all variables. The number of clusters is determined via the BIC. The number of clusters in GDP per worker, according to the BIC, is two for the period 1960-1970 and three for 1980-2007. The movement from two groups to three is consistent with the previous literature[10] and also is indicative of the movement to the upper end of the income distribution by the Asian tigers. These results also highlight the growing divergence in per capita incomes, with the difference in the means of the high and low groups expanding from just under 14,000 in 1960 to almost 50,000 in 2000.

Table 3 about here

Focusing our attention on the traditional Solow determinants we notice in Table 1 both σ and γ convergence do not appear to occur. Human capital shows a quite stable picture, always implying two groups whose means are roughly the same distance apart overtime. TFP, while showing two

[9] In a mixture regression context, Hawkins, Allen & Stromberg (2001) found similar results.

[10] This result is consistent with the global three components solution found in Pittau et al. (2010), who used a larger sample than that employed here.

groups in 1960, most likely due to the presence of outliers, bifurcated at some point in the mid to later 1980s.[11] Physical capital displays the most divergent behavior, progressing from two groups in 1960 to 4 in 2000[12]. Looking towards the disparity in mean between the highest capital accumulation cluster and the lowest we see that the difference is increasing over the decades, moving from just under 4500 to well over 15000 by 2000. A consistent theme with the description of the global economy presented in Table 1 is that the upper segments, for all the variables, contain the majority of the OECD countries. Moreover, we also see that no variable displays a similar pattern with the clustering of GDP per worker. Thus, it is likely that the evolution of these variables are not directly causing the behavior we observe and most likely feedback effects are present.

Before considering our multivariate clustering results we more carefully assess transition and convergence results within and across clusters. These results are summarized in Tables 4 and 5. If we focus on income convergence across the decades for the clusters analyzed via our econometric procedures, we notice a time dynamic for polarization indicating the group means moving in opposite directions. The second result is the asymmetric behaviour of convergence within the clusters: our measures of convergence show that countries in the 'richer' group appear to be converging whereas countries in the 'poorer' group do not display any type of convergent behavior. In the period from the 1980s there is the wave of polarization and the favoured solution is now a distribution composed of a mixture of three gaussian densities.[13]

Table 4 about here

Once again all of the convergence measures show that in 1990 convergence inside groups is confirmed, while the aggregate measures highlight a greater spread among countries. In this way, the divergence highlighted in table 1 from 1990 to 2000, veils a variegated situation that is convergence or stability for the 'poorest' and 'richest' segments and divergence within the 'middle' segment. The scope of our results show that it is misleading to look at aggregate indexes of convergence without taking into account the fact that the distributions are changing over time and that they are comprised of subsegments which draws into question the homogeneity of the sample. This lends

[11]This is consistent with both Feyrer (2008) and Henderson et al. (2008).

[12]Also in the decomposition approach of Beaudry, Collard & Green (2005) physical capital accumulation is found to be a critical factor in changes of the income distribution between 1980 and 2000.

[13]Pittau et al. (2010) also find evidence supporting a distribution composed of three groups for this period as well.

further support for the use of mixture analysis prior to assessing convergence across any group of countries.

Table 5 about here

Table **??** shows the complete polarization from 1960 to 2000, by using the cluster membership of 2000. In this way we are assuming that polarization has taken place as a continuous pattern where each country belonging to a cluster began this process in 1960 and ended in 2000 (we used percentage changes in variance so to compare effects for all the variables). It is equivalent to say that polarization ended in 2000, but countries in 1960 already belonged to the groups they are in 2000. In this way σ-convergence inside groups shows the subdistributions becoming less dispersed over 1960-2000, and it is a direct measure of polarization. This style of analysis is similar to that of Feyrer (2008) and Henderson et al. (2008) by focusing attention not only on the evolution of income per capita over time but also on the factors which are believed to influence it.

Here, instead of direct assessment of these factors as in Henderson et al. (2008), we focus attention on the behavior of each of these variables using our club assignments dictated from our mixture density estimates described earlier. In addition, instead of focusing on modality of the distribution of human capital, we focus on the classical measures of convergence for human capital for the countries of each cluster. Tests of modality are quite limited in explaining the behavior of one variable in response to another. This is because the univariate tests do not recognize the location of a variable within a density. Here, since we are *a priori* placing our canonical variables in a specific group, our convergence measures afford us more information.

Table 6 about here

5.2. **Multivariate clustering results.** While our univariate results suggest that the distribution of output per worker and the corresponding Solow determinants are all indicative of segmented processes, the impact that one has the other cannot be understood in this type of analysis. To more clearly grasp the evolution of the 'Solow' model over time we employ bivariate and multivariate mixture models to determine if distinct clusters emerge across all (or some) variables considered. While presenting results for multivariate cluster is difficult when considering more than two variables simultaneously, we can provide two-fold plots that show how any set of variables are clustered together. An example of this is provided in Figure 1.

