Income and poverty in a developing economy

A. K. Chattopadhyay\textsuperscript{1,2(a)}, G. J. Ackland\textsuperscript{2} and S. K. Mallick\textsuperscript{3}

\textsuperscript{1} Aston University, School of Engineering and Applied Sciences - Aston Triangle, Birmingham B4 7ET, UK, EU
\textsuperscript{2} SUPA, School of Physics and Astronomy, The University of Edinburgh - Mayfield Road, Edinburgh EH9 3JZ, UK, EU
\textsuperscript{3} School of Business and Management, Queen Mary University of London - Mile End Road, London E1 4NS, UK, EU

received 1 April 2010; accepted in final form 16 August 2010
published online 22 September 2010

PACS 89.65.Gh – Economics; econophysics, financial markets, business and management
PACS 02.50.-r – Probability theory, stochastic processes, and statistics

Abstract – We present a stochastic agent-based model for the distribution of personal incomes in a developing economy. We start with the assumption that incomes are determined both by individual labour and by stochastic effects of trading and investment. The income from personal effort alone is distributed about a mean, while the income from trade, which may be positive or negative, is proportional to the trader's income. These assumptions lead to a Langevin model with multiplicative noise, from which we derive a Fokker-Planck (FP) equation for the income probability density function (IPDF) and its variation in time. We find that high earners have a power law income distribution while the low-income groups have a Levy IPDF. Comparing our analysis with the Indian survey data (obtained from the world bank website: http://go.worldbank.org/SWgZB45DN0) taken over many years we obtain a near-perfect data collapse onto our model’s equilibrium IPDF. Using survey data to relate the IPDF to actual food consumption we define a poverty index (Sen A. K., Econometrica., 44 (1976) 219; Kakwani N. C., Econometrica, 48 (1980) 437), which is consistent with traditional indices, but independent of an arbitrarily chosen “poverty line” and therefore less susceptible to manipulation.

Copyright © EPLA, 2010

Poverty has been a feature of all human societies throughout time. The underlying cause is the unequal distribution of personal incomes which is an emergent feature of a free economy, invariably resulting in extreme wealth for a few and relative poverty for many.

Since the work of Pareto [1], the distribution of incomes has been known to have a power law tail at the high end [2]. There have been many models of the dynamic process [1,3–7] by which a power law tail can develop for high incomes. Yet in an interlinked economy the low-income distribution emerges from the same dynamics as the high income.

Significantly less effort has been applied to study the distribution of low incomes, but to study poverty this is the critical part of the income probability distribution function (IPDF). Empirical data shows that low-income distributions are not well described by a Pareto-style power law with a sharp cutoff, as is typically introduced to obtain a normalisable IPDF. Rather than curve-fitting to data, we seek to model the most elementary processes of economic activity, and to find the distribution which emerges.

While it is easy to postulate reasonable-looking intuitive theories for income distribution, there are no known fundamental laws, and so empirical verification is essential [8]. The largest dataset available for personal income in a developing economy is that collected by the Indian National Sample Survey Organisation (NSSO) covering incomes of millions of people for almost 40 years [9]. The same survey reveals the fraction of income spent on staple food (cereals). Since food is the absolute minimum necessity for survival, we will base our measure of poverty on expenses related to its consumption.

The raw NSSO data comprises income bands ("expenditure classes") of irregular size. By regularising the data to account for inflation, and scaling by the measured mean income, we can generate the cumulative distribution functions (CDFs), and the probability distribution functions (PDFs). Once this rescaling has been done, this data exhibits a remarkable data collapse (points in fig. 1). The existence of a distribution which is both time and policy independent strongly suggests that it can be described by a relatively simple fundamental process.
Interacting systems are well described using methods in statistical physics [10–12]. A decentralised free economy can be regarded as such an interacting system. Our basic idea is to represent each individual as an “agent”, generating income through personal labour and trade. We describe the income above starvation level, $y_i(t)$, of each agent, $i$, with a stochastic dynamical equation which describes both labour and trade. It will turn out that trade is the crucial feature for high-income groups, while labour is important for lower-income groups.

