Boats and Tides and "Trickle Down" Theories: What Economists Presume about Wellbeing When They Employ Stochastic Process Theory in Modeling Behavior

Gordon Anderson
University of Toronto

Abstract Aphorisms that "rising tides raise all boats" or that material advances of the rich eventually "trickle down" to the poor are really maxims regarding the nature of stochastic processes that underlay the income/wellbeing paths of groups of individuals. This paper looks at the implications for the empirical analysis of wellbeing of conventional assumptions regarding such processes which are employed by both micro and macro economists in modeling economic behavior. The implications of attributing different processes to different groups in society following the club convergence literature are also discussed. Various forms of poverty, inequality, polarization and income mobility structures are considered and much of the conventional wisdom afforded us by such aphorisms is questioned. To exemplify these ideas the results are applied to the distribution of GDP per capita in the continent of Africa.

Correspondence Gordon John Anderson, University of Toronto, Economics Department, Max Gluskin House, 150 St George St, Ontario, Toronto, M5S 3G7, Canada;
e-mail: anderson@chass.utoronto.ca

1 Introduction

The aphorism that "a rising tide raises all boats" and the theory that advances in economic well-being of the rich ultimately "trickle down" to the poor have frequently been cited as reasons for believing that growth elevates poor from poverty.[1] These are essentially notions regarding the nature of income or consumption processes as stochastic processes. Economists interested in growth, consumption and convergence issues of various forms have a long tradition of modeling income or consumption as a stochastic process (either of a stationary or of a random walk variety), presumably because such processes provide simple and effective descriptions of income and consumption paths for modeling purposes but also because such formulations, in the form of consumption and growth regressions, provided a useful way of relating consumption trends and growth rates to initial conditions.

A microeconomic literature that built on Modigliani and Brumberg (1954) and Friedman (1957) developed models of agents who maximized the present value of lifetime happiness $(_0\int^T U(C(t))e^{-r^*t}dt)$ subject to the present value of lifetime wealth $(_0\int^T Y(t)e^{-rt}dt)$ where $U(\ )$ is an instantaneous felicity function, Y is income, r^* is the individuals rate of time preference and r is the market lending rate. Browning and Lusardi (1996) show that this taken together with the assumption of a constant relative risk aversion and no bequest motive preference structure leads to a consumption smoothing model of the form:

$$C(t) = e^{(r-r^*)t/\varsigma}C(0)$$

where ζ is the risk aversion coefficient and by implication $g = (r{-}r^*)t/\zeta$ is the consumption growth rate. The empirical counterpart of this equation is the familiar (non-stationary) random walk model:

$$\ln C(t) = \ln(C(t-1)) + g + e(t)$$

These types of formulations are also close to those used by macro modelers in developing savings and income equations in a growth literature that yields a

---

[1] Recently there has been considerable interest in the "rising tides" notion (Freeman 2001; Hines et al. 2001; Burgess et al. 2001) which the Oxford Dictionary of Quotations (Knowles 2004) attributes to the Kennedy family. Anderson (1964) is responsible for the "trickle down" theory.

stationary process. Here the usual empirical model of conditional convergence in the growth literature comes from a first-order Taylor series approximation to the dynamic path of capital and/or output around its steady-state growth, it is a proposition about mean reversion, not about unit roots (see e.g. Barro and Sala-i-Martin 1991, Mankiw et al. 1992 and Temple 1999). So, given a Cobb–Douglass technology $Y=K^{\alpha}H^{\beta}(AL)^{(1-\alpha-\beta)}$ where Y is output, H is human capital stock, A is technology and L is labour letting $y(t) = Y(t)/L(t)$ the first order Taylor Series expansion of $lny(t)$ around its steady state lny^* yields:

$$dlny(t)/dt = \lambda[lny^*-lny(t)]$$

where $\lambda=(n+g+\delta)(1-\alpha-\beta)$ with n and g being respectively exogenous labour and technology growth rates and δ is the physical capital depreciation rate. This leads to a stationary version of the above per capita consumption equation (in terms of income) of the form:

$$lny(t)-lny(0)=(1-e^{-\lambda t})lny^*- (1-e^{-\lambda t})lny(0)$$

Surprisingly, for the link does not appear to have been made very often in the income size distribution and economic well-being literatures,[2] such models have implications for, and provide predictions as to, the progress of inequality, poverty and polarization that would be of interest to those interested in various aspects of empirical well-being.

Stochastic process theory also provides a motivation for fitting particular size distributions of income or consumption (since the nature of the stochastic process has strong implications for the nature of the size distribution of income). There are advantages associated with fitting size distributions parametrically. Poverty calculations of the non-parametric variety can be difficult, especially when sample sizes are small or the poverty group is small in number relative to the size of the population because information on the relevant tail of the distribution is sparse and changes in the tails of distributions can be very difficult to get a handle on (see Davidson and Duclos (2013) for a discussion that highlights this problem). A parametric distribution that fits the data well can provide substantive information about the nature of such tails. Furthermore if incomes are truly governed by such

---

[2] Battistin et al. (2009), Deaton and Paxson (1994), Meghir and Pistaferri (2004), Meyer and Sullivan (2003), Neal and Rosen (2000), O'Neill (2005) and Osberg (1977) are exceptions.

processes poverty, inequality or polarization policies need to focus on changing the structure of the processes or at least mitigating their effects and obviously a full understanding the processes and their implications will help in this regard. Indeed some policies, such as defining a poverty frontier (social security net) or lower boundary below which incomes are not permitted to fall, can become part of the process, changing its structure and the nature of the resultant size distribution of income. This in turn provides a powerful test of the effectiveness of such a policy in terms of the extent to which the distribution conforms to that predicted by the process structure.

Alternatively one may construe the population as a collection of subgroups each with their own process with the poor as a particular sub group, an entity in itself, with a unique stochastic process defining its path as opposed to the paths of the other presumably more advantaged groups in society. The societal income distribution then in effect becomes a mixture distribution governed by the variety of processes defining the separate groups and the mixture coefficients which define the respective memberships. This imposes additional strictures on rising tides and trickle down theories in the way they impact the separate groups. Anti-poverty policies can then focus on the changing the nature of the processes governing the poorest groups and ideas from the convergence literature become relevant in understanding the relative poverty process. In these circumstances the way poverty is measured also needs to be reviewed since poverty is now about the changing membership of a class and the way the stochastic process describing that class proceeds through time.

Before examining what such models imply for the progress of wellbeing one needs to be clear as to what sort of poverty or inequality it is that is in question. There has been considerable debate about the nature of poverty measurement as to whether it should be an absolute or a relative measure. The issue entertained the minds of the founders of the discipline. Adam Smith (1776) can be interpreted to have had a relative view of poverty viz: "….. By necessaries I understand, not only the commodities which are indispensably necessary for the support of life, but whatever the custom of the country renders it indecent for creditable people, even the lowest order, to be without." Similarly Ferguson (1767) states "The necessary of life is a vague and relative term: it is one thing in the opinion of the savage; another in that of the polished citizen: it has a reference to the fancy and to the habits of living". Marshall (1890) on the other hand had a very clear idea of

income poverty as an absolute concept, his comment on poverty in the introduction to The Principles being "...for with £150 the family has, with £30 it has not, the material conditions of a complete life.". More recently Townsend (1985) (the major advocate of the relative measure in recent times) and Sen (1983) (who favours a basic needs formulation) have lead the debate regarding the two approaches, both claim consistency with the intent of Smith's thoughts largely via different interpretations of the words decent, creditable etc.

Interestingly enough no such debate seems to have taken place regarding relative versus absolute inequality, though invariably relative inequality measures (Coefficient of Variation, Gini and Shutz coefficients for example) seem to have been favored and recently the concept of polarization which is related to, but distinctly different from, relative inequality has gained favour (see Duclos, Esteban and Ray (2004) for details). Some absolute inequality measures (variance levels and quantile differences for example) have currency and absolute measures of polarization are also a possibility.

Here the theoretical implications of stochastic processes for absolute and relative wellbeing measures will be outlined and the results employed in looking at the stochastic processes underlying the per capita GDP of African nations and considering what they imply for the progress of poverty and inequality on that continent. After a consideration of the implications of some aspects of relatively simple stochastic processes for the progress of poverty and inequality in Section 2 mixtures are considered in Section 3. These ideas are considered in the light of data on per capita GDP for African nations over the period 1985 to 2005 in Section 4 and conclusions are drawn in Section 5.

2 Gibrat's Law, Kalecki's Law, The Pareto Distribution and Notions of Absolute and Relative Poverty and Inequality

Two early front runner's for describing the size distribution of income or consumption were the Pareto distribution and the Log-normal distribution,[3] subsequently it has been learned that they are linked via stochastic process theory.

