epl draft

Income and Poverty in a Developing Economy

Amit K Chattopadhyay[1,2] [(a)], Graeme J Ackland[2] and Sushanta K Mallick[3]

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

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) [1] 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 [2, 3], which is consistent with traditional indices, but independent of an arbitrarily chosen "poverty line" and therefore less susceptible to manipulation.

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 [4], the distribution of incomes has been known to have a power law tail at the high end [5]. There have been many models of the dynamic process [4, 6–10] 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 [11]. 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 [1]. 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 [12, 13, 22]. A decentralised free economy can be regarded as such an interacting system. Our

[(a)]Email: A.K.Chattopadhyay@aston.ac.uk

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:

$$\frac{dy_i}{dt} = C(t) - My_i + \eta_i(t)y_i \tag{1}$$

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 [14–17] in economics, producing anomalous diffusion from the noise itself, not fractional dynamics [17, 27]. Equation 1 does not map on to 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 Mézard [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 Ito 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).

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

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

$$\hat{f}(y, t \to \infty) = \frac{C_0^{M+1}}{\Gamma(M+1)}\,\frac{\exp(-C_0/y)}{y^{M+2}} \tag{3}$$

This income distribution is in a form suitable for comparison with the NSSO data. The parameter C_0 is set by
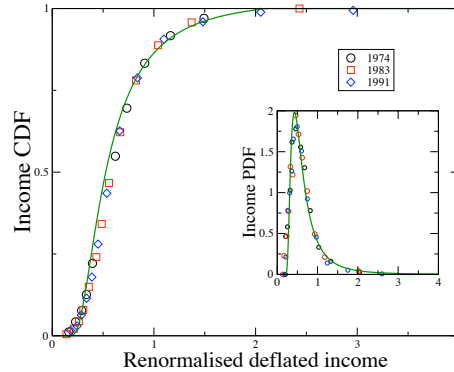

Fig. 1: Plots of the cumulative distribution functions (CDFs) against deflated income for selected years, with inflation independently sourced from the consumer price index (CPI) and renormalized to the 1974 mean income in rupees ($\sim$73 INR). The green line is our theoretical curve, taking y_i as 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 $M = 1.6$

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 4-6) between the years 1959-1991. 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. [1] 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 equation 1 and equation 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.

---

[1] The NSSO data is available in selective bins, called "expenditure classes", which tells us what number of surveyed population falls within a specified income sector eg. between 50-100 INR.
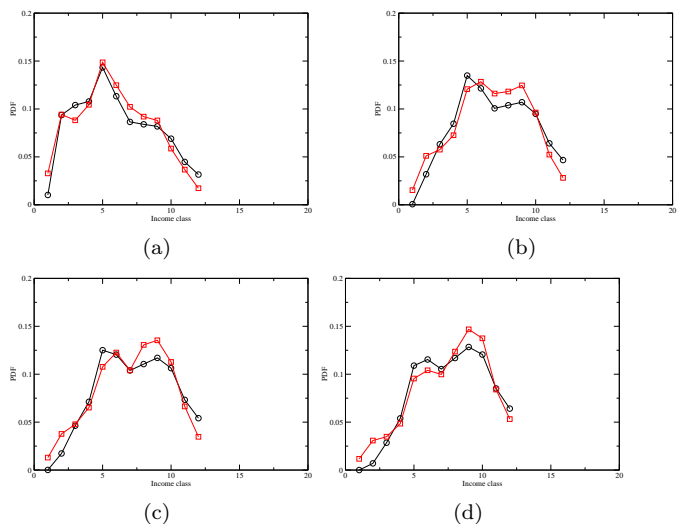
Fig. 2: Graph showing the NSSO rural income raw data (red) and our model data(black). Data periods shown cover Jun-Jul for years 1963-68. The points on the x-axes are twelve income classes into which the NSSO data is distributed.
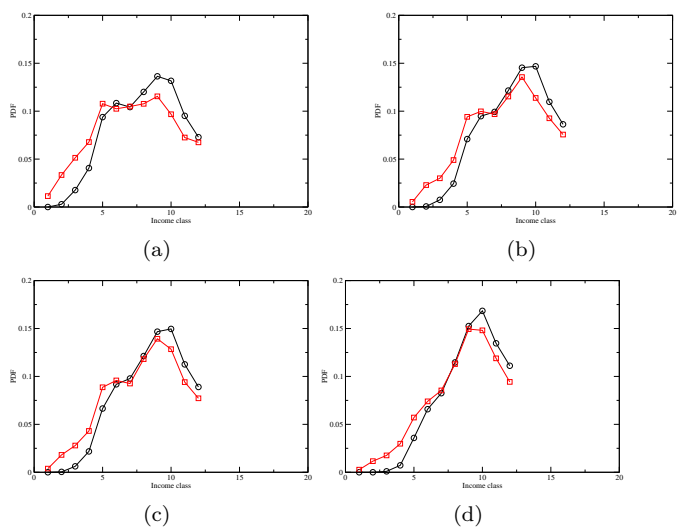