We postulate that the time variation of agent income has the form of a Langevin equation:

$$\frac{dy_i}{dt} = C(t) - My_i + \eta_i(t)y_i,$$

(1)

where $C$ represents the rate of increase in income possible from labour, $My_i$ represents the increasing difficulty to maintain a high income for agent $i$, and $\eta_i(t)$, a random variable with zero mean, represents the stochastic effects of trading. $C(t)$ is a property of the economy as a whole and is slowly varying in time. Possible gains from employment depend on how the economy as a whole is performing. $M$ is a constant which we shall later determine from empirical data. Note that a non-zero mean for the noise term would be equivalent to a smaller value of $M$, so no assumption is being made about net benefits of trade.

It can be seen that income from labour alone is the same for each person, however the value of trading is proportional to an individual’s current wealth. This mix introduces a multiplicative noise which is in contrast with previous dynamical approaches [13–16] in economics, producing anomalous diffusion from the noise itself, not fractional dynamics [16,17]. Equation (1) does not map onto any well-known physical system, however, there is increasing evidence for such non-linear risk taking [18–20]. This equation is qualitatively similar to the one obtained by Bouchaud and M´ezard [21] through a “mean-field”-type model. That work has been widely used to describe the distribution of incomes in the high-income range; we will show that our equation is equally valid for low incomes, and therefore a sensible basis for studies of poverty.

We assume that the trading decisions are made before their outcome is known, which indicates that we should use Itô calculus, as opposed to Stratonovich calculus used in [21]: had we assumed mid-term review of trading strategy it would imply Stratonovich calculus, which leads to an equivalent equation with a rescaled $M$. These assumptions lead to a Fokker-Planck equation derived from the Langevin model (eq. (1)):

$$\frac{\partial f}{\partial t}(y,t) = \frac{\partial}{\partial y} \left\{ [(M+2)y - C(t)] \frac{\partial f}{\partial y} + y^2 \frac{\partial^2 f}{\partial y^2} \right\}.$$  

(2)

In the steady state ($C(t) = C_0$), this would give us the income distribution

$$\hat{f}(y,t \to \infty) = \frac{C_0^{M+1}}{\Gamma(M+1)} \frac{\exp(-C_0/y)}{y^{M+2}}.$$  

(3)

This income distribution is in a form suitable for comparison with the NSSO data. The parameter $C_0$ is set by the mean income: because we have been using deflated income this has already been factored into the data. Thus the only fitting parameters are $M$ and the zero for $y_i$. This latter quantity is non-zero because a minimum level of income is needed to avoid death by starvation. We find the best fit to the data for $M = 1.6$.

Figure 1 shows three typical graphs out of 21 surveys across more than a million households (household size varies from 4 to 6) between the years 1959 to 1991. As previously described, the data collapse indicates that a single functional form could describe the data for all years. The graph, and the inset, shows that the IPDF emerging from our model is also in excellent approximation to this functional form, with the same parameter value applicable to all years.

To further illustrate the data collapse and goodness of fit, we use our analytic PDF function to describe the raw data using the binning methods actually employed in the NSSO survey. The results in fig. 2 and fig. 3 show normalized plots where the red lines represent the NSSO data while the black ones are our theoretical fits using results from eq. (1) and eq. (2). These plots indicate how effectively the existence of a universal PDF is concealed by the data collection methodology. Further plots for other years can also be generated from available data.

Fig. 1: (Colour on-line) Plots of the cumulative distribution functions (CDFs) against deflated income for selected years, with inflation independently sourced from the consumer price index (CPI) and renormalized to the 1974 mean income in rupees ($\sim$73 INR). The green line is our theoretical curve, taking $y_i$ as income above a non-zero level below which agents would die of starvation (set at 0.15 in renormalized units). Inset shows the IPDF which is the differential of the CDF, evaluated from the data by interpolation. The points are the data from the NSSO, the lines are our analytic function for the steady-state distribution, the only fitting parameters are the power law tail exponent $M = 1.6$.