Figure 1 about here

We can see that the TFP/GDP distribution is composed of many segments in 1960 and by 2000 has settled into a very distinct, two cluster distribution. We see a high income, high TFP distribution and a low income, low TFP distribution. This split is consistent with both the insights of Zeira (2007) and Galor & Mountford (2006). The polarization within the TFP/GDP distribution is by no means unique. Figures 2 and 3 present similar bivariate clustering results for GDP and human capital accumulation and GDP and capital accumulation, respectively. Again, we see the same features we saw in the univariate results, over the 1960 to 2000 period there is an evolution of the joint distribution to a bimodal shape composed of a high income-high capital group and a low income-low capital group. The theoretical model of Galor & Mountford (2006) which suggests that gains from trade are channelled towards population growth seem to resonate here. Almost uniformly, the countries in each of the low and high clusters are the same across Figures 1 through 3, suggesting that population expansion negates any opportunities for sustained growth via reductions in per capita capital stocks and total factor productivity.

Figures 2 and 3 about here

To further illustrate the insights provided by multivariate clustering and the robustness of the polarization of joint distribution of GDP per worker and the Solow determinants, Figure 4 plots bivariate results based on a full cluster analysis for the four-variate distribution. We see that progressing from 1960 to 2000 there is polarization between GDP and each determinant, which confirms both the univariate and the bivariate results presented earlier. These results suggest a world of haves and have-nots not just in income per capita (or worker) but also in education, total factor productivity and physical capital accumulation.

In addition to this we may observe at least two features arising from multivariate clustering, apart small groups comprising outliers: the first one is that distance among clusters is growing over time, especially after 1980, while bivariate relationship for both the 1960s and 1970s are partially overlapping among clusters. The second one is that if we interpret these relationships as depicting production functions based on country characteristics we see that while is usually steeper for the lower clusters in figure 3 that is GDP versus physical capital per capita, the opposite happens in figure 1 that is GDP versus human capital. Figure 4 summarizes this time pattern and this asymmetric productive factors' behaviour (note that axis are inverted).

We notice several features that are missing in the univariate clustering results presented earlier. First, we see that our clusters are diverging not only over income but also over our three additional variables. This suggest something more complex than just a great income divergence. It points to something deeper that is causing an entire shift in the main resources that promote output. Furthermore, if we look at the orientation of the clusters over time we notice that the slope of human capital and total factor productivity (with respect to output per capita) for the 'rich' clusters are steeper than the 'poor' clusters, suggesting greater returns within the production function. This hints that achieving a developed economy is quite difficult and is not as simple as increasing human or physical capital, but deep structural changes of an entire economy must be undertaken to parlay those increase in capital stocks into long term sustained output.

One interpretation of this result is that for less developed countries returns from either TFP of human capital accumulation do not materialize to the same degree than physical capital accumulation does but that without contributions from TFP and human capital, an economy cannot grow the way that it does once these returns from TFP and human capital are realized. Further work in this direction could be linked to an interpretation of the determinants of the spread among clusters.

Figure 4 about here

6. Conclusions

This work aimed to make a joint assessment of various measures of convergence, also taking into account the fact that the income distribution became polarized over the later half of the 20th century. The apparent divergence hides local phenomena that is consistent with 'within-group' or 'club' convergence. Evidence provided here shows that beginning from a near homogeneous distribution there is the presence of bimodality until 1980, when the rich group bifurcated into an intermediate and very rich group of countries. The number of transitions among clusters is greater in the period 1960-1980 (which has two segments) and lower in the 1980-2000 period (which has three segments). An interpretation of this phenomenon is that countries exchange their relative position among clusters until polarization has not emerged, then the situation is very stable.

It is interesting to note that γ-convergence depicts the opposite situation with more changes in the second period, while these changes are "local" ones, that is, inside the same group. Changes of clusters where countries are getting to be closer is much more relevant than simple changes in the ranks of the distribution, when there is no reason to think that the sample observations obey the same probability law. Considered globally, the appearance of divergence is due to the subsequent strong polarization of the world income distribution after the second world war coupled with clubs of countries behaving in a similar fashion with respect to the classic production inputs. Standard convergence tools *within* endogenously determined clusters highlight a quick convergence and the traditional MRW variables work in a different way inside each group, with for instance the human capital proxy having an important impact for the richest countries, but proving to have less of an impetus for the group of poor countries. This is consistent with the finding of Maasoumi et al. (2007) that human capital mattered for OECD nations but not for those countries outside of the OECD. Finally, we found that also multivariate clusters are similar to univariate ones, but from them we have insights about the relationships among growth determinants and income that highlight the relative roles in different groups of countries.

