The Pareto Law of Incomes – an Explanation and an Extension.

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

May, 1999. Revised July 2000.

Abstract

A stochastic model for the generation of observed income distributions is used to provide an explanation for the Pareto law of incomes. Analysis of the model also yields a prediction of Paretian (power law) behaviour in the lower tail of the distribution and this is shown to occur in a number of empirical distributions. A computable four-parameter distribution is derived, and shown to fit extremely well to a number of different empirical income distributions.
Keywords: Pareto Law, income distribution, GBM, mixture model, double Pareto-lognormal distribution.

1 INTRODUCTION.

The purpose of this paper is to demonstrate that many income distributions share a common underlying structure which can be explained using a stochastic model for the evolution and observation of individual incomes. The explanation yields a parametric distribution which has the property of power-law behaviour in both tails. Thus, as well as explaining the Pareto law, it suggests the possibility of a similar law relating to cumulative frequencies of incomes at the lower end of the distribution.

Modelling of income distributions began over 100 years ago, with the work of Vilfredo Pareto (Pareto, 1897), who observed that for many populations, a plot of the logarithm of the number of incomes above a level x against the logarithm of x, yielded points close to a straight line of slope $-\alpha$. This suggests a distribution (now known as the *Pareto distribution*) with density function proportional to $x^{-\alpha-1}$. It has long been realized that the Pareto distribution provides a poor fit to observed frequency data over the whole range of incomes, but it is generally accepted that it provides a good fit to the distribution of high incomes. There have been a number of attempts to explain this Paretian tail behaviour, the principal ones being due to Champernowne (1953), using a Markov Chain to model the evolution of individual incomes and Mandelbrot (1960) using stable Lévy distributions.

Champernowne's simplest model, in which individual incomes were assumed to follow a random walk in the logarithmic scale, yielded a Pareto distribution over the whole range of incomes. However he observed that actual income distributions were very different from this in the lower tail, and that in fact something like a two-tailed Pareto distribution held *i.e.* the density of incomes at a low levels x was proportional to x raised to a positive power; while at high levels it was proportional to x raised to a negative power. In order to ensure that his model produced an equilibrium with such characteristics, Champernowne had to make the unrealistic assumptions that on average the drift in large incomes was negative and that for low incomes was positive. Mandelbrot's (1960) explanation for the Paretian (upper-) tail behaviour was based on the assumption that an individual's income comprises the sum of many independent random components or "shocks". Using results from the theory of stable distributions he showed that the distribution of the sum (suitably normed) of shocks with infinite variance could converge to a distribution which, in the upper tail, behaves like a Pareto distribution, with exponent $\alpha < 2$. The fact that estimates of α, from empirical income distributions, are frequently greater than 2 seems to cast considerable doubt on this theory. Another model, more along the lines of that of Champernowne, was presented by Rutherford (1955). This model included entry and exit from the population of income recipients. However the resulting distribution obtained by Rutherford was not of the Pareto form.

Although there have been a number of other refinements and alternative models proposed (for summaries see Ord 1975; Dagum 1983), none seems to have provided much of an improvement in terms of explanatory power over the model of Champernowne. Attention has focussed more 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 1983). Recently Parker (1999) has 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).

In this paper a model is developed which predicts Paretian behaviour in both tails and which fits a variety of observed earnings and income distributions extremely well. Like the Champernowne model it is based on a stochastic model for the evolution of individual incomes. However unlike Champernowne's model it does not assume equilibrium conditions. Rather it explicitly incorporates the fact that the observed distribution of income over a population (or over a random sample from it) will depend on the age profile of the population. The age distribution will in turn depend on the demographic dynamics, which may not be in equilibrium. Thus for example for a growing population there will be more younger workers than older ones in the workforce (with on average lower incomes) and conversely for a declining population *etc.*

While there are several distributions derived in the paper corresponding to divers dynamic behaviour, the most important one is a four parameter distribution which I call the *double Pareto-lognormal* distribution. It exhibits Paretian behaviour in both tails, and is shown to provide an excellent fit to observed data on income and earnings, ranging over a a number of jurisdictions and times.

2 CHARACTERISTICS OF EMPIRICAL INCOME DISTRIBUTIONS.

Fig. 1 displays plots of four income and earnings distributions which are quite typical. They are for U. S. household income, 1997 (U.S Bureau of Census 1998); Canadian personal earnings, 1996 (Statistics Canada 1997); Sri Lankan 6-months household income, 1981 (Central Bank of Ceylon 1984) and Bohemian personal income, 1993 (Rutherford 1955). The plots are of density (frequency per unit of income) *vs.* income, both on logarithmic scales. Since income data is typically grouped into classes, what is actually plotted is frequency divided by class width *vs.* class mid-point. Also since the top income class is open-ended, the observed density for this class is not well defined, and therefore does not appear in the plots. In all the plots the right hand part appears to be asymptotic to a straight line of negative slope, in agreement with Pareto's law for high incomes. In addition, there is apparently linear behaviour at the left-hand end of the plots, in agreement with Champernowne's observation of Paretian tail behaviour for low incomes. This is further illustrated in Fig. 2, which shows (plotted on logarithmic axes) cumulative frequencies above and below various levels of income (or earnings).

The Canadian data is different from the rest in that the slope at the left-hand end is negative. However this distribution is for earnings, while the others are for income, either personal or household. Many distributions appear to follow a form similar to one or other of those displayed in Fig. 1. In the remainder of this paper I will present a stochastic model which predicts distributions like those displayed. A glance ahead at Figs. 4-7 which show the fitted distributions obtained from maximum likelihood estimates of the parameters, should suggest to the reader something of the explanatory power of the model.

3 THE MODEL.

The current distribution of incomes over a population is the same as the probability distribution of the income of an individual randomly selected from that population. Thus a stochastic model for the generation of the income of such an individual can be used to explain the observed distribution of incomes in a population or random sample. To this end assume that individual incomes follow *Geometric Brownian Motion* (GBM):

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

where $X(t)$ denotes an individual's income at time t, 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). This model is the continuous-time analogue of the proportional effects model of Gibrat (1931) and of the logarithmic random walk model used by Champernowne (1953). Individuals will enter the workforce at different times and with different starting incomes. (Note that the expression "enter the workforce" is appropriate for discussion of earned income. When discussing income from all sources it would be more accurate to use the expression "begin receiving income". For the sake of simplicity the former expression will be used throughout. Similarly the word "individual" should be interpreted as "household", where appropriate). Suppose that the distribution of starting incomes, $X_0(t)$ say, at time t, is lognormally distributed and also that it evolves as GBM:

$$dX_0 = \mu_0 X_0 dt + \sigma_0 X_0 dw$$

so that

$$\log(X_0(t)) \sim N(a + (\mu_0 - \frac{\sigma_0^2}{2})t, b^2 + \sigma_0^2 t)$$

where a and b^2 are the mean and variance of log(starting income) at some initial reference time $t = 0$, τ years ago, say. With these assumptions it can easily be shown that the current income, $X = X|T$, of a randomly selected individual, who entered the workforce T years ago (at time $\tau - T$) will be log-normally distributed with $\log(X)$ having mean

$$E(\log(X)) = a + (\mu_0 - \frac{\sigma_0^2}{2})(\tau - T) + (\mu - \frac{\sigma^2}{2})T = A_0 + (\tilde{\mu} - \frac{\tilde{\sigma}^2}{2})T$$

and variance

$$V(\log(X)) = b^2 + \sigma_0^2(\tau - T) + \sigma^2 T = B_0^2 + \tilde{\sigma}^2 T$$

where $A_0 = a + (\mu_0 - \frac{\sigma_0^2}{2})\tau$ and $B_0^2 = b^2 + \sigma_0^2\tau$ are the mean and variance of the log of current starting income (i.e at time $t = \tau$), and

$$\tilde{\mu} = \mu - \mu_0; \qquad \tilde{\sigma}^2 = \sigma^2 - \sigma_0^2.$$

If an individual is randomly selected from the current workforce, the time T that he or she has been in the workforce will be a random variable, whose distribution will reflect the the history of entry and exit from the workforce. Thus the distribution of current income, $\bar{X}$, say, of the randomly selected individual can be regarded as a mixture of lognormal distributions of the above type, with the mixing parameter T. If the distribution of T is known, or can be specified parametrically, one can in principle find the distribution of $\bar{X}$ (*e.g.* by integrating the density of the above lognormal distribution with respect to the density of T). This will provide a parametric form for the current distribution of incomes.

