On Pareto's Law and the determinants of Pareto exponents

William J. Reed
Department of Mathematics and Statistics
University of Victoria
PO Box 3045
Victoria B.C.
Canada V8W 3P4.
e-mail: reed@math.uvic.ca

November, 2002.

Abstract

A stochastic model for the generation of observed income distributions is used to provide an explanation for the Pareto law of incomes. The basic assumptions of the model are that the evolution of individual incomes follows Gibrat's law and that the population or workforce is growing at a fixed (probabilistic) rate. Analysis of the model suggests that Paretian behaviour can occur in either or both tails of an income distribution. Examination of the empirical distribution of total money income in the USA confirms Paretian behaviour in both tails. What determines the magnitude of the power-law (Pareto) exponents is discussed.

Keywords: Pareto Law, power law, income distribution, Gibrat's law, killed stochastic process

1 Introduction.

The discovery more than 100 years ago by V. Pareto (1897) that many income distributions exhibited power-law behaviour in the upper tail[1] has led over the years to many attempts to explain this phenomenon (Pareto's law of incomes) and to find a satisfactory parametric form which can be fitted to any income distribution. During the 1950's and 60's there was a lot of activity along these lines using stochastic modelling of the evolution of incomes. The papers of Champernowne (1953), Rutherford (1955) and Mandelbrot (1960) are the most noteworthy of this work. Although there have been a number of other refinements and alternative models proposed (for summaries see Ord, 1975 and Dagum, 1983), none seems to have provided much of an improvement in terms of explanatory power over the model of Champernowne. More recent work has eschewed process modelling, instead focusing on finding alternative parametric distributions which provide a good fit to a wide range of observed data, and for which the parameters have a meaningful economic interpretation (see *e.g.* Bordley *et al.*, 1996; Dagum, 1996). Exceptions are the papers of Parker (1999) which presented a model which is claimed to explain why incomes should follow a generalised beta distribution – a distribution which has been shown to provide a good fit to much empirical income data (McDonald, 1984; Bordley *et al.*, 1996); and of Solomon and Richmond (2001) which offered an explanation for the power-law behaviour of wealth distributions and other phenomena based on

[1] Pareto plotted the the number of incomes above a level x against x, both on logarithmic scales for many jurisdictions and found that the plotted points lay close to a straight line of negative slope

the Lotka-Volterra model of population biology.

In this article, in Sec. 2, a model in the spirit of the earlier work of Champernowne (1953) and others is presented which provides an explanation of why Pareto's law should hold. In addition the model predicts that a Pareto-type law may hold in the lower-tail of income distributions (*i.e.* that a logarithmic plot of the number people earning *less than* x against x, should yield points close to a straight line. Examination of a number of empirical income distributions indicates that this often appears to be the case.

A discussion of what determines the magnitude of the power-law (Pareto) exponents in both tails is presented in Sec. 3 and an analogous continuous-time model briefly discussed in Sec. 4.

2 The model and analysis.

In the tradition of Champernowne (1953), Rutherford (1955) and other earlier authors, we consider a discrete-time stochastic model for the evolution of the income of an individual. Let X_n denote the income of an individual in period n, and suppose that it is related to the individual's income in the previous period by

$$X_n = (1 + r_{n-1})X_{n-1} \tag{1}$$

where $\{r_n\}_{n=0,1,\ldots}$ are the per-period rates of growth in the individual's income. Letting $Z_n = 1 + r_n$ denote per-reriod growth factors we have

$$X_n = Z_{n-1}X_{n-1}. \tag{2}$$

We shall assume that $\{Z_n\}_{n=0,1,...}$ are independent, identically distributed, continuous, non-negative random variables. This model is simply Gibrat's law of proportional effects (Gibrat, 1931) and is compatible with Champerenowne's (1953) logarithmic random walk model.

Clearly by iterating (2) the state the individual's income in period N is

$$X_N = X_0 Z_0 Z_1 Z_2 \dots Z_{N-1} = X_0 \prod_{i=0}^{N-1} Z_i, \tag{3}$$

where X_0 represents the starting income of the individual. Gibrat showed that for large fixed N, the distribution of X_N should be approximately lognormal. However it is well-known that empirical income distributions are poorly approximated by a lognormal distribution. The reason for this is that over the whole population or workforce, the parameters N (time since entry into the workforce[2]) and X_0 (starting income) vary. To take this into account we need to consider a mixture of (approximately lognormal) distributions (of $X_N = X_0 \prod_{i=0}^{N-1} Z_i$) with mixing parameters X_0 and N.

The distribution of N over the whole workforce (*i.e.* of the time since entry of individuals) will reflect the history of the entry and exit from the workforce via its current 'age' profile. For example if the workforce has been growing at a fixed rate there will be more younger workers than older ones. In fact in this case the time in the workforce of an individual will have a truncated exponential distribution; this distribution also results if the workforce is growing *stochastically* (in a Yule process) at a fixed average rate (see Reed and Hughes, 2002a).

[2]Note that the expression "entry into the workforce" is appropriate for discussion of earned income. When discussing income from all sources it would be more accurate to use the expression "first receipt of income". For the sake of simplicity the former expression will be used throughout.

To a first approximation we can ignore the truncation, and assume an exponential distribution, or its equivalent in discrete time, a geometric distribution. Thus we shall assume that N follows a geometric distribution with probability mass function (pmf)

$$f_N(n) = \Pr(N = n) = pq^{n-1}, \quad \text{for } n = 1, 2, \dots \tag{4}$$

where $0 < p < 1$ and $q = 1 - p$. This model corresponds to a workforce growing at an average (proportional) rate $\lambda = p/q$ per period.

Now let $\bar{X}$ be a random variable denoting the current income of an individual *i.e.* $\bar{X} = X_0 \prod_{i=0}^{N-1} Z_i$, where N is now considered a random variable with pmf (4). We show in the Appendix that the distribution of $\bar{X}$ will exhibit power-law (Paretian) behaviour in the upper tail if $\Pr(Z_i > 1)$ is non-zero; and will exhibit power-law (Paretian) behaviour in the lower-tail if $\Pr(Z_i < 1)$ is non-zero. Precisely if we denote the probability density function (pdf) of $\bar{X}$ by $f_{\bar{X}}(x)$, and its cumulative distribution function (cdf) by $F_{\bar{X}}(x)$, then, if an individual's income can increase (*i.e.* there is a non-zero probability that the growth rate r_i is positive)

$$f_{\bar{X}}(x) \sim c_1 \; x^{-\alpha-1}; \quad 1 - F_{\bar{X}}(x) = \Pr(\bar{X} > x) \sim c_2 \; x^{-\alpha}, \quad \text{as } x \to \infty \tag{5}$$

where α, c_1 and c_2 are positive constants; and if there if an individual's income can decrease (*i.e.* there is a non-zero probability that the growth rate in any period, r_i, is negative) then

$$f_{\bar{X}}(x) \sim c_3 \; x^{\beta-1}; \quad F_{\bar{X}}(x) = \Pr(\bar{X} \le x) \sim c_4 \; x^{\beta}, \quad \text{as } x \to 0 \tag{6}$$

where β, c_3 and c_4 are positive constants. The result (5) is the usual Pareto law; while the result (6) is a lower-tail Pareto law.

Although Champernowne (1953) pointed out that many empirical income distributions exhibited Paretian lower-tail behaviour (*i.e.* logarithmic plots of number of incomes less than x against x have points close to a straight line) the fact doesn't seem to be widely acknowledged. The model of this paper predicts such behaviour. The plot in Fig 1 for money incomes in the USA (and many other plots not shown) confirm its existence.

3 Determinants of the magnitude of the Pareto exponents.

Details of 'comparative statics' for the Pareto exponents α and β in the upper and lower tail of the distribution of current incomes, under the model of Sec. 2 are given in the Appendix. A small value of the upper-tail exponent α corresponds to a very long upper tail (extreme inequality). In a similar fashion we shall refer to a small β as corresponding to a distribution with a long lower tail. Of course the lower tail is always bounded by zero, but for small β there will be more probability concentrated near zero. Indeed if β is less than one the pdf will be decreasing at the lower end.

The results can be summarized as follows:

- the upper-tail exponent α will be smaller (longer upper tail) in cases where (a) the workforce is growing slower on average (more older workers); and (b) the growth of individual incomes is faster on average, and exhibits more variability, than in cases where the opposite of (a) and (b) hold.

- The lower-tail exponent β will be smaller (long lower tail) in cases where (c) the workforce is growing slower on average; and (d) the growth of individual incomes is slower on average and exhibits more variability, than in cases where the opposite of (c) and (d) hold.

Thus rapid growth in the workforce and low variability in the growth of individual incomes cause *both* exponents to be large - reduce inequality The effect of the average rate of growth for individual incomes has opposite efefcts on the upper and lower tail exponents. When a high average growth rate in individual incomes prevails there will be a longer upper tail and a shorter lower tail, than in the corresponding case with a lower positive, or even negative, average growth rate.