References

Alfo, M., Trovato, G. & Waldmann, R. (2008), 'Testing for country heterogeneity in growth models using a finite mixture approach', *Journal of Applied Econometrics* **23**, 487–514.

Azariadis, C. & Drazen, A. (1990), 'Threshold externalities and economic development', *Quarterly Journal of Economics* **105**, 501–526.

Banfield, J. & Raftery, A. (1993), 'Model-based gaussian and non-gaussian clustering', *Biometrics* **49**, 803–821.

Barro, R. J. (1991), 'Economic growth in a cross section of countries', *Quarterly Journal of Economics* **106**(2), 407–443.

Barro, R. J. & Sala-i-Martin, X. (1991), 'Convergence across states and regions', *Brookings Papers on Economic Activity* **1**, 107–158.

Battisti, M. & Di Vaio, G. (2008), 'A spatially-filtered mixture of β-convergence regressions for EU regions', *Empirical Economics* **34**(1), 105–121.

Beaudry, P., Collard, F. & Green, D. A. (2005), 'Changes in the world distribution of output per worker, 19601998: How a standard decomposition tells an unorthodox story.', *Review of Economics and Statistics* **87**(4), 741–753.

Bianchi, M. (1997), 'A contribution to the empirics of economic growth.', *Journal of Applied Econometrics* **12**(4), 393–409.

Bourguignon, F. & Morrisson, C. (2002), 'Inequality among world citizens 1820-1992', *American Economic Review* .

Boyle, G. E. & McCarthy, T. G. (1997), 'A simple measure of β-convergence', *Oxford Bulletin of Economics and Statistics* **59**(2), 257–264.

Boyle, G. E. & McCarthy, T. G. (1999), 'Simple measures of convergence in per capita gdp: a note on some further international evidence', *Applied Economics Letters* **6**, 343–347.

Brock, W. A. & Durlauf, S. N. (2001), 'Growth empirics and reality', *The World Bank Economic Review* **15**(2), 229–272.

Canova, F. (2004), 'Testing for convergence clubs in income per capita: a predictive density approach.', *International Economic Review* **45**, 49–77.

Dalgaard, C. & Vastrup, J. (2001), 'On the measurement of σ-convergence.', *Economics Letters* **70**, 283–287.

Dempster, A., Laird, N. M. & Rubin, D. (1977), 'Maximum likelihood from incomplete data via the em algorithm.', *Journal of the Royal Statistical Society, B* **39**, 1–38.

Desdoigts, A. (1999), 'Patterns of economic development and the formation of clubs.', *Journal of Economic Growth* **4**, 305330.

Durlauf, S. N. & Johnson, P. A. (1995), 'Multiple regimes and cross-country growth behavior', *Journal of Applied Econometrics* **10**.

Durlauf, S. N., Johnson, P. & Temple, J. (2005), Growth econometrics, *in* P. Aghion & S. N. Durlauf, eds, 'Handbook of Economic Growth', North Holland, Amsterdam.

Feyrer, J. (2008), 'Convergence by parts', *The B.E. Journal of Macroeconomics* **8**(1).

Fraley, C. & Raftery, A. (2002), 'Model-based clustering, discriminant analysis, and density estimation.', *Journal of the American Statistical Association* **97**, 611–631.

Fraley, C. & Raftery, A. (2007), 'Mclust version 3 for r: Normal mixture modelling and model-based clustering', *Technical Report Department of Statistics, University of Washington* (504).

Fraley, C. & Raftery, A. E. (1998), 'How many clusters? Which clustering method? Answers via model-based cluster analysis', *The Computer Journal* **41**, 578–588.

Galor, O. (1996), 'Convergence? Inferences from theoretical models', *Economic Journal* **106**, 1056–1069.

Galor, O. (2007), 'Multiple growth regimes - insights from unified growth theory.', *Journal of Macroeconomics* **29**, 470–475.

Galor, O. & Mountford, A. (2006), 'Trade and the great divergence: The family connection', *American Economic Review* **96**, 229–303.

Grün, B. & Leisch, F. (2007), 'Fitting finite mixtures of generalized linear regressions in r.', *Computational Statistics and Data Analysis* **51**(11), 5247–5252.

Hall, R. & Jones, C. (1999), 'Why do some countries produce so much more output per workers than others?', *Quarterly Journal of Economics* **114**(1).

Hawkins, D., Allen, D. M. & Stromberg, A. (2001), 'Determining the number of components in mixtures of linear models.', *Computational Statistics and Data Analysis* **38**(1), 323–341.

Henderson, D. J. & Zelenyuk, V. (2007), 'Testing for (efficiency) catching-up', *Southern Economic Journal* **73**(4), 1003–1019.

Henderson, D., Parmeter, C. F. & Russell, R. R. (2008), 'Convergence clubs: Evidence from calibrated modality tests', *Journal of Applied Econometrics* **23**, 607–638.