The simplest case considered will be for a workforce or population growing at a fixed (proportional) rate. In this case to a first approximation the time T since recruitment of a randomly selected individual will have an exponential distribution, with probability density function (p.d.f.)

$$f_T(t) = \lambda e^{-\lambda t}, \qquad t > 0.$$

In this case it can easily be shown (see Appendix) that the distribution of $\bar{Y} = \log(\bar{X})$ is that of the sum of two random variables, one with a normal distribution, and the other with a two-tailed exponential distribution and hence that the distribution of current income, $\bar{X}$, is that of the product of independent random variables Z and V say, where Z is log-normally distributed and V has a *double Pareto distribution*, with p.d.f.

$$f_V(v) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta}\, v^{\beta-1}, & \text{for } v \le 1 \\ \frac{\alpha\beta}{\alpha+\beta}\, v^{-\alpha-1}, & \text{for } v > 1 \end{cases} \tag{1}$$

where α and β are two positive parameters. From this the distribution of $\bar{X}$ can be determined. In the Appendix it is shown that it has density

$$\begin{aligned} f_{\bar{X}}(x) = \tfrac{\alpha\beta}{\alpha+\beta} \Big[& x^{-\alpha-1} \exp\{\alpha A + \alpha^2 B^2/2\} \Phi\left(\tfrac{\log x - A - \alpha B^2}{B}\right) + \\ & x^{\beta-1} \exp\{-\beta A + \beta^2 B^2/2\} \Phi^c\left(\tfrac{\log x - A + \beta B^2}{B}\right)\Big] \end{aligned} \tag{2}$$

on $x > 0$ where Φ is the cumulative distribution function of the standard normal distribution; $\Phi^c = 1 - \Phi$, and $A = A_0$ and $B^2 = B_0^2$ are the mean and variance parameters of the log-normally distributed component Z.

This distribution, which I shall call the *double Pareto-lognormal* (or *dPlN*) distribution is displayed in Fig. 3, which shows the density $f_{\bar{X}}(x)$ plotted against x in the two cases $\beta > 1$ and $\beta < 1$, as well as the logarithm of the density $f_{\bar{Y}}(y)$ plotted against y, where $\bar{Y} = \log(\bar{X})$. Note the similarity to the logarithmic plots for the empirical distributions displayed in Fig. 1. Also the Paretian behaviour of the distribution (2) in both tails is easily established *i.e.*

$$f_{\bar{X}}(x) \sim x^{-\alpha-1} \quad (x \to \infty); \qquad f_{\bar{X}}(x) \sim x^{\beta-1} \quad (x \to 0)$$

and can be seen in the logarithmic plots of Fig. 3. In terms of probabilities

$$\mathrm{P}(\bar{X} \ge x) \sim x^{-\alpha} \quad (x \to \infty); \qquad \mathrm{P}(\bar{X} \le x) \sim x^{\beta} \quad (x \to 0).$$

While the first of these results (limit as $x \to \infty$) is the usual Pareto law, the second (limit as $x \to 0$) corresponds to a form of the Pareto law in the lower tail, which seems to hold empirically in many cases (Fig. 2).

A distribution similar to the *dPlN* was proposed by Colombi (1990), who considered the distribution resulting from the product of independent Pareto and lognormal random variables, and named it the *Pareto-lognormal distribution.* This corresponds to the limiting form of the above *dPlN* distribution as $\beta \to \infty$. While Colombi showed that his distribution provided a very good fit to empirical income data, its genesis was essentially *ad hoc* and not based on an underlying stochastic model.

In order to fit the *dPlN* model, by maximum likelihood, to empirical income distribution data, which is typically grouped into disjoint cells, one needs to compute from (2) the probability θ_i of income $\bar{X}$ belonging to a cell i ($i.e$ to integrate $f_{\bar{X}}(x)$ between lower and upper cell boundaries). Expressions for the probabilities θ_i are given in the Appendix. The log-likelihood is

$$\ell = \sum_{i=1}^{m} f_i \log(\theta_i)$$

where $f_i \quad (i = 1, \ldots, m)$ are the observed frequencies in the m cells. Maximum likelihood (ML) estimates of the model parameters can be found by numerically maximizing ℓ over α, β, A and B.

If the observations do not correspond to a true random sample, the log-likelihood will not be as above. If there is information on the way the data were obtained, it may be possible to determine the correct log-likelihood. In the absence of such information one can regard the above procedure as providing what Barndorff-Nielsen (1977) has termed *maximum likeness estimates* and which minimize a directed measure of dissimilarity (the Kullback-Liebler information) between the observed frequency distribution and the estimated *dPlN* parametric distribution.

4 RESULTS.

The top panels in Fig. 2 exemplify the lower-tail Pareto law predicted from the model for the four empirical income distributions introduced in Sec 2 (and displayed in Fig.1). While it apparently only holds over a limited range for the Sri Lankan and Bohemian data (with few data points), the same could be said of the familiar upper-tail Pareto law (lower panels of Fig. 2) *e.g.* for Canadian and Sri Lankan data.

Figs. 4-7 show the the double Pareto-lognormal distribution fitted to the four empirical income distributions. The top two panels in each show the fitted density in natural and logarithmic scales superimposed on the observed distributions. The bottom two panels show the observed cell frequencies plotted against fitted frequencies and a Q-Q plot of quantiles of fitted distribution *vs.* quantiles of empirical distribution. (Note that because of the grouped nature of the data one cannot calculate exact quantiles of the empirical distributions. Here the upper boundary of each cell is used as the quantile corresponding to the proportion of observations at or below that level. Thus it provides only an upper bound of the true quantile. However for large datasets this value should be close to the true quantile). Clearly the model provides a very good fit to the data. Table 1 displays the ML estimates of the four parameters and of the Gini ratio as well as the chi-square goodness of fit statistic and the deviance (using percentage frequencies).

Note that for Canadian 1996 personal earnings, the ML estimate of β is less than one. This corresponds to the negative slope of left-hand arm in the log-log plot of density (Fig.1 top right-hand panel and Fig. 3 bottom right-hand panel), which in turn corresponds to a monotone decreasing distribution of income, with the modal class being the lowest income class (Fig. 4 top left-hand panel). In the Appendix it is shown that $\beta < 1$ if and only $\mu - \mu_0 + \lambda < \sigma^2 - \sigma_0^2$. Rates of growth for individual earnings could be lower on average over the population, than those of income (personal or household) because of the inclusion in the latter of social welfare payments, rental and interest income etc. Thus one might expect a smaller μ for earnings. How the variability in growth (σ^2) would compare for earnings and incomes is less clear, since the inclusion of social welfare payments would tend to lower the variability, while that of rental and interest income would tend to increase it.

Note also that $\hat{\alpha} - \hat{\beta} > 1$ for the more recent North American data (U.S. 1997 household income and Canadian 1996 individual earnings) but that for the Sri Lankan and Bohemian data $\hat{\alpha} - \hat{\beta} < 0$. In the Appendix it is shown that $\alpha - \beta < 1$ if and only if $\mu > \mu_0$ *i.e.* the growth rate of individual incomes on average exceeds that of starting incomes. This result would suggest that this condition holds for the Sri Lankan and Bohemian populations, but not for the Canadian or American ones, although of course it could be due to sampling error, with the likelihood being flat in the region of its maximum.

It seems implausible that in Canada and the U.S.A starting incomes should be growing faster than individual incomes or that the mean starting incomes should be respectively CAD\$43,200 and US\$68,900, as would be indicated from the ML estimates $\exp(\hat{A} - \hat{B}^2/2)$, (on the other hand the corresponding ML estimates for Sri Lanka (5,400 Rps.) and Bohemia (7.72 thousand Kr.) seem much more plausible). The reason for this inconsistency is probably related to the fact that the assumption of a workforce or population, growing at a fixed rate, is more closely approximated in the Sri Lankan and Bohemian cases than in the Canadian and American ones. Indeed under the assumption of a declining workforce (see following Section), the lognormal component of the distribution of $\bar{X}$ no longer represents the current distribution of starting incomes, and so μ no longer represents the mean of the log(starting income).

The four empirical distributions used above for illustration were chosen partly because of the excellence of the fit of the *dPlN* distribution, but also because of the fact that they cover a variety of types of income, countries and times. There are of course other examples that could have been used, where the fit would not have been quite so impressive. However by and large the *dPlN* distribution has proved to provide a very good fit.

5 EFFECTS OF CHANGING THE ASSUMPTION OF A CONSTANTLY GROWING POPULATION.

The double Pareto-lognormal distribution used in Sec.4 is based on the assumption that the distribution of the time, T, that an individual has been in the workforce is exponential. This is justified, at least to a first approximation, if the workforce or population is growing at a fixed (proportional) rate. Even though the double Pareto-lognormal distribution yields an extremely good fit to the four empirical distributions considered, it is doubtful whether the constant growth rate assumption is valid in all cases. An obvious question is whether a better fit can be obtained using other, perhaps more plausible, assumptions concerning workforce dynamics?