4 A continuous-time model.

Reed (2003) considers a continuous-time version of the model presented here, in which the income of an individual is assumed to follow *geometric Brownian motion* (GBM):

$$dX = \mu X dt + \sigma X dw$$

where μ and σ are mean drift and variance parameters (rate and volatility of growth) and dw is white noise (the increment of a Wiener process with $dw \sim$ N$(0, dt)$). In addition starting incomes are assumed to follow another GBM with different parameters (so that the starting incomes at any fixed time are lognormally distributed). Thus the model is both more general than the model of this paper in that it allows for variations in starting incomes, but less general in that it does not have the flexibility of allowing different distributions for the random per-period growth factors for individual incomes (the Z_i in (2)). For example it does not include the possibility that an individual's income be monotonically increasing, with probability one, as would occur in the model (2) if $\Pr(Z_i > 1) = 1$.

For the GBM model Reed (2003) shows that if the distribution over the workforce of the time since entry of individuals is exponential, then power-law behaviour can be expected in both tails of the distribution of current incomes. Furthermore the distribution of current incomes for this model is shown to be the product of independent 'double Pareto' and lognormal components. Reed (2003) names it the *double Pareto-lognormal* distribution and derives expressions for its pdf and cdf in terms of four parameters α, β, ν and τ^2. Specifically

$$f_{\bar{X}}(x) = \tfrac{\alpha\beta}{\alpha+\beta}\left[x^{-\alpha-1}A(\alpha,\nu,\tau)\Phi\left(\tfrac{\log x-\nu-\alpha\tau^2}{\tau}\right) + x^{\beta-1}A(-\beta,\nu,\tau)\Phi^c\left(\tfrac{\log x-\nu+\beta\tau^2}{\tau}\right)\right] \tag{7}$$

where

$$A(\theta,\nu,\tau) = \exp(\theta\nu + \alpha^2\tau^2/2)$$

and

$$F_{\bar{X}}(x) = \Phi\left(\tfrac{\log x-\nu}{\tau}\right) - \tfrac{1}{\alpha+\beta}\left[\beta x^{-\alpha}A(\alpha,\nu,\tau)\Phi\left(\tfrac{\log x-\nu-\alpha\tau^2}{\tau}\right) + \alpha x^{\beta}A(-\beta,\nu,\tau)\Phi^c\left(\tfrac{\log x-\nu+\beta\tau^2}{\tau}\right)\right] \tag{8}$$

where Φ is the cumulative distribution function of a standard normal, $\mathrm{N}(0,1)$, random variable and $\Phi^c = 1 - \Phi$. These closed-form expressions permit model fitting for both ungrouped and grouped data by maximum likelihood. Properties of this distribution and methods of fitting are discussed in Reed and Jorgensen (to appear). Reed (2003) gives examples of the excellent fit to a number of empirical income and earnings distributions. Thus the double Pareto-lognormal distribution appears to offer an attractive new distribution for the modelling of income distributions, and has the added benefit of being derived from a theoretical model for the way in which empirical income distributions arise.

5 Discussion.

This article has presented an explanation of the Pareto law of incomes, based on a model which uses Gibrat's law to model the growth of individual incomes. In this respect the model is like many earlier attempts to explain Pareto's law (see references in the Sec. 1). However unlike many of these earlier attempts, the analysis of the model in this article does not assume a population in equilibrium. In fact it specifically assumes a population or workforce which is growing (possibly stochastically) at a fixed mean rate. It thus recognizes the fact that different individuals will have been in the workforce for different lengths of time. It is this fact, coupled with the multiplicative (*i.e* geometric or exponential) nature of growth under Gibrat's law, which gives rise to power-law behaviour in one or both tails of the current income distribution.

The model is a discrete-time analogue of a model presented by Reed (2003), based on a geometric Brownian motion model for individual income growth. The discrete model however has, in some respects, more flexibility than the continuous-time model, in that it allows for different possible distributions for the per-period growth factors, rather than exclusively a lognormal distribution as in the GBM model *e.g.* growth rates can be assumed to be always positive. With this further flexibility it is possible to show that a Pareto upper tail will occur only if individual incomes can increase, and a Pareto lower tail will occur only if incomes can decrease.