Fig. 3: Histogram showing the NSSO urban income raw data (red) and our model data(black), details as in fig 3.

We now return to a mathematical analysis of the process. 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: $\frac{C(t)}{M} = \int_0^\infty dy\, y\hat{f}(y,t)$ which can then be solved [2] to get $C(t) = C_0 \exp(-Mt)$. This means that

---

[2]The trick is differentiating the C(t) - equation with respect to time t.

the time decay of the mean income occurs at a rate $1/M$ which is faster than the rate of change of the global mean income. In this theoretical construct, equation (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:

$$\hat{f}(y,t) = \sum_{n=0}^{n=\infty} \exp(-\omega_n t)\, g_n(y) \tag{4}$$

where $\omega_n = 2\pi n$ and

$$
\begin{aligned}
g_n(y) &= A_1 \left(\frac{C_0}{y}\right)^{\alpha_-} F(\alpha_-, \beta_-, -\frac{C(t)}{y}) \\
&+ A_2 \left(\frac{C_0}{y}\right)^{\alpha_+} F(\alpha_+, \beta_+, -\frac{C(t)}{y}) \tag{5} \\
\alpha_\pm &= \frac{3 + M \pm \sqrt{(1+M)^2 + 4\omega_n}}{2} \tag{6} \\
\beta_\pm &= 1 \pm \sqrt{(1+M)^2 + 4\omega_n} \tag{7}
\end{aligned}
$$

where constants A_1 and A_2 depend on initial conditions.

The data collapse in figure 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 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 $\frac{C_0}{M}$ is the mean income.

Fig 1 has shown that $M = 1.6$ gives extremely good agreement with the NSSO data, and it also agrees with the theoretical prediction of Bouchaud-Meézard [21]. The exact results obtained from our model can be mapped on 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 $\frac{J}{\sigma^2}$ 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 *approximately* taken as $\sim$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 of poverty, a semantic which is still argued over. Sen has defined the so-called "axioms of poverty" [2].

- 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 equation 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 [24–26]. 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

$$CD(y) = \frac{VK}{K+y} \qquad (8)$$

where the parameters V and K are time dependent [25]. 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_{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 [2,3], eliminates the arbitrary "poverty line", *and* makes explicit the meaning of "given other things". Using the NSSO data, we can fit an analytic form [25] to the ratio of grain expenditure to income. The CD-index is then defined by the relation

$$P_{CD}(t) = \int \frac{V(t)K(t)}{K(t)+y} \hat{f}(y,t) dy \qquad (9)$$

where parameters V, K are obtained from NSSO data while $\hat{f}(y,t)$ is the solution of equation (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 $P_{CD}(t)$ 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 1978-84 (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 (Fig. 5(a)), while the relative price of cereals ($V(t)$) has generally dropped (Fig.5(b)). This helps to reduce poverty, although more direct targeting [29] may be even more effective.