Johnson, P. (2005), 'A continuous state space approach to 'convergence by parts'', *Economics Letters* **86**, 317–321.

Maasoumi, E., Racine, J. S. & Stengos, T. (2007), 'Growth and convergence: A profile of distribution dynamics and mobility', *Journal of Econometrics* **127**(2), 483–508.

McLachlan, G. J. & Peel, D. (2000), *Finite Mixture Models*, John Wiley, New York.

Owen, A., Videras, J. & Davis, L. (2009), 'Do all countries follow the same growth process?', *Journal of Economic Growth* **14**(4), 265286.

Paap, R., Franses, P. H. & Van Djik, D. (2005), 'Does africa grow slower than asia, latin america and the middle east? evidence from a new data-based classification method', *Journal of Development Economics* **77**(2), 553–570.

Paap, R. & Van Dijk, H. K. (1998), 'Distribution and mobility of wealth of nations', *European Economic Review* **42**, 1269–1293.

Pittau, M. G., Zelli, R. & Johnson, P. A. (2010), 'Mixture models and convergence clubs', *Review of Income and Wealth* **56**(1), 102–122.

Psacharopoulos, G. (1994), 'Returns to investment in education: a global update', *World Development* **22**(9).

Quah, D. T. (1993*a*), 'Empirical cross-section dynamics in economic growth', *European Economic Review* **37**, 426–434.

Quah, D. T. (1993*b*), 'Galton's fallacy and tests of the convergence hypothesis', *The Scandinavian Journal of Economics* **95**, 427–443.

Quah, D. T. (1996*a*), 'Empirics for economic growth and convergence', *European Economic Review* **40**, 1353–1375.

Quah, D. T. (1996*b*), 'Twin peaks: Growth and convergence in models of distribution dynamics', *Economic Journal* **106**, 1045–1055.

Symons, M. (1981), 'Clustering criteria and multivariate normal mixtures', *Biometrics* **37**, 35–81.

Zeira, J. (2007), 'Wage inequality, technology, and trade', *Journal of Economic Theory* **137**(1).

Table 1. Convergence Measures Across Decades.

Decade	β	MRW-β	$\Delta\sigma$	γ
60s	0.003	-0.008***	0.117	0.971
70s	-0.003	-0.013***	0.008	0.967
80s	0.001	-0.009***	0.067	0.972
90s	0.005	-0.002	0.135	0.977
00s	-0.001	-0.013***	0.013	0.985

*,** ,*** indicate significant values respectively at 90, 95, 99%.

Table 2. Convergence in conditioning variables over decades.

Variable		1960 vs 1970	1970 vs 1980	1980 vs 1990	1990 vs 2000
Human Capital					
	γ	0.976	0.977	0.978	0.983
	$\Delta\sigma\%$	0.146	0.167	0.008	0.031
Capital per worker					
	γ	0.937	0.931	0.936	0.967
	$\Delta\sigma\%$	-0.123	-0.038	0.013	0.118
Capital per capita					
	γ	0.947	0.942	0.948	0.963
	$\Delta\sigma\%$	-0.089	-0.013	0.087	0.113
TFP per worker					
	γ	0.848	0.840	0.939	0.927
	$\Delta\sigma\%$	0.107	-0.065	0.068	0.114
TFP per capita					
	γ	0.866	0.858	0.963	0.953
	$\Delta\sigma\%$	0.154	0.131	0.182	0.124

Table 3. GDP and growth variables clusters composition over time

	GDP	Human Capital	Physical Capital	TFP
1960	2	2	2	2
Number of countries	30-44	28-46	33-41	72-2
Cluster means	5041-18951	1.24-1.92	569-4722	16947-57536
1970	2	2	3	1
Number of countries	30-44	54-20	16-26-32	74
Cluster means	6103-28098	1.48-2.45	273-1584-8324	22401
1980	3	2	3	1
Number of countries	15-29-30	50-24	14-23-37	74
Cluster means	3911-14935-40573	1.60-2.58	1198-8778	23724
1990	3	2	4	2
Number of countries	34-17-23	48-28	13-23-11-27	40-34
Cluster means	12624-48626	1.71-2.69	333-1682-4335-12122	14460-35433
2000	3	2	4	2
Number of countries	40-12-22	38-36	11-20-17-26	41-33
Cluster means	10297-32135-61341	1.78-2.72	241-1438-4383-15507	14456-38382

Table 4. Convergence and transitions among clusters 1960-80.

	1960-70		1970-80	
	1^{st} Cluster	2^{nd} Cluster	1^{st} Cluster	2^{nd} Cluster
β	-0.008	-0.017**	-0.007	-0.037***
MRW-β	-0.009*	-0.028***	-0.016*	-0.038***
$\Delta\sigma^2$	-0.03	-0.07	+0.004	-0.42
γ	0.94***	0.88***	0.87***	0.82***
Mean Log GDP	8.54	10.21	8.80	10.37
Number obs.	33	41	44	30
Transitions	1	1	12	12

*,** ,*** Indicate significant values respectively at 90, 95, 99%.

Table 5. Convergence and transitions among clusters 1980-00.

	1980-90			1990-00		
	1^{st} Cluster	2^{nd} Cluster	3^{rd} Cluster	1^{st} Cluster	2^{nd} Cluster	3^{rd} Cluster
β	-0.01*	-0.056***	-0.054***	-0.005	-0.047*	-0.039**
MRW-β	-0.011	-0.064***	-0.050***	-0.004	-0.063***	-0.040*
$\Delta\sigma^2$	-0.010	-0.52	-0.49	-0.03	-0.14	+0.05
γ	0.90***	0.67***	0.67***	0.93***	0.60**	0.61***
Mean Log GDP	8.83	10.07	10.81	9.00	10.37	11.02
Number obs.	35	16	23	40	12	22
Net transitions	+19	-17	-7	+5	-4	-1

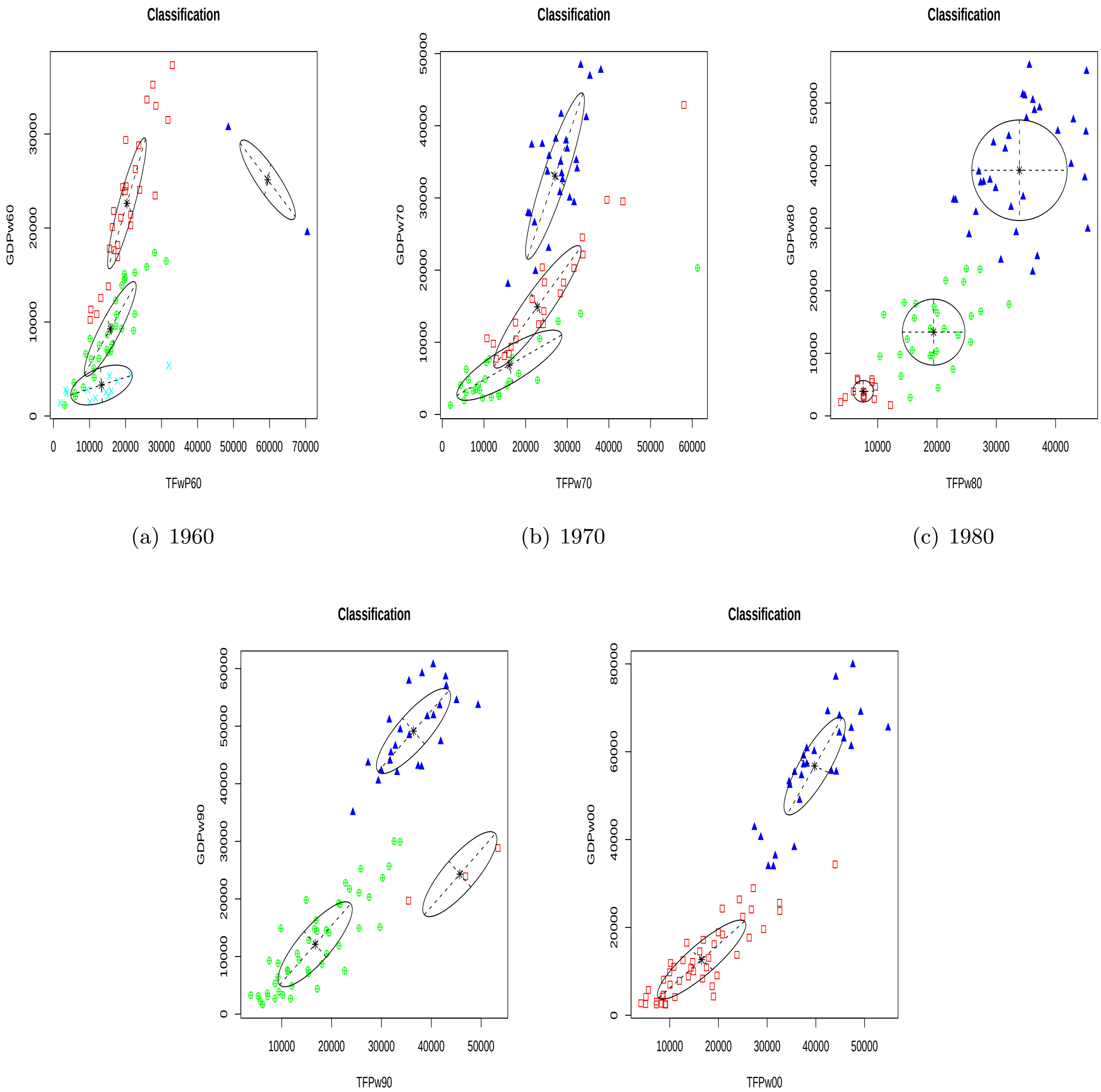
*,** ,*** Indicate significant values respectively at 90, 95, 99%.

Table 6. Convergence inside clusters 1960-2000

	1^{st} Cluster	2^{nd} Cluster	3^{rd} Cluster		1^{st} Cluster	2^{nd} Cluster	3^{rd} Cluster
β	-0.008**	-0.025***	-0.023***	MRW-β	-0.009**	-0.028***	-0.021***
$\Delta\sigma^2_{GDP}$	-0.09	-0.58	-0.88	γ_{GDP}	0.74***	0.09	0.17
$\Delta\sigma^2_{HumanCapital}$	1.81	-0.20	-0.38	$\gamma_{HumanCapital}$	0.78***	0.74***	0.69***
$\Delta\sigma^2_{PhysicalCapital}$	0.36	0.63	1.82	$\gamma_{PhysicalCapital}$	0.73***	0.23	0.10
$\Delta\sigma^2_{TFP}$	-0.64	-0.60	-0.19	γ_{TFP}	0.55***	-0.27	0.17

*,** ,*** Indicate significant values at the 90, 95, and 99% level, respectively.

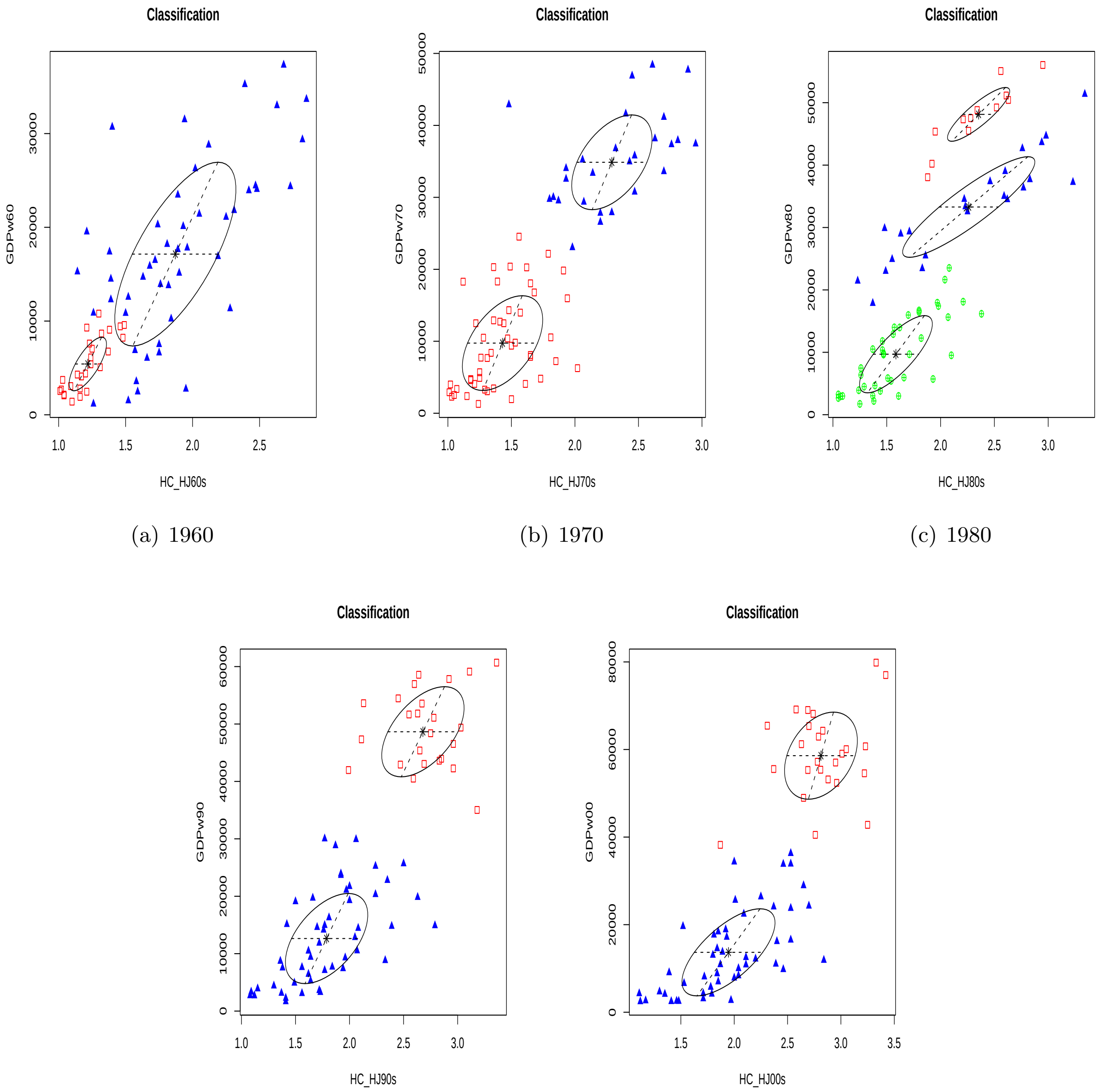
FIGURE 1. **Bivariate mixture analysis for GDP per worker and Total Factor Productivity.**