To this end suppose the number of recruits entering the workforce in time t to $t + dt$ was $R(t)dt$ and that at that time recruitment was growing at the rate $\lambda(t)$ *i.e.* $\lambda(t) = R'(t)/R(t)$. It follows then that $R(t) = R(0)\exp\left(\int_0^t \lambda(s)ds\right) = R(0)\exp\left(\Lambda(t)\right)$, say. If all participants were to leave the workforce K years after recruitment, the number of workers currently in the workforce (at time τ) would be $N(\tau) = R(0)\int_{\tau-K}^{\tau} \exp\left(\Lambda(s)\right) ds$. Of these the number who entered the workforce between t and $t+dt$ years ago is $R(0)\exp\left(\Lambda(\tau - t)\right) dt$. Thus under these assumptions, the probability density of the time T since recruitment of a randomly selected member of the current workforce is

$$f_T(t) = \frac{\exp\left(\Lambda(\tau - t)\right)}{\int_{\tau-K}^{\tau} \exp\left(\Lambda(s)\right) ds}, \qquad 0 < t < K$$

(More generally, one could consider the time K that an individual spends in the workforce as a random variable. In this case the marginal density of T could in principle be found by integrating the above expression with respect to the density of K).

It is possible to obtain closed-form expressions for the density $f_T(t)$ in some simple cases:

• *Constant Recruitment.* In this case $\lambda(t) \equiv 0$ and $f_T(t) = 1/K$ for $0 < t < K$. *i.e.* T is uniformly distributed on $(0, K)$. The resulting distribution of $\bar{X}$ can be shown to have p.d.f.

$$f_{\bar{X}}(x) = \frac{1}{\mu K}\left[x^{2\mu/\sigma^2 - 1} \exp\left(\frac{2\mu}{\sigma^2}(\mu\frac{B^2}{\sigma^2} - A)\right) J_1 - J_2\right]$$

where

$$J_1 = \Phi\left(\frac{\log(x) - A + \mu K + 2\mu B^2/\sigma^2}{\sqrt{K\sigma^2 + B^2}}\right) - \Phi\left(\frac{\log(x) - A + 2\mu B^2/\sigma^2}{B}\right)$$

and

$$J_2 = \Phi\left(\frac{\log(x) - A + \mu K}{\sqrt{K\sigma^2 + B^2}}\right) - \Phi\left(\frac{\log(x) - A}{B}\right)$$

where to avoid clutter the subscripts and tildes have been omitted from A_0, B_0, $\tilde{\mu}$ and $\tilde{\sigma}^2$. To calculate the log-likelihood for grouped data, the p.d.f. $f_{\bar{X}}(x)$ must be integrated numerically to obtain the cell probabilities θ_i, since closed-form expressions are not attainable. The p.d.f. of $\bar{X}$ is unimodal and looks somewhat like the *dPlN* density with $\beta > 1$ (Fig. 3 right hand panels), although it actually exhibits Paretian behaviour in at most one tail (the upper tail if $\mu < 0$; the lower tail if $\mu > 0$ and neither tail if $\mu = 0$. Nonetheless for large K the departures from Paretian behaviour may be slight. (NB: The result for $\mu < 0$ corresponds essentially to Champernowne's (1953) equilibrium result with negative mean drift).

• *Recruitment growing at a constant rate.* In this case $\lambda(t) \equiv \lambda$, a positive constant, and $f_T(t) = \frac{\lambda e^{-\lambda t}}{1-e^{-\lambda K}}$ for $0 < t < K$. *i.e.* T has a truncated negative exponential distribution on $(0, K)$. If K is considered to be large, one has essentially the exponential distribution considered in the previous section, leading to the double Pareto-lognormal distribution of observed incomes.

• *Recruitment declining at a constant rate.* In this case $\lambda(t) \equiv -\rho$, a negative constant, and $f_T(t) = \frac{\rho e^{\rho t}}{e^{\rho K}-1}$ for $0 < t < K$. For ρK large and $\mu > 0$ it can be shown that the distribution of $\bar{X}$ is approximately that of the product of a lognormal random variable with a random variable U, with density

$$f_U(u) = \frac{\alpha\beta}{\alpha - \beta}\left[u^{\beta-1} - u^{\alpha-1}\right], \quad 0 < u < 1$$

and zero elsewhere, where α and β are positive parameters ($\alpha \neq \beta$). The lognormal component no longer represents the current distribution of starting income - rather its logarithm has mean $A_0 + K(\tilde{\mu} - \tilde{\sigma}^2/2)$ and variance $B_0^2 + \tilde{\sigma}^2$. Using this approximation, closed-form expressions for the p.d.f. of $\bar{X}$ and the cell probabilities θ_i can be obtained. The p.d.f of $\bar{X}$ exhibits lower-tail Paretian behaviour. For $\mu < 0$ a similar result holds only now $\alpha, \beta < 0$;

the density of the component U has support on $\{1 < u < \infty\}$ and Paretian behaviour occurs only in the upper tail.

The constant recruitment and declining recruitment models were fitted to the US and Canadian data. In both cases the fit of the constant recruitment model provided a poorer fit than the double Pareto-lognormal model. However while the fit of the declining recruitment model was also, in both cases, inferior to that of the double Pareto-lognormal distribution, it was only marginally so (difference only in the third decimal of the chi-square and deviance statistics).

Another alternative would be to use actual data on growth in the workforce to determine the empirical distribution of the random variable T. One could then use a Monte Carlo procedure to determine the distribution of $\bar{X}$ by sampling from the distribution of T and for each value generating a lognormal deviate with the appropriate mean and variance parameters. In principle one could numerically calculate the likelihood (and hence obtain approximate ML estimates) by approximating the cell probabilities θ_i using the above Monte Carlo procedure with given values of the parameters. However in view of the added computational complexity and the loss of analytic simplicity this procedure was not carried out.

6 CONCLUDING REMARKS.

The main results of this paper are:

• to emphasize that many empirical income distributions exhibit Paretian behaviour in both tails. Although the lower-tail Paretian behaviour has been identified before (*e.g.* Champernowne 1953), unlike the upper-tail Paretian behaviour (Pareto's Law), it has not been regarded as a phenomenon requiring explanation.

• To demonstrate that a distribution of incomes with such two-tailed Paretian behaviour can be explained by a simple stochastic model for the evolution of individual incomes which recognizes the fact that individuals have been in the workforce for different lengths of time.

• To derive a new, flexible, four-parameter probability distribution (the double Pareto-lognormal distribution), which has a theoretical basis and which can be used to describe many income distributions. It is shown to fit extremely well to four different empirical income distributions.

Like any good model, the one used in this paper leaves out more than it includes. Nonetheless I claim that it captures the essence of the process underlying Pareto's law. The main novelty in the analysis is the recognition of the fact that an observed income distribution will depend on the "age" profile of the workforce, which in turn will depend on recruitment dynamics. Mathematically this involves determining the distribution of the state of a stochastic process after a random length of time. This differs from the approach used by Champernowne (1953) and others, who sought a stationary distribution, corresponding to a population in equilibrium.

The stochastic model used for the evolution of individual incomes is based on geometric Brownian motion (GBM) and is, in essence, a formulation of

Gibrat's (1931) law of proportional effects. It is analogous to Chapernowne's (1953) logarithmic random walk model. While in reality an individual's income would not change continuously (as does GBM) but rather would change only at discrete times, one could think of GBM as approximating a geometric Poisson jump process in which (proportional) changes in income are independent random variables occurring at random times (in a Poisson process). For such a process the mean and variance of the logarithm of the income of an individual t years after entering the workforce are both proportional to t, just as for GBM. While no evidence is presented to justify the use of GBM to model the distribution of starting incomes, it is convenient and probably provides a reasonable approximation – starting incomes are likely to have a skewed distribution and to evolve over time. While the GBM specification implies a variance which increases over time, this can be avoided if so desired by setting the variance parameter $\sigma_0^2 = 0$. The use of other specifications for starting incomes has not been considered, although it may be possible to do so and maintain analytic tractability.

The model used in the paper allows for differences between individuals (or households *etc.*) only through stochastic effects. Thus differences in educational background, age, sex *etc.* are not explicitly included. They can be thought of as entering the model through the distribution of starting incomes, and through the way in which the incomes of different individuals evolve stochastically. A subject for future research would be to treat these characteristics as covariates, and to relate the model parameters to them through some sort of generalized linear model (McCullagh and Nelder 1989).

Different assumptions on workforce dynamics have been shown to give rise to different observed income distributions, with Paretian behaviour in one or other tail. The key fact concerning Paretian tail behaviour relates to singularities in the moment generating function (m.g.f.) of $\bar{Y}$, the logarithm of observed income. If, as with an exponential mixing distribution (for T) the m.g.f. of $\bar{Y}$ has simple poles on opposite sides of the origin, then Paretian behaviour will occur in both tails. Upper- (respectively lower-) tail Paretian behaviour is associated with simple poles which are positive (respectively negative). If there are poles of higher order (as would arise for example if T followed an Erlang distribution), then true Paretian behaviour would not occur, although the density of $\bar{X}$ would differ from a power law in the tail only by a function exhibiting so-called regular variation (Bingham *et al.* 1987). In this case departures from Paretian behaviour might not be apparent in observed distributions.

The double Pareto-lognormal distribution, derived under the assumption of a workforce or population growing at a constant rate, provides a good fit to the empirical income distributions considered, even in cases when such an assumption is of questionable validity. This distribution satisfies the three conditions spelt out in the Encyclopedia of Statistical Sciences (Dagum 1983) as desirable properties of an income distribution model, *viz.* (a) stochastic foundation - it is deduced from an *a priori* set of probability assumptions; (b) convergence to the Pareto law; and (c) goodness of fit. In view of this it should be given serious consideration as an income distribution model. A subject for future research is to compare its performa nce in goodness of fit (over a range of empirical income distributions) to other proposed parametric forms, *e.g.* the Dagum (1996) distributions and the generalized beta distributions (McDonald 1984).

ACKNOWLEDGEMENTS.

I would like to thank Dr. Harry Clarke (La Trobe University), Professor David Giles (University of Victoria) and Professor Camilo Dagum (Università di Bologna) for their helpful comments and suggestions. Financial support was provided by NSERC grant OGP 7252.

APPENDIX: DERIVATIONS AND PROPERTIES OF THE DOUBLE PARETO-LOGNORMAL DISTRIBUTION.

Under the assumptions of the model of Sec. 3, the logarithm $Y = Y|T$ of the current income $X = X|T$ of an individual who entered the workforce T years ago has a lognormal distribution:

$$Y = \log(X) \sim N\left(A_0 + (\tilde{\mu} - \frac{\tilde{\sigma}^2}{2})T,\ B_0^2 + \tilde{\sigma}^2 T\right),$$

which has moment generating function (m.g.f.)

$$M_{Y|T}(\theta) = \exp\left(A_0\theta + B_0^2\theta^2/2 + \left[(\tilde{\mu} - \frac{\tilde{\sigma}^2}{2})\theta + \tilde{\sigma}^2\theta^2/2\right]T\right)$$

The distribution of the logarithm of current income of a randomly selected individual, $\bar{Y}$, say, will have m.g.f.

$$M_{\bar{Y}}(\theta) = \mathrm{E}\left(\exp(\theta\bar{Y})\right) = \mathrm{E}_T\left(\mathrm{E}_{Y|T}\left(\exp(\theta Y)\right)\right) = \mathrm{E}_T\left(M_{Y|T}(\theta)\right)$$

which from the above can be written

$$M_{\bar{Y}}(\theta) = \exp\left(A_0\theta + B_0^2\theta^2/2\right)\ M_T\left((\tilde{\mu} - \frac{\tilde{\sigma}^2}{2})\theta + \tilde{\sigma}^2\theta^2/2\right).$$

If T is exponentially distributed with parameter λ, then the m.g.f. of T is $M_T(\theta) = \frac{\lambda}{(\lambda-\theta)}$ from which it follows that

$$\begin{aligned} M_{\bar{Y}}(\theta) &= \frac{\lambda\exp\left(A_0\theta + B_0^2\theta^2/2\right)}{\lambda - (\tilde{\mu} - \frac{\tilde{\sigma}^2}{2})\theta - \frac{\tilde{\sigma}^2}{2}\theta^2} \\ &= \exp\left(A_0\theta + B_0^2\theta^2/2\right)\frac{\alpha\beta}{(\alpha-\theta)(\beta+\theta)} \end{aligned}$$

where α and $-\beta$ $(\alpha, \beta > 0)$ are the two roots of the characteristic equation

$$\frac{\tilde{\sigma}^2}{2}z^2 + (\tilde{\mu} - \frac{\tilde{\sigma}^2}{2})z - \lambda = 0. \qquad (3)$$

Now it is easily confirmed that $\frac{\alpha\beta}{(\alpha-\theta)(\beta+\theta)}$ is the m.g.f. of the *double exponential distribution* with density function

$$f(x) = \begin{cases} \frac{\alpha\beta}{\alpha+\beta}e^{\beta x} & \text{if } x < 0 \\ \frac{\alpha\beta}{\alpha+\beta}e^{-\alpha x} & \text{if } x \geq 0 \end{cases}$$

Also since $\exp\left(A_0\theta + B_0^2\theta^2/2\right)$ is the m.g.f of an $N(A_0, B_0^2)$ random variable, it follows that the distribution of $\bar{Y}$ can be represented as that of the sum of independent normal and double exponential random variables. Since the logarithm of a double exponential random variable has a double Pareto distribution as defined in Sec.3, it follows that the income $\bar{X}$ of a randomly selected individual can be represented as the product of independent lognormal and double Pareto random variables. The p.d.f. can be obtained from the p.d.f. of $\bar{Y} = \log(\bar{X})$ which in turn can be found by convolving a double exponential density with a normal density. The details are tedious and are omitted. The result is

$$f_{\bar{Y}}(y) = \tfrac{\alpha\beta}{\alpha+\beta}\Big[e^{-\alpha(y-A)+\alpha^2B^2/2}\Phi\left(\tfrac{y-A-\alpha B^2}{B}\right) + e^{\beta(y-A)+\beta^2B^2/2}\Phi^c\left(\tfrac{y-A+\beta B^2}{B}\right)\Big] \quad (4)$$

from which the p.d.f of $\bar{X}$ in Sec. 3 follows. Note that to avoid clutter the subscripts have been dropped from A_0 and B_0.

To fit the model by maximum likelihood to grouped income data, one can maximize the multinomial log-likelihood $\ell = \sum_{i=1}^{m} f_i \log(\theta_i)$, where θ_i is the probability (expressed parametrically in terms of α, β, A and B) of the income of a randomly selected income individual falling in cell i. The θ_i can be obtained by integrating the density $f_{\bar{Y}}(y)$ between the logarithms of upper and lower cell limits. With some work it can be shown that

$$\theta_i = \Phi\left(\frac{y_{i+1}-A}{B}\right) - \Phi\left(\frac{y_i - A}{B}\right) - \frac{\beta}{\alpha+\beta}I_1 + \frac{\alpha}{\alpha+\beta}I_2$$

where

$$I_1 = \exp\left(-\alpha(y_{i+1}-A-\alpha B^2/2)\right)\Phi\left(\frac{y_{i+1}-A-\alpha B^2}{B}\right) - \exp\left(-\alpha(y_i - A - \alpha B^2/2)\right)\Phi\left(\frac{y_i - A - \alpha B^2}{B}\right);$$

$$I_2 = \exp\left(\beta(y_{i+1}-A+\beta B^2/2)\right)\Phi^c\left(\frac{y_{i+1}-A+\beta B^2}{B}\right)$$

$$- \exp\left(\beta(y_i - A + \beta B^2/2)\right)\Phi^c\left(\frac{y_i - A + \beta B^2}{B}\right);$$

and y_{i+1} and y_i are respectively the logarithms of the upper and lower income limits for cell i.

With these formulas the log-likelihood, $\ell(\alpha, \beta, A, B)$, can be coded and then maximized numerically, using a routine that does not require derivatives, to find ML estimates of the parameters. The results in Sec. 4 (Table 1) were obtained using the S-Plus routine nlminb (Splus 4 1997).

The random variable $\bar{X}$ (the current income of a randomly selected individual) can be represented as

$$\bar{X} = e^Z V$$

where $Z \sim N(A_0, B_0^2)$ and V has the double Pareto distribution with p.d.f. (1). The lognormal component e^Z represents the distribution of current starting incomes, while the double Pareto component V represents the income of the selected individual relative to current starting incomes. One can easily show that $P(V > 1) > 1/2$ if and only if $\alpha < \beta$. In other words, loosely put, the distribution of current incomes across the whole workforce will on the whole be "above" that of current starting incomes if $\alpha < \beta$. The mean of the double Pareto distribution (1) can easily be shown to be

$$E(X) = \frac{\alpha\beta}{(\beta+1)(\alpha-1)}.$$

for $\alpha > 1$ (the expectation does not converge for $\alpha \leq 1$). From this it follows $E(\bar{X}) > E(e^Z)$ ($i.e$ the mean of current incomes over the whole workforce exceeds the mean current starting income) if and only if $\alpha < \beta + 1$. Examining the roots of the characteristic equation (3) this condition can be expressed in terms of the original model parameters. Precisely it is $\mu > \mu_0$ or in other words that the growth rate of individual incomes is on average higher than that of starting incomes. One would expect this to hold in most circumstances.