1The NSSO data is available in selective bins, called “expenditure classes”, which tells us what number of surveyed population falls within a specified income sector, e.g., between 50 and 100 INR.
data is distributed. Incomerawdata (red) and our model data (black), details as Fig. 3: (Colour on-line) Histogram showing the NSSOurban shown cover June to July for years 1963 to 1968. The pointson rawdata (red) and our model data (black). The dataperiods definition of mean:

\[ C = \frac{\int y f(y,t) \, dy}{\int f(y,t) \, dy} \]

where \( f(y,t) \) is the PDF at time \( t \). The timescale of fluctuations in individual processes. Eigenvalue analysis shows that the steady-state PDF (eq. (3)) is stable against perturbations, with \( C(t) \) decaying at the timescale of fluctuations in individual incomes. This can be easily seen by using the probabilistic definition of mean:

\[ C(t) = \frac{\int_0^\infty dy \exp(-\omega_n t) g_n(y)}{\int_0^\infty dy \exp(-\omega_n t)} \]

where \( \omega_n = 2\pi n \) and

\[ g_n(y) = A_1 \left( \frac{C_0}{y} \right)^{\alpha_-} F \left( \alpha_-, \beta_-, -\frac{C(t)}{y} \right) + A_2 \left( \frac{C_0}{y} \right)^{\alpha_+} F \left( \alpha_+, \beta_+, -\frac{C(t)}{y} \right), \]

which is faster than the rate of change of the global mean income. In this theoretical construct, eq. (2) can be exactly solved in the limit \( C(t) \sim C_0 \), using a combination of harmonic solutions together with Laplace transforms. This solution can then be extrapolated in the slow decaying limit to obtain the full time-dependent solution as a sum of confluent hypergeometric functions \( F(a, b, z) \) with slowly varying time-dependent coefficients:

\[ \dot{f}(y, t) = \sum_{n=0}^{\infty} \exp(-\omega_n t) g_n(y), \]

where \( \omega_n = 2\pi n \) and

\[ g_n(y) = A_1 \left( \frac{C_0}{y} \right)^{\alpha_-} F \left( \alpha_-, \beta_-, -\frac{C(t)}{y} \right) + A_2 \left( \frac{C_0}{y} \right)^{\alpha_+} F \left( \alpha_+, \beta_+, -\frac{C(t)}{y} \right), \]

\[ \alpha_\pm = 3 + M \pm \sqrt{(1 + M)^2 + 4\omega_n}, \]

\[ \beta_\pm = 1 \pm \sqrt{(1 + M)^2 + 4\omega_n}, \]

where constants \( A_1 \) and \( A_2 \) depend on initial conditions.

The data collapse in fig. 1 suggests that the relaxation time (\( \sim 1/M \)) is short, meaning the actual Indian economy has never strayed far from the equilibrium state. The one free parameter, \( M \), incorporates the difficulty of maintaining high income, mean benefit of trade and any possible misconception from our choice of Ito calculus. \( M = 1.6 \) is the exponent in the power law tail (the Pareto part of the distribution) and \( C_0 \) is the mean income.

Figure 1 has shown that \( M = 1.6 \) gives extremely good agreement with the NSS data and it also agrees with the theoretical prediction of Bouchaud-Mézard [21]. The exact results obtained from our model can be mapped to the asymptotic results obtained in this paper. On a one-to-one correspondence, our \( C_0 \) represents their \( \mu = 1 \) while our \( M \) is equivalent to the \( \frac{\mu}{2} \) of Bouchaud-Mézard [21]. The comparison ends there, though, because as opposed to the Bouchaud-Mézard model, our analysis is completely data based, thereby lending credence to any hypothesis made overall; the other major digression being the relaxation time which is \( \approx 100 \) years in [21] whereas real data and consideration of the human-scale stochastic processes involved suggests this time scale \( \sim \) a few years. The fact that our theory naturally agrees with such a time scale is borne out by fig. 4. If not for an exact time scale, the crests and troughs of the “theoretical” (i.e. model-based) poverty index would not have almost exactly matched the raw data.

All this gives us great confidence that our simple, intuitive model describes the coarse features of the Indian economy, and allows us to proceed with the main motivation behind our work, our model of poverty.