(a) 1960 (b) 1970 (c) 1980

(d) 1990 (e) 2000

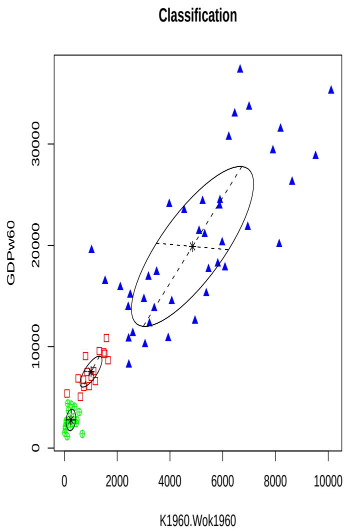
FIGURE 2. **Bivariate mixture analysis for GDP per worker and Human Capital Accumulation.**



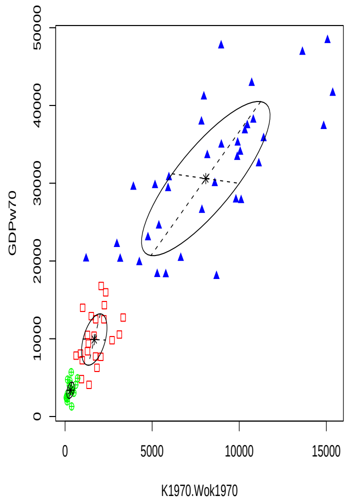
(a) 1960 (b) 1970 (c) 1980

(d) 1990 (e) 2000

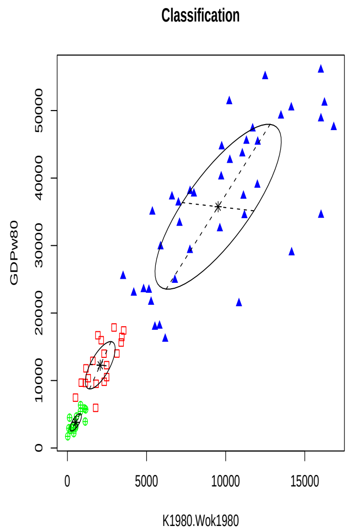
FIGURE 3. **Bivariate mixture analysis for GDP per worker and Capital Accumulation.**