Returning to the stationary IPDF, the power law exponent M is seen to be a crucial component in quantifying the mean income: $\frac{C_0}{M}$. Critically, since we have shown that if trading is, on average, beneficial rather than neutral, it will reduce M. Small M increases both the mean income and the level of inequality - it transfers 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

(a)   (b)

Fig. 5: (a) Deflated mean income plotted against respective years (b) the falling (deflated) price of cereals over those years.
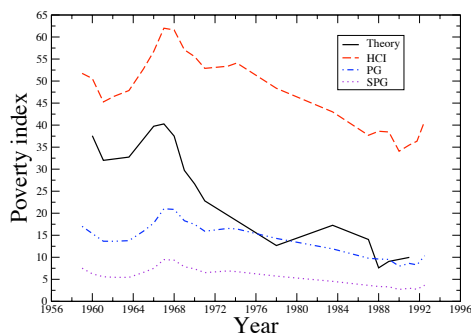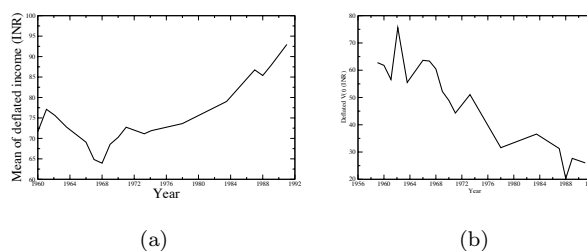
Fig. 4: Plots of the CD-index, a measure of poverty [25], against time. The broken lines represent the official head-count poverty index, poverty gap index and squared poverty gap index; while the continuous line shows the poverty index arising from our IPDF. To generate the latter we have taken the function $C(t)$ in equation 1 to be a piecewise-linear interpolation of the NSSO-measured mean income at each round and assumed a relaxation time less than a year. $V(t)$ and $K(t)$ are defined from the NSSO data:

things". Although one can devise pathological cases, what we find here is that the effect of increasing trade ($M \approx 1$) is to reduce absolute poverty provided the mean income is above the "poverty line" for a headcount index or K for P_{CD}. 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], [3] 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 (as opposed to our ab initio model) of functions to NSSO data and to show which gains credence, we show in Figure 6 fits to three representative datasets using these alternative distributions, each of which has two "fitting"

parameters. 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 head-count 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.

$$* * *$$

[3]Measured on a scale between 0-10, larger the CPI value, lower is the corruption.

REFERENCES

[1] WWW.WORLDBANK.ORG/POVERTY/DATA/INDIADATA.HTM, *National Sample Survey, NSSO, Government of India*, edited by WORLD BANK, Vol. **Online** (NSSO) 2003, p. .

[2] SEN A. K.., *Econometrica*, **44** (1976) 219.

[3] KAKWANI N. C., *Econometrica*, **43** (1980) 437.

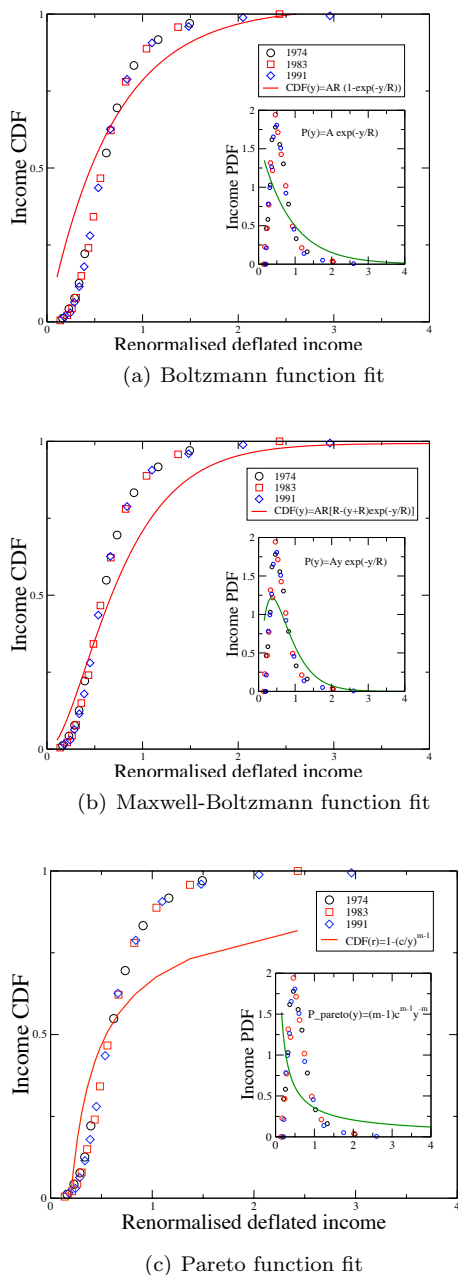[4] MANDELBROT B. B., *International Economic Review*, **1** (1960) 79.

