University of Toronto
Department of Economics

Working Paper 421

Partially Identified Poverty Status: A New Approach to Measuring Poverty and the Progress of the Poor.

By Gordon Anderson, Maria Grazia Pittau and Roberto Zelli

January 20, 2011

Partially Identified Poverty Status: A New Approach to Measuring Poverty and the Progress of the Poor.

Gordon Anderson, *
Maria Grazia Pittau and Roberto Zelli †

November 4, 2010

Abstract

Poverty measurement and the analysis of the progress (or otherwise) of the poor is beset with difficulties and controversies surrounding the definition of a poverty line or frontier. Here, using ideas from the partial identification literature and mixture models, a new approach to poverty measurement is proposed which avoids specifying a frontier, the price is that an agent's poverty status is only partially identified. Invoking variants of Gibrat's law to give structure to the distribution of outcomes for homogeneous subgroups of a population within the context of a finite mixture model of societal outcomes facilitates calculation of the probability of an agent's poverty status. From this it is straightforward to calculate all the usual poverty measures as well as other characteristics of the poor and non poor subgroups in a society. These ideas are exemplified in a study of 47 countries in Africa over the recent quarter century which reveals among other things a growing poverty rate and a growing disparity between poor and non poor groups not identified by conventional methods.

Keywords: Poverty Frontiers, Mixture Models, Gibrat's law, Partial Identification
JEL classification: C14; I32; O1.

*Department of Economics, University of Toronto, email: anderson@chass.utoronto.ca
†Department of Statistical Science, Sapienza University of Rome,
email: grazia.pittau@uniroma1.it, roberto.zelli@uniroma1.it.

1 Introduction

Conventional poverty analysis identifies the poor by defining a poverty frontier in terms of a variable or variables that reflect individual wellbeing so that individuals, households or countries below the frontier are poor with probability one. Societal poverty is then measured using some aggregating function of agent distances below the poverty frontier in terms of those variables. The definition of the frontier has long been a matter of considerable debate across a spectrum of literatures[1] and debates about the progress of the poor have largely been about how the frontier is defined (Rogers and Rogers, 1993; Sala-i-Martin, 2006; Slesnick, 1993), indeed this has been extended to the notion of defining the poverty frontier as the boundary of a fuzzy set (Betti *et al.*, 2004). Furthermore, the debate has become more extensive as the dimensions in which wellbeing, and consequently the poverty frontier, has been measured increased[2]. Defining wellbeing in terms of what are often fundamentally unobservable concepts of functionings and capabilities (see Sen, 1992 and chapters in Grusky and Kanbur, 2006) has complicated the situation even further. Atkinson (1987) and Duclos *et al.* (2006) afforded some relief by developing stochastic dominance techniques which could establish changes in poverty for any definition of the poverty frontier within a region. Unfortunately these techniques do not appear to have been taken on to a great extent in the literature (or by practitioners) partly because the approach provides only a partial ordering and thus precludes the comparison of all states which a poverty index facilitates. Also policy makers like a "number" to hang their hat on which dominance techniques do not provide. Here indices for analyzing the plight of the poor are developed without resort to a contentious, sometimes unobservable or hard to calculate poverty frontier by a quite different approach to identifying poverty status. As indices however they do represent a complete ordering of poverty states and will thus be of use to analysts, practitioner's and policy makers alike. The approach is based on the presumption that a society is a collection of subgroups of agents where each subgroup is governed by distinct sets of circumstances making some groups follow "poor" paths

[1]See Sen (1983) versus Townsend (1985) for examples in a relative versus absolute debate. Citro and Michael (1995).

[2]See Bouguignon and Chackravarty (2003) for the union or intersection of sets in the multi-dimensional case.

and some follow "non-poor" paths. The problem is that these circumstances are not directly observed for each individual, all that is observed is some economic outcome (their income or consumption level for example). Thus the size distribution of the economic outcome in the population is a mixture of the corresponding size distributions of the subgroups. When these paths (subgroup distributions) overlap the group status of an individual is not directly discernable but resort can be made to concepts in the partial identification literature (Manski, 2003) to estimate the "chance" that an individual could belong to a particular group which in turn will permit more standard measures of the status of the poor, such as poverty rates or depth of poverty measures to be calculated.

Partial identification requires additional information in the form of theoretical restrictions or additional data in order to calculate the probability that an agent is a member of a particular group, here the predictions of some elementary stochastic process theory are employed. Their use can be justified as follows. The functionings and capabilities approach (Sen, 1992) defines deprivation (poorness) in terms of the boundaries set by the inherent circumstances of an agent (health, education, location, freedoms, genetic endowments etc). This is undoubtedly appropriate conceptually but frequently these boundaries and the circumstance characteristics are fundamentally unobservable. However if the poor are characterized by a particularly limited set of such circumstances (which the non poor are not), poor and non-poor observed economic behaviors (consumption or income for example) will follow distinct stochastic processes which generate distinct behavior distributions, and any population distribution will be a mixture of these. Poverty measurement then becomes a matter of identifying and measuring aspects of the subdistributions driven by these processes.

Perhaps the best known relationship between a process and the distribution it engenders is Gibrat's law (Gibrat, 1930; 1931) [3] recently employed in an application to individual consumption and income patterns in Battistin *et al.* (2009). This very powerful law, which is essentially a strong form of a statistical central limit theorem, tells us that a starting value that is subjected to a sequence of independent proportionate shocks over time will ultimately have a log normal size distribution, regardless of the

[3] See Sutton (1997) for a comprehensive discussion of the law and Kalecki (1945) for a modification of the law pertinent to the application herein.

nature of the distribution from which the shocks are drawn. If the measured wellbeing of a group of agents is governed by such a process, the starting value and the average shock (essentially the growth rate) will be determined by the groups' circumstances (essentially the limitations in their functionings and capabilities). Such processes will differ between poor and rich groups rendering different log normal distributions for the different groups. At a particular point in time the observed wellbeing size distribution in the population will be a mixture of these sub distributions with the mixing coefficient on the poor distribution corresponding to the poverty rate.

Articulated this way poverty measurement is about estimating various aspects of the subdistributions in the mixture together with their mixing coefficients. The price paid is that individual poverty status is no longer identified with probability one, all that can be done is to attach a probability to (or partially identify in the sense of Manski, 2003) individual poverty status[4]. The partial identification issue arises when the estimated poor and non poor distributions overlap and some agents cannot be unequivocally attributed to either the poor or non poor distributions. If the non-poor distribution had a well defined lower bound those whose incomes are below this bound can be definitively identified as poor. Those with incomes above the lowest non-poor income and below the highest poor income can only be associated with a probability of being poor. Those whose incomes are above the highest poor income can be definitively identified as non-poor. The middle grouping, where the distributions overlap, have an only partially identified status in that only the probability of being poor can be calculated. However as will be seen this does not inhibit estimating various aspects of the status of the poor (or other groups for that matter).

In what follows Section 2 develops the probability of being poor and how it can be applied in calculating various poverty measures. Section 3 outlines the versions of Gibrat's law to be employed in specifying the mixture distribution and the technique for estimating the mixtures is explained in Section 4. Section 5 reports an illustrative application to African countries over the period 1985-2008 and Section 6 offers some conclusions.

[4]Note that an advantage of this technique is that it can readily handle poverty measurement couched in multivariate contexts without resort to defining a contentious poverty frontier–a subject for future work.