The double Pareto (and double Pareto-lognormal) distributions are qualitatively different in the two cases $\beta > 1$ and $\beta < 1$. In the former case both are unimodal distributions; in the latter the densities are decreasing. From the characteristic equation (3) it can be shown that $\beta < 1$ if and only if $\lambda < \sigma^2 - \sigma_0^2 - \mu + \mu_0$. Thus one might expect a decreasing income distribution (with its mode at zero) when λ and/or $\mu - \mu_0$ are small, and/or $\sigma^2 - \sigma_0^2$ is large. The combination of a slowly growing workforce with slow but variable increases in income for workers as they progress, could result in this condition being met.

REFERENCES.

Barndorff-Nielsen, O. (1977). "Exponentially Decreasing Distributions for the Logarithm of Particle Size," *Proceedings of the Royal Society , London, A* 353, 401-419.

Bingham, N. H., Goldie, C. M. and J. L. Teugels (1987) *Regular Variation,* Cambridge: Cambridge University Press.

Bordley, R.F., McDonald, J.B. and A. Mantrala (1996). " Something New, Something Old: Parametric Models for the Size Distribution of Income," *Journal of Income Distribution* 61, 91-103.

Central Bank of Ceylon (1984) *Report on Consumer Finances and Socio-Economic Survey 1981/82 Sri Lanka.* Colombo: Author, Statistics Department.

Champernowne, D.G. (1953) "A Model of Income Distribution," *Economic Journal* 23, 318-351.

Colombi, R. (1990) "A New Model of Income Distribution: The Pareto-Lognormal Distribution," *Studies in Contemporary Economics.* eds. C. Dagum and M. Zenga, Heidelberg: Springer-Verlag.

Dagum, C. (1996). "A Systematic Approach to the Generation of Income Distribution Models," *Journal of Income Distribution* 61, 105-126.

———(1983) "Income Distribution Models," *Encyclopedia of Statistical Sciences,* Vol 4 eds. Kotz, S. and N.L.Johnson, New York: Wiley, pp. 27-34.

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

Mandelbrot, B. (1960) "The Pareto-Lévy Law and the Distribution of Income," *International Economic Review* 1, 79-106.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models, 2nd. ed.* London: Chapman and Hall.

McDonald, J.B. (1984). "Some Generalized Functions for the Size Distribution of Income," *Econometrica* 52, 647-663.

Ord, J.K. (1975). "Statistical Models for Personal Income Distributions," in *Statistical Distributions in Scientific Work,* Vol. 2, ed. by Patil, G.P, Kotz, S. and J. K. Ord. Boston: Reidel, pp. 151-158

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

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

Rutherford, R.S.G. (1955). Income Distributions: a New Model. *Econometrica,* 23, 277-294.

S-PLUS 4. (1997) *Guide to Statistics.* Seattle: MathSoft Data Analysis Products.

Statistics Canada (1997) *Catalogue 13-217-XPB.* Ottawa: Statistics Canada.

United States Bureau of the Census (1998): Current Population Reports, Series P60-200. *Money Income in the United States:1997* Washington, D.C.: U.S. Government Printing Office.

FIGURE CAPTIONS.

Fig.1 Plots of income density against income both on logarithmic scales for four empirical income distributions. The Pareto Law is exemplified by the asymptotic linearity in the right-hand tail. Notice also the apparent linearity in the left-hand tail.

Fig.2 Lower-tail and upper-tail Pareto laws. These are demonstrated, for the four empirical distributions discussed in Sec. 2. by the linear nature of the relationship between cumulative frequencies below (top row) and above (bottom row) specified levels of income or earnings, when both are plotted on logarithmic scales.

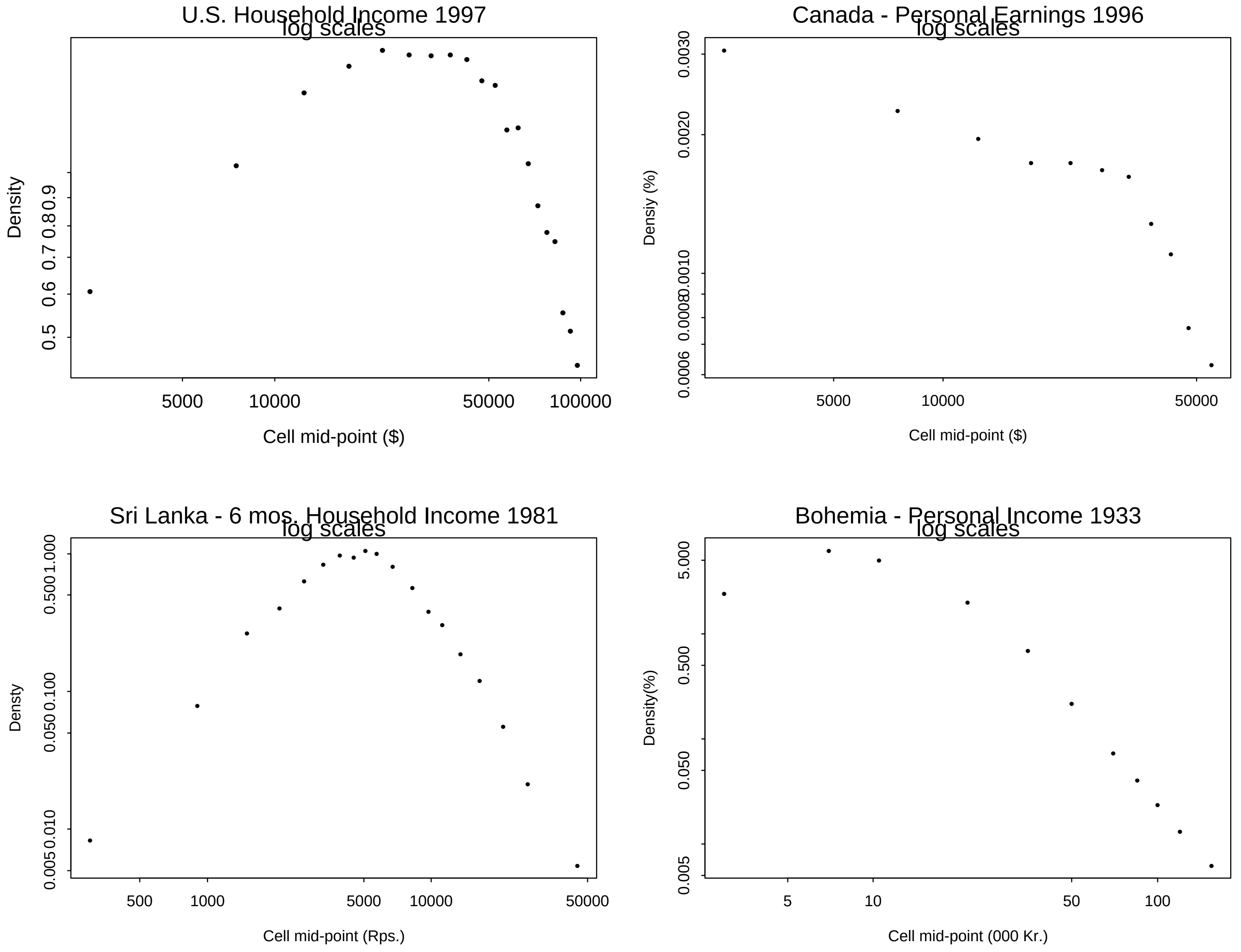
Fig.3 Plots of the shape of the density function of the double Pareto-lognormal distribution for the cases $\beta < 1$ and $\beta > 1$. The lower panels show the corresponding density of $\log(X)$ on the logarithmic scale. Observe the asymptotic linearity in both tails.