The way in which the 'length' of the Pareto tails (the magnitude of the Pareto exponents) depends on the mean and variance of the growth rate of individual incomes, and on the growth rate of the size of the workforce are discussed. In many ways *a posteriori* these results seem obvious, but apparently have not been recognized before. Analysis of the model also leads to the prediction of a lower-tail Pareto law. Although the existence of such for empirical distributions was discovered by Champernowne fifty years ago, it does not appear to be widely known. Both of the above findings demonstrate the value of modelling in the discovery process.

The model of course is a gross oversimplification of reality. However simplification is at the heart of modelling, and the fact that the model can provide an explanation of an observed phenomenon, as well as predicting another unexpected one, suggests that the important components of the mechanism behind Pareto's law have been retained in the model. Pareto's law seems to be the consequence of a mathematical mechanism, which operates not only in the field of incomes (see Reed and Hughes, 2002b, for further examples in biology, geography, geneology and internet ecology). Thus although Pareto's law operates in the economic sphere, it does not require an explanation in economic terms.

Appendix

We derive the Paretian tail behaviour for the model described in Sec. 2. The derivation uses generating functions – the *probability generating function* (pgf), which for a discrete random variable X with pmf $f_X(x)$, is defined as

$$G_X(s) = \mathrm{E}(s^X) = \sum f_X(x) s^x$$

and the *moment generating function* (mgf) which for any random variable X is defined as

$$M_X(s) = \mathrm{E}(e^{sX}),$$

provided the expectations exist. For a random variable N with the geometric distribution (4) the pgf is

$$G_N(s) = \frac{ps}{1 - qs} \tag{9}$$

Now let $\bar{Y} = \log(\bar{X})$ where $\bar{X}$ is a random variable denoting the current income of an individual($\bar{X} = X_0 \prod_{i=0}^{N-1} Z_i$) as described in Sec. 2. Then

$$\bar{Y} = Y_0 + \sum_{i=0}^{N-1} U_i \tag{10}$$

where $Y_0 = \log(X_0)$ and $U_i = \log(Z_i)$ for $i = 0, 1, \ldots, N-1$. The mgf of $\bar{Y}$ is

$$M_{\bar{Y}}(s) = E(e^{s\bar{Y}}) = \mathrm{E}\left(\exp\left[Y_0 s + \sum_{i=0}^{N-1} U_i s\right]\right) \tag{11}$$

where the expectation is taken with respect to the random variables Y_0, N and $U_0, U_1, \ldots, U_{N-1}$, assumed to be independent. Using conditional expectation this can be written

$$M_{\bar{Y}}(s) = M_0(s)\mathrm{E}\left([M_U(s)]^N\right) \tag{12}$$

where $M_0(s)$ is the mgf of Y_0; $M_U(s)$ is the (common) mgf of the U_i and the expectation is taken with respect to the random variable N. Now $\mathrm{E}\left([M_U(s)]^N\right) = G_N(M_U(s))$ so that the mgf of $\bar{X}$ can be written

$$M_{\bar{Y}}(s) = M_0(s)\,\frac{pM_U(s)}{1 - qM_U(s)} \tag{13}$$

Now from standard results, the tail behaviour of the pdf of $\bar{Y}$ can be determined from the singularities of its mgf. These occur at solutions (in s) to $M_U(s) = 1/q$. We examine these in the three cases:

- (a) *incomes always increasing:* $Z_i > 1$ (so that $U_i > 0$) with probability one;
- (b) *incomes always decreasing:* $Z_i < 1$ (so that $U_i < 0$) with probability one; and
- (c) *incomes able to increase and decrease:* $\Pr(Z_i > 1) > 0$ and $\Pr(Z_i < 1) > 0$ so that $\Pr(U_i > 0) > 0$ and $\Pr(U_i < 0) > 0$

In all cases $M_U(0) = 1$ and $M_U''(s) > 0$, so that real zeros of $M_U(s) - 1/q$ are simple zeros. In case (a) $M_U(s)$ is increasing in s with $M_U(s) \to \infty$ as $s \to \infty$ and $M_U(s) \to 0$ as $s \to -\infty$. It follows that there is a simple zero of $M_U(s) - 1/q$ (so that there is a simple pole of $M_{\bar{Y}}(s)$) at $\alpha > 0$ (NB $1/q > 1$). This implies that the pdf of $\bar{Y}$ behaves asymptotically (as $y \to \infty$) like $c_1 e^{-\alpha y}$; and therefore after a change of variable that the pdf of $\bar{X}$ behaves asymptotically as $f_{\bar{X}}(x) \sim c_1 x^{-\alpha-1}$ (as $x \to \infty$).

In case (b) $M_U(s)$ is decreasing in s with $M_U(s) \to 0$ as $s \to \infty$ and $M_U(s) \to \infty$ as $s \to -\infty$. It follows that there is a simple pole of $M_{\bar{Y}}(s)$ at $-\beta < 0$. This implies that the pdf of $\bar{Y}$ behaves asymptotically (as $y \to -\infty$) like $c_3 e^{\beta} y$; and therefore after a change of variable that the pdf of $\bar{X}$ behaves asymptotically as $f_{\bar{X}}(x) \sim c_3 x^{\beta-1}$ (as $x \to 0$).

In case (c) $M_U(s) \to \infty$ as $s \to \infty$ and as $s \to -\infty$. From this fact and the convexity of $M_U(s)$ it follows that there are two simple poles of $M_{\bar{Y}}(s)$, one positive (α, say) and the other negative ($-\beta$, say). This implies $f_{\bar{X}}(x) \sim c_1 x^{-\alpha-1}$ (as $x \to \infty$) and $f_{\bar{X}}(x) \sim c_3 x^{\beta-1}$ (as $x \to 0$).

The asymptotic behaviour of the cdf $F_{\bar{X}}(x)$ or the complementary cdf, $1 - F_{\bar{X}}(x)$, follows from integration.

The way in which the exponents α and β depend on other quantities can be examined by re-writing the equation $M_U(s) = 1/q$ in terms of the cumulant generating function

$$K_U(s) = \log\left(M_U(s)\right) = \mu s + \frac{\sigma^2}{2!}s^2 + \frac{\kappa_3}{3!}s^3 + \frac{\kappa_4}{4!}s^4 + \dots$$

where μ and σ^2 are mean and variance of the $U_i = \log Z_i$, and κ_3, κ_4 etc. are higher order cumulants. The quantities α and $-\beta$ are the positive and negative roots (if they exist) of

$$K_U(s) = -\log q$$

For simplicity we consider only the case (c) above where both roots exist. Similar arguments apply in cases (a) and (b). In case (c) when incomes can both increase and decrease, $K_U(0) = 0$ with K_U decreasing for $s < 0$ and increasing for $s > 0$. From this it follows that α and β both increase with $-\log q$. In a similar way it can be established that increasing σ^2 results in a decrease of both α and β

(since $K_U(s)$ increases for all s). For an increase in μ, $K_U(s)$ decreases if $s < 0$ and increases if $s > 0$. From this it follows that α decreases and β increases with an increase in μ. An increase in the third cumulant κ_3 (skewness) with μ and σ^2 staying constant, has the same effect as an increase in μ.

The expected growth rate per period of the workforce is $\lambda = p/q$, and the quantity $\theta = -\log q$ is related to this as $\theta = \log(1+\lambda)$, which is an increasing function. It follows that both α and β increase with λ. Thus in a fast growing workforce one would expect large Pareto exponents α and β in both tails, and for a slower growing workforce the exponents would be smaller. In fact in the limit as the growth rate $\lambda \to 0$ (so that the distribution of time in the the workforce tends to an improper uniform distribution) both $\alpha, \beta \to 0$, so that the pdf $f_{\bar{X}}(x)$ tends to an extremely long-tailed (improper) distribution behaving like x^{-1} in both tails.

The parameters μ and σ^2 are the mean and variance of $U_i = \log Z_i = \log(1+r_i) = r_i - r_i^2/2 + r_i^3/3 - \ldots$, where the r_i are random variables representing the per period growth rates for the income of an individual. Approximately one can consider μ and σ^2 to be the mean and variance of the growth rate of an individual's income (for short periods the r_i will be small and higher order terms in the power series can be ignored). Thus if individual incomes grow rapidly on average (large μ) one would expect α to be small and β to be large, and vice versa if individual incomes grow slowly, or even negatively. If growth of individual incomes is highly variable (large σ^2) one would expect both Pareto exponents to be smaller than in the case of less variability in growth.

References.

Bordley, R.F., McDonald, J.B. and A. Mantrala. “Something new, something old: parametric models for the size distribution of income”, *J. Income Dist.* **61** (1996) 91-103.

Champernowne, D.G. “A model of income distribution”, (1953) *Econ. J.* **23** (1953) 318-351.

Dagum, C. “Income distribution models”, *Encyclopedia of Statistical Sciences, Vol 4*, eds. Kotz, S. and N.L.Johnson. (1983), 27-34. Wiley, .

Dagum, C. “A systematic approach to the generation of income distribution models”, *J. Income Dist.* **61** (1996) 105-126.

Gibrat R. *Les Inégalités éEconomiques*, (1931), Sirely, Paris.

Mandelbrot, B. “The Pareto-Lévy law and the distribution of income”, *Int. Econ. Rev.* **1** (1960) 79-106.

McDonald, J.B. “Some generalized functions for the size distribution of income,”*Econometrica* **52** (1984) 647-663.

Ord, J.K. ‘Statistical models for personal income distributions,”*Statistical Distributions in Scientific Work, Vol. 2*, eds. Patil, G.P, Kotz, S. and J. K. Ord. (1975), 151-158, Reidel.

Pareto, V. *Cours d’Economie Politique, Vol. 2*, (1897), F. Pichou, Paris.

Parker, S.C. “The Generalised Beta as a Model for the Distribution of Earnings”, *Econ. Letters* **62** (1999) 197-200.

Reed, W. J. "The Pareto law of incomes - an explanation and an extension", To appear *Physica A* **318** (2003).

Reed, W. J. and B. D. Hughes. "On the size distribution of live genera", *J. Theor. Biol.* (2002a), **217**, 125-135.

Reed, W. J. and B. D. Hughes. "From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature", *Phys. Rev. E* **66** (2002b), to appear.

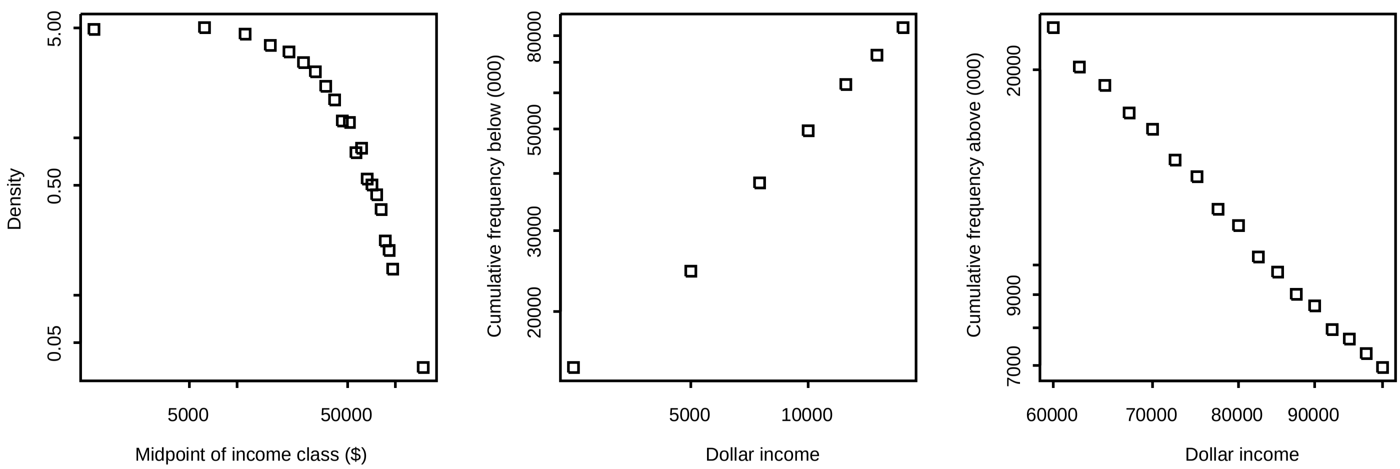
Reed, W. J. and M. Jorgensen. "The double Pareto-lognormal distribution - a new parametric model for size distributions", in preparation.

Rutherford, R.S.G. "Income distributions: a new model", *Econometrica,* **23** (1955) 277-294.

Solomon, S. and P. Richmond. "Power laws of wealth, market order volumes and market returns", *Physica A* **299** (2001)188-197.

Figure legends.

Fig. 1 Distribution of total money income of 216 million people in USA in 2000. The left-hand panel is a plot (on logarithmic axes) of the frequency density distribution. The centre panel (again on logarithmic axes) is a plot of the cumulative frequency distribution in the lower tail. The linearity of the plot confirms the lower-tail Pareto law. The right hand panel is a plot (on logarithmic axes) of the cumulative frequency distribution in the upper tail and reveals the familiar upper-tail Pareto law.


Density
5.00
0.50
0.05
5000
50000
Midpoint of income class ($)
Cumulative frequency below (000)
80000
50000
30000
20000
5000
10000
Dollar income
Cumulative frequency above (000)
20000
9000
7000
60000
70000
80000
90000
Dollar income