_____

[3] Conventional wisdom was that Pareto fit well in the tails whereas the log-normal fit well in the middle (Harrison 1981; Johnson et al. 1994).

Pareto (1897) felt that his distribution was a law which governed the size distribution of incomes, Gibrat (1930)(1931), working with firm sizes, used statistical central limit theorem type arguments to demonstrate that a sequence of successive independent proportionate "close to one" shocks to an initial level of a variable would yield an income the log of which was normally distributed regardless of the distribution governing the shocks.[4] Gabaix (1999), working with city size distributions, highlighted the link in showing that if a process such as that proposed by Gibrat was subjected to a reflective lower boundary, bouncing back the variable should it hit the boundary from above, the resulting distribution would be Pareto.[5] Obviously a social security net of some kind, such as a legislated low income cut-off below which no one was permitted to fall, would constitute such a boundary[6] for an income process. Both of these notions regarding the shape of income size distributions draw on theories of stochastic processes which, if empirically verified, will also tell us much about the progress of poverty and or inequality however defined.

Starting off with Gibrat's law of proportionate effects in a discrete time paradigm 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:

$$x_t = (1+\delta_{t-1})x_{t-1}; \quad and \quad x_T = x_0 \prod_{i=1}^{T-1}(1+\delta_i) \quad [1]$$

Assuming the $\delta's$ 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:[7]

––––––––––––––––––––––––––

[4] Kalecki (1945) showed that Gibrat's result could be obtained from a stationary process as well.

[5] This has been established before, Harrison (1985) and Champernowne (1953) demonstrate a somewhat similar result. Reed (2001) provides an alternative link between the Log-normal and Pareto and provides rationales from stochastic process theory for more complex size distributions.

[6] The Millenium goals and $1 and $2 poverty frontiers may be construed as such potential frontiers.

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

$Ln(x_t) \sim N((ln(x_0)+T(\mu+0.5\sigma^2)), T\sigma^2)$ [2]

These types of models are consumption process literature 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 [2] could also be the consequence of a process of the form:

$ln(x_t) = ln(x_{t-1}) + \psi + e_t$

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

Clearly for a needs based (absolute) poverty line (say x^*) and growth exceeding $-0.5\sigma^2$ the poverty rate would be 0 in the limit (i.e. $lim_{T->\infty}$ $\Phi([(ln(x^*/x_0)-T(\mu+0.5\sigma^2))/(\sigma\sqrt{T})]$ where $\Phi(z)$ is the cumulative density of the standard normal distribution) and for growth less than $-0.5\sigma^2$ the poverty rate would be 1. For a relative poverty line, for example 0.6 of median income (note median income will be $exp(ln(x_0)+T(\mu+0.5\sigma^2))$ and the poverty cut-off will be .6 of that value), the poverty rate would be $\Phi([ln(0.6)/(\sigma\sqrt{T})])$ which obviously increases with time reaching .5 at infinity. The income quantiles in such an income process will not have common trends and, provided growth is sufficiently small,[8] such a society exhibits increasing inequality by most measures that are not location normalized (hereafter referred to as absolute inequality). For aficionados of the

---

$dx = \mu x dt + \sigma x dw$

Where μ is the mean drift σ is a variance factor and dw is the white noise increment of a Weiner process.

[8] Many inequality measures are location normalized measures of dispersion they are in effect relative inequality measures, (for example the Coefficient of Variation and Gini) if the location is increasing slow enough and the dispersion is increasing fast enough inequality by any measure will be increasing

Gini what really matters is the growth rate, Lambert (1993) shows that for the Log Normal Distribution with mean and variance θ, γ respectively and with a distribution Function $F(z \mid \theta, \gamma)$ the Gini coefficient may be written in the present context as:

$$2F(exp(ln(x_0)+T(\mu+0.5\sigma^2)) \mid exp(ln(x_0)+T(\mu+0.5\sigma^2)),T\sigma^2)-1$$

This will tend to zero as $T => \infty$ when $\mu < -0.5\sigma^2$ and will tend to 1 otherwise, note particularly for zero growth Gini will tend to 1. Note this is purely a function of the Gini being a relative inequality index that is normalized on the mean (it may be interpreted as the average distance between individual incomes normalized by the mean income). If the average distance between individuals were normalized by one of the income quantiles the results would be quite different with a greater propensity to increasing inequality when normalized by a low income quantile and a smaller propensity when normalized by a high income quantile.

The Polarization index proposed by Esteban and Ray (1994) may be seen as closely related to the discrete version of the Gini Index since, given π_i is the probability of being in the i'th cell, it is of the form:

$$P_\alpha = K \sum_{i=1}^{n} \sum_{j=1}^{n} | x_i - x_j | \pi_i^{1+\alpha} \pi_j$$

A continuous version of this index is given in Duclos Esteban and Ray (2004) as:

$$P_\alpha = K \int f(x)^\alpha \mid y - x \mid dF(x)dF(y)$$

where in each case $\alpha > 0$ is the index of polarization aversion (when $\alpha = 0$ we have the Gini index) and K is a scale factor. In the Gini version K is the inverse of mean income giving it its relative flavour thus the effects on this index wil be much the same as for the Gini coefficient.

With respect to relative poverty measures should a "civil society" protect its poor in maintaining its "relative status", for example by defining a poverty cutt-off such that the poorest 20% of society were considered the poor, then the cutt-off would exhibit a lower growth rate than mean income. One may thus engage propositions such as those mooted in Friedman (2005) by considering the dynamics of the poverty cut-off relative to the mean income.

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

$lnx_t - lnx_{t-1} = \lambda(f(w_t) - lnx_{t-1}) + e_t$ [3]

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). These models are close to the cross-sectional growth (or Barro) regressions familiar in the growth and convergence literature (see Durlauf et al. (2005) for details) 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 $e_t \sim N(0,\sigma^2)$ in the long run $ln(x_t) \sim N(f(w_t), \sigma^2/\lambda^2)$. 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 log-normality 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) [3] 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. The distinction has major implications for the progress of poverty and inequality.

Clearly for a needs based (absolute) poverty line (say x^*) the poverty rate will depend upon the time profile of $f(w_T)$ in the limit (i.e. $\lim_{T\to\infty} \Phi([(x^* - f(w_T)))/(\sigma/\lambda)])$ for positive growth it will be 0 and for negative growth it will be 1. For a relative poverty line, 0.6 of median income for example (note median income will

be *exp(f(w_T))* and the poverty cut-off will be .6 of that value), the poverty rate would be *Φ([ln(0.6)/(σ/λ)])* which obviously remains constant over time. Inequality measures that are not mean income normalized will remain constant over time location normalized inequality measures will diminish with positive growth and diminish with negative growth since the Gini coefficient may be written as:

$$2F(exp(\ln(x_0)+T\mu),\sigma^2/\lambda^2)-1$$

which will be 0 for negative growth, 1 for positive growth and constant for zero growth.

3   Where Does Pareto's Law Fit In?

Suppose the income process is governed by [1] but now, should x_t fall below x^* which is a lower reflective boundary (such as an enforced poverty frontier for example a mandated social security benefit payment), then the process is modified to [1] plus:

$$x_t = x^* \ \ if\ (1+\delta_{t-1})x_{t-1} < \ x^* \ \ \ [1a]$$

Gibrat's Law will no longer hold, in fact after a sufficient period of time the size distribution of x would be Pareto ($F(x) = 1-(x^*/x)^\theta$) with a shape coefficient $\theta=1$. In the literature on city size distributions this distribution is known as Zipf's Law and in that literature Gabaix (1999) showed that Zipf's law (Zipf (1949)) follows from a Gibrat consistent stochastic process (essentially a random walk) that is subject to a lower reflective boundary. In fact this phenomena, that a random walk with drift that is subject to a lower reflective boundary generates a Pareto distributed variable, has been known in the statistical process literature for some time (see for example Harrison 1985).[9] In the present context this has many implications, the Pareto distribution has a very different shape from the log normal and it would be constant through time, all relative poverty measures, absolute

_____

[9] Champernowne (1953) discovered as much in the context of income size distributions.

poverty measures and inequality measures[10] would be constants over time so that Pareto based predictions provide very powerful tests of the effectiveness of a mandated social security safety net.