2 The Probability of Being Poor.

For expositional convenience suppose there are two groups in society each with distinct limitations on their functionings and capabilities which govern the processes under which an observable characteristic x proceeds. These two processes result in a poor wellbeing distribution $f_p(x)$ of the observable characteristic x (again for expositional convenience the analysis will be performed in terms of a univariate distribution but it should be stressed that the analysis can be readily performed in a multivariate environment), and a rich wellbeing distribution $f_r(x)$, the unobservable proportions of agents under these distributions are w_p and $1 - w_p$ respectively so that the observable size distribution of incomes in this society is:

$$f(x) = w_p f_p(x) + (1 - w_p) f_r(x) \tag{1}$$

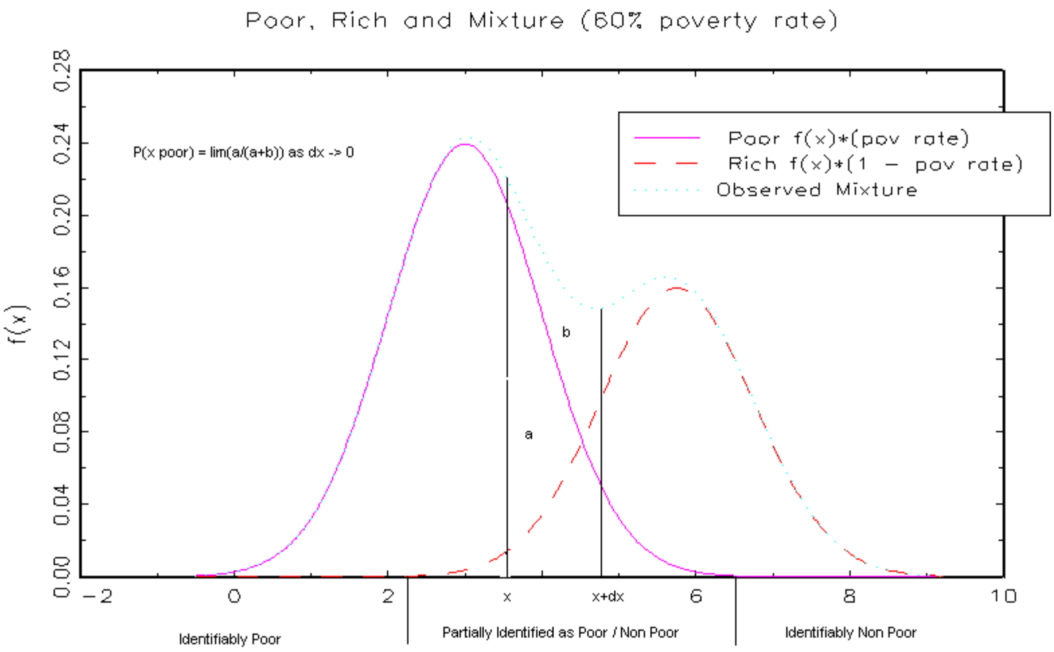
For understanding the plight of the poor the components of interest to be estimated are w_p (as the proportion of people governed by the poor process it can be viewed as the real poverty rate) and the nature of the distribution of incomes among the poor, $f_p(x)$ and the distributions of incomes among the non-poor $f_r(x)$. The mean of $f_p(x)$ yields the average incomes of the poor, its variance gives us a measure of inequality amongst the poor and the distribution itself will permit generation of indices akin to FGT2 and FGT3 indices (Foster *et al.*, 1984) for particular choices of a poverty reference point. The diagram in Figure (1) illustrating equation (1) highlights the identification problem.

For x's in the partially identified range all that can be established is $P(x)$, the probability that someone with x is poor since:

$$P(x) = lim_{dx \longrightarrow 0}(a/(a + b)) = w f_p(x)/(w f_p(x) + (1 - w) f_r(x)) \tag{2}$$

Given estimates of the sub distributions and the weights, this can be estimated for each agent. When the partially identified range is empty we have complete identification which Yitzhaki (1994) refers to this as perfect stratification, i.e. no overlapping between the two groups. When the overlap is complete and the two subdistributions have common modes even partial identification becomes difficult if not impossible. It

Figure 1: Diagram representing a society characterized by groups: poor and rich sub-groups.



follows that one minus a measure of the overlap (Anderson *et al.*, 2009; 2009a) provides an indicator of the extent of identification. In the present context:

$$1 - \frac{1}{w} \int \min\left(w f_p(x), (1-w) f_r(x)\right)\, dx \tag{3}$$

provides an index between 0,1 of the extent to which identification prevails.

One attraction of $P(x_i)$ is that it affords us a poverty ranking of agent i, just as $1 - P(x_i)$ yields a wellbeing ranking, this is especially useful when the analysis is multivariate in nature, i.e. x is a vector of attributes. Note also $E(P(x)) = \int P(x) f(x)\, dx = \int w f_p(x)\, dx = w$, thus, given $P(x)$ a whole range of sample equivalents for estimation purposes is possible. Given a sample x_i $i = 1, \cdots, n$, the mean incomes of the poor (M_p) and non poor (M_r) are $\sum P(x_i)\, x_i / w$ and $\sum (1 - P(x_i))\ x_i/(1-w)$, respectively. Variances of poor and non poor incomes may be similarly respectively estimated as $\sum P(x_i)\,(x_i - M_p)^2/w$ and $\sum (1 - P(x_i))\ (x_i - M_r)^2/(1-w)$. Various inequality and polarization measures follow in a quite natural fashion. Suppose there is an announced poverty line at z, FGT measures for $I = 1, \cdots$ with respect to the poor and non-poor may be contemplated for example:

$$\mathrm{FGT(I)}_{\mathrm{poor}} = \frac{1}{w} \int_{-\infty}^{z} \left((z-x)/z\right)^{I-1} P(x) f(x)\, dx$$

$$\mathrm{FGT(I)}_{\mathrm{rich}} = \frac{1}{1-w} \int_{-\infty}^{z} \left((z-x)/z\right)^{I-1} (1 - P(x)) f(x)\, dx.$$

A variety of relative poverty measures of the poor could easily be generated by making z some function of the rich distribution (e.g. some particular quantile of the rich distribution). The sample analogues are obvious and simple to calculate.

Other population characteristics of the poor and non-poor are similarly easy to identify for poor and rich groups using $P(x)$ and $(1 - P(x))$. Consider the characteristic h (suppose it to be a "health" index for example) associated with income level x then respective poor and rich means and variances of the poor and rich health status would be given by:

$$\mathrm{E(H_{poor})} = \frac{1}{w} \int_{0}^{\infty} h(x) P(x) f(x)\, dx$$

$$\mathrm{E}(\mathrm{H}_{\text{rich}}) = \frac{1}{1-w}\int_0^\infty h(x)(1-P(x))f(x)\,dx$$

$$\mathrm{V}(\mathrm{H}_{\text{poor}}) = \frac{1}{w}\int_0^\infty \Big(h(x) - \mathrm{E}(\mathrm{H}_{\text{poor}})\Big)^2 P(x)f(x)\,dx$$

$$\mathrm{V}(\mathrm{H}_{\text{rich}}) = \frac{1}{1-w}\int_0^\infty \Big(h(x) - \mathrm{E}(\mathrm{H}_{\text{rich}})\Big)^2 (1-P(x))\,f(x)\,dx.$$

All of which can be estimated by the corresponding sample equivalents.