Whether poverty reduction measures are regarded as successful or not often depends on the precise definition

---

\[ \frac{C(t)}{M} = \int_0^\infty dy \, y \dot{f}(y, t) \]

\[ \text{Fig. 2: (Colour on-line) Graph showing the NSSO rural income raw data (red) and our model data (black). The data periods shown cover June to July for years 1963 to 1968. The points on the x-axes are the twelve income classes into which the NSSO data is distributed.} \]

\[ \text{Fig. 3: (Colour on-line) Histogram showing the NSSO urban income raw data (red) and our model data (black), details as in fig. 3.} \]

---

\[ C(t) \sim C_0 \exp(-M t) \]

\[ \text{The trick is differentiating the } C(t) \text{ equation with respect to time } t. \]
of poverty, a semantic which is still argued over. Sen has defined the so-called “axioms of poverty” [22–24]:

- Given other things, a reduction in income of a person below the poverty line must increase poverty.
- Given other things, a pure transfer of income from a person below the poverty line to anyone who is richer must increase poverty.

While these seem to be self-evident, they are based on the ill defined notion of “other things” being unchanged. This is troublesome: a dynamical system will have some non-trivial response to any income reduction of transfer. For example, increasing the mean income, or a flat-rate tax, corresponds to increasing \( C(t) \) in eq. (1). The effect of such changes will induce both a transient response and a steady-state change in IPDF.

Three conventional poverty measures involve defining a certain income as the “poverty line”, and counting

i) the fraction of the population with incomes below it (headcount index, HCI);

ii) the mean percentage below the poverty line (poverty gap index, PG);

iii) the mean percentage squared below the poverty line (squared poverty gap index, SPG).

A difficulty with such measures is to define the “poverty line”, a somewhat arbitrary level of income which also changes with time due to inflation. The successive definitions of poverty measures above reduce the sensitivity of the poverty index to this choice, but do not eliminate it, and pathological cases can easily be derived, especially in practice where NSSO data is discretised into expenditure classes.

To define a more robust poverty measure, we apply the idea of consumption deprivation (CD) for a specific resource [25–27]. This uses the fact that expenditure on cereals is monotonically increasing with income, but flattens above a certain income, reflecting the saturation of demand for cereal once one is well fed.

Correlating the NSSO income data with that for cereal expenditure, we find a good fit to a Monod relationship:

\[
CD(y) = \frac{VK}{K + y},
\]

where the parameters \( V \) and \( K \) are time dependent [26]. Broadly, \( K \) can be taken as a “poverty line” which accounts indirectly for cereal price inflation as opposed to general inflation. \( V \) measures the deflated price of cereals.

A more intuitive measure of deprivation is the quantity \( \frac{CD(y)}{V} \), which is the fraction of the maximum desirable cereal consumption actually consumed by someone of income \( y \).

The advantage of this measure is that it is based on people’s actual choices and not on the price of an arbitrarily chosen “basket of goods”. So, for example, increasing housing or clothing costs may affect CD even when cereal prices are steady, as income has to be moved from one commodity to another to balance the overall budget. Similarly CD is not affected by changes in the CPI due to price shifts of luxury goods purchased only by the wealthy. Perhaps most importantly, cereal consumption is directly measured by the NSSO. This allows us to assign a level of poverty to each such NSSO “expenditure classes”. By summing this measure, a poverty index based on actual consumption deprivation may be evaluated. We refer to this as the CD-index of poverty, \( P_{\text{CD}} \).

Our model allows us to quantify this CD-index of poverty. The model definition of the CD-index satisfies the standard axioms of a poverty index [22,28], eliminates the arbitrary “poverty line”, and makes explicit the meaning of “given other things”. Using the NSSO data, we can fit an analytic form [26] to the ratio of grain expenditure to income. The CD-index is then defined by the relation

\[
P_{\text{CD}}(t) = \int \frac{V(t)K(t)}{K(t) + y} \hat{f}(y, t) dy,
\]

where parameters \( V, K \) are obtained from NSSO data while \( \hat{f}(y, t) \) is the solution of eq. (2). The income data used to parameterize our model is independent of the consumption data used to measure CD directly. In fig. 4, we compare the \( P_{\text{CD}}(t) \) evaluated directly from the NSSO consumption data, and indirectly from our income-data–based model. We also show the PG and SPG indices. All indices show poverty declining in time, with a peak due to sharp drops in income in the 1960s. However the CD-index shows the effect of increasing cereal prices between 1978 to 1984 (see fig. 5(b)) as causing an increase in poverty, an effect which cannot be captured in the standard indices.