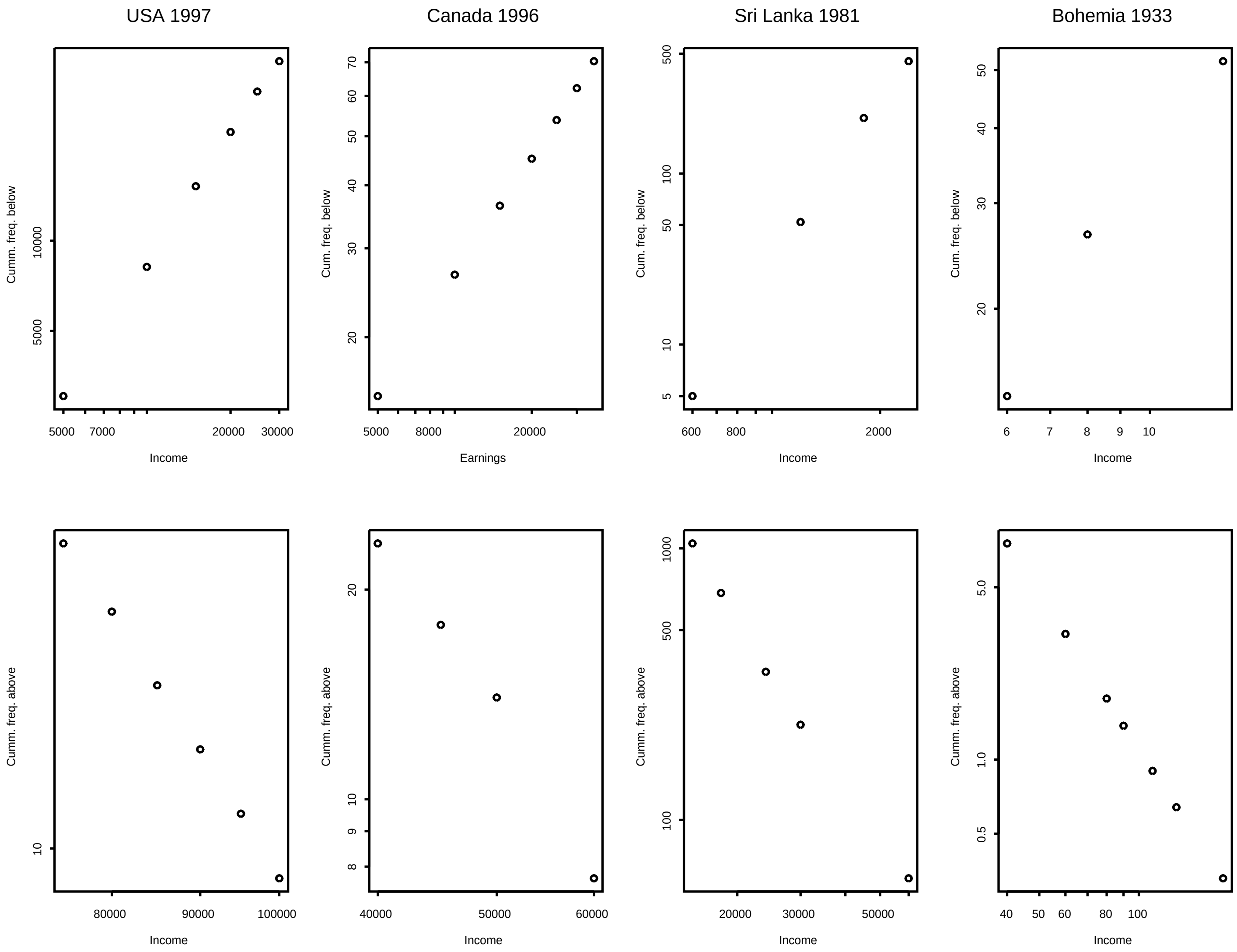
Fig.4 The double Pareto-lognormal distribution fitted by maximum likelihood to U.S household income (1997) data.

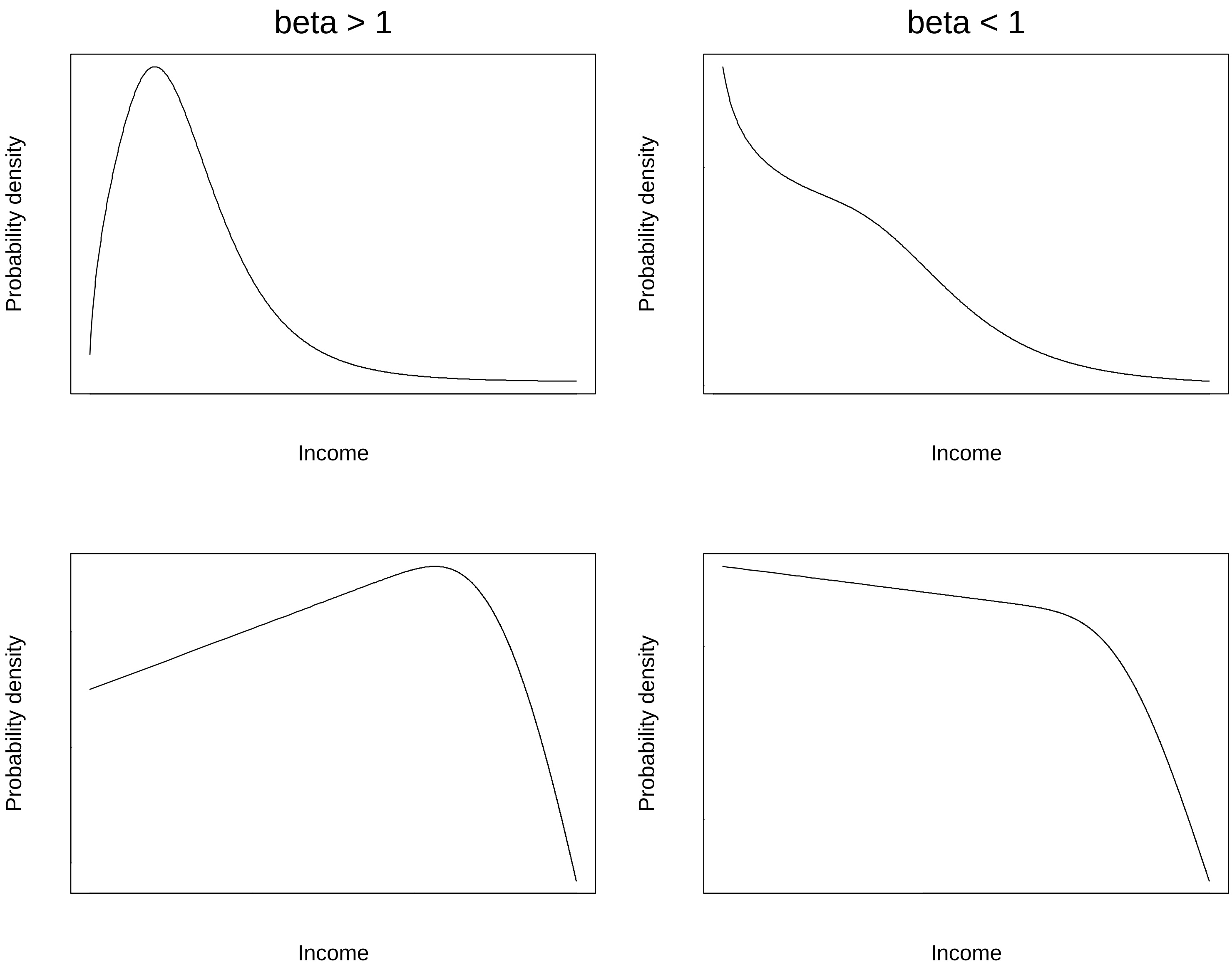
Fig.5 The double Pareto-lognormal distribution fitted by maximum likelihood to Canadian personal earnings (1996) data.