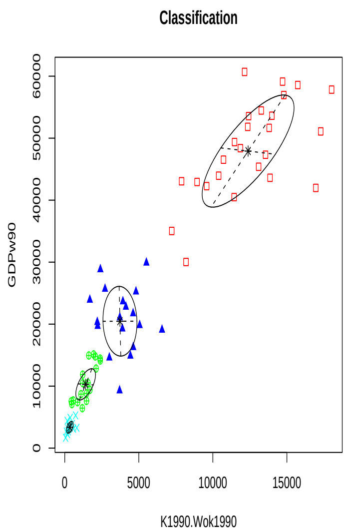
(a) 1960



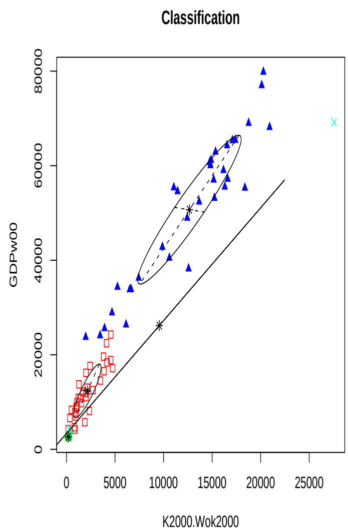
(b) 1970



(c) 1980

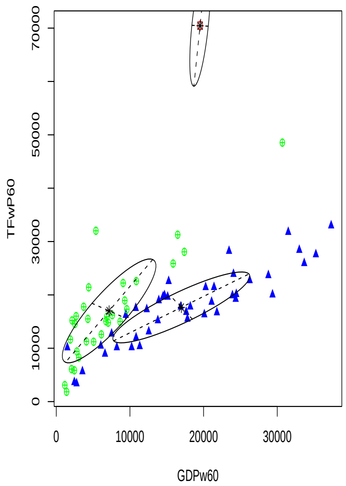


(d) 1990

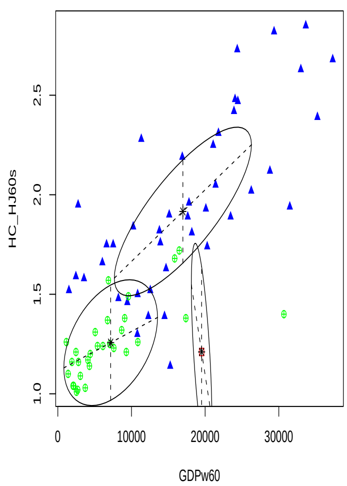


(e) 2000

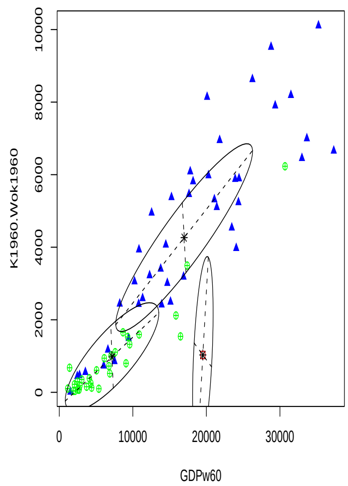
Figure 4. **Multivariate mixture analysis for GDP per worker and Solow Determinants.**



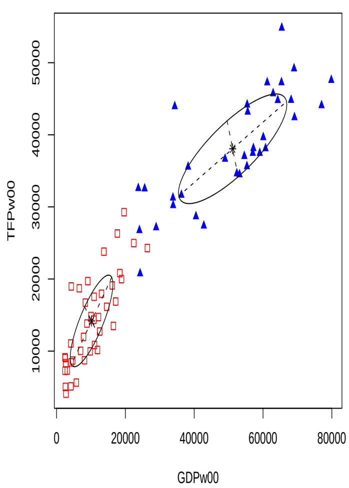
(a) 1960, TFP & GDP



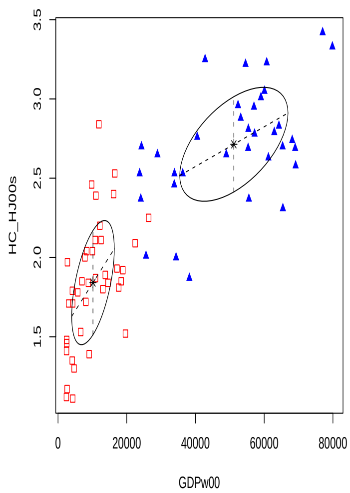
(b) 1960, HC & GDP



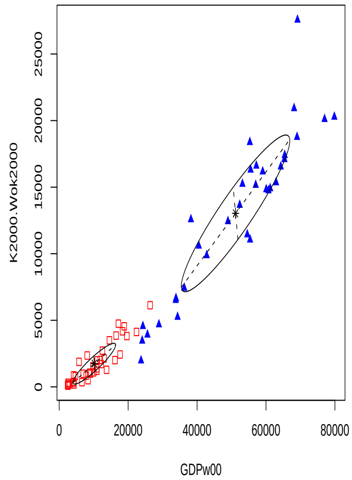
(c) 1960, K & GDP



(d) 2000, TFP & GDP



(e) 2000, HC & GDP



(f) 2000, K & GDP

Table 7. Country names and codes.

Argentina	ARG	Australia	AUS	Austria	AUT
Belgium	BEL	Bolivia	BOL	Brazil	BRA
Barbados	BRB	Canada	CAN	Switzerland	CHE
Chile	CHL	Cameroon	CMR	Colombia	COL
Costa Rica	CRI	Denmark	DNK	Dominican Republic	DOM
Algeria	DZA	Ecuador	ECU	Spain	ESP
Finland	FIN	France	FRA	Ghana	GHA
Greece	GRC	Guatemala	GTM	Hong Kong	HKG
Honduras	HND	Indonesia	IDN	India	IND
Ireland	IRL	Iran	IRN	Iceland	ISL
Israel	ISR	Italy	ITA	Jamaica	JAM
Jordan	JOR	Japan	JPN	Kenya	KEN
Sri Lanka	LKA	Lesotho	LSO	Mexico	MEX
Mali	MLI	Mozambique	MOZ	Mauritius	MUS
Malawi	MWI	Malaysia	MYS	Niger	NER
Nicaragua	NIC	Netherlands	NLD	Norway	NOR
Nepal	NPL	New Zealand	NZL	Pakistan	PAK
Panama	PAN	Peru	PER	Philippines	PHL
Portugal	PRT	Paraguay	PRY	Romania	ROM
Senegal	SEN	Singapore	SGP	El Salvador	SLV
Sweden	SWE	Syria	SYR	Togo	TGO
Thailand	THA	Trinidad & Tobago	TTO	Turkey	TUR
Uganda	UGA	Uruguay	URY	United Kingdom	GBR
United States	USA	Venezuela	VEN	South Africa	ZAF
Zambia	ZMB	Zimbabwe	ZWE		