In a situation where there is an announced poverty cutoff, relationships between the truly poor and the identified poor can be established. Suppose identification of the poor was pursued by employing an arbitrarily determined poverty cutoff c, then:

$\mathrm{P}_{rmp} = \int_{-\infty}^{c} f_y(x)\,dx$ proportion of the rich miss-identified as poor

$\mathrm{P}_{pmr} = \int_{c}^{-\infty} f_p(x)\,dx$ proportion of the poor miss-identified as rich

and the calculated poverty rate would be $w_p(1-P_{pmr}) + (1-w_p)P_{rmp} \neq w_p$ (unless the proportion of the misidentified that are rich identified as poor is equal to the real poverty rate i.e. $w_p = P_{rmp}/(P_{pmr} + P_{rmp})$. What may clearly be seen is that tracking the progress of the poor defined by a poverty cutoff will result in part of the rich group being tracked as a proxy for part of the poor group who are not being tracked. Conditional on the extent of identification of the poor group this may be used as a means of tracking how well frontier based poverty measures track the progress of the poor.

3 The Stochastic Processes.

Without knowledge of which observations are in which group the subdistributions have to be estimated, unfortunately all that is observed is the mixture distribution. Partial identification (estimation of $P(x)$) can be achieved by employing some sort of theory regarding the nature of the subdistributions. Here versions of Gibrat's law are invoked. Suppose that x_t, the income of the representative agent at period t, follows the law of proportionate effects with δ_t its income growth rate in period t, T the elapsed time period of earnings with x_0 the initial income. Thus:

$$x_t = (1 + \delta_{t-1})\, x_{t-1}; \;\text{ and }\; x_T = x_0 \prod_{i=1}^{T-1} (1 + \delta_i) \tag{4}$$

Assuming the δ's to be independent identically distributed random variables with a small (relative to one) mean μ and finite variance σ^2 it may be shown that for an agents life of T years with starting income x_0 the log income size distribution of such agents would be linked systematically from period to period in terms of means and variances in the form[5] :

$$\ln(x_T) \sim \mathrm{N}\left(\left(\ln(x_0) + T(\mu + 0.5\sigma^2)\right), T\sigma^2\right) \tag{5}$$

Note that the distribution is governed by the initial condition $\ln(x_0)$ and the growth rate μ which in turn are dependent on the circumstances of the agent. These types of models are very close to the cross - sectional growth (or Barro) regressions familiar in the growth and convergence literature (see Durlauf *et al.* (2005) for details) except that the properties of the error processes they engender are usually ignored in cross-sectional comparisons, in particular the variance of the process is heteroskedastic increasing in a cumulative fashion through time implying increasing absolute inequality. Note that 5 could also be the consequence of a process of the form:

$$\ln(x_y) = \ln(x_{t-1}) + \psi + e_t \tag{6}$$

which had started at $t = 0$ and had run for T periods where e_t was an i.i.d. $\mathrm{N}(0, \sigma^2)$ and where $\psi = \mu + 0.5\,\sigma^2$. Indeed the i.i.d. assumption regarding the δ's s is much stronger than needed, under conditions of 3^{rd} moment boundedness, log normality can be established for sequences of non-independent, heteroscedastic and heterogeneous δ (see Gnedenko, 1962) where the variance of the process still grows as $O(T)$. The power

[5]The same result can be achieved in the continuous time paradigm by assuming a Geometric Brownian Motion for the x process of the form:

$$dx = \mu_x\, dt + \sigma_x\, dw$$

Where μ_x is the mean drift, σ_x is a variance factor and dw is the white noise increment of a Weiner process.

of the law, like all central limit theorems, is that a log normal distribution prevails in the limit almost regardless of the underlying distribution of the δ's (or e's).

To somewhat muddy the waters Kalecki (1945) generated a lognormal size distribution from a stationary process of the form:

$$\ln(x_t) - \ln(x_{t-1}) = \lambda\left(f(w_t) - \ln(x_{t-1})\right) + e_t \tag{7}$$

with $0 < \lambda < 1$ this corresponds to a partial adjustment model to some equilibrium $f(w_t)$, (which in the context of incomes would be a "fundamentals" notion of long run log incomes). This is essentially a reversion to mean type of process where the mean itself could be a description of the average income level at time t (which incidentally may well be trending through time) but here the variance of the process (and concomitantly absolute inequality) stays constant over time. For $e_t \sim \mathrm{N}(0, \sigma^2)$ in the long run $\ln(x_t) \sim \mathrm{N}(f(w_t), \sigma^2/\lambda)$. There are several observations to be made.

Firstly the pure integrated process story associated with Gibrat's law is not even a necessary condition for lognormality of the income size distribution, such distributions can be obtained from quite different, more generally integrated or non-integrated processes. Secondly stationary processes are in some sense memory-less in that the impacts of the initial value of incomes $f(w_0)$ and the associated shock e_0 disappear after a sufficient lapse of time. On the other hand integrated processes never forget, the marginal impact of the initial size and subsequent shocks remain the same throughout time. Thirdly, if $f(w_t)$ were itself an integrated process (if the $w's$ were integrated of order one and $f(w)$ was homogenous of degree one for example) 7 would correspond to an error correction model and incomes would still present as an integrated process in its own right with x and the function of the w's being co-integrated with a co-integration factor of 1. This is the key to distinguishing between Kalecki's law and Gibrat's law, the cross-sectional distribution of the former only evolves over time in terms of its mean $f(w_t)$, its variance (written as σ^2/λ^2) is time independent, whereas the cross distribution of the latter evolves in terms of both its mean and its variance overtime.

4 Finite mixture models and the EM algorithm.

Mixture models can be viewed as a semi-parametric alternative to the non-parametric densities and, by controlling for the number of components, we can obtain a suitable compromise between the efficiency of parametric models and the flexibility of non-parametric methods. We are interested in a parametric family of *finite mixture densities*, i.e., a family of probability density functions of the form:

$$f(x_i, \mathbf{\Psi}) = \sum_{j=1}^{g} \pi_j f_j(x_i, \theta_j) \tag{8}$$

where the vector $\mathbf{\Psi} = (\pi_1, \cdots, \pi_{g-1}, \Theta')'$ contains all the unknown parameters in the mixture model; $\pi_j,\ j = 1, \cdots, g$ represent the mixing proportions and the vector Θ contains all the parameters $(\theta_1, \cdots, \theta_g)$ known *a priori* to be distinct; $f_j\ (x_i, \theta_j)$ denotes the values of the univariate density specified by the parameter vector θ_j and g the number of mixing components.

The *mixture density estimation problem* (Redner and Walker, 1984) consists in estimating Ψ in (8) given the number g of component populations in the mixture. The method of Maximum Likelihood (ML) has become the most commonly approach for solving the the mixture density estimation problem. As is well known, a maximum-likelihood estimate is a choice of parameters which maximizes the probability density function of the sample and ensures important properties of the parameters. However, there are two practical difficulties associated with ML estimation for mixture densities. The first difficulty is that the log-likelihood function increases without bound and, as first pointed out by Kiefer and Wolfowitz (1956), the global maximizer of the likelihood function, might not exist for a mixture of two univariate normal distributions with unequal variances. The second difficulty is related to the question of which root of the likelihood equation corresponding to a local maximum of the likelihood function (i.e. which local maximizer) to choose as the estimate of the vector of unknown parameters Ψ. The log-likelihood function can, and in fact often does have, local maxima which are not necessarily the global maxima. However, there is little one can do about this problem and the non-existence of a global maximizer of the likelihood function estimate does not place a *caveat* on the proceedings, as the essential aim of the likelihood estimation is to find a sequence of roots of the likelihood equation that is consistent, and hence efficient if the usual regularity conditions hold. These conditions should hold for many parametric families.