These stochastic theories also have something to say about societal mobility. From a somewhat different perspective than is usual, mobility in a society may be construed as its agents opportunity for changing rank. Suppose that opportunity is reflected in the chance that two agents change places and consider two independently sampled agents $x_{it} = x_{it-1} + e_{it}$ and $x_{jt} = x_{jt-1} + e_{jt}$, so that $E(e_{it}-e_{jt}) = 0$ and $V(e_{it}-e_{jt}) = 2\sigma^2$. For the Gibrat model the probability that agents switch their relative ranks in period t is given by:

$$P(x_{it} > x_{jt} \mid x_{it-1} < {}_{xjt-1}) \; = \; P(e_{it}-e_{jt} > x_{jt-1} - x_{it-1})$$

By noting that the Gini coefficient is one half the relative mean difference between agents and that, for the log normal distribution, this may be written as twice the integral of a standard normal curve over the interval *[0, ($\sqrt{(V(x)/2)}$)]*, the average distance between two agents $= 4E(x)(\Phi(\sqrt{(V(x)/2)})-\Phi(0))$ and this probability may be written as:

$$= \; P((e_{it}-e_{jt})/(\sigma/\sqrt{2}) > 4exp(ln(x_0)+T(\mu+0.5\sigma^2))(\Phi(T^{0.5}\sigma/\sqrt{2})-\Phi(0))/(\sigma/\sqrt{2}))$$

$$= \; P(Z > 4exp(ln(x_0)+T(\mu+0.5\sigma^2))(\Phi(T^{0.5}\sigma/\sqrt{2})-\Phi(0))/(\sigma/\sqrt{2}))$$

The point is this probability diminishes over time (the intuition being that under constant population size the agents are growing further and further apart on average) so that mobility diminishes over time. For Kalecki's law note that the independently sampled agents processes may be written as $x_{it} = f(w_t) + (1-\lambda)x_{it-1} + e_{it}$ and $x_{jt} = f(w_t) + (1-\lambda)x_{jt-1} + e_{jt}$, so that $E(e_{it}-e_{jt}) = 0$ and $V(e_{it}-e_{jt}) = 2\sigma^2$, with the inequality being written as:

$$P(x_{it} > x_{jt} \mid x_{it-1} < x_{jt-1}) \; = \; P((e_{it}-e_{jt})/(1-\lambda) > x_{jt-1} - x_{it-1})$$

In this case using the Gini relationship to the population mean and the mean difference this probability may be written as:

$$P((1-\lambda)(e_{it}-e_{jt})/(\sigma/\sqrt{2}) \; > (1-\lambda)4exp(f(w_t))(\Phi(\sigma/(\lambda\sqrt{2}))-\Phi(0))/(\sigma/\sqrt{2}))$$

---

[10] The Gini for a Pareto distribution is $1/(2\theta-1)$ which is 1 when the shape coefficient is one because in this case the Pareto distribution has no moments or an infinite mean.

$$= P(Z > (1–\lambda)4exp(f(w_t)) (\Phi(\sigma/((\lambda\sqrt{2}))–\Phi(0))/(\sigma/\sqrt{2}))$$

so that as long as the fundamentals process $f(w_t)$ is constant then so will mobility be in that society. A similar result may be established for Pareto's law. Table 1 summarizes all these results.

Table 1. Summary of Results

Wellbeing Type	Type of Stochastic Process		
	Gibrat's Law	Kalecki's Law	Pareto's Law
Absolute Poverty	Increasing or decreasing dependent upon growth rate	Increasing or decreasing dependent upon growth rate	Constant if the reflective boundary is at the poverty cut-off.
Relative Poverty	Increasing with time	Constant	Constant if the reflective boundary is at the poverty cut-off.
Non-Normalized Inequality	Increasing	Constant	Constant
Location Normalized Inequality and Polarization	Decreasing with +ve growth rate	Decreasing	Constant
Mobility	Diminishing	Constant (if the fundamentals are constant)	Constant (if the fundamentals are constant)

4   Mixture Distributions and Trickle Down Theories (Anderson 1964))

The popularity of the "Rising Tide Raises All Boats" argument for the alleviation of poverty through growth has already been alluded to. For basic needs based definitions of the poverty cut-off this is no doubt true though it is not true if relative poverty is the measurement criterion. A slightly more sophisticated development of this argument is the "Trickle Down" effect (Anderson 1964). The idea is that it is necessary for economic growth to initially benefit the higher income groups (because they make the marginal product of labour enhancing investments that increased the incomes of the poor) but it transits downward to the

lower income groups over time. However this idea is predicated on the notion of economically different groups in society (with different investment behaviours for example) and it is not unreasonable to presume that different stochastic processes govern their respective behaviours.[11] In effect one is "modeling" different groups, one of which would be the poor group, which will perhaps calls for a different approach to measuring poverty and inequality. The poor are identified by the extent to which their income processes are noticeably different from the income processes of other groups in society rather than because their income is less than some pre-specified boundary. It follows that some identifiably "rich" individuals may have, at least temporarily, incomes that are lower than some of the members in the poor group.

To explore the implications of this structure imagine the societal income process to be that of a mixture of K normal distributions corresponding to the K income classes so that:

$$f_k(\ln(x_{kT})) = N((\ln(x_{k0}) + T(\mu_k + 0.5\sigma_k^2)), T\sigma_k^2) \quad k = 1, .., K$$

The classes are distinguished by their initial conditions in the sense that $\ln(x_{k0}) > \ln(x_{j0})$ for all j < k where j, k = 1,..,K and the proportion of the population in each class is given by w_k. Note that:

$$\int_0^x (f_i(z) - f_j(z))dz \leq 0 \ \forall \ x > 0 \ and \ i > j.$$

The class k=1 may be thought of as "the Permanently or Chronically Poor" so that CP, the Chronic Poverty rate is w_1. In this form there will be chronically poor agents whose incomes will be, at least temporarily, higher than some non-poor agents since the poor and non-poor distributions will overlap. The extent to which the Permanently Poor distribution overlaps the distribution of the other classes those members of those classes may be considered the transitorily poor, so that the Transitorily Poor rate TP may be written as:

_____

[11] Sen's functionings and capabilities approach (see Grusky and Kanbur (2006) for an extensive discussion) can also be construed as arguing for different processes describing rich and poor group behaviours distinguished by their different sets of circumstances.

$$TP = \sum_{k=2}^{K} w_k \int_0^\infty f_1(x_k)dx_k$$

For momentary convenience assume $\mu_k = \mu$ and $\sigma_k = \sigma$ for all k (i.e. the various classes are distinguished by their initial conditions alone). So here the poor are class k=1 and the rich are class k=K. Assume now that at period T_1 society moves to a new higher growth rate $\mu^* > \mu$, if all classes move together the income groups would have size distributions of the form:

$$\ln(x_{kT}) \sim N((\ln(x_{k0}) + T_1(\mu - \mu^*) + T(\mu^* + 0.5\sigma^2)), T\sigma^2) \; k = 1,.., K, \, and \, T > T_1$$

All classes retain their mean differences over time, the extent to which class A first order dominates class B remains constant and there is no change in the degree of polarization between the classes. However if in the first period only the highest income group (K) moves, j periods later only the next highest group (K–1) moves, j periods later only the next highest group (K–2) moves etc… Then the new societal income process will be the same mixture of K normal distributions corresponding to the K income classes but with:

$$\ln(x_{kT}) \sim N((\ln(x_{k0}) + (T_1 + jK - jk)(\mu - \mu^*) + T(\mu^* + 0.5\sigma^2)), T\sigma^2)$$
$$for \, k = 1,.., K, \, and \, T > T_1 + K$$

There are several observations to make. The distribution of average log incomes of the classes will be more widely spread initially and these initial differences will only be dissipated asymptotically (If the growth rates differ between the classes with the differences increasing with income class the differences will not dissipate asymptotically). In the short term the income size distribution of high income groups will more strongly first order dominate that of lower income groups increasing the polarization or lack of identification between poor and rich groups and the effect will be larger the longer is the lag in the trickle down effect (j). In effect there will be greater absolute inequality in the short run. When poverty lines are defined relative to an income quantile this will increase the probabilities of both transient and chronic poverty for the lower income classes. All of this is predicated on all income groups differing only in their initial incomes, should there be heterogeneity in growth rates and variances as well as starting incomes then anything is possible.

It should be said that in this model structure so-called poverty cutoffs are superfluous since the poverty group is better defined as those agents governed by the process that is dominated by all other processes. Issues concerning measuring the plight of the poor would centre upon w_1, the mixture coefficient for the poor group, and measuring the differences between the poor group sub-distribution and the other distributions in the mixture. Inequality can also be measured in terms of these concepts or it can be measured in terms of the general variability characteristics (variance or coefficient of variation for example) of the overall mixture distribution. Similarly issues concerned with addressing the plight of the poor may then be seen in terms of influencing the weights attached to each class as well as changing the nature of the process that governs the poor group outcomes.

For expositional convenience suppose there are 2 groups in society governed by processes which dictate their poor or non-poor status as in diagram 1, these two processes result in a poor wellbeing distribution $f_p(x)$ of the measurable characteristics x (for expositional convenience the analysis will be performed in terms of a univariate distribution but is should be stressed that the analysis can be readily performed in a multivariate environment), and a rich wellbeing distribution $f_r(x)$, the proportions of agents under these distributions are w_p and $1-w_p$ respectively so that the size distribution of incomes in this society is:

$$f(x) = w_p f_p(x) + (1 - w_p) f_r(x) \quad [4]$$

For understanding the plight of the poor the components of interest are w_p (which is essentially the proportion of people governed by the poor process which can be viewed as the real poverty rate) and the nature of the distribution of incomes among the poor, $f_p(x)$ (its mean gives us the average incomes of the poor, its variance gives us a measure of inequality amongst the poor and will permit generation of indices akin to FGT2 and FGT3 indices (Foster Greer Thorbeke 1984).

Suppose identification of the poor was pursued by employing an arbitrarily determined poverty cutoff c, then:

$$P_{rmp} = \int_{-\infty}^{c} f_r(x)dx \quad proportion \ of \ the \ rich \ miss-identified \ as \ poor$$

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

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. $w_p = P_{rmp}/(P_{pmr}+P_{rmp})$).

To the extent that the distributions overlap, for the sake of argument just in the region [a,b], so that both distributions have support in that interval, then there is potential for agents to not be point identified as either poor or rich. Essentially the interval contains those people who got a bad draw from the rich distribution and those people who got a good draw from the poor distribution. The overlap of the two weighted distributions, a measure of the extent of polarization of the two groups (Anderson et al. 2009), is given by $\int min((w_p f_p(x),(1–w_p)),f_r(x))dx$ and here corresponds to a measure of the potential for the lack of point identification. Suppose we could estimate the mixture distribution and by so doing identify w_p, $f_p(\ )$, $f_r(\ )$, a and b then we could identify some of the rich and some of the poor. We could assert that all agents with x below "a" would be point identified as poor, all with x above "b" would be point identified as rich. All agents with x in the interval [a,b] we would not know for sure, however more could be said in the sense that in this segment agents are partially identified, if the sub distributions could be estimated it would be possible to attach to each agent a probability that they were in one of the particular groups, very much in the spirit of the partial identification literature (see for example Manski 2002).

An allocation rule.

Consider the interval x+dx, the probability that someone from this interval is poor is given by $\theta_p(x,dx)$ where:
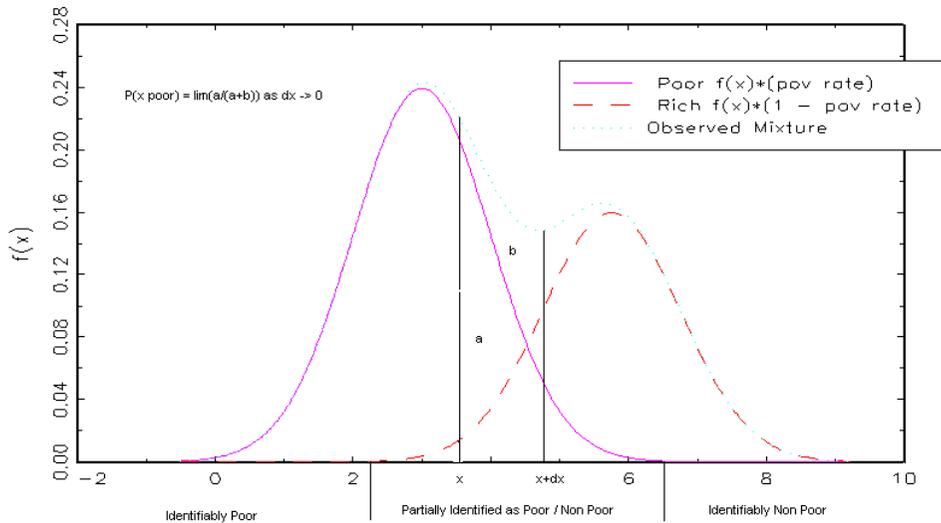
$$\theta_p(x,dx) = \int_x^{x+dx} w_p f_p(z)\,dz \ / \int_x^{x+dx} \left(w_p\ f_p(z) + \left(1 - w_p\right)f_r(z)\right)dz$$

Note that:

$$\lim_{dx\to 0}\theta_p(x,dx) \ = \ \theta_p(x) = \frac{w_p f_p(x)}{\left(w_p\ f_p(x) + \left(1 - w_p\right)f_r(x)\right)} \quad [5]$$

This corresponds to the probability that a person with x is poor so that *(1–w_p)$f_r(x) / (w_p f_p(x)+ (1–w_p)f_r(x))$* corresponds to the probability that a person with x is rich. Given estimates of w_p, $f_p(x)$ and $f_r(x)$, these probabilities can be estimated.

Diagram 1. Poor, Rich and Mixture (60% poverty rate)

Note also that, given a poverty line c, analogues to the FGT indices for the "poor" who are below the poverty line can be generated, thus:

$$\int_{-\infty}^{\infty} \theta_p(x) f(x) dx = w, \frac{1}{w} \int_{-\infty}^{\infty} x \theta_p(x) f(x) dx = \mu_p \text{ and}$$

$$\frac{1}{1-w} \int_{-\infty}^{\infty} x(1-\theta_p(x)) f(x) dx = \mu_r$$

are respectively the poverty rate, the mean income of the poor and the mean income of the non-poor which may respectively be estimated by:

$$\sum_{i=1}^{n} \hat{\theta}(x_i) / n, \quad \frac{1}{\sum_{i=1}^{n} \hat{\theta}(x_i) / n} \sum_{i=1}^{n} x_i \hat{\theta}(x_i) / n \text{ and } \frac{1}{1-\sum_{i=1}^{n} \hat{\theta}(x_i) / n} \sum_{i=1}^{n} x_i (1-\hat{\theta}(x_i)) / n$$

Furthermore

$$\frac{1}{w}\int_{-\infty}^{\infty}(x-\mu_p)^2\theta_p(x)f(x)dx = \sigma_p^2 \ and$$

$$\frac{1}{1-w}\int_{-\infty}^{\infty}(x-\mu_r)^2(1-\theta_p(x))f(x)dx = \sigma_r^2$$

correspond to the poor and non-poor income variances respectively which may be respectively estimated by:

$$\frac{1}{\sum_{i=1}^{n}\hat{\theta}(x_i)/n}\sum_{i=1}^{n}(x_i-\hat{\mu}_p)^2\hat{\theta}(x_i)/n \ and \ \frac{1}{1-\sum_{i=1}^{n}\hat{\theta}(x_i)/n}\sum_{i=1}^{n}(x_i-\hat{\mu}_p)^2(1-\hat{\theta}(x_i))/n$$

Finally the j'th order Foster Greer Thorbeke index for the poor and its estimator is given by:

$$FGT(j) = \int_{0}^{c}(\frac{c-x}{c})^j\theta_p(x)f(x)dx / w \ and$$

$$\sum_{\hat{i}=1}^{n}I(x_i<c)\widehat{(\theta_p(x)}(\frac{c-x_i}{c})^j / n\hat{w}, \ j=1,..$$

5   The Experience of Africa 1985–2005

To illustrate these issues data on per capita GDP for 47 African countries[12] together with their populations were drawn from the World Bank African Development Indicators CD-ROM for the years 1985, 1990, 1995, 2000, 2005 were used. An issue immediately arises as to whether the raw data or population weighted data should be employed. At the statistical level the parameters of interest will be estimated as though the data were an independent random sample

---

[12] The countries in the sample were: Algeria, Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Congo, Rep., Cote d'Ivoire, Egypt, Arab Rep., Equatorial Guinea, Ethiopia, Gabon, Gambia, The, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Morocco, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Seychelles, Sierra Leone, South Africa, Sudan, Swaziland, Togo, Tunisia, Uganda, Zambia, Zimbabwe.

and the properties of those estimators predicated upon that assumption. If the population of interest is that of Africa then this is not so and sample weighting is necessary to adjust for the under sampling of highly populated countries and over sampling of sparsely populated countries. A similar argument prevails at the economic theoretic level if some sort of representative agent model is presumed and the wellbeing of all Africans is of interest. For the purposes of comparison, and to highlight the substantive differences the distinction makes both will be reported here.

Table 2 reports summary statistics (means and variances) for the sample years and clearly indicates the implicit over sampling of higher income nations in the unweighted estimates. Notice the unweighted estimates record a growth of 14% over the period whereas the weighted results report less than 8% growth, a substantial difference.

Diagrams 2 and 3 present the beginning of period and end of period size distributions[13] of ln(*GDP per capita*) again in unweighted and weighted form. The shifts in location and spread discerned in Table 1 can readily be perceived in these diagrams.

Pearson Goodness of Fit Tests (Pearson (1900) of the hypothesis that these distributions are Log Normal Or Pareto, performed for both weighted and unweighted samples, are reported in Table 3. At the 1% critical value there is a preponderance of evidence favouring the Log Normal formulation (it only gets rejected twice in the un-weighted sample and once in the weighted sample and pretty marginally so at that) whereas the Pareto gets solidly rejected in every

Table 2. ln*(GDP per capita)*

	Unweighted		Weighted	
Year	Means	Variances	Means	Variances
1985	6.1820771	0.92778351	6.0692785	0.90744347
1990	6.1784034	0.98733961	6.0827742	0.88023763
1995	6.1113916	1.1495402	6.0053685	0.95556079
2000	6.2174786	1.2386242	6.0488727	1.0276219
2005	6.3223021	1.3481553	6.1461927	1.0300315

_____

[13] These are essentially Epetchanikov kernel estimates of the respective size distributions.

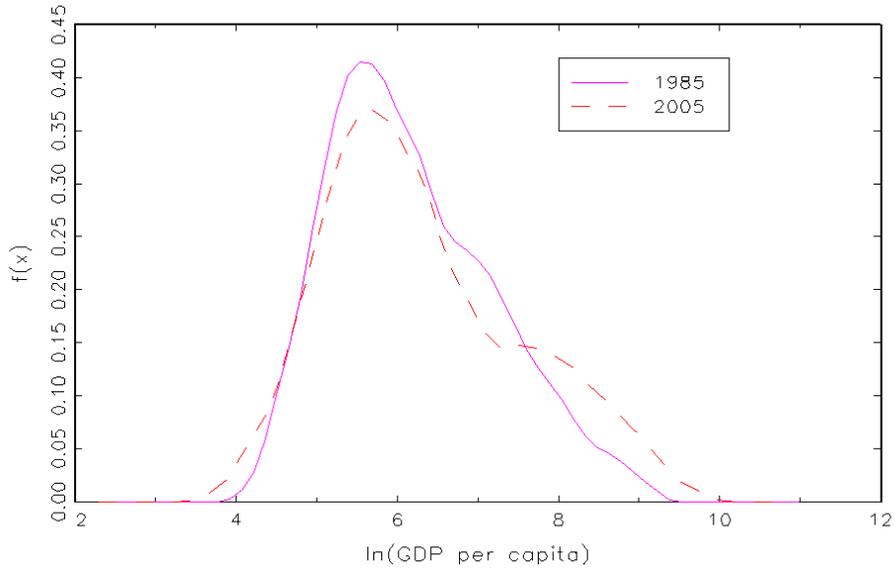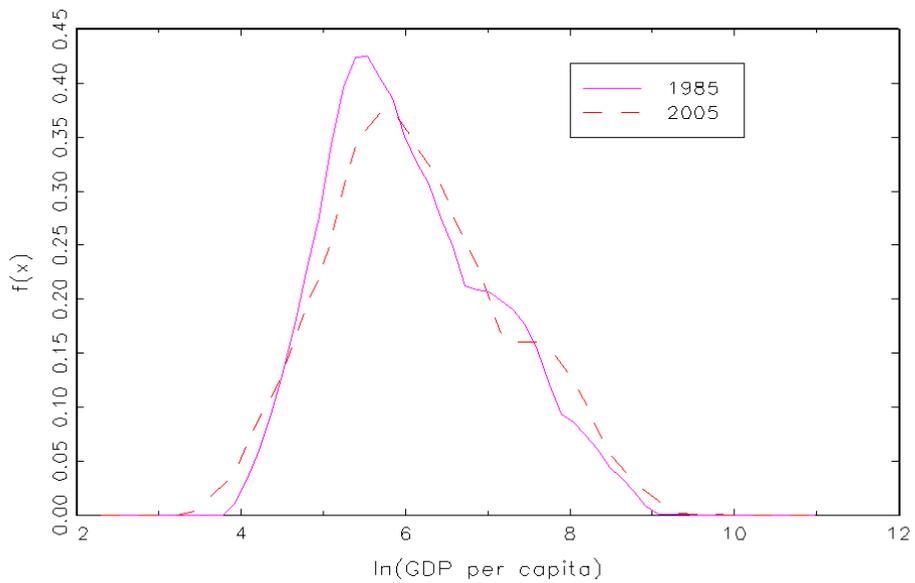*Diagram 2*. Africa GDP per capita distributions 1985–2005

Diagram 3. Africa GDP per capita distributions 1985–2005 (population weighted)

instance. This is slightly surprising since, to the eye, Diagrams 1 and 2 suggest mixtures of 2 normals (one large poor group and a much smaller rich group), but the evidence does not appear strong enough in the data to really reject pure normality. Indeed the joint test of normality over the 5 observation periods does not reject normality at the 1% level for the weighted sample. Interestingly enough the populations are clearly log-normally distributed so that it may be inferred that gdp's are themselves log-normally distributed since the difference or sum of two normally distributed variables is also normally distributed.

Table 3. Tests of Norma land Pareto Distributional Specifications

Ln GNP per capita				
Unweighted	Normal $\chi^2(4)$, [P(Upper Tail)]		Pareto $\chi^2(4)$, [P(Upper Tail)]	
1985	12.320748	[0.015118869]	599.37137	[2.1196982e–128]
1990	18.172132	[0.0011420720]	648.71719	[4.4178901e–139]
1995	7.2879022	[0.12143386]	254.88561	[5.7677072e–054]
2000	10.035066	[0.039841126]	413.23135	[3.8477939e–088]
2005	19.439149	[0.00064419880]	418.01949	[3.5518925e–089]
All years $\chi^2(20)$	67.254997	[5.0783604e–007]	2334.2250	[0.00000000]
Weighted	Normal $\chi^2(4)$, [P(Upper Tail)]		Pareto $\chi^2(4)$, [P(Upper Tail)]	
1985	5.4247526	[0.24642339]	850.67992	[8.0715931e–183]
1990	10.882507	[0.027916699]	890.92447	[1.5416785e–191]
1995	3.7072456	[0.44707297]	389.07173	[6.3881404e–083]
2000	2.8078575	[0.59047714]	869.74574	[5.9768551e–187]
2005	13.992590	[0.0073187433]	935.78289	[2.9403608e–201]
All years $\chi^2(20)$	36.814952	[0.012314284]	3936.2047	[0.00000000]
Ln Population				
Unweighted	Normal $\chi^2(4)$, [P(Upper Tail)]		Pareto $\chi^2(4)$, [P(Upper Tail)]	
1985	2.6979621	0.60957127	194.39270	6.0291654e–041
1990	3.4815355	0.48069145	193.84403	7.9101293e–041
1995	2.5274974	0.63971883	308.86671	1.3243132e–065
2000	2.9924681	0.55908664	250.28037	5.6645137e–053
2005	3.7924813	0.43481828	245.77727	5.2865977e–052

Given the joint normality of the 5 observation periods is accepted, the restrictions implied by [2] can be examined. Under the heroic assumption that the 5 year periods are independent the Likelihood for the sample may be written as:

$$L = \prod_{j=0}^{4} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(\theta+j)\sigma^2}} e^{-\frac{(x_{ij}-\ln x_0-(\theta+j)(\mu+0.5\sigma^2))^2}{2(\theta+j)\sigma^2}} ,$$

and the null and alternative hypotheses may be written as:

H_o: $f(x_{T+i}) \sim N((ln(x_o)+(\theta+i)\mu+0.5\sigma^2),(\theta+i)\sigma^2)$ versus H_1: $f(x_{T+i}) \sim N(\mu_i,\sigma_i^2)$

for i=0, 1, …,4

Here $\ln(x_0)$ and θ are parameters thus implying 6 restrictions on the alternative. Table 4 reports the estimates of $\ln(x_0)$, θ, μ and σ^2 for the weighted and un-weighted samples, together with the test of the restrictions. In both cases the restrictions are not rejected implying that, conditional on the underlying normality of the distributions, Gibrat's law is an adequate description of the data.

To examine what these different approaches imply for poverty measurement headcount measures of both the absolute and relative type are considered in Table 5, the former based upon the popular dollar and 2 dollar a day cutt-offs (which for our sample become 5.8749307 and 6.5680779 respectively) and the latter based upon the also popular 50% and 60% of the median cutoffs. As expected in general absolute poverty measures decline with economic growth and relative poverty measures decline.

Table 4. Estimates and Tests of Restrictions

	Unweighted Sample	Weighted Sample
Ln(x_o)	5.7827258	4.0511799
θ (# of five year intervals)	9.9694699	103.01656
μ	–0.012055759	0.014659728
σ^2	0.094224026	0.0091376682
$X^2(6)$	0.19646781	0.78654360
$P(>\chi^2(6))$	0.99985320	0.99242963

Table 5. Poverty Estimates

Year	Year by year normality		Under Gibrats Law	
Absolute (1 and 2 dollar a day cut-off) Unweighted poverty rate estimates				
1985	0.37491020	0.65569496	0.39532698	0.67353927
1990	0.38002613	0.65253180	0.38684395	0.65329716
1995	0.41272336	0.66492689	0.37893064	0.63472718
2000	0.37912189	0.62362764	0.37150161	0.61758498
2005	0.35000783	0.58381961	0.36449051	0.60167400
Absolute (1 and 2 dollar a day cut-off) Weighted poverty rate estimates				
1985	0.41916948	0.69972853	0.43567992	0.70969353
1990	0.41233910	0.69751462	0.42823308	0.70198016
1995	0.44692413	0.71757254	0.42088097	0.69425410
2000	0.43188059	0.69573737	0.41362387	0.68651911
2005	0.39462744	0.66118137	0.40646200	0.67877884
Relative (50% and 60% median cut-off) Unweighted Estimates				
1985	0.23588005	0.29794024	0.23725235	0.29907775
1990	0.24272087	0.30359444	0.24768529	0.30767299
1995	0.25898028	0.31687954	0.25697797	0.31525458
2000	0.26670464	0.32312080	0.26532206	0.32200680
2005	0.27526233	0.32998688	0.27286739	0.32807031
Relative (50% and 60% median cut-off) Weighted Estimates				
1985	0.23341745	0.29589478	0.23748342	0.29926912
1990	0.23001494	0.29305958	0.23854875	0.30015083
1995	0.23913688	0.30063716	0.23960137	0.30102105
2000	0.24706083	0.30716108	0.24064154	0.30188005
2005	0.24731361	0.30736834	0.24166951	0.30272805

As for mobility the transitions can be evaluated for each of the five year intervals via the distance of the joint density matrix from that of a diagonal. The statistic

$$\sum_{ij} min(p_{ij}, diag(p_{i.})).$$

provides an index of immobility, where p_{ij}, is the probability of a country being in category i in period t and in category j in period $t+1$, diag($p_{i.}$), a square matrix

with the vector of probabilities of being in category i in period t on the on diagonal corresponds to complete immobility (Anderson, Ge and Leo (2010) establish that this is asymptotically normal). Splitting this sample into 5 equal sized categories the four 5 year transitions generate statistics (standard errors) 0.87234 (0.04868) for the first three transitions and 0.82979 (0.05482) for the fourth. This corresponds to very high immobility consistent with Gibrats law though it was expected to be increasing rather than staying constant.

6   The African Distribution as a Mixture of Normals

As observed earlier the diagrams are very suggestive of a mixture of normals one largish poor group and one smaller rich group and it is of interest to see the consequences of modelling the processes under this structure. First it is appropriate to examine the degree of mobility within the distribution over time. Table 6 reports the 20 year transition probability matrix and indicates that in essence there appears to be very little mobility over the period between 5 rank groups (the five rank cells were 1–10 11–19 20–28 29–37 38–47). Essentially 5 countries moved up from cell 1 to cell 2 and 4 moved down from cell 2 to cell 1, 1 moved up from cell 2 to cell 3, two moved up from cell 3 to cell 4 and one moved down from cell 4 to cell 3. The only change of more than one cell was Liberia who dropped from cell 4 to cell 1 over the period and one other original cell 4 Member moved up to cell 5 and one cell 5 member moved down to cell 4. In sum there appears to be some deal of mobility at the lowest end of the spectrum but very little elsewhere, certainly it is reasonable to assume that memberships of the large poor and small rich groups apparent in Diagrams 1 and 2 (and hence the mixture coefficients) appear to be relatively constant.

Techniques for estimating mixtures of normals are available (see for example Johnson et al. 1994) but tend to be complex and depend upon fairly large numbers of observations. Here, since there are a limited number of observations, an ad hoc method is used for simplicity and convenience, but it turns out to be quite successful in terms of replicating the empirical distribution. Given the evidence is that the membership of the groups is very stable over the period, countries are allocated into rich and poor groups as follows. Inspection of the 2005 distribution in Diagram 1 suggests that the modal values of the respective poor and rich groups

Table 6. 20 year Transition Probability Matrix

1985 Cell	2005 Cell				
	1–10	11–19	20–28	29–37	38–47
1–10	0.10638298	0.10638298	0.00000000	0.00000000	0.00000000
11–19	0.085106383	0.085106383	0.021276596	0.00000000	0.00000000
20–28	0.00000000	0.00000000	0.14893617	0.042553191	0.00000000
29–37	0.021276596	0.00000000	0.021276596	0.12765957	0.021276596
38–47	0.00000000	0.00000000	0.00000000	0.021276596	0.19148936

Immobility Index   0.65957447   Standard Error 0.069118460

are approximately 6 and 7. Observations below 6 can be almost all be attributed to the poor group and similarly observations above 7 can be similarly attributed to the rich group and were allocated accordingly. Given the symmetry of the underlying log-normals around their modes, the relative size of the below six and above 7 observations can be used to establish the poor and rich group weights w_r (20/47) and w_p (27/47). Observations between 6 and 7 were allocated randomly according to these weights to rich and poor groups. After an initial fit one below median poor country was switched with an above median rich country[14] improving the fit and establishing the following two rich and poor subgroups dispositions.

Table 7. Disposition of Poor and Rich Countries

Poor Group	Rich Group
Benin, Burkina Faso, Burundi, Cape Verde, Central African Republic, Chad, Democratic republic of the Congo, Cote d'Ivoire, Ethiopia, The Gambia, Ghana, Guinea-Bissau, Kenya, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Niger, Rwanda, Sierra Leone, Sudan, Togo, Uganda, Zambia, Zimbabwe.	Algeria, Angola, Botswana, Cameroon, Comoros, Republic of the Congo, Egypt, Equatorial Guinea, Guinea, Lesotho, Mauritius, Morocco, Namibia, Nigeria, Senegal, Seychelles, South Africa, Swaziland, Tunisia.

———————————————

[14] Relative to a normal distribution the initial poor country distribution appeared attenuated in the upper tail and the rich country distribution appeared attenuated in the lower tail.

Having partitioned the sample in this fashion estimation of the mixture distribution is quite simple in both un-weighted and population weighted modes. Tables 8 and 9 and Diagrams 4 and 5 report the results. In both cases the fits are extremely good and correspond to a more than adequate description of the data. The poor group has enjoyed zero economic growth and the rich group has enjoyed a steady one per cent annual growth rate over the period. Differences between the un-weighted and weighted cases emerge when GDP per capita levels and variances are concerned. In the un-weighted case income levels are generally higher and variances are lower but increasing over time whereas in the weighted case incomes are lower and variances are higher but diminishing over time. The restrictions implied by Gibrat's law for the separate poor and rich groups in both weighted and unweighted samples are rejected in all cases (frequently resulting in nonsense estimates such as negative variances and negative time parameters) though basic log normality is not rejected in any case suggesting that Kalecki's Law is the best description of the data for the individual groups. Poor and rich groups appear to be moving apart, Table 10 reports a trapezoidal measure of bi-polarization (Anderson 2010; Anderson, Leo and Linton 2012) for both weighted and unweighted samples illustrating the point and Table 11 reports differences over time. All of which may be interpreted as the poor becoming relatively poorer.