Against the CPI-deflated data, we see that mean incomes have generally risen over the last forty years...
... above the "poverty line" for a headcount index or reduce absolute poverty provided the mean income is here is that the effect of increasing trade (Although one can devise pathological cases, what we find on poverty depend on the definition of "other things". In this world view, the effect of beneficial trade transferring income to higher-income groups increases mean incomes given other things reduces poverty, while transferring capital from lower- to higher-income groups.

This illustrates a problem with Sen’s axioms. Raising mean incomes “given other things” reduces poverty, while transferring income to higher-income groups increases poverty. In this world view, the effect of beneficial trade on poverty depends on the definition of “other things”. Although one can devise pathological cases, what we find here is that the effect of increasing trade (M ≈ 1) is to reduce absolute poverty provided the mean income is above the “poverty line” for a headcount index or K for PCD. However, it also has the effect of increasing measures of “relative poverty” where the “poverty line” is a fixed fraction of the mean income.

Comparing our results with some earlier works by Dragulescu and co-workers [30,31], we note major differences between the distribution patterns of “developed” and “developing” economies: we have a Levy distribution as opposed to a sharp rising curve [30] in the low-income regimes. The high-income sectors in both cases, though, scale as power laws. Essentially, this amounts to an alteration of the Boltzmannian relaxation as observed by Dragulescu, et al. to a Levy distribution for the developing economies. Focusing on a different aspect of the problem, Ivanov, Stanley and co-workers [32,33] studied the relevance of negative-growth factors, like corruption, in overall economic growth. The basic idea can be directly related to our case by defining our poverty index as a measure of negative growth. It would be most interesting to see if such a negative-growth coefficient, like the corruption perceptions index (CPI) [33], would still retain its universal character, especially in the low-income sectors.

In order to make this comparison more quantitative as well as allay any confusion regarding random empirical fits

\[ P_C(y) = A \exp(-y/R) \]

\[ P_C(y) = A y \exp(-y/R) \]

\[ P_C(y) = A y \exp(-y/R) \]

(2) Boltzmann function fit

(3) Maxwell-Boltzmann function fit

(c) Pareto function fit

Fig. 6: (Colour on-line) Comparison of the data collapsed NSSO data, suitably deflated, with three different theoretical functions, each with two adjustable parameters. The insets represent the probability distribution functions: (a) for the Boltzmann function \( P(r) = \exp(-r/R)/R \) [30]; (b) for the Maxwell-Boltzmann function \( P(r) = Ay \exp(-r/R) \) as studied in [31]; (c) the Pareto function [30], while the insets represent the cumulative distribution function derived from these aforementioned PDFs. The PDF least-square fit parameter values used were the following: (a) \( A = 9.07, R = 0.36 \), correlation coefficient = 0.8613; (b) \( A = 1.62, R = 0.84 \), correlation coefficient = 0.7587; (c) \( m = 1.145, c = 14.26 \), correlation coefficient = 0.5040.

(as opposed to our ab initio model) of functions to NSSO data and to show which gains credence, we show in fig. 6 fits to three representative datasets using these alternative distributions, each of which has two “fitting” parameters.

---

Fig. 5: (a) Deflated mean income plotted against respective years; (b) the falling deflated price of cereals over those years.
The qualitative contrast with fig. 1 is striking. A further comparison of the best-fit Gaussian (to the NSSO data) shows that the error margin is about 8 times larger than the fit of our theoretical model to the same data set.

In summary, we have postulated a stochastic model for the evolution of the income distribution in a developing economy. The steady state of the distribution is stable and robust, and in excellent agreement with the massive NSSO data set for Indian incomes over many years. The existence of an underlying probability distribution function parameterized by mean income makes it much easier to estimate poverty than existing measures such as the headcount index. Under this measure the poverty index is completely specified by the data, without recourse to defining a “poverty line”. Moreover, the measure is less susceptible to manipulation by distortions to the income distribution around the poverty line: “lifting people out of poverty” (just). Another strength of this theory lies in its potential power of predicting changes in the IPDF, and hence poverty, with external perturbative effects, and as a framework on which to build more complicated models for non-perturbative effects.

***

AKC and GJA thank EPSRC for support of the NANIA project and discussions with J. Tailleur. AKC is grateful to the University of Delhi for partial research support.

REFERENCES

[1] Mandelbrot B. B., Int. Econ. Rev., 1 (1960) 79.
[2] Mantegna R. N. and Stanley H. E., Nature, 376 (1995) 46.
[3] Montroll E. W. and Shlesinger M. F., J. Stat. Phys., 32 (1983) 209.
[4] Bouchaud J.-P. and Potters M., Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management (Cambridge University Press) 2003, p. 87.
[5] Mantegna R. N. and Stanley H. E., An Introduction to Econophysics: Correlations and Complexity in Finance (Cambridge University Press) 1999.
[6] Chatterjee A. and Chakrabarti B. K., Eur. Phys. J. B, 60 (2007) 135.
[7] Mohanty P. K., Phys. Rev. E, 74 (2006) 01117.
[8] Bouchaud J.-P., Nature, 455 (2008) 1181.
[9] World Bank Organization (NSSO), National Sample Survey Organization, NSSO, Government of India, http://go.worldbank.org/SWGZB45SN0 (2003).
[10] Challet D., Marsili M. and Zecchina R., Phys. Rev. Lett., 84 (2000) 1824.
[11] Jeffries P., Hart M. L., Hui P. M. and Johnson N. F., Eur. Phys. J. B, 20 (2001) 493.
[12] During B., Matthies D. and Toscani G., Phys. Rev. E, 78 (2008) 056103.
[13] Bouchaud J.-P. and Cont R., Eur. Phys. J. B, 6 (1998) 543.
[14] Takayasu H., Sato A. H. and Takayasu M., Phys. Rev. Lett., 79 (1997) 996.
[15] Richmond P. and Sabatelli L., Physica A, 36 (2004) 27.
[16] Picozzi S. and West B. J., Phys. Rev. E, 66 (2002) 046118.
[17] Metzler R. and Klaffer J., J. Phys. A, 37 (2004) 161.
[18] Glimcher P. W. and Rustichini A., Science, 306 (2004) 447.
[19] Delgado M. R., Schotter A., Ozbay E. Y. and Phelps E. A., Science, 321 (2008) 1849.
[20] Duarte Queiros S. M., EPL, 80 (2007) 30005.
[21] Bouchaud J.-P. and Mézard M., Physica A, 282 (2000) 536.
[22] Sen A. K., Econometrica, 44 (1976) 219.
[23] Palmer-Jones R. K. and Sen A. K., J. Dev. Stud., 40 (2003) 1.
[24] Atkinson A. B., Econometrica, 55 (1987) 749.
[25] Pradhan M. and Ravallion M., Rev. Econ. Stat., 82 (2000) 462.
[26] Kumar T. K., Gore A. P. and Sitaraman V., J. Stat. Plann. Inference, 49 (1996) 53.
[27] Sitaraman V., Paranjpe S. A., Kumar T. K., Gore A. P. and Sastry J. G., Econ. Polit. Wkly., Special Number, 31 (1996) 2499.
[28] Kakwani N. C., Econometrica, 43 (1980) 437.
[29] Kaiser J., Science, 319 (2008) 754.
[30] Dragulescu A. A. and Yakovenko V. M., Eur. Phys. J. B, 17 (2000) 723; Physica A, 299 (2001) 213; Eur. Phys. J. B, 20 (2001) 585.
[31] Yakovenko V. M. and Rosser J. B. Jr., Rev. Mod. Phys., 81 (2000) 1703.
[32] Shao J., Ivanov P. Ch., Podobnik B. and Stanley H. E., Eur. Phys. J. B, 56 (2007) 157.
[33] Podobnik B., Shao J., Njavro J., Ivanov P. Ch. and Stanley H. E., Eur. Phys. J. B, 63 (2008) 547.