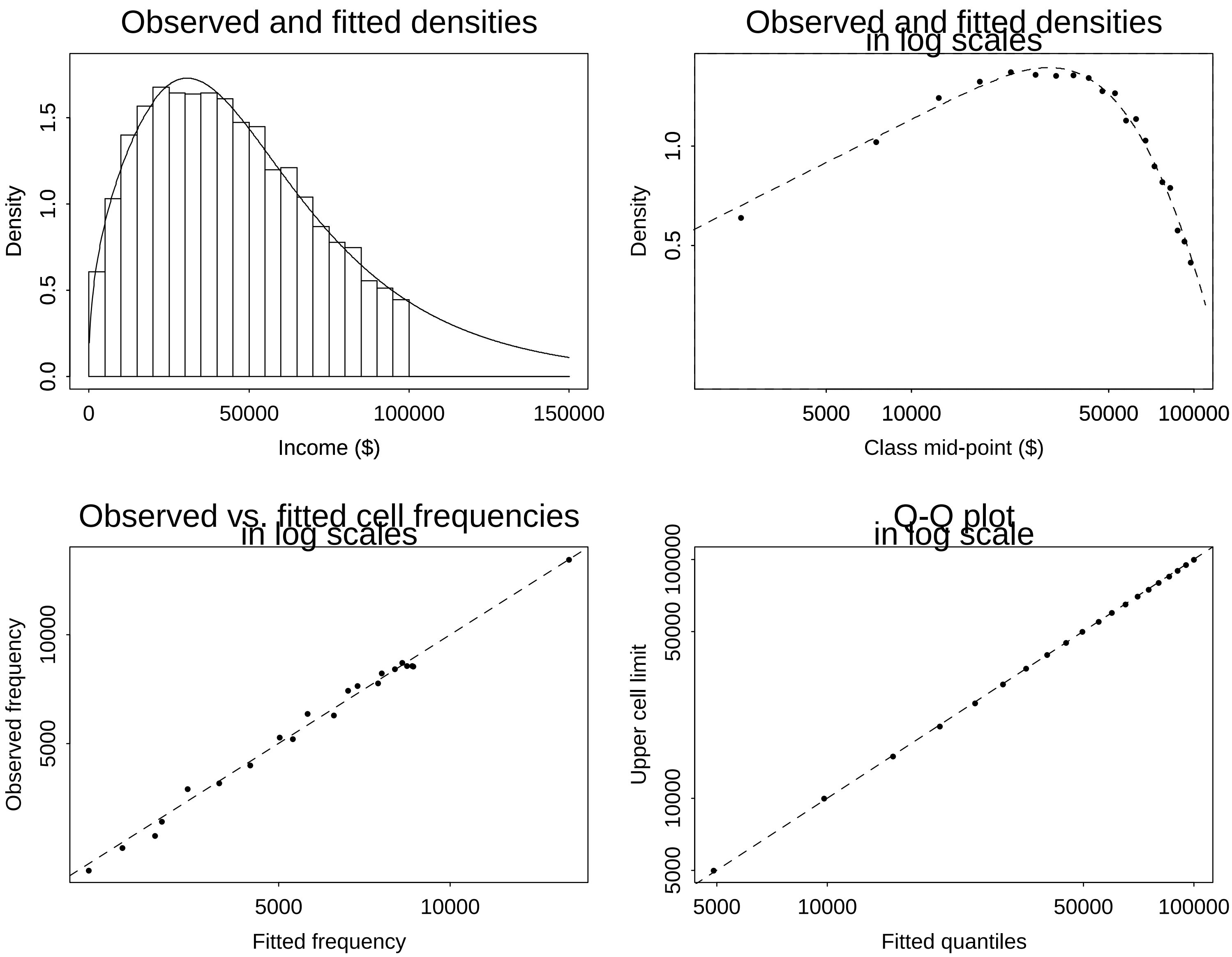
Fig.6 The double Pareto-lognormal distribution fitted by maximum likelihood to Sri Lankan six-month household income (1981) data.

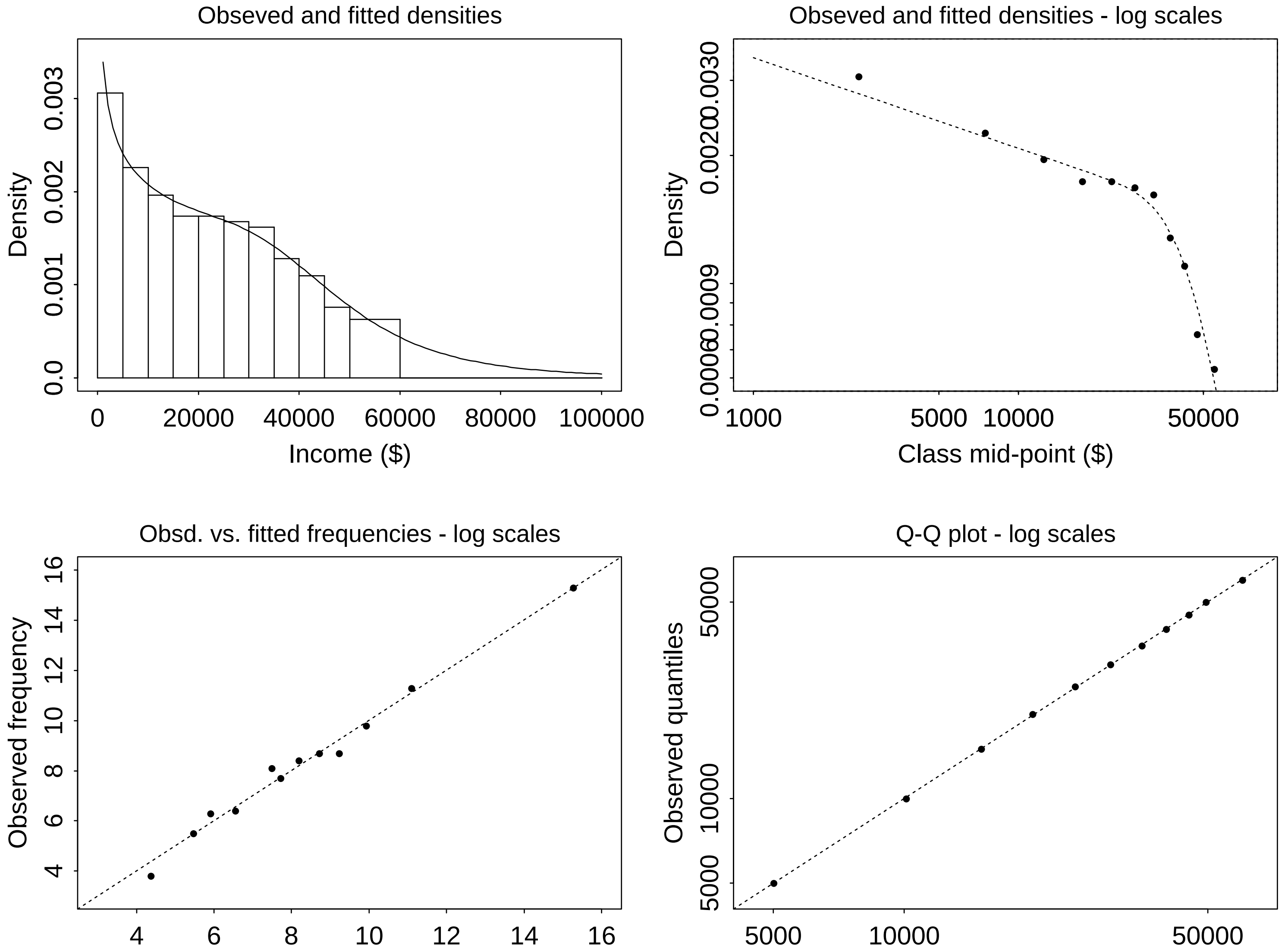
Fig.7 The double Pareto-lognormal distribution fitted by maximum likelihood to Bohemian personal income (1933) data.


U.S. Household Income 1997
log scales
Density
0.5 0.6 0.7 0.8 0.9
5000 10000 50000 100000
Cell mid-point ($)
Canada - Personal Earnings 1996
log scales
Densiy (%)
0.0006 0.0008 0.0010 0.0020 0.0030
5000 10000 50000
Cell mid-point ($)
Sri Lanka - 6 mos. Household Income 1981
log scales
Densty
0.005 0.010 0.050 0.100 0.500 1.000
500 1000 5000 10000 50000
Cell mid-point (Rps.)
Bohemia - Personal Income 1933
log scales
Density(%)
0.005 0.050 0.500 5.000
5 10 50 100
Cell mid-point (000 Kr.)

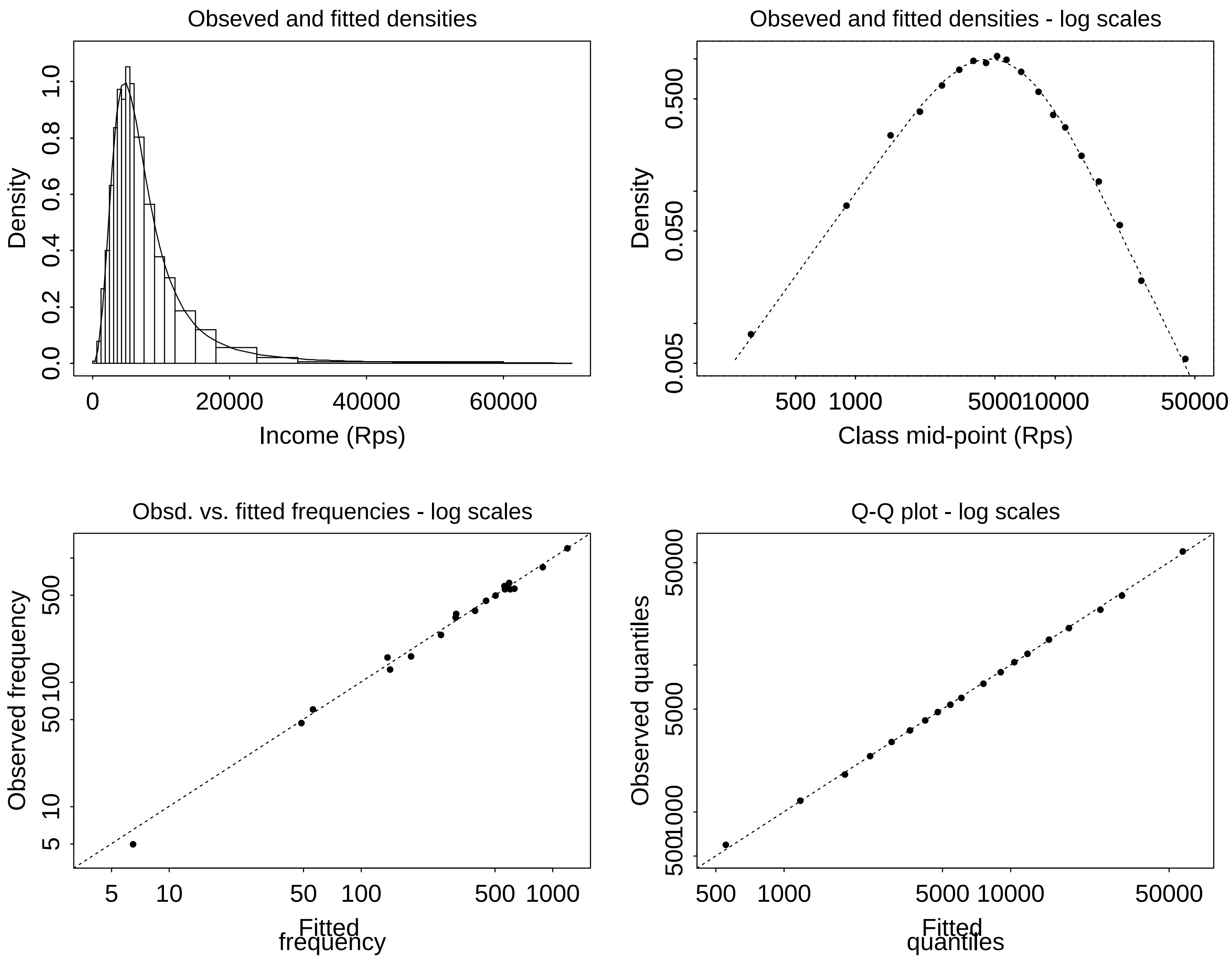

USA 1997
Canada 1996
Sri Lanka 1981
Bohemia 1933
Cumm. freq. below
Cum. freq. below
Income
Earnings
Cumm. freq. above
Income


beta > 1
beta < 1
Probability density
Income
Probability density
Income
Probability density
Income
Probability density
Income

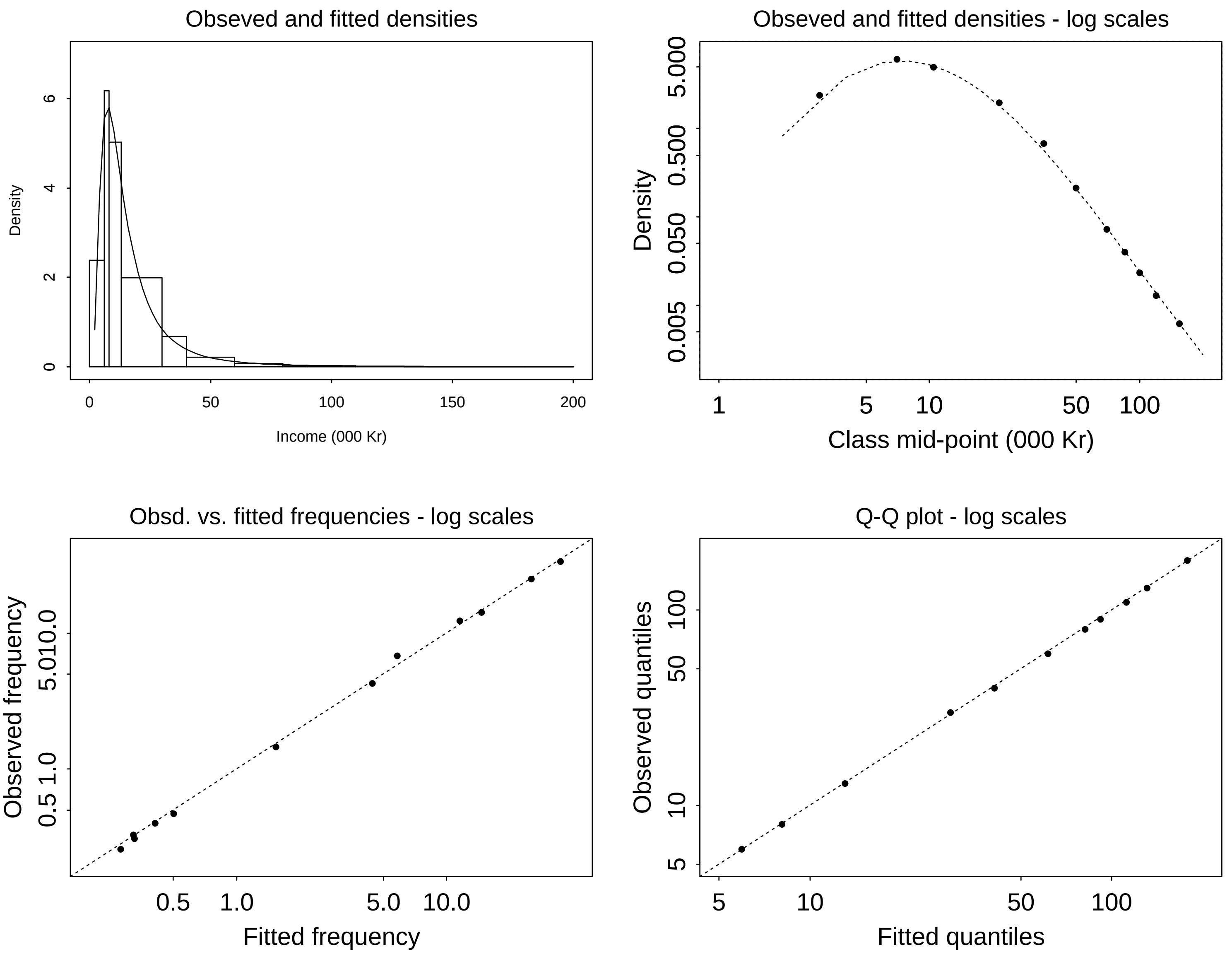

Observed and fitted densities
Density
Income ($)
Observed and fitted densities
in log scales
Density
Class mid-point ($)
Observed vs. fitted cell frequencies
in log scales
Observed frequency
Fitted frequency
Q-Q plot
in log scale
Upper cell limit
Fitted quantiles



Obseved and fitted densities
Density
0.0
0.001
0.002
0.003
0
20000
40000
60000
80000
100000
Income ($)
Obseved and fitted densities - log scales
Density
0.0006
0.0009
0.0020
0.0030
1000
5000
10000
50000
Class mid-point ($)
Obsd. vs. fitted frequencies - log scales
Observed frequency
4
6
8
10
12
14
16
Fitted frequency
Q-Q plot - log scales
Observed quantiles
5000
10000
50000
Fitted quantiles



Obseved and fitted densities
Density
Income (Rps)
Obseved and fitted densities - log scales
Density
Class mid-point (Rps)
Obsd. vs. fitted frequencies - log scales
Observed frequency
Fitted frequency
Q-Q plot - log scales
Observed quantiles
Fitted quantiles


Obseved and fitted densities
Density
Income (000 Kr)
0
2
4
6
0
50
100
150
200
Obseved and fitted densities - log scales
Density
Class mid-point (000 Kr)
0.005
0.050
0.500
5.000
1
5
10
50
100
Obsd. vs. fitted frequencies - log scales
Observed frequency
Fitted frequency
0.5 1.0
5.010.0
0.5
1.0
5.0
10.0
Q-Q plot - log scales
Observed quantiles
Fitted quantiles
5
10
50
100
5
10
50
100

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{A}$	$\hat{B}$	Gini	χ^2	Deviance
U.S. Hshd. 97	22.43	1.43	11.27	0.51	0.39	0.16	0.16
Can. Ind. 96	4.16	0.79	10.71	0.27	0.45	0.19	0.18
S.L. Hshd. 81	2.09	3.09	8.68	0.41	0.42	0.40	0.40
Bohemia Ind. 33	2.15	8.40	2.23	0.61	0.55	0.28	0.29

Table 1: Maximum likelihood estimates of the four parameters (α, β, A and B) and of the Gini ratio for the double Pareto - lognormal distribution, along with the Pearson χ^2 goodness-of-fit statistic and the deviance. The four cases considered are those discussed in Sec. 2