There are also computational difficulties associated with obtaining maximum-likelihood (ML) estimates in the mixture models framework, essentially related to the strong dependence of the likelihood function on the parameters to be estimated. For mixture density problems, the likelihood equations are nonlinear and, consequently, we cannot find analytical solutions but rather seek for approximate solutions via iterative procedures. A particular iterative procedure for numerically approximating maximum-likelihood estimates of the parameters in mixture density is the expectation-maximization (EM) algorithm. The EM algorithm for mixture density estimation problems can be viewed as a specialization of the general version as originally formalized by Dempster *et al.* (1977) used to approximate ML estimates of the parameters for incomplete data problems. In the mixture density estimation problem the incompleteness refers to the assignments of data points to mixture components. At present, in the mixture-density parameter estimation problem, the EM is being widely applied because of its superior performances over other estimation procedures in finding a local maximum of the likelihood function (McLachlan and Peel, 2000). In fact, one of its more attractive feature is that it produces sequences of iterates on which the log-likelihood function increases monotonically. This monotonicity is the basis for producing iteration sequences with good global convergence characteristics (Redner and Walker, 1984). The algorithm is, however, guaranteed to converge to a local maximum of the likelihood function.

4.1 Estimation of mixture models.

In this paper we fit mixtures of Gaussians components, i.e. incomes (in logs) come from a mixture of g normal distributions, with unrestricted variances. The vector $\theta_j = \left(\mu_j, \sigma_j^2\right)$ denotes the mean and the variance of the univariate normal component j. The mixing proportions $\pi_1, \cdots, \pi_g$, that are nonnegative and sum to one, give the prior probability that a agent belongs to the gth component of the mixture, representing an endogenous parameter which determines the relative importance of each component in the mixture. The fitting of the mixture model provides a probabilistic clustering of n agents in terms of their estimated *ex post* probabilities of membership of the individual

g components of the mixture of distributions. The posterior or conditional probability τ_{ij} is given by:

$$\tau_{ij} = \texttt{Prob}\{C(i) = g \mid (x_i; \boldsymbol{\Psi})\} = \frac{\pi_j f_j(x_i)}{\sum_{h=1}^{g} \pi_h f_h(x_i)} \tag{9}$$

where $C(i)$ indicates the component to which agent i belongs, $i = 1, \ldots, n$.

The estimation procedure throughout the EM algorithm provides ML estimation of the parameters of the constituent populations and the proportions in which they are mixed. Given current estimates of the model parameters, in the Estimation step we used Bayes' rule to calculate for each component j and for each data point i the expected posterior or conditional probabilities τ_{ij}. In the Maximization step we maximized the likelihood function under the assumption that the missing data are known and we ML-estimate each density f_j weighing the data by the marginal probabilities π_j. The pairs of E and M steps are applied iteratively[6]. Convergence problems are encountered with the EM algorithm when the fraction of the information missing because of the unobserved $C(i)$ is substantial, or the log-likelihood has ridges and similar features. The number of iterations tends to increase with the complexity of the model (number of components). In our analysis, we have had no problems with convergence or multiple extremes.

We dealt with the problem of seeking for the largest local maxima by applying a range of starting values. Typical initial values for a mixtures of Gaussian are equal mixtures coefficients, components variances equal to that of data, components means drawn from a normal distribution estimated from the data. In our analysis, we chose different ways of specification of initial values, like random starts and k-means (crisp and fuzzy) clustering based starting values.

4.2 Assessing the number of mixture components.

The mixture density estimation problem discussed in section 4.1 assumes g, the number of components, to be known. Unless there is strong *a priori* information on g it becomes a matter of choice and the choice of the "optimal" number of component densities is a difficult problem which has not been fully solved (Aitkin and Wilson, 1980; Richardson

[6] All the computing was carried out in R. The software developed (functions) can be obtained from the authors on request.

and Green, 1997; McLachlan and Peel, 2000). Here we test for the minimum number of components that are simultaneously compatible with the data, where compatibility refers to the goodness of fit of the estimated model. An empirical comparison with other well-known procedures is in Pittau *et al.* (2010). We compare the shape of the hypothesized mixture distribution with the true unknown density, consistently estimated by a kernel estimator. The two densities are compared by measuring the integrated distance between the estimated mixture and the kernel estimated density. Since the exact shape of the mixture density varies with the number of components, we start comparing kernel density with a single normal distribution and we keep adding constituents until the integrated distance between the estimated mixture and the kernel density is not significantly different from zero.

Specifically, the goodness of fit of the mixture model is assessed by a measure of global distance between the estimated true density and the estimated parametric density:

$$J = \int_x [\widehat{f}(x) - f(x; \widehat{\boldsymbol{\Psi}}_0)]^2 dx. \tag{10}$$

$\widehat{f}(x)$ is the estimated kernel density of the unknown density function $f(x)$ obtained by the kernel estimator:

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} \left(\frac{x - X_i}{h} \right) \tag{11}$$

where h is the bandwidth that governs the degree of smoothing and $K(\cdot)$ is a kernel function; $f(x; \widehat{\boldsymbol{\Psi}}_0)$ is the mixture of normal components ML-estimated via the EM algorithm under the null hypothesis of $g = g^*$ components.

The smoothing operation in kernel estimation produces bias. To avoid this smoothing induced bias, following Bowman (1992), at each point x the kernel density is compared with the kernel-smoothed parametric estimate under the assumption of a mixture of g normal components. Formally, the test contrasts the density estimate with its estimated mean under the assumption of mixture, that is the expected value under the null hypothesis of a mixture of g^* components:

$$\widehat{J} = \int_x \left[\widehat{f}(x) - \widehat{E}\widehat{f}(x) \right]^2 dx. \tag{12}$$

The test statistic is a measure of agreement between these two densities, the estimated integrated squared error (ISE) statistic. The hypothesis to be tested is: H_0:

$g = g^*$ against H_1: $g \neq g^*$. Using the test sequentially starting from $g^* = 1$ the alternative becomes $g > g^*$. The estimated mean $\widehat{E}\widehat{f}(x)$ under the null hypothesis is equal to a convolution of a kernel with a mixture of normal components.

When the kernel function is Gaussian, it is easy to show that the convolution collapses into a mixture of normal densities with means equal to μ_j and variances equal to $(\sigma_j^2 + h^2)$, $j = 1, \cdots, g$, where h is the bandwidth used to compute $\widehat{f}(x)$. Under the null hypothesis of correct parametric specification, Fan (1994) proves that a center-free test statistic based on $\widehat{J}$ is asymptotically normally distributed (see also Fan and Ullah, 1999, for the properties of the test when observations are weakly dependent). However, convergence is rather slow and the normal distribution is an accurate approximation of the test statistic only for very large samples. A more powerful solution for small sample is to calculate critical values for the test statistic by simulations. As in Fan (1995), we use a parametric bootstrap procedure (with 1000 replications).

Starting from the null hypothesis of $g = 1$, the rejection of the null implies the implementation of the test under the null hypothesis that the population density is a mixture of $g = 2$ normal distributions, and so on until we cannot reject the null of $g = g^*$, finding the smallest number of components compatible with the data.

The implementation of the procedure requires a good choice of the smoothing parameter h. Data-based smoothing value that achieves good estimation of the kernel density irrespective of the null or the alternative hypothesis is true was considered and the bandwidth obtained by using the Sheather and Jones method (1991) was preferred because of its overall good performance (Jones *et al.*, 1996)[7].

Table 1 reports the ISE statistics and the corresponding bootstrapped p-values for testing the null hypothesis of $g = g^*$ components for g^* ranging from 1 to 4. In each year the value of the ISE statistic implies rejection, at least at 10% significance level, of the null hypotheses $g = 1$, $g = 2$ and $g = 3$, but it never rejects $g = 4$. That is, the null of $g = 4$ is selected over the alternative that $g = 2$ and $g = 3$. These results are corroborated by the good fitting provided by the four-components fitted semi-parametric density versus the non-parametric smoothed density.

As a result of the curve fitting procedures Figure 2 and Figure 3 show the estimated

[7] These results are robust to the selection of the smoothing parameter h used in the kernel estimation of the unknown density function. The results are available upon request.

Figure 2: Kernel density estimation and the two-components mixture model fit over the period 1985–2008.

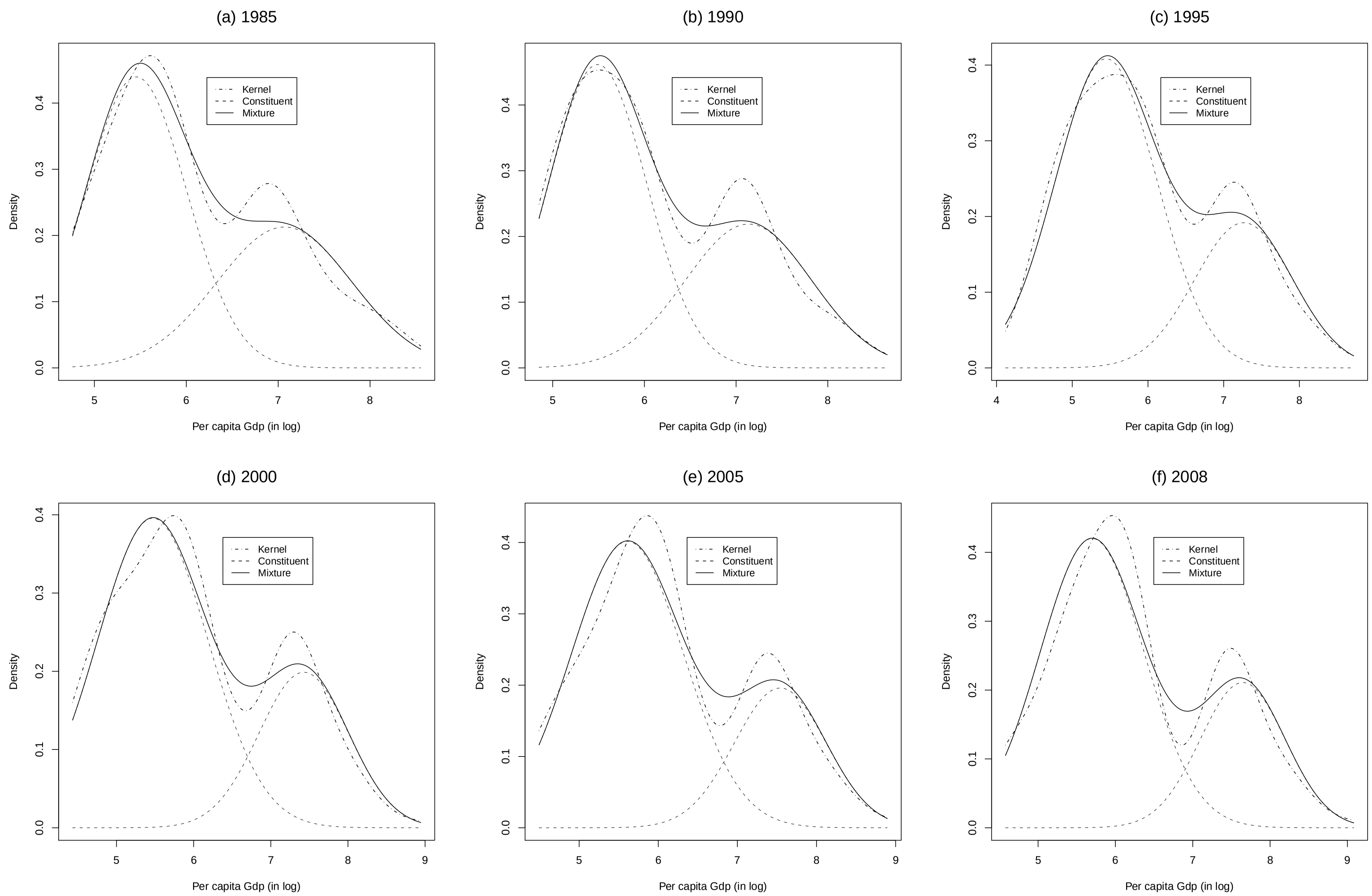


Figure 3: Kernel density estimation and the four-components mixture model fit over the period 1985–2008.

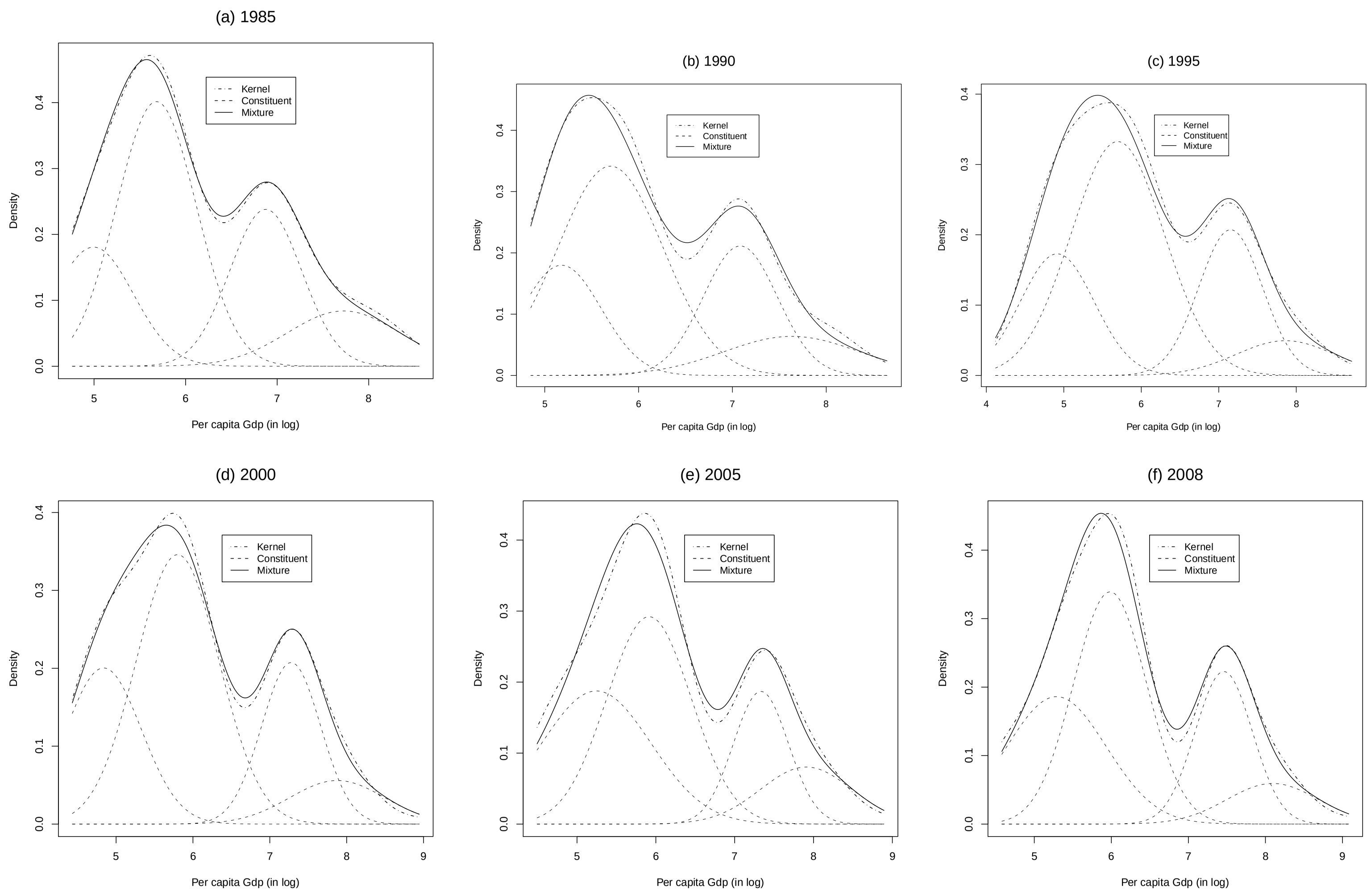


Table 1: The choice of the number of components according to the goodness of fit test.

	m^*=1		m^*=2		m^*=3		m^*=4	
year	$\widehat{J}$	p-value	$\widehat{J}$	p-value	$\widehat{J}$	p-value	$\widehat{J}$	p-value
1985	3.48	0.00	0.24	0.30	0.25	0.21	0.02	0.80
1990	2.96	0.00	0.24	0.32	0.24	0.09	0.06	0.17
1995	2.51	0.00	0.16	0.40	0.11	0.12	0.05	0.34
2000	4.12	0.00	0.28	0.24	0.15	0.08	0.05	0.38
2005	6.18	0.00	0.46	0.12	0.14	0.19	0.05	0.42
2008	8.22	0.00	0.60	0.12	0.20	0.15	0.06	0.48

The estimated ISE, $\widehat{J}$, is multiplied by 100

kernel and the bias-adjusted mixture densities per capita GDP (in $US 2000 constant prices) for 47 African countries over the period 1985–2008 as well as the constituents of a two and a four-component estimated mixture, respectively. Since the test statistic is a simple measure of the distance between these two densities it facilitates a picture on a density scale of how well the model with the minimum number of components fits the data.

5 An Example: Africa 1985-2008.

As a result of the curve fitting procedures, a mixture distribution of four log normal size distributions of per capita GDP (in $US 2000 constant prices) was estimated for 47 African countries (appendix 1 lists the countries, along with the ex-post estimated probabilities to belong to the very Poor and to the Poor group) over the period 1985–2008 for which data was available. The 47 countries are not equally sized in terms of population, with some remarkable differences. In the year 2008, for example, the population in Nigeria was over 151 million inhabitants, while the Seychellois citizens were only 86 thousands in number. The variability across African countries is notable, though not comparable with differences in world population distribution and, therefore, population-based weighting schemes were integrated into all the estimation procedures. Figure 3 provides a comparison of the chosen mixture distribution and its sub components with a kernel estimate of the overall size distribution and Table 2 reports the means, standard deviations and mixture coefficients for the four component distributions which may be associated with "very poor" (or "chronic"), "poor", "middle income" and "rich" groups. As is obvious from the diagram the overall distribution is bi-modal permitting the interpretation of the overall distribution being a mixture of mixture distributions. This affords us two possible analyses one with the poverty group being the poorest single component, the other with the poverty group being a mixture of two lower distributions.

The four groups have economic growth rates of 1.27%, 1.31%, 2.54% and 1.6% respectively indicative of a growing apart of the respective poor and rich groups however they are defined. The standard deviations of the two poor components are growing over time (consistent with Gibrat's law) and the standard deviations of the non-poor groups have stayed roughly constant over time (consistent with Kalecki's law). The mixing coefficients, reflective of proportionate group membership, have increased substantially for the poorest group (20% to 30%) declined slightly for the next poorest group (44% to 40%) and declined solidly for the two top groups so that defining the poor as the bottom two groups it may be seen that the poverty rate has increased from 64% to 70%. Defining the poor as the bottom group the growth of the poverty rate over the period has been 2.32% per annum, defining the poor as the bottom two groups the growth in the poverty rate has been 0.45% per annum.

How well identified are these effects? Table 3 reports the identification index (equation (3)), i.e. one minus the overlap measure for the poorest group and the rest and for the lower modal group and the upper modal group. The overlap measure of two Normal distributions can be written as following. Letting x^* be the intersection point of two normals $\mathrm{N}\left(\mu_p, \sigma_p^2\right)$ and $\mathrm{N}\left(\mu_r, \sigma_r^2\right)$ withe respective weights w and $1-w$ where $\mu_p < x^* < \mu_r$, then the Overlap Measure (OV) is given by:

$$OV = (1 - \Phi\left((x^* - \mu_p)/\sigma_p\right)) + (1-w)\Phi\left((x^* - \mu_r)/\sigma_r\right)/w \tag{13}$$

being Ψ the cumulative normal distribution function and x^* the solution to

$$w\,\frac{\mathrm{e}^{\frac{-(x^*-\mu_p)^2}{2\sigma_p^2}}}{\sqrt{2\pi\sigma_p^2}} = (1-w)\,\frac{\mathrm{e}^{\frac{-(x^*-\mu_r)^2}{2\sigma_r^2}}}{\sqrt{2\pi\sigma_r^2}}$$

Table 2: Means, Standard Deviations and Mixing Proportions of the 4-components: Very Poor, Poor, Middle-income, Rich.

		Means		
Year	Very Poor	Poor	Middle	Rich
1985	4.996	5.681	6.875	7.726
1990	5.179	5.700	7.084	7.632
1995	4.915	5.695	7.153	7.861
2000	4.838	5.801	7.277	7.878
2005	5.245	5.914	7.333	7.908
2008	5.289	5.984	7.459	8.094
		Standards deviation		
Year	Very Poor	Poor	Middle	Rich
1985	0.248	0.251	0.167	0.498
1990	0.279	0.467	0.235	0.649
1995	0.287	0.463	0.155	0.509
2000	0.335	0.401	0.138	0.504
2005	0.601	0.408	0.025	0.467
2008	0.552	0.333	0.157	0.489
		Components		
Year	Very poor	Poor	Middle	Rich
1985	0.197	0.439	0.235	0.129
1990	0.191	0.484	0.210	0.116
1995	0.207	0.500	0.214	0.079
2000	0.247	0.467	0.200	0.087
2005	0.326	0.393	0.164	0.117
2008	0.302	0.403	0.207	0.088

which may be written as the following quadratic form in x^* with the root between μ_p and μ_r providing the intersection point:

$$\left(\frac{1}{\sigma_r^2} - \frac{1}{\sigma_p^2}\right) x^{*2} - 2\left(\frac{\mu_r}{\sigma_r^2} - \frac{\mu_p}{\sigma_p^2}\right) x^* + \left(\frac{\mu_r^2}{\sigma_r^2} - \frac{\mu_p^2}{\sigma_p^2}\right) - 2\ln\left(\frac{(1-w)\sigma_p}{w\sigma_r}\right) = 0$$

It is evident that identification is weak in the poorest group case but very strong in the lower mixture subgroup case, not surprising given the almost perfect segmentation in the latter case (see diagram 1).

The probability of being poor can also be used as an identifier to evaluate other aspects of the poor and non-poor groups. Here as an illustration a health outcome–life

Table 3: Identification index for the poorest group and the rest as well as for the lower modal group and the upper modal group

Identification Index		
Year	Poorest group	Lower modal group
1985	0.753	0.996
1990	0.171	0.958
1995	0.504	0.987
2000	0.734	0.994
2005	0.471	0.991
2008	0.548	0.995

Table 4: Life Expectancy for the chronic poor and for the poor group, years: 1985–2005)

Life Expectancy (Chronic Poor)					
	1985	1990	1995	2000	2005
Poor Mean	48.589	48.183	44.960	41.748	50.244
Non-Poor Mean	52.935	54.379	55.954	57.646	56.337
Poor Std Deviation	3.753	3.967	4.262	7.141	4.659
Non- Poor Std Deviation	8.394	10.237	13.368	16.441	11.235
Life Expectancy (Poor)					
	1985	1990	1995	2000	2005
Poor Mean	50.078	50.982	51.390	51.622	53.001
Non-Poor Mean	55.575	57.788	59.208	58.956	57.798
Poor Std Deviation	4.453	5.338	5.322	5.047	5.426
Non- Poor Std Deviation	10.295	12.191	13.666	14.351	14.107

expectancy is considered (note only data through 2005 was available at time of writing). Table 4 reports the comparisons. As will be seen the outcomes are substantially different for the two poor group definitions but notwithstanding these differences life expectancy for the poor groups is decidedly inferior to that of the non poor groups however defined with very little change in the life expectancy for the poorest group. There also appears to be some growth in the variability of life expectancy for both groups in both definitions.

It is also possible to examine the extent to which current practice of identifying the poor by employing a poverty cut off actually captures the poor group. Two examples,

Table 5: Poverty line based on 0.9 × overall mean poverty cut off, years: 1985–2005

		Chronic Poverty Group			Poor Group		
Year	Identified poverty rate	% Poor Identified as Poor	% Poor Identified as Rich	% Rich Identified as Poor	% Poor Identified as Poor	% Poor Identified as Rich	% Rich Identified as Poor
1985	29%	98%	3%	12%	45%	55%	0%
1990	33%	88%	12%	20%	49%	51%	0%
1995	33%	96%	4%	17%	47%	53%	0%
2000	32%	96%	4%	11%	45%	55%	0%
2005	30%	69%	31%	11%	41%	59%	0%
2008	28%	74%	26%	9%	40%	60%	0%

a poverty cut off at 0.9 of mean income (which may be thought of as a relative poverty line) and a \$2 per day cut off (which may be thought of as an absolute poverty line), are reported in Tables 5 and 6 respectively. When the lowest component in the mixture is considered the poor, both poverty lines (with a couple of exceptions), capture the poor group pretty well, but they do miss-identify a substantial proportion of the non-poor group in the calculus. When the two lowest components are considered the truly poor the relative poverty line identifies less than 50% of the poor correctly whereas the absolute poverty line does a pretty good job except that its effectiveness is declining (largely due to the growth in incomes and the relatively stable standard deviation of this grouping). This is because the \$2 a day poverty line is close to the separation point of the poor and non-poor distributions in this structure.

It is instructive to compare the growth in poverty rates obtained when identification of the poor is achieved by employing a poverty line with those obtained in the partially identified case. In both the relative poverty (-0.01%) and absolute poverty (0.27%) cases the growth rates of the poor group are understated, this is largely due to substantial portions of the rich being miss-identified as poor and substantial portions of the poor being miss-identified as rich. The comparison that comes closest is that between the absolute poverty line and the lower modal grouping of the poor but this is because the absolute poverty line is very close to where the almost perfect segmentation appears so that there is very little miss-identification in either direction.

Table 6: Poverty line based on $2 a day poverty cut off, years: 1985–2005

		Chronic Poverty Group			Poor Group		
Year	Identified poverty rate	% Poor Identified as Poor	% Poor Identified as Rich	% Rich Identified as Poor	% Poor Identified as Poor	% Poor Identified as Rich	% Rich Identified as Poor
1985	65%	100%	0.00%	56%	99%	0.01%	3.38%
1990	67%	100%	0.00%	59%	98%	2.00%	3.12%
1995	69%	100%	0.00%	62%	98%	1.90%	0.18%
2000	70%	100%	0.00%	61%	98%	1.57%	0.16%
2005	70%	99%	1.24%	56%	97%	3.17%	0.10%
2008	69%	99%	0.91%	56%	98%	2.32%	0.03%

6 Conclusions

Drawing on ideas from the partial identification literature and using predictions from various versions of Gibrat's law a new approach to poverty measurement has been proposed which avoids the many difficulties associated with defining poverty lines. The cost is that no longer are the poor perfectly identified, individuals in the poor group are only partially identified with a probability being attached to their poor status. However this does not impede calculation of the usual poverty statistics such as the poverty rate, the average income of the poor and the variability of the incomes of the poor, nor does it impede the calculation of statistics relating to other characteristics of the poor. The probability of being poor can also be used to elicit other characteristics of the poor versus the non poor. Furthermore it is possible to calculate an index of the extent to which identification of the poor group has been achieved.

The approach is illustrated in an application to 47 African countries over the period 1985–2008. Identification of the poorest group is limited in some years but identification of the lowest modal group appears quite sound. In essence the incomes of the poor countries are shown to be growing more slowly than those of the non-poor countries, membership of the poor group is increasing (and that of the rich country group is declining) and, though the life expectancies of the poor groups are increasing, the gap between the life expectances of the chronically poor and the rest is widening. Conventional methods of identifying the poor by specifying a poverty line appear to understate the growth in the poor group in all instances.

Appendix A: estimated ex post probabilities for each country of being chronically poor and poor[8]

Year	1985		1990		1995		2000		2005		2008	
Country	Chronic Poor	Poor	Chronic Poor	Poor	Chronic Poor	Poor	Chronic Poor	Poor	Chronic Poor	Poor	Chronic Poor	Poor
Algeria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Angola	0.00	0.00	0.00	0.31	0.00	1.00	0.00	0.99	0.15	0.88	0.00	0.01
Benin	0.00	1.00	0.11	1.00	0.01	1.00	0.01	1.00	0.26	1.00	0.21	1.00
Botswana	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Burkina Faso	0.76	1.00	0.56	1.00	0.40	1.00	0.21	1.00	0.44	1.00	0.46	1.00
Burundi	0.95	1.00	0.62	1.00	0.79	1.00	0.96	1.00	0.97	1.00	1.00	1.00
Cameroon	0.00	0.00	0.00	0.57	0.00	1.00	0.00	0.99	0.15	0.99	0.13	1.00
Cape Verde	0.00	0.00	0.00	0.17	0.00	0.10	0.00	0.01	0.03	0.09	0.00	0.00
Cent.Afr.Rep.	0.01	1.00	0.18	1.00	0.08	1.00	0.10	1.00	0.56	1.00	0.62	1.00
Chad	0.48	1.00	0.54	1.00	0.51	1.00	0.69	1.00	0.38	1.00	0.52	1.00
Comoros	0.00	1.00	0.01	0.99	0.00	1.00	0.00	1.00	0.22	1.00	0.20	1.00
CongoDemRep	0.10	1.00	0.46	1.00	0.82	1.00	0.99	1.00	0.99	1.00	1.00	1.00
CongoRep	0.00	0.00	0.00	0.02	0.00	0.05	0.00	0.10	0.08	0.33	0.02	0.05
Cote Ivoire	0.00	0.00	0.00	0.77	0.00	0.99	0.00	0.99	0.16	1.00	0.12	1.00
EgyptArabRep	0.00	0.00	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Equat.Guinea	0.00	0.31	0.00	0.95	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
Ethiopia	1.00	1.00	0.63	1.00	0.82	1.00	0.92	1.00	0.86	1.00	0.84	1.00
Gabon	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gambia The	0.00	1.00	0.04	1.00	0.01	1.00	0.01	1.00	0.26	1.00	0.19	1.00
Ghana	0.43	1.00	0.39	1.00	0.12	1.00	0.08	1.00	0.34	1.00	0.26	1.00
Guinea	0.00	1.00	0.04	1.00	0.00	1.00	0.00	1.00	0.21	1.00	0.16	1.00
Guinea-Bissau	0.82	1.00	0.54	1.00	0.38	1.00	0.67	1.00	0.93	1.00	0.99	1.00
Kenya	0.00	1.00	0.00	0.98	0.00	1.00	0.00	1.00	0.19	1.00	0.13	1.00
Lesotho	0.01	1.00	0.05	1.00	0.00	1.00	0.00	1.00	0.18	1.00	0.12	1.00
Liberia	0.00	0.77	0.47	1.00	0.83	1.00	0.36	1.00	0.92	1.00	0.97	1.00
Madagascar	0.01	1.00	0.12	1.00	0.08	1.00	0.08	1.00	0.46	1.00	0.43	1.00
Malawi	0.93	1.00	0.64	1.00	0.70	1.00	0.79	1.00	0.90	1.00	0.93	1.00
Mali	0.41	1.00	0.41	1.00	0.19	1.00	0.11	1.00	0.36	1.00	0.34	1.00
Mauritius	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Morocco	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Mozambique	0.96	1.00	0.53	1.00	0.38	1.00	0.15	1.00	0.30	1.00	0.20	1.00
Namibia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Niger	0.43	1.00	0.50	1.00	0.52	1.00	0.70	1.00	0.79	1.00	0.88	1.00
Nigeria	0.00	1.00	0.02	0.99	0.00	1.00	0.00	1.00	0.19	1.00	0.13	1.00
Rwanda	0.05	1.00	0.32	1.00	0.31	1.00	0.22	1.00	0.41	1.00	0.29	1.00
Senegal	0.00	0.99	0.00	0.98	0.00	1.00	0.00	1.00	0.16	1.00	0.12	1.00
Seychelles	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sierra Leone	0.03	1.00	0.26	1.00	0.33	1.00	0.79	1.00	0.49	1.00	0.47	1.00
South Africa	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sudan	0.09	1.00	0.21	1.00	0.02	1.00	0.01	1.00	0.19	1.00	0.12	1.00
Swaziland	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Togo	0.01	1.00	0.17	1.00	0.09	1.00	0.08	1.00	0.46	1.00	0.54	1.00
Tunisia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Uganda	0.81	1.00	0.54	1.00	0.19	1.00	0.08	1.00	0.34	1.00	0.22	1.00
Zambia	0.00	1.00	0.02	0.99	0.01	1.00	0.02	1.00	0.25	1.00	0.18	1.00