(a) Boltzmann function fit

(b) Maxwell-Boltzmann function fit

(c) Pareto function fit

Fig. 6: 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: Fig. 6(a) for the Boltzmann function $P(r) = \exp(-y/R)/R$ [30]; Fig. 6(b) for the Maxwell-Boltzmann function $P(r) = Ay\exp(-y/R)$ as studied in [31]; Fig. 6(c) the pareto function [30], while the outsets represent the cumulative distribution function derived from these aforementioned PDFs. The PDF least square fit parameter values used were the following: Fig. 6(a) $A = 9.07, R = 0.36$, correlation coefficient $= 0.8613$; Fig. 6(b) $A = 1.62, R = 0.84$, correlation coefficient $= 0.7587$; Fig. 6(c) $m = 1.145, c = 14.26$, correlation coefficient $= 0.5040$.

[5] MANTEGNA R. N. and STANLEY H. E., *Nature*, **376** (1995) 46.

[6] MONTROLL E. W. and SHLESINGER M. F., *Journal of Statistical Physics*, **32** (1983) 209.

[7] BOUCHAUD J. -P. and POTTERS M., *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, edited by , Vol. (Cambridge University Press) 2003, p. 87.

[8] MANTEGNA R. N. and STANLEY H. E., *An Introduction to Econophysics: Correlations and Complexity in Finance*, edited by , Vol. (Cambridge University Press) 1999, p. .

[9] CHATTERJEE A. and CHAKRABARTI B. K., *European Physical Journal B*, **60** (2007) 135.

[10] MOHANTY P. K., *Physical Review E*, **74** (2006) 01117.

[11] BOUCHAUD J. -P., *Nature*, **455** (2008) 1181.

[12] CHALLET D., MARSILI M. and ZECCHINA R., *Physical Review Letters*, **84** (2000) 1824.

[13] JEFFRIES P., HART M. L., HUI P. M. and JOHNSON N. F., *European Physical Journal B*, **20** (2001) 493.

[14] BOUCHAUD J. -P. and CONT R., *European Physical Journal B*, **6** (1998) 543.

[15] TAKAYUSU H., SATO A. H. and TAKAYUSU M., *Physical Review Letters*, **79** (1997) 996.

[16] RICHMOND P. and SABATELLI L., *Physica A*, **36** (2004) 27.

[17] PICOZZI S. and WEST B. J., *Physical Review E*, **66** (2002) 046118.

[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., *Europhysics Letters*, **80** (2007) 30005.

[21] BOUCHAUD, J.-P. and MÉZARD M., *Physica A*, **282** (2000) 536.

[22] DURING B., MATTHES D. and TOSCANI G., *Physical Review E*, **78** (2008) 056103.

[23] PALMER-JONES R. K. and SEN A. K., *Journal of Development Studies*, **40** (2003) 1.

[24] PRADHAN M. and RAVALLION M., *The Review of Economics and Statistics*, **82(3)** (2000) 462.

[25] KUMAR T. K., GORE A. P. and SITARAMAN V., *Journal of Statistical Planning and Inference*, **49** (1996) 53.

[26] SITARAMAN V., PARANJPE S. A., KUMAR T. K., GORE A. P. and SASTRY J. G., *Economic and Political Weekly, Special Number*, **31(35-37)** (1996) 2499.

[27] METZLER R. and KLAFTER J., *Journal of Physics A*, **37** (2004) 161.

[28] ATKINSON A. B., *Econometrica*, **55** (1987) 749.

[29] KAISER J., *Science*, **319** (2008) 754.

[30] DRAGULESCU A. A. and YAKOVENKO V. M., *European Physical Journal B*, **17** (2000) 723; *Physica A*, **299** (2001) 213; *European Physical Journal B*, **20** (2001) 585

[31] YAKOVENKO, V. M. and ROSSER JR., J. B., *Reviews of Modern Physics*, **81** (2000) 1703.

[32] SHAO J., IVANOV, P. CH., PODOBNIK B. and STANLEY H. E., *European Physical Journal B*, **56** (2007) 157.

[33] PODOBNIK B., SHAO J., NJAVRO J., IVANOV P. CH. and STANLEY H. E., *European Physical Journal B*, **63** (2008) 547.