Table 8. A Mixture of 2 Log Normals (Poor group and Rich group)

	Poor Mean	Rich Mean	Poor Variance	Rich Variance	$\chi^2(4)$	[P(Upper Tail)]
1985	5.5869	6.9856	0.2691	0.6950	3.6312	0.4582
1990	5.5425	7.0369	0.2350	0.7183	5.9701	0.2014
1995	5.4325	7.0278	0.3461	0.7705	0.2800	0.9911
2000	5.5006	7.1852	0.3258	0.8370	0.4526	0.9780
2005	5.5680	7.3406	0.3219	0.9234	1.9169	0.7510

Table 9. Sample Weighted Mixtures

	Poor Mean	Rich Mean	Poor Variance	Rich Variance	X²(4) Tail)]	[P(Upper
1985	5.4115	6.7584	0.6262	1.6988	3.0473	0.5499
1990	5.4146	6.7975	0.5422	1.4666	1.6053	0.8078
1995	5.3109	6.7532	0.5743	1.4099	4.3005	0.3669
2000	5.3423	6.8289	0.5671	1.2868	8.2530	0.0827
2005	5.4340	6.9527	0.4943	1.1327	5.9860	0.2002

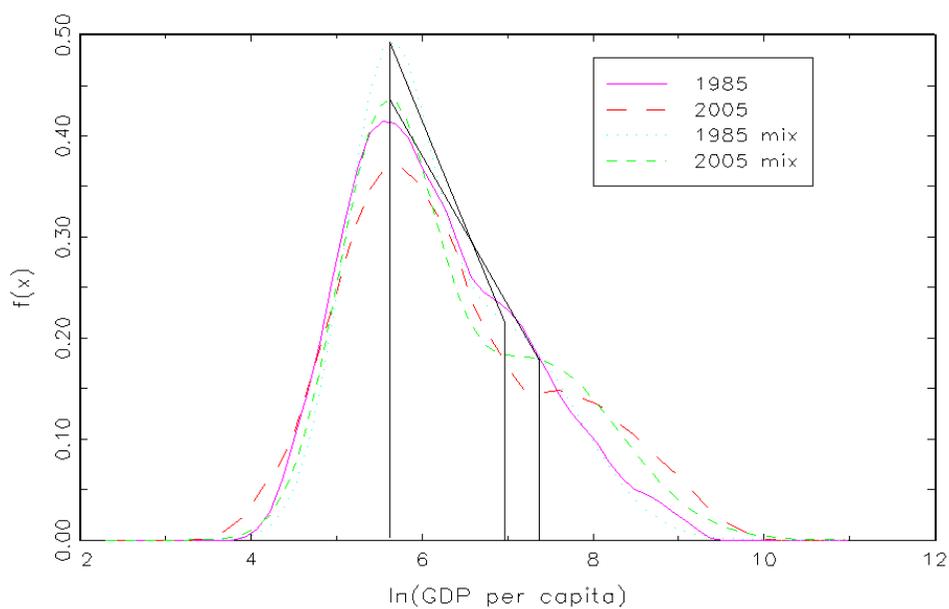*Diagram 4*. Africa GDP per capita distributions 1985–2005

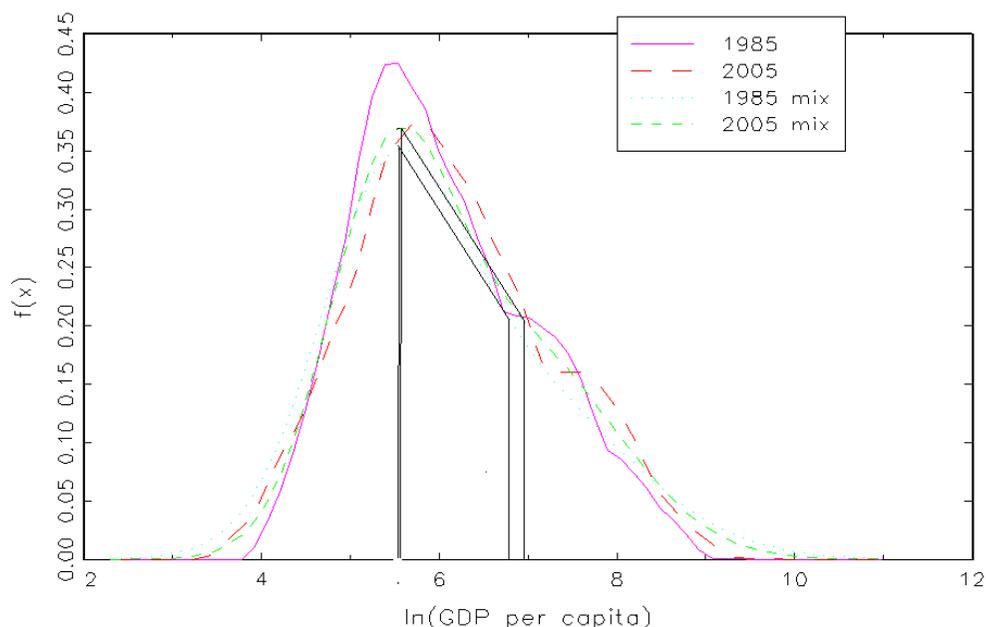*Diagram 5*. Africa GDP per capita distributions 1985–2005 (population weighted)

Table 10. Bi-Polarization Index = $0.5(f_p(x_{pmode}) + f_r(x_{rmode}))(x_{rmode} - x_{pmode})$

Year	Unweighted Sample		Weighted Sample	
1985	0.87249993	(0.027374440)	0.54564641	(0.046480295)
1990	0.96662956	(0.028177909)	0.60239981	(0.042441864)
1995	0.90342925	(0.028177909)	0.62192838	(0.041522365)
2000	0.95600414	(0.030424419)	0.65517946	(0.039377304)
2005	0.99116013	(0.032321510)	0.71551955	(0.036396440)

In this circumstance the issue of population weighting makes a big difference, with no population weighting the poor group are becoming absolutely poorer and exhibiting diminishing within group association. The source of polarization is the increased alienation or distance between the two groups. With population weighting they are not becoming absolutely poorer but are exhibiting increased within group association, there is a small amount of between group alienation but a

Table 11. Polarization Tests[*]

Unweighted Sample

Comparison Years	Difference	("t"test)	{P(T<t)}
1990–1985	0.094129622	(1.6144785)	{0.94678816}
1995–1985	0.030929320	(0.53083245)	{0.70223256}
1995–1990	−0.063200302	(−1.0761795)	{0.14092349}
2000–1985	0.083504205	(1.4159301)	{0.92160202}
2000–1990	−0.010625417	(−0.1787883)	{0.42905197}
2000–1995	0.052574885	(0.88520127)	{0.81197595}
2005–1985	0.11866019	(1.9819244)	{0.97625615}
2005–1990	0.024530571	(0.40667537)	{0.65787679}
2005–1995	0.087730873	(1.4553093)	{0.92720818}
2005–2000	0.035155988	(0.57662100)	{0.71790225}

Weighted Sample

Comparison Years	Difference	("t"test)	{P(T<t)}
1990–1985	0.056753398	(0.79846297)	{0.78769906}
1995–1985	0.076281972	(1.0823101)	{0.86044263}
1995–1990	0.019528574	(0.28349312)	{0.61160057}
2000–1985	0.10953305	(1.5752869)	{0.94240488}
2000–1990	0.052779657	(0.77714642)	{0.78146381}
2000–1995	0.033251083	(0.49415344)	{0.68940109}
2005–1985	0.16987315	(2.4790814)	{0.99341394}
2005–1990	0.11311975	(1.6913645)	{0.95461639}
2005–1995	0.093591173	(1.4127973)	{0.92114233}
2005–2000	0.060340090	(0.92496252)	{0.82250730}

* Tests are based on the trapezoid measure being asymptotically normally distributed with a variance $\approx (f(x_{1m})+f(x_{2m}))^2 (f(x_{1m})/[f'(x_{1m})]^2 + f(x_{2m})/[f'(x_{2m})]^2) \|K'\|_2^2$ where x_{mj} j = 1,2 are the modes of the respective distributions, where f() is the normal and K is the Gaussian kernel (Anderson, Linton and Wang (2012)).

substantial increase in within group association. The only significant changes in polarization in both comparison types were increases in polarization over time. In both cases the poor and rich groups are following distinct stochastic processes and there is no sense in which "the rising tide is raising all boats" or improvements in the well-being of the rich African countries are trickling down to the poor countries. For the sample to hand with population weighted data we have poverty rates of:

1985	1990	1995	2000	2005
0.51162747	0.51682920	0.51850631	0.52472944	0.53106760

This indicates a poverty group over the period that is not only declining relatively in its income status but is growing relatively in its size.

7   Conclusions

It is not at all clear that boats and tides aphorisms and trickle down theories apply either in theory or practice when the well-being indicator is well described by some sort of stochastic process, especially when the process is one that is frequently observed in practice. It really depends on the nature of poverty or inequality being considered as well as the precise nature of the stochastic process(es) involved. Stochastic processes that are non-stationary engender distributions whose dispersion (absolute inequality) increases over time, whether or not relative inequality increases depends upon the nature of the growth process. Similar statements can be made about poverty, but here the nature of the growth process affects both absolute and relative poverty. What is clear is that it is not unequivocally the case that rising tides raise all boats or that wellbeing unequivocally trickles down even in the simplest of circumstances. This is even more so the case when the progress of the poor and the non-poor are described by different stochastic processes.

In the case of Africa when GNP per capita is modelled over the recent two decades as a singular stochastic process the prediction of Gibrat's law appears to hold true regardless of whether the analysis is performed under a population weighting scheme or a non-weighted scheme in the sense that the distribution is

log normal. Under this description absolute poverty is diminishing and relative poverty is increasing and absolute inequality is increasing and relative inequality is diminishing. Kernel estimates of the density indicated some evidence of bimodality suggesting a mixture of at least two distributions. When log GNP per capita is described by a mixture of two normals (which was not rejected by the data), one describing the poor country process and the other describing the rich country process, it is apparent that the two groups are polarizing, with the poor group in this sense becoming relatively poorer. In this circumstance the issue of population weighting made a big difference, with no population weighting the poor group are becoming absolutely poorer and exhibiting diminishing within group association, with population weighting they are not becoming absolutely poorer but are exhibiting increased within group association.

References

Anderson, W.H.L. (1964). Trickling Down, the Relationship between Economic Growth and the Extent of Poverty of American Families. *Quarterly Journal of Economics,* 78: 511–524. http://www.jstor.org/stable/1879653

Anderson, G.J., O. Linton, and Y. Wang (2012). Non-Parametric Estimation of a Polarization Measure. *Journal of Econometrics,* 171 (1): 1–23. http://www.sciencedirect.com/science/article/pii/S0304407612001108

Anderson, G.J., Y. Ge, and T.W. Leo (2010). Distributional Overlap: Simple Multivariate Para-metric and Non-Parametric Tests for Alienation, Convergence and General Distributional Difference Issues. *Econometric Reviews,* 29: 1532–4168. http://www.tandfonline.com/doi/abs/10.1080/07474930903451532

Anderson, G.J., T.W. Leo, and O. Linton (2012). A Polarization-Cohesion Perspective on Cross Country Convergence. *Journal of Economic Growth,* 17: 49–69 http://ideas.repec.org/a/kap/jecgro/v17y2012i1p49-69.html

Anderson, G.J. (2010). Polarization of the Poor: Multivariate Relative Poverty Measurement Sans Frontiers. *Review of Income and Wealth,* 56: 84–101 http://ideas.repec.org/a/bla/revinw/v56y2010i1p84-101.html

Barro, R.J., and X. Sala-i-Martin (1992). Convergence. *Journal of Political Economy,* 100 (2): 223–251. http://ideas.repec.org/a/ucp/jpolec/v100y1992i2p223-51.html

Battistin, E., R. Blundell, and A. Lewbel (2009). Why Is Consumption More Log Normal than Income? Gibrat's Law Revisited. *Journal of Political Economy,* 117: 1140–1154. http://ideas.repec.org/a/ucp/jpolec/v117y2009i6p1140-1154.html

Browning M., and A. Lusardi (1996). Household Saving Micro Theories and Micro Facts. *Journal of Economic Literature,* 34: 1797–1855. http://ideas.repec.org/a/aea/jeclit/v34y1996i4p1797-1855.html

Burgess S., K. Gardiner, and C. Propper (2001). "Why Rising Tides Don't Lift All Boats: An Explanation of the Relationship Between Poverty and Unemployment in Britain". CASE Working Paper no 46, 2001. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1158937

Champernowne, D.G. (1953). A Model of Income Distribution. *Economic Journal*, 63: 318–351. http://www.jstor.org/stable/2227127

Davidson, R., and J.-Y. Duclos (2013). Testing for Restricted Stochastic Dominance. *Econometric* Reviews, 32: 84–96. http://www.tandfonline.com/doi/abs/10.1080/07474938.2012.690332

Deaton A., and C. Paxson (1994). Inter-temporal Choice and Inequality. *Journal of Political Economy*, 102: 384–394. http://www.jstor.org/stable/2138618

Duclos J-Y., J. Esteban, and D. Ray (2004). Polarization: Concepts, Measurement and Estimation. *Econometrica,* 72: 1737–1772. http://ideas.repec.org/a/ecm/emetrp/v72y2004i6p1737-1772.html

Durlauf, S. N., P.A. Johnson, and Temple, J. R. W. (2005). Growth Econometrics. In P. Aghion and S. N. Durlauf (eds.), *Handbook of Economic Growth*, Volume 1A, North-Holland, Amsterdam.

Ferguson, A. (1767). *A History of Civil Society*. Dublin.

Freeman, R. (2001). The Rising Tide Lifts…? NBER Working Paper no 8155. http://ideas.repec.org/p/nbr/nberwo/8155.html

Friedman, M. A Theory of the Consumption Function. Princeton University Press. http://ideas.repec.org/b/nbr/nberbk/frie57-1.html

Gabaix, X., (1999). Zipf's Law for Cities: An Explanation. *Quarterly Journal of Economics,* 114: 739–767. http://ideas.repec.org/a/tpr/qjecon/v114y1999i3p739-767.html

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

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

Gnedenko, B.V. (1962). *Theory of Probability.* Chelsea, New York.

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

Harrison, A. (1981). Earnings by Size: A Tale of Two Distributions, *Review of Economic Studies*, 48: 621–631. http://ideas.repec.org/a/bla/restud/v48y1981i4p621-31.html

Harrison T. (1985). *Brownian Motion and Stochastic Flow Systems*. Malabar, F.C. Krieger.

Hines J., H. Hoynes, and A. Krueger (2001). "Another Look at Whether a Rising Tide Lifts All Boats". NBER Working Paper no. 8412. http://ideas.repec.org/p/nbr/nberwo/8412.html

Johnson, N.L., S. Kotz, and N. Balakrishnan (1994). *Continuous Univariate Distributions*. Volume 1. 2nd Edition. Wiley Series in Probability and Mathematical Statistics. Wiley-Interscience.

Kalecki, M. (1945). On the Gibrat Distribution. *Econometrica,* 13: 161–170. http://ideas.repec.org/p/nbr/nberwo/8412.html

Knowles, E. (ed.) (2004). *Oxford Dictionary of Quotations. 6th Edition*. Oxford: Oxford University Press.

Mankiw, G., D. Romer, and D.N. Weil (1992). A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics,* 107 (2): 407–437. http://ideas.repec.org/a/tpr/qjecon/v107y1992i2p407-37.html

Marshall, A., (1890). *Principles of Economics*. Macmillan

Meghir, C., and C. Pistaferri (2004). Income Variance, Dynamics and Heterogeneity, *Econometrica,* 72: 1–32. http://ideas.repec.org/a/ecm/emetrp/v72y2004i1p1-32.html

Meyer, B., and J. Sullivan (2002). Measuring the Wellbeing of the Poor Using Income and Consumption. *Journal of Human Resources,* 1180–1220. http://www.jstor.org/stable/3558985

Modigliani, F., and R. Brumberg (1954). Utility Analysis and the Consumption Function: An Interpretation of Cross Section Data. In K.K.Kurihara (ed.). *Post Keynesian Economics*. Ruthers University Press, New Brunswick, New Jersey.

Neal D., and S. Rosen (2000). Theories of the Distribution of Earnings. Chapter 8 in: *Handbook of Income Distribution*, Volume 1. A.B. Atkinson and F. Bourguignon (eds). North Holland. http://ideas.repec.org/h/eee/incchp/1-07.html

O'Neill, D. (2005). The Welfare Implications of Growth Regressions. Finance and Accounting Department Working Paper Series 1570505. Department of Economics, Finance and Accounting, National University of Ireland - Maynooth. http://ideas.repec.org/p/may/mayecw/n1570505.html

Osberg, L. (1977). Stochastic Process Models and the Distribution of Earnings. *Review of Income and Wealth*, 23: 205–215. http://ideas.repec.org/a/bla/revinw/v23y1977i3p205-15.html

Pareto, V. (1897). *Cours d'Economie Politique*. Rouge et Cie, Paris.

Pearson, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling. *Philosophical Magazine,* 50, 157–175. http://www.economics.soton.ac.uk/staff/aldrich/New%20Folder/kpreader1.htm

Reed, W. J. (2001). The Pareto, Zipf and Other Power Laws. *Economics Letters,* 74: 15–19. http://ideas.repec.org/a/eee/ecolet/v74y2001i1p15-19.html

Sen, A.K. (1983). Poor Relatively Speaking. *Oxford Economic Papers,* 35: 153–169 http://ideas.repec.org/a/oup/oxecpp/v35y1983i2p153-69.html

Smith, A. (1776). *An Enquiry into the Nature and Causes of the Wealth of Nations*. Liberty Classics

Temple, J. (1999). The New Growth Evidence. *Journal of Economic Literature*, 37 (1): 112–156. http://ideas.repec.org/a/aea/jeclit/v37y1999i1p112-156.html

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

Zipf, G., (1949). *Human Behavior and the Principle of Least Effort*. Addison Wesley, Cambridge MA.

Please note:

You are most sincerely encouraged to participate in the open assessment of this article. You can do so by either recommending the article or by posting your comments.

Please go to:

The Editor