[8] The first columns reports the probability of being in the very poor or chronic poverty group, the second the probability of being poor, i.e. the probability of belonging to the lowest components of the mixture.

References

Aitkin, M. and Wilson, G. T. (1980). "Mixture models, outliers, and the EM algorithm",. *Technometrics*, 22, 325–331.

Anderson, G.J., Linton, O. and Wang, Y. (2009). "Non-Parametric Estimation of Polarization Measures", Mimeo LSE.

Anderson, G.J., Leo, T.W., and Linton, O. (2009a). "Making Inferences About Rich Country-Poor country Convergence: The Polarization Trapezoid and Overlap Measures". Mimeo, University of Toronto, Economic Department.

Battistin, E., Blundell, R. and Lewbel, A. (2009). "Why Is Consumption More Log Normal than Income? Gibrat's Law Revisited". *Journal of Political Economy*, 117, 1140–1154.

Betti, G., Cheli, B., Lemmi, A. and Verma, V. (2006). "On the construction of fuzzy measures for the analysis of poverty and social exclusion". *Statistica & Applicazioni*, 4, 77–97.

Bourguignon, F. and Chakravarty, S.R. (2003). "The measurement of multidimensional poverty". *Journal of Economic Inequality*, 1, 25–49.

Bowman, A.W. (1992). "Density based tests of goodness-of-fit". *Journal of Statistical Computation and Simulation*, 40, 1–13.

Citro, C.F. and Michael, R.T. (1995). *Measuring Poverty: A New Approach.* National Academy Press, Washington, DC.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). "Maximum likelihood from incomplete data via EM algorithm". *Journal of the Royal Statistical Society*, B, 69, 1–38.

Duclos, J-Y., Sahn, D. and Younger, S.D. (2006). "Robust Multidimensional Poverty Comparisons". *The Economic Journal*, 116, 943–968.

Durlauf, S., Johnson, P.A. and Temple, J. (2005). "Growth Econometrics". forthcoming in *Handbook of Growth Economics*, P. Aghion and S. Durlauf editors.

Fan, Y. (1994). "Testing the goodness of fit of a parametric density function by kernel method", *Econometric Theory*, 10, 316–356.

Fan, Y., (1995). "Bootstrapping a Consistent Non Parametric Goodness-of-Fit Test". *Econometrics Reviews*, 14, 367–82.

Fan, Y. and Ullah, A. (1999). "On Goodness-of-fit Tests for Weakly Dependent Processes Using kernel Method". *Journal of Nonparametric Statistics*, 11, 337–360.

Foster, J., Greer, J. and Thorbecke, E. (1984). "A Class of Decomposable Poverty Measures". Econometrica, 42, 761–766.

Gibrat, R., (1930), "Une Loi Des Repartitions Economiques: L'effet Proportionelle". *Bulletin de Statistique General*, France, 19, 469.

Gibrat, R., (1931). *Les Inegalites Economiques.* Libraire du Recueil Sirey, Paris.

Grusky, D., and Kanbur, R. (eds.) (2006). *Poverty and Inequality.* Stanford University Press, Stanford, California.

Jones, M.C., Marron, J. S. and Sheather, S. J. (1996)."A Brief Survey of Bandwidth Selection for Density Estimation". *Journal of the American Statistical Association*, 91, 401–407.

Kalecki, M. (1945). "On the Gibrat distribution". *Econometrica*, 13, 161–170.

Kiefer, J. and Wolfowitz, J. (1956). "Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters". *Annals of Mathematical Statistics*, 27, 887–906.

Manski, C.F. (2003). *Partial Identification of Probability Distributions.* New York, Springer-Verlag.

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

Pittau M.G., Zelli, R. and Johnson P.A. (2010). "Mixture Models, Convergence Clubs and Polarization". *Review of Income and Wealth*, 56, 102–122.

Redner, R.A. and Walker, H.F. (1984). "Mixture densities, maximum likelihood, and the EM algorithm". *SIAM Review*, 26, 195–239.

Richardson, S. and Green, P.J. (1997). "On Bayesian analysis of mixtures with unknown number of components". *Journal of Royal Statistical Society*, B, 731–792.

Rodgers, J.R. and Rodgers, J.L. (1993). "Chronic poverty in the United States". *Journal of Human Resources*, XXVIII, 25–54.

Sala-i-Martin, X. (2006). "The World Distribution of Income: Falling Poverty and...Convergence, period". *Quarterly Journal of Economics*, 121, 351–397.

Sen, A.K. (1983). "Poor Relatively Speaking", *Oxford Economic Papers.* 35, 153–169.

Sen, A. K. (1992). *Inequality Reexamined.* Oxford, Oxford University Press.

Slesnick, DJ. (1993). "Gaining ground: poverty in the postwar United States". *Journal of Political Economy*, 101, 1–38.

Sutton, J. (1997). "Gibrat's Legacy". *Journal of Economic Literature* 35, 40–59.

Townsend, (1985). "A Sociological Approach to the Measurement of Poverty - A Rejoinder to Professor Amartya Sen". *Oxford Economic Papers*, 37, 659–668.

Yitzhaki, S. (1994). "Economic Distance and overlapping of distributions". *Journal of Econometrics*, 61 147–159.