METHODS FOR SUMMARIZING AND COMPARING WEALTH DISTRIBUTIONS

Stephen P. Jenkins and Markus Jäntti

ISER Working Paper
Number 2005-05

Institute for Social and Economic Research

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It was responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1980 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national datasets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Acknowledgement:

Paper presented at the Luxembourg Wealth Study Workshop, Perugia, 27–29 January 2005. Jenkins's research was supported by a Nuffield Foundation Small Grant and core funding of ISER from the UK Economic and Social Research Council and the University of Essex. We are grateful to Frank Cowell, Philippe Van Kerm, and the Editors for their comments.

Readers wishing to cite this document are asked to use the following form of words:

> **Jenkins, Stephen P. and Jäntti, Markus (May 2005) 'Methods for summarizing and comparing wealth distributions', ISER Working Paper 2005-05. Colchester: University of Essex, Institute for Social and Economic Research.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

This paper reviews methods for summarizing and comparing wealth distributions. We show that many of the tools commonly used to summarize income distributions can also be applied to wealth distributions, albeit adapted in order to account for the distinctive features of wealth distributions: zero and negative wealth values; spikes in density at or around zero; right-skewness with long and sparse tails combined with non-trivial prevalence of extreme values. Illustrations are provided using data for Finland.

Contents

1. Introduction

Given data on one or more distributions of wealth, what methods should analysts use to describe and compare them? Perhaps a prior question is whether or not one needs any special methods since there is, after all, a huge literature on summarizing income distributions, measuring inequality, and so on. (Cowell (2000) is a recent survey.) In this paper, we argue that many standard tools can be used, but there are particular features of wealth distributions that make empirical analysis non-standard in several ways.

Our target reader is an empirical researcher who is interested not only in describing distributions in a relatively simple way, but also in making distributional comparisons that have some normative interpretation. To put things another way, for the purposes of some analysts, simple descriptions of distributions in terms of, say, average wealth and the shares of total wealth held by the richest 1%, the richest 5%, and the richest 10%, are sufficient. Other analysts may wish to make more sophisticated comparisons of wealth inequality based on the complete distribution in the same way that is routine now in income inequality analysis, and wonder whether one can undertake comparisons of wealth data using standard tools such as Lorenz curves, Gini coefficients, and so on.

We review issues and methods for the second type of analyst, beginning in Section 2 with a review of what is distinctive about wealth distributions in contrast to income distributions. The next four sections consider methods per se. Estimation of distribution, quantile, and density functions, is discussed in Section 3. Section 4 considers Lorenz curves, not only the conventional relative Lorenz and generalized Lorenz curves, but also the absolute Lorenz curve (Moyes 1987) which has some advantages in this context. In Section 5, we discuss summary indices of inequality, including the Gini coefficient. Section 6 considers how to summarize wealth distributions using parametric functional forms, moving beyond the commonly-used Pareto distribution to more general finite mixture models. Section 7 contains concluding remarks. To maximize dissemination of the methods described, we illustrate them using wealth data for Finland and provide the Stata scripts ('do' files) used to derive the estimates and draw the graphs. The data are described in Appendix 1 and the scripts are set out in Appendix 2.

We assume throughout that the wealth data to hand are unit record ('micro') data – as with all the data included in the Luxembourg Wealth Study – and so do not consider methods for summarizing wealth data in grouped (banded) form, as presented in publications such as

Inland Revenue (2003) for Britain, for example. We also assume that any definitional issues have been resolved, for example the appropriate unit of assessment, whether wealth should be equivalized, the precise definition of wealth itself and its components, the appropriate deflators for cross-time and cross-national comparisons, and so on. Sierminska and Smeeding (2005) review several of these issues.

We focus on summaries of cross-sectional distributions, and do not consider methods for analysis of the longitudinal dynamics of wealth distributions. (See Klevmarken (2004) for a recent study of wealth dynamics.) Nor do we examine other types of multivariate distributions such as the joint distribution of income and wealth (cf. Mosler 2002). We give no attention to issues of statistical inference.

2. What is different about wealth distributions?

For our purposes, it shall be sufficient to distinguish between three dimensions of financial wealth. They characterize the aspects of wealth that are most interesting from a normative perspective and that have been studied most:

1. gross wealth, G_i, which is the aggregate for each unit $i = 1,\ldots, N$ of its financial assets (for example cash, money in bank accounts, stocks, bonds, property, and so on);
2. debt, D_i, which is the (negative of the) aggregate debt held by each unit; and
3. net wealth, $W_i = G_i – D_i$, which is gross wealth minus debt for each unit.

What are the distinctive features of these variables that analysis must take account of? First, there is their support. Gross wealth may have a value of zero, and positive values above that. Debt is negative wealth, and ranges from zero to large negative values. However it may be summarized in the same manner as gross wealth if we examine the distribution of the negative of debt values. (The analogy is with the approach to the measurement of poverty that analyzes the distribution of individual deprivation – typically assumed to be some function of an income shortfall from the poverty line (cf. Atkinson 1987) – and less deprivation corresponds to an improvement in social welfare.) Net wealth may legitimately take on negative, zero, and positive values, and the prevalence of negatives and zeros is relatively high (see the illustrations below). Indeed, mean net wealth may be negative.

This situation is quite different from that for income. It is assumed in the formal literature on income inequality and measurement that incomes can only take on positive values. And although zero and negative values for income are found in survey data, they are usually treated as nuisance observations. The convention is to either omit them from the analysis or to include them but censor them at zero (or a very small positive value). The prevalence of such observations is low, so the choice between the alternatives is unlikely to substantially affect a comparison of two distributions. Their exclusion allows one to use standard inequality measurement tools. So, can one apply these tools to analyze distributions of wealth?

Amiel et al. (1996) answer persuasively in the affirmative. They argue that it is appropriate to continue to make the assumption of monotonicity when there are negative values for the economic variable of interest. Suppose that aggregate social welfare is a function of the wealth of each wealth-holding unit within the population. Then it is appropriate to suppose that social welfare is increased by a increase in the wealth of any one unit, *ceteris paribus*, regardless of whether the unit's wealth is initially negative (or zero or positive). In addition, there appears to be no strong reason why the principle of transfers should not hold when there are negative values for the variable of interest, i.e. that a mean-preserving spread in wealth may be supposed to reduce social welfare, even if the mean of the wealth distribution is negative. The upshot is that non-crossing Lorenz curves may be given their standard interpretations even if, as discussed below, the shapes and positions of the curves are not always the ones that conventionally arise in the analysis of incomes.

A second distinctive feature of wealth distributions concerns the concentration of density mass. There is often a marked spike at zero because a relatively large fraction of the population has no financial wealth or debt. (Similar spikes do not occur with income distributions except, perhaps, in countries in which a relatively large fraction of the population receives the same social assistance benefit.) Spikes such as these complicate the estimation of frequency density functions (see below). In addition, it is often the case that there are many units with very small positive wealth holdings (for example some cash, or a little money in the bank), or with a small amount of negative net wealth. Thus, in addition to there being a spike exactly at zero, a substantial fraction of the density mass lies close to zero.

A third feature of wealth distributions is that they are right skewed with long and sparse right-hand tails, as are income distributions. A related feature, also shared with income distributions, is that there is a non-trivial prevalence of extreme values. These observations may be 'dirt' – error-ridden values that contaminate the data in the sense of Cowell and Victoria-Feser (1996) – or they may be genuine observations with high 'leverage', in the sense of providing valuable information. In either case, estimates of distributional summary statistics may be unduly sensitive to the inclusion of these observations. The precise impact depends on both the measure used and the shape of the particular distribution in question. We draw attention to the prevalence of such observations in our data, and explore the sensitivity of some estimates to their exclusion.

Let us now turn to our data for Finland to illustrate the points made. The data are derived from surveys of wealth in 1994 and 1998. The unit of assessment is the household and all wealth variables are expressed in '2000 international dollars' (and are not equivalized for differences in household size or composition). There are 5,210 households in the 1994 survey, and 3,893 in the 1998 survey. All calculations use the sampling weights produced by the data provider. Appendix 1 describes the data in more detail. For a more detailed analysis of them, see Jäntti (2004). We consider the distributions of gross wealth, debt, and net wealth, as defined earlier, focusing for the most part on net wealth as it is that which raises the most new issues.

Table 1 provides a number of summary statistics. Consider first the prevalence of zero and negative values. In both 1994 and 1998, just under 1½% of households had no gross wealth and fewer than 1% had no net wealth, but more than one third had no debt (35% in 1994, 39% in 1998). More than one in ten Finnish households had negative net wealth: 12.7% in 1994; 10.5% in 1998. The picture for income is quite different. Virtually none of the same Finnish households had zero disposable income – just 0.1% in 1994 and 0% in 1998 (Jäntti 2004). Over the four-year period, average (mean) debt rose slightly, but this was offset a relatively large increase in mean gross wealth; as a result, mean net wealth rose by some 29%, from $65,066 to $83,046. By contrast, average disposable income increased by 15%.

Table 1 also provides some information about the extreme values, reporting the four highest values for each wealth variable, and also the four lowest values for net wealth. It is at the top of range that there appears the most scope for extreme observations to influence calculations: note the large differences between each of the four richest gross wealth and four richest net wealth values.

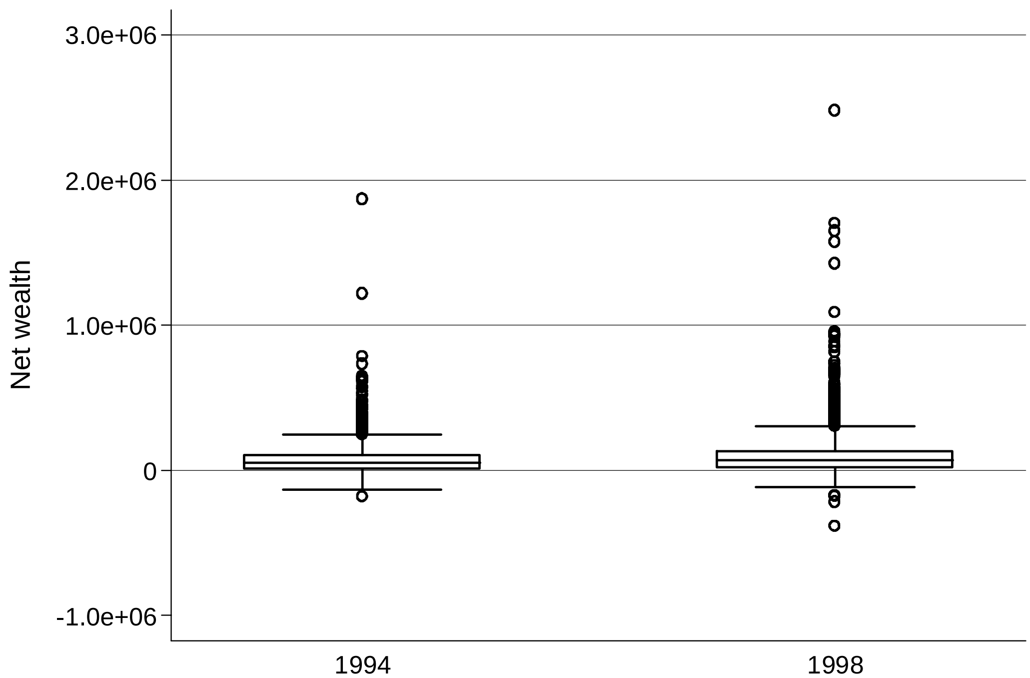
Table 1
Wealth in Finland, 1994 and 1998: summary statistics

	1994	1998
Gross wealth (G)		
% with $G = 0$	1.4	1.3
% with $G > 0$	98.6	98.7
Four largest values	732,380	1,573,263
	786,716	1,646,121
	1,217,829	1,270,825
	1,873,044	2,476,660
Mean	82,408	101,720
Mean if $G > 0$	83,614	103,064
Debt (D)		
% with $D = 0$	34.9	39.3
% with $D > 0$	65.1	60.7
Four largest values	189,403	283,434
	192,339	313,153
	205,088	415,940
	210,446	442,866
Mean	17,342	18,673
Mean if $D > 0$	26,625	30,758
Net wealth (W)		
% with $W < 0$	12.7	10.5
% with $W = 0$	0.9	0.8
% with $W > 0$	86.4	88.8
Four smallest values	–182,839	–385,016
	–132,234	–219,123
	–127,487	–176,739
	–125,596	–176,346
Four largest values	729,948	1,573,263
	786,716	1,646,121
	1,217,829	1,700,825
	1,869,533	2,476,660
Mean	65,066	83,046
Mean if $W > 0$	77,177	95,422
Mean if $W < 0$	–12,684	–16,010

Note. All money values are expressed in ‘2000 international dollars’.

Additional information about the long tails of the distribution of net wealth, and the extreme values in particular, is provided by the boxplots shown in Figure 1. The top and bottom of each dark rectangular area (the ‘box’) mark the upper and lower quartiles for a given year; the median is marked by the horizontal line through the box. The ‘T’ above the box and the '⊥' below it (the 'whiskers') show the 'adjacent values', i.e. the upper quartile plus 1.5 times the inter-quartile range and lower quartile minus 1.5 times the inter-quartile range, respectively. There are additional extreme observations outside the wide range spanned by the adjacent values, shown by the circles above and below the whiskers, with apparently greater prevalence and range in 1998 compared to 1994. One might suspect that these changes were partly responsible for the increase in mean net wealth. Indeed if one trims the richest 1% and poorest 1% of net wealth values, the estimated increase in the mean is 24% rather than 29%. In Section 4, we examine how sensitive estimates of inequality indices are to the exclusion of extreme values.

Figure 1
Boxplots for net wealth in Finland, 1994 and 1998

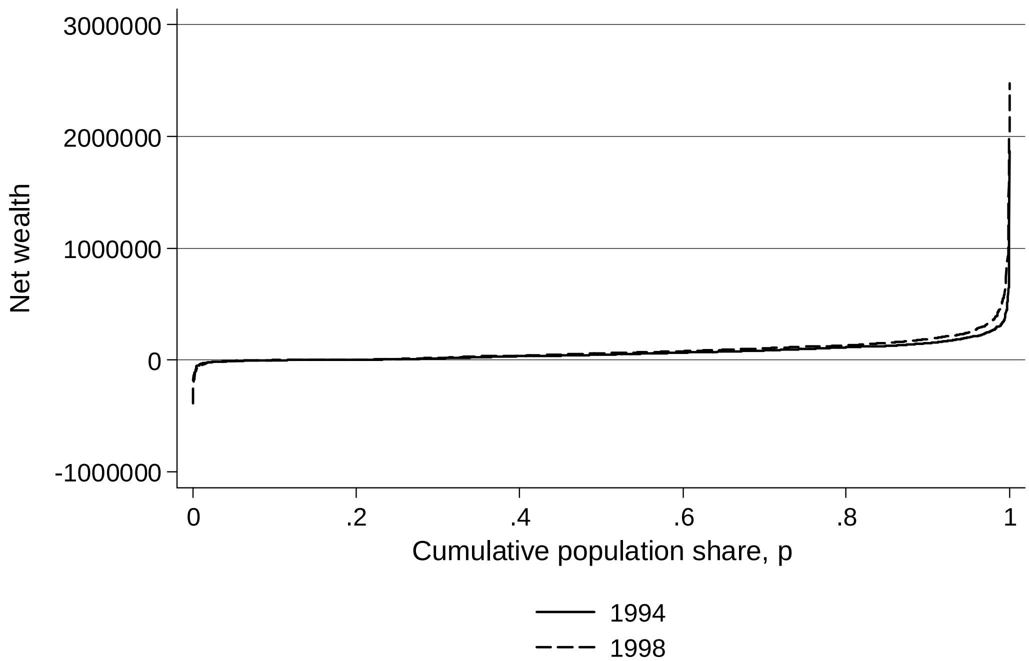


3. Distribution and density functions

Jan Pen's (1972) Parade of Dwarfs and a few Giants is a famously evocative description of the distribution function for income: the parade's silhouette is the shape of the graph of $p = F(W)$ against W). The parade concept can be applied to wealth as well as income: the cumulative distribution function for wealth and its inverse (the quantile function) is well-defined, regardless of whether there are negative values for wealth or not.

Pen's Parades for net wealth in Finland are shown in Figure 2. The pictures for 1994 and 1998 are similar in shape, except that the tall and the real giants (those in the richest tenth) became somewhat taller over the four year period. It is striking how tall the giants are in both years, relative to the majority of other households. Observe too that is not until between one tenth and one one fifth of the parade has passed by that we see a household with revealing any height above ground (i.e. with positive net wealth). As in Pen's own parade for incomes, there are households in the net wealth parade that are upside down, but there are more of them here.

Figure 2
Pen's Parades (CDFs) for net wealth in Finland, 1994 and 1998



The quantiles of the gross wealth, debt, and net wealth distributions are summarized numerically in Table 2. The estimates reveal a similar trend for all three wealth variables, viz distinct growth in real terms for the top percentiles combined with little change or even some falling back for the smallest percentile. In other words, there was an anti-clockwise twist to the shape of the wealth parade over the four year period. This suggests that inequality grew, a hypothesis that we examine in more detail below.

Table 2
Wealth in Finland, 1994 and 1998: quantiles

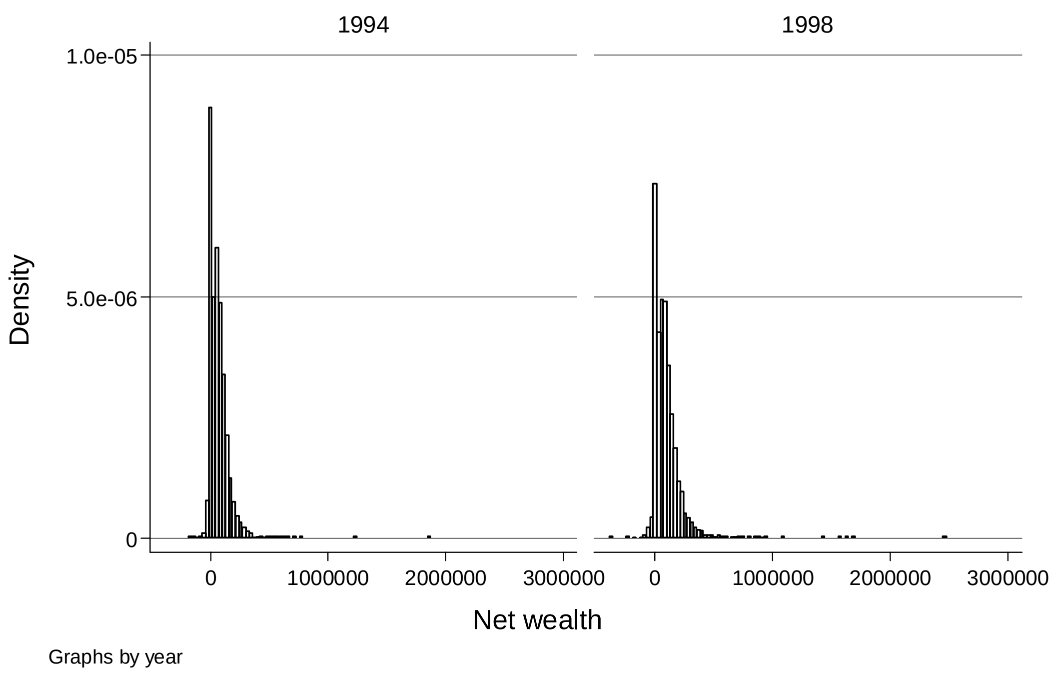
	1994		1998	
	Quantile	(as % of median)	Quantile	(as % of median)
Gross wealth (G)				
1	0	(0)	0	(0)
5	333	(0.5)	301	(0.4)
10	1,526	(2.1)	1,166	(1.4)
20	8,776	(12.2)	9,034	(10.9)
30	36,990	(51.2)	37,024	(44.7)
40	57,369	(79.5)	64,659	(78.1)
50 (median)	72,206	(100.0)	82,783	(100.0)
60	87,504	(121.2)	102,568	(123.9)
70	104,826	(145.2)	126,306	(152.6)
80	127,751	(176.9)	156,558	(189.1)
90	171,531	(237.6)	208,238	(251.5)
95	215,250	(298.1)	265,366	(320.6)
99	343,929	(476.3)	513,194	(620.0)
Debt (D)				
1	0	(0)	0	(0)
5	0	(0)	0	(0)
10	0	(0)	0	(0)
20	0	(0)	0	(0)
30	0	(0)	0	(0)
40	579	(13.1)	266	(6.7)
50 (median)	4,416	(100.0)	3,986	(100.0)
60	9,977	(225.9)	9,920	(248.9)
70	18,606	(421.3)	20,939	(525.3)
80	34,357	(778.0)	35,429	(888.9)
90	54,838	(1,241.8)	56.687	(1,422.2)
95	72,705	(1,646.4)	79,716	(2000.0)
99	111,043	(2,514.6)	125,774	(3,155.6)
Net wealth (N)				
1	–30,381	(–60.5)	–44,287	(–71.3)
5	–8,998	(–17.9)	–6,430	(–10.4)
10	–1,857	(–3.7)	–404	(–0.6)
20	2,069	(4.1)	3,011	(4.9)
30	13,857	(27.6)	18,600	(30.0)
40	32,326	(64.3)	40,929	(65.9)
50 (median)	50,239	(100.0)	62,081	(100.0)
60	67,845	(135.0)	79,158	(127.5)
70	85,730	(170.6)	104,274	(167.96)
80	112,189	(223.3)	134,853	(217.22)
90	152,154	(302.9)	186,745	(300.8)
95	201,051	(400.2)	246,765	(397.5)
99	321,809	(640.6)	476,259	(767.2)

Note. All money values are expressed in '2000 international dollars'.

The parade draws attention to the extremes of the distribution, but provides little detail about the wealth of the dwarf households who comprise the vast majority of the population. The same may be said of the boxplots for net wealth shown in Figure 1. The whiskers and the extreme observations beyond them dominate the picture, obscuring the relatively large changes that occurred over much of the wealth range, and that are summarized by the changes in percentiles shown in Table 2. The frequency density function is a device which reverses this emphasis.

The simplest approach to density estimation is to use histograms: Figure 3 shows them for Finnish net wealth in 1994 and 1998. The distinctive features of wealth distributions that were cited earlier are clearly apparent, including the large spike at zero. There also appears to be an additional mode close to zero in both years.[1] However, it is not obvious whether this is an artefact of the histogram construction or whether, more generally, the nature of the picture derived is sensitive to the number of bins used and their positioning along the support of the distribution.

Figure 3
Histograms for net wealth in Finland, 1994 and 1998



Note: fixed bin-width histogram, 100 bins.

[1] The economic explanation for the second mode is not clear. It might be accounted for by small amounts of cash and money in the bank, such as the remainder of that month's pay. The mode is also apparent in the density for gross wealth, but not in the density for debt.

Kernel density estimation methods are designed to address these and other issues.[2] To account for right skewness in distributions combined with sparse tails, analysts typically either (i) transform the variable of interest (for example taking logarithms), estimate the density of the transformed variable, and then reverse the transformation to derive the density estimates in the original metric, or (ii) use an adaptive kernel density estimator, which uses wider bandwidths in sparse regions of the support and narrower bandwidths in less sparse regions of the support, or (iii) both of these methods. Application of method (i) is problematic with wealth data because the prevalence of zero and negative values means that the standard transformations are not defined for all observations. Basing estimation on the subset of observations with positive values may omit a significant part of the story to be told. Application of method (ii) may address issues associated with having negative values, but it also does not solve the concomitant issue of the spike at zero. By their very nature, kernel density estimators smooth data within a window of observation, and hence inevitably transfer some density mass from a (genuine) spike to neighbouring values.

The statistical literature suggests some transformations that can be applied to variables that can take values along the whole real line. For instance, Burbidge et al. (1988) use the inverse hyperbolic sine and generalized Box-Cox transformations.[3] Although these transformations were developed for use in a regression context in order to render residuals more normally distributed, they could also be applied in a density estimation framework. We leave this as a topic for future research, and rely here on method (ii).

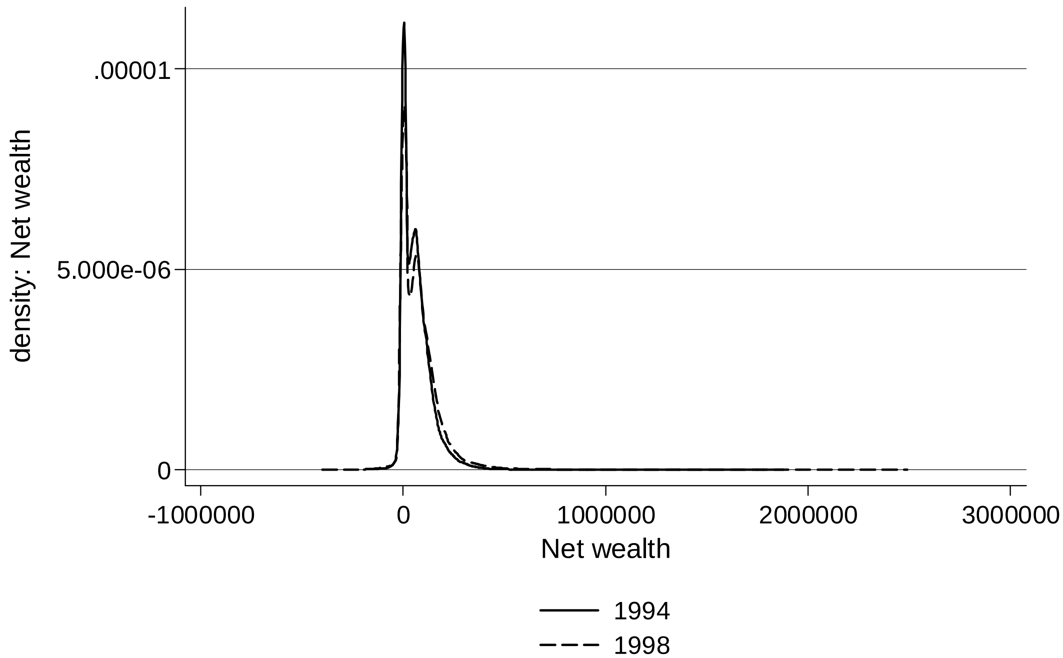
It turns out that kernel density estimates for the Finnish net wealth distribution convey a similar picture to the corresponding histograms (Figure 3). Figure 4 shows estimates derived using an adaptive kernel density estimator; similar pictures arose for kernel density estimates based on a fixed-bandwidth (the conventional 'optimal' bandwidth). The estimates point to a large concentration of density mass at zero and a very sharp falling away in mass at values below that. There is a second mode relatively close to zero, but again concentration declines fairly rapidly at values above this. The main change between 1994 and 1998 was a reduction in mass at each of the modes, shifted rightwards to the range between about \$200,000 and \$500,000.

---

[2] See Cowell et al. (1996), for example, for a non-technical review of kernel density estimation, and applications to UK income data for the 1980s.

[3] We are grateful to Arthur Kennickell for pointing us to the inverse hyperbolic sine transformation.

Figure 4
Density estimates for net wealth in Finland, 1994 and 1998



Note. Adaptive kernel density estimator, Epanechnikov kernel, 1000 data points.

4. Lorenz curves

The most common method of summarizing wealth distributions is in terms of the shares in total wealth of the richest *x*% where *x* equals 1, 5, or 10, for example. This is equivalent to reporting selected ordinates of the Lorenz curve (in this case, the shares of the poorest 99%, 95% and 90% respectively). What if we wish summarize the Lorenz curve as a whole, including the wealth shares of the poorest households whose wealth may be zero or negative?

To examine the nature of the Lorenz curve in this case, it is instructive to note that its slope at each $p = F(W)$ is equal to W/μ_W where μ_W is mean wealth.[4] Consider first the case when mean wealth is positive. Then, starting from the poorest unit, the Lorenz curve has a negative slope, lying below the horizontal axis, over the range of negative wealth values. Then the curve is horizontal, corresponding to the population subgroup that has zero wealth, and has the conventional positive slope over the remaining units (with positive wealth values). The

Lorenz curve takes on an even more non-standard shape in the case when $\mu_W < 0$. Starting from the poorest unit, the curve has a positive slope, and may lie above the 45° line representing perfect equality. The Lorenz curve is horizontal where wealth is zero and then has a negative slope over the remaining wealth units. As Amiel et al. (1996, p. S65) put it, relative to the conventional picture of a Lorenz curve, the curve in this case appears to be 'flipped vertically'.

Lorenz curves for Finnish net wealth are shown in Figure 5 (this is a case in which $\mu_W > 0$). The curve hangs beneath the horizontal axis up to almost the poorest 40% of the population in both years: in fact, the share in total net wealth of the least wealthy 40% was only 2.2% in 1994 and 2.9% in 1999. In contrast, the wealthiest tenth received more than one third of total wealth in 1994 (35.5%) and almost four-tenths in 1998 (38.2%).

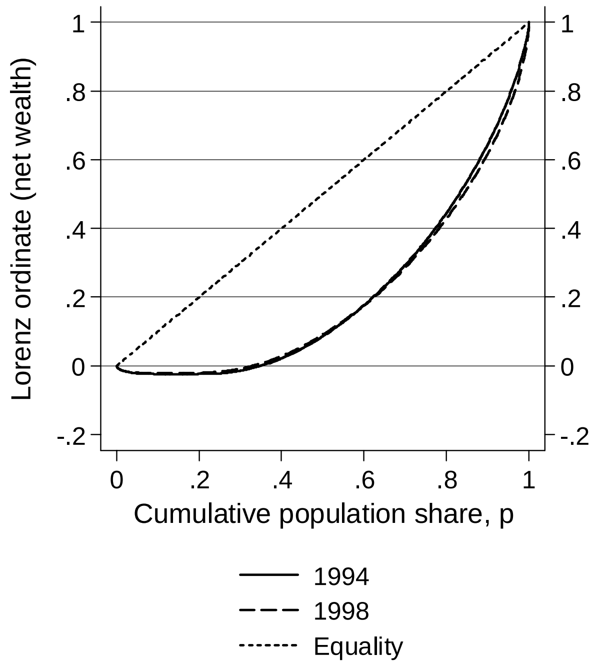
Taking the Lorenz curves as a whole, we see that the 1998 curve lies slightly above the 1994 curve up until a population share of about 65% and lies below the 1994 curve thereafter. What conclusions can be drawn from this configuration (issues of statistical inference aside)? Here we recall the discussion of Section 2, in particular the arguments of Amiel et al. (1996). The important and practical conclusion is that, as long as one is prepared to accept the assumptions of monotonicity and the principle of transfers in situations when there are negative wealth values, the conventional interpretations also apply. In particular, if two Lorenz curves do not cross, then the curve further away from the line of perfect equality represents a distribution with greater inequality according to all standard relative inequality measures (Atkinson 1970; Foster 1985).[5] In addition, with a single-crossing configuration as in Figure 5, then the 1998 distribution is more unequal than the 1994 distribution according to all standard transfer-sensitive relative inequality measures if and only if the coefficient of variation for the 1998 distribution is greater than that for the 1994 distribution (Shorrocks and Foster 1987). This was in fact the case (see below). A comparison of the Lorenz curves for disposable income suggests that income inequality increased unambiguously between 1994 and 1998: the two curves do not intersect (Jäntti 2004).

---

[4] The next two sections draw heavily on Amiel et al. (1996, p. S65).

[5] By standard inequality measures, we mean measures satisfying the strong principle of transfers, anonymity, and population replication. Relative inequality measures are those that are invariant to equi-proportionate changes in all wealth values. Absolute inequality measures (considered below) are invariant to equal absolute increments (or decrements) in all wealth values.

Figure 5
Lorenz curves for net wealth in Finland, 1994 and 1998



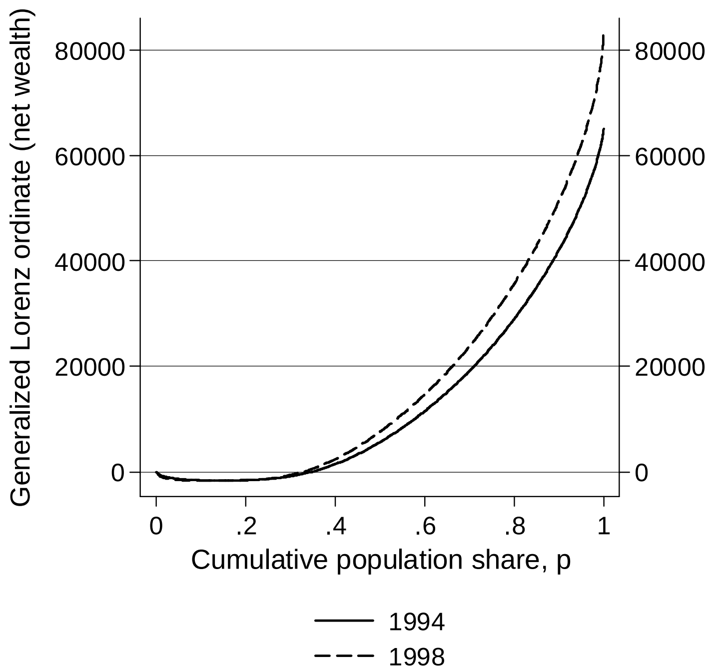
In the case when $\mu_W = 0$, then the Lorenz curve is not well-defined, and if $\mu_W \approx 0$, estimates may be numerically unstable. Alternative representational devices are required. One approach is to consider social welfare rather than (relative) inequality, i.e. one summarizes distributions using Generalized Lorenz curves (Shorrocks 1983). The Generalized Lorenz curve – the Lorenz curve scaled up by mean wealth at each point – is well-defined for all values of wealth along the real line. Its slope at each wealth value W is equal to W itself. The curve is therefore negatively sloped as long as wealth is negative, horizontal where wealth is zero, and positively sloped over the units with positive wealth. The right-hand intercept of the curve is μ_W, and so lies below the horizontal axis if $\mu_W < 0$.

Generalized Lorenz curves for Finnish net wealth are shown in Figure 6. The 1998 curve lies just below the 1994 curve initially – when both lie below the horizontal axis – but then crosses it and lies distinctly above it thereafter, reflecting the rise in wealth in the upper regions of the distribution. Because the curves cross, there is no unambiguous social welfare ordering.[6] The Figure emphasizes how reaching a conclusion about welfare change requires

[6] Because the 1998 curve crosses the 1994 curve from below rather than above, transfer-sensitivity considerations are not applicable.

the trading-off of the substantial gains for the wealthiest against the lower wealth for the lowest percentiles.

Figure 6
Generalized Lorenz curves for net wealth in Finland, 1994 and 1998



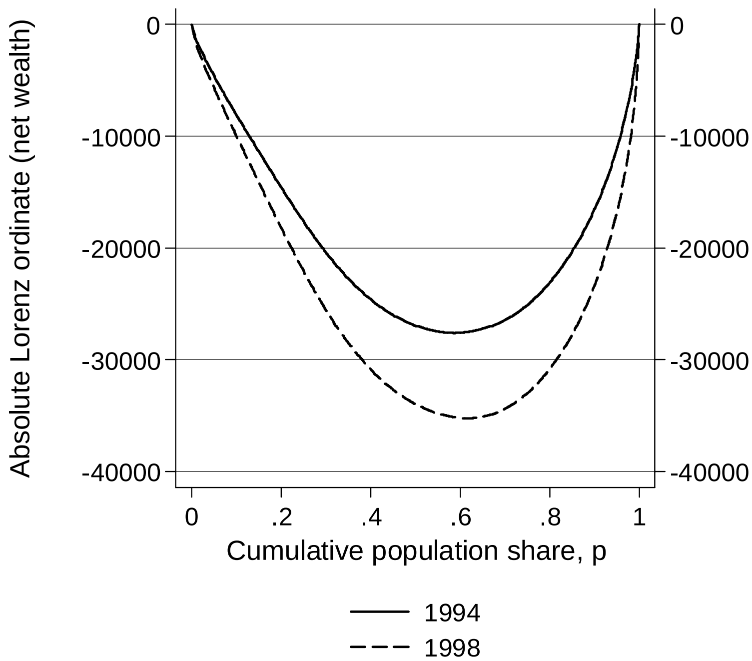
An alternative graphical device is the Absolute Lorenz curve (Moyes 1987), configurations of which are closely related to orderings of distributions according to absolute inequality measures. The abscissa of the Absolute Lorenz curve is the cumulative population share p multiplied by average wealth among the poorest p of the population minus population average wealth.[7] The slope of the curve at each $p = F(W)$ is $W - \mu_W$. As long as $\mu_W \geq 0$, then the curve has the shape of a tear drop hanging below the horizontal axis defined by p (the line of perfect absolute inequality), connected to the axis at $p = 0$ and $p = 1$. Non-standard shapes arise if mean wealth is sufficiently negative, in which case the curve is 'flipped vertically' relative to the case when $\mu_W \geq 0$, lying above the horizontal axis.

Absolute Lorenz curves for Finnish net wealth are shown in Figure 7. The 1998 curve clearly lies below the 1994 curve over the complete range of wealth values, in which case we

[7] Cf. the abscissa of the standard Lorenz curve which equals p times average wealth among the poorest p divided by population average wealth.

may conclude that the 1998 distribution was more unequal than the 1994 distribution according to all standard absolute inequality measures (Moyes 1987).

Figure 7
Absolute Lorenz curves for net wealth in Finland, 1994 and 1998



One cautionary note is in order concerning the use of Absolute and Generalized Lorenz curves for comparisons of wealth distributions. Their ordinates are not unit-free (as for the standard Lorenz curve), but in the units of wealth. Comparisons may therefore be sensitive to the choice of the price deflator and exchange rate.

5. Inequality indices

As Amiel et al. (1996) have pointed out,

> many standard aggregative inequality measures are undefined for negative incomes, and a substantial class of these measures will not work even for zero incomes, in the sense that they are either undefined, or are unbounded, or attain their maximum value at any income distribution that has one or more zero incomes. (1996, p. S65.)

Most measures are built up from evaluations of individual wealth values, W_i. For Generalized Entropy inequality measures, the individual evaluation is based on a power function, W_i^{θ}, where θ may take any real value; for Atkinson (1970) inequality measures, the evaluation function is W_i^{ε}, where $\varepsilon > 0$. If wealth is negative, then the evaluation functions and hence measures are well-defined only if $\theta > 1$ and $\varepsilon > 1$.

The good news, therefore, is that various standard inequality measures can be calculated; the bad news, as Amiel et al. (1996) remind us, is that these measures are ones that are particularly sensitive to extreme values. For example, inequality comparisons based on even the coefficient of variation – ordinally equivalent to the Generalized Entropy measure with $\theta = 2$ – may be substantially affected by the inclusion or exclusion of just one very high value. Descriptive indices such as the ratio of the 90th percentile to the 10th percentile ($p90/p10$) may be problematic. They are undefined if the percentile in the numerator of the calculation is equal to zero, and problems of interpretation arise in cases where the percentile in numerator is negative and that in the numerator is positive. Both situations are possible for net wealth variables.

The Gini coefficient is a measure that is well-defined when wealth values are negative.[8] Because the Gini is a function of absolute differences between all possible pairs of wealth values (suitably normalized), it does not matter that some values may be negative or zero. Observe, however, that when there are negative values, estimates of the Gini may be greater than one (cf. the upper bound of one in the standard case). The reason is that the Lorenz curve lies below the horizontal axis in this case (see above), and so twice the area between the curve and the ray of perfect equality (equal to the Gini) may be greater than one. For this situation, Chen et al. (1982) proposed a renormalization of the usual Gini formula to ensure that the index value was bounded between zero and one.

The discussion in this section so far has assumed that mean wealth is positive. If the mean is negative, then estimates of indices such as the coefficient of variation and Gini will also be negative (cf. the lower bound of zero in the standard case). A mean-preserving spread of wealth in this case leads to smaller values of the inequality index, as in the standard case, except that here smaller means more negative rather than less positive.

---

[8] In fact, so too are all members of the generalized Gini class of indices (Donaldson and Weymark 1980, Yitzhaki 1983). Other less commonly used indices that are also well-defined are the relative mean deviation and the Pietra ratio (the former is equal to half the latter).

If mean wealth equals zero, then relative inequality measures such as the coefficient of variation and the Gini are not well-defined and, if the mean is close to zero, they are well-defined but estimates may be numerical unstable. Absolute inequality indices may be used in this case: for example, the absolute Gini (the standard Gini times the mean) or the Kolm (1976) class of indices. Each measure in the latter family is built up from evaluations of individual wealth values using evaluation functions of form $\exp(-\kappa W_i)$, and these are well-defined for negative, zero, or positive values of W_i.

The utility of the absolute indices is limited by two factors. First, as with Absolute and Generalized Lorenz curves, the measures are not unit-free. It becomes particularly important to have an appropriate price deflator and exchange rate when making comparisons across time or countries. Second, the Kolm measures are relatively unfamiliar, which means that gaining a feel for the implications of differences in key sensitivity parameters is more difficult. In this regard, Atkinson and Brandolini (2004) helpfully pointed out that

> the marginal value of income accruing to person i is equal to κy_i, which provides a guide to interpreting the value of κ in the context of a specific choice of units for income. If κ were to equal the reciprocal of mean income, then the elasticity of the marginal valuation of income would be equal to 1 at the mean (and equal to 0.5 at half the mean income). (2004, pp. 6–7.)

If the mean is used to benchmark the κ parameter in this way, there remains the issue that the mean changes over time or differs between countries, so a choice has to be made about which mean.[9]

Estimates of the degree of inequality in gross wealth, debt, and net wealth in Finland are shown in Table 3. We report estimates for both relative and absolute inequality measures. Of the former, we report the $p90/p10$ percentile ratio, the relative mean deviation, half the coefficient of variation (CV) squared, and the Gini index. The absolute indices are the absolute Gini index and three Kolm absolute indices. For the latter, we use three values of κ: the reciprocal of the 1994 mean, the reciprocal of the 1998 mean, and a simple average of those two parameter values. The Gini and half the CV squared are standard relative measures (cf footnote 6) and the absolute Gini and Kolm indices are each standard absolute inequality measures.

Table 3
Wealth inequality in Finland, 1994 and 1998: relative and absolute indices

	1994	1998
Gross wealth (G)		
Percentile ratio, *p*90/*p*10	112.4	178.6
Relative mean deviation	0.336	0.360
Half of CV squared	0.466	0.678
Gini	0.476	0.511
Absolute Gini	39,227	51,990
Kolm (a)	23,932	37,461
Kolm (b)	20,535	32,891
Kolm (c)	22,275	35,248
Debt (D)		
Percentile ratio, *p*90/*p*10	–	–
Relative mean deviation	0.563	0.574
Half of CV squared	1.141	1.306
Gini	0.705	0.717
Absolute Gini	12,219	13,396
Kolm (a)	9,350	10,648
Kolm (b)	9,031	10,314
Kolm (c)	9,194	10,485
Net wealth (W)		
Percentile ratio, *p*90/*p*10	–81.9	–462.4
Relative mean deviation	0.424	0.424
Half of CV squared	0.736	0.971
Gini	0.591	0.599
Absolute Gini	38,439	49,714
Kolm (a)	28,037	45,574
Kolm (b)	23,461	37,064
Kolm (c)	25,794	41,096

Notes. For the Kolm indices, parameter κ is the reciprocal of the 1994 mean in case (a), the reciprocal of the 1998 mean in case (b), and a simple average of these two parameter values in case (c). The *p*90/*p*10 indices are undefined for Debt (*p*10 = 0). All money values are expressed in ‘2000 international dollars’.

The Lorenz curve configurations shown in Figures 5 and 7 showed that household net wealth inequality increased between 1994 and 1998 in Finland according to all standard

[9] Cf. the Atkinson (1970) indices of relative inequality for which the marginal elasticity of income is a constant, equal to the inequality aversion parameter ε.

inequality measures, and the estimates confirm this, while also showing how much inequality increased by. For example, the Gini coefficient for net wealth increased from 0.591 to 0.599, a relatively small change of just over 1 percent. This can be compared to an increase of around 19% in the Gini for disposable income, from 0.212 to 0.252 (Jäntti 2004). The increase registered by half the CV squared was much larger, almost one third, from 0.736 to 0.971.

Observe that the relative mean deviation was unchanged (to 4 d.p.), which highlights the fact that the Lorenz ordering result referred to standard inequality measures. (These measures do not include the relative mean deviation: it is insensitive to transfers on the same side of the mean.) All the absolute inequality indices for net wealth increased substantially between 1994 and 1998, by 59% according to Kolm index (c), for example. The top two panels of Table 3 show that there was also an increase in the inequality of gross wealth and of debt according to all the indices calculated.

Are the inequality estimates sensitive to the inclusion of extreme observations? To explore this we recalculated estimates first excluding the largest observation in each of the net wealth distributions and, second, excluding the top and bottom percentile groups: see Table 4.

Table 4
Sensitivity of inequality indices to different treatments of extreme values: Finland, 1994 and 1998

	½ CV^2	Gini
All obs (as in Table 3)		
1994	0.736	0.591
1998	0.971	0.599
% increase	*31.9*	*1.4*
Drop richest one		
1994	0.701	0.590
1998	0.920	0.597
% increase	*31.2*	*1.2*
Trim top and bottom 1%		
1994	0.526	0.553
1998	0.548	0.549
% increase	*4.3*	*–0.7*

In the first case, the net wealth Gini coefficient for 1994 was 0.590 and half the CV squared was 0.701; for 1998 the estimates were 0.597 and 0.920, respectively. Thus removal of the largest observation had little effect on the Gini, or on its change over time. For half the CV squared, the impact was larger (as expected): the new estimates for each year were about 5% smaller than their counterparts in Table 4, and the change between 1994 and 1998 was 24% (compared to 32%). Trimming had a much more substantial impact on the estimates. In this case, the estimated Ginis were 0.553 for 1994 and 0.549 for 1998, and the estimates of half the CV squared were 0.526 and 0.548, implying a decline in the former index of –0.7% and an increase in the latter one of only 4.3%. It seems that the robustness of the inequality indices in the sense analyzed by Cowell and Victoria-Feser (1996) may be an even more important issue for wealth distributions than for the income distributions that they studied.

6. Fitting parametric size distributions

In the analysis of income distributions, analysts have found it useful to complement the non-parametric methods discussed so far with distributional summaries based on estimates of specific parametric functional forms: 'some standard functional forms claim attention, not only for their suitability in modelling some features of many empirical income distributions, but also because of their role as equilibrium distributions in economic processes' (Cowell 2000, p. 145). Fitting of parametric functional forms has also been common for wealth distributions. The most commonly-used has been the single-parameter Pareto distribution, which provides a description of the density for wealth values above some lower bound, $W_0 > 0$. See, for example, Atkinson and Harrison (1978, especially Appendices IV and IX) and Kleiber and Kotz (2003, chapter 3).

If one focuses on the distribution amongst those with wealth greater than W_0, there are simple expressions for the moments which depend only on the Pareto parameter α and W_0. Moreover, the expressions for most common inequality measures depend only on α, so that the (inverse of) α may also be considered as an inequality measure. However, the apparent attractions of the Pareto distribution evaporate somewhat when one considers its implications for the distribution of wealth amongst the population as a whole, i.e. including units with wealth less than W_0. Atkinson and Harrison (1978, Appendix IV) show how expressions for the Gini and the relative mean deviation depend on assumptions about the size of 'excluded population' (the proportion of the population with wealth below W_0) and their average wealth. In particular α no longer has such a straightforward interpretation. For example, an increase in α may be associated with an increase in inequality according to the Gini, but a decrease according to the coefficient of variation.

This suggests fitting of parametric models for the distribution of wealth as a whole. The income distribution literature suggests a large number of candidates, including two-parameter models such as the log-normal and gamma, three-parameter distributions such as the Singh-Maddala and Dagum I, and four-parameter distributions such as the generalized beta distributions of the first and second kind. See the comprehensive survey by Kleiber and Kotz (2003). The problem for the wealth researcher is that virtually all of these distributions are defined for variables taking only strictly positive values. If the functional forms are defined also for values of zero, the density typically has zero mass at that point, and so cannot capture any spike at that point. One could of course fit a model to the positive observations only, but that may omit a significant part of the story.

The number of models for wealth with the real line as support appears to be small. One is the three-parameter log-normal. Compared to the usual log-normal distribution, there is an additional and estimable parameter characterizing a threshold below which the probability of observing wealth is zero. However, there are problems in fitting the model by maximum likelihood methods – the likelihood may be unbounded (Kleiber and Kotz, 2003, p. 122). In any case, the log-normal shape may not be suitable for wealth distributions.

More promising alternatives are provided by finite mixture models, in which the overall distribution function is a population-share-weighted sum of distribution functions characterising wealth over different regions of the support, including negative and zero values as well as positive ones.[10] The four-parameter Dagum Type II distribution (Kleiber and Kotz, 2003, pp. 219–220) has a parameter that characterizes a (discrete mass) probability that a unit has a wealth value equal to zero. Positive wealth values in this model are described by the three-parameter Dagum I (Burr Type 3) distribution, which has a CDF given by $F(W) = [1 + (b/W)^a]^{-p}$, where parameters a, b, $p > 0$. The b is a scale parameter; a and p are shape parameters. Dagum (1990) extended his model to incorporate a third mixture component: an exponential distribution to describe negative values ($F(W) = \exp(\theta W)$, $W < 0$, $\theta > 0$).

We provide estimates for Finland of this 'Dagum III' model, partly motivated by the fact that there are no applications other than Dagum's (1990) one to Italian wealth data for 1977, 1980, and 1984, that we are aware of. Maximum likelihood estimates for the 1994 and

1998 Finnish net wealth distributions are shown in Table 5. All the parameters were very precisely estimated. The mixture proportions (the λ) correspond exactly to the sample estimates shown in Table 1, and the increase in the scale parameter (b) between 1994 and 1998 reflects the increase in average net wealth over the period. However, the other parameters (a, p, θ), characterizing distributional shape, are intrinsically difficult to interpret, as the effect of changing one of them is contingent on the values of the other parameters.

Table 5
Estimates of 'Dagum III' finite mixture model for net wealth in Finland, 1994 and 1998

	1994		1998	
	Estimate	($\|t\|$)	Estimate	($\|t\|$)
a	3.916	(30.0)	3.428	(32.8)
b	159,355	(52.4)	189,723	(43.9)
p	0.168	(23.9)	0.182	(26.2)
θ ($\times 10^4$)	0.788	(25.7)	0.625	(54.2)
λ_1 (fraction with $W < 0$)	0.127	(27.5)	0.105	(28.1)
λ_2 (fraction with $W = 0$)	0.009	(7.0)	0.008	(6.6)
λ_3 (fraction with $W > 0$)	0.864	(181.9)	0.888	(229.4)
Log-likelihood	–48,616		–64,045	
Mean (predicted)	62,843		78,230	
Mean (sample)	65,066		83,046	
Median (predicted)	43,050		50,575	
Median (sample)	50,239		62,081	
Gini (predicted)	0.560		0.572	
Gini (sample)	0.591		0.599	

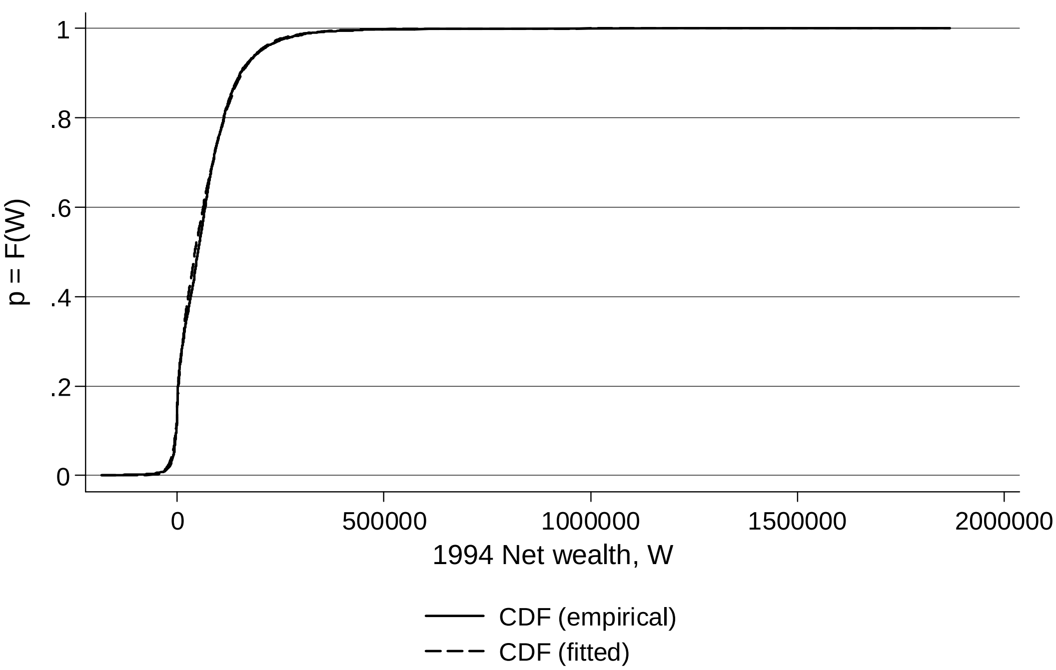
Notes. Maximum likelihood estimates of finite mixture model described in text. $|t|$ is the absolute value of the asymptotic t-ratio. All money values are expressed in '2000 international dollars'.

It is easiest to interpret parameter estimates, and to assess overall goodness of fit, by comparing predicted values for key distributional summary measures with their sample counterparts. Figure 8 compares the fitted and sample estimates of the CDF for net wealth in 1998 (the picture for 1994 was similar), and suggests that the model fits well. However, what one sees partly depends on the lens used. Figure 9 shows the fitted probability density function and, although it captures the shape at zero and negative values relatively well

[10] Finite mixture models have also been fitted to income distributions: see for example Paap and van Dijk (1998). The models assume that incomes take on positive values only.

(compared to Figures 3 or 4), it is too convex to the origin over positive wealth values. This pattern is reflected in the other summary statistics shown in the bottom panel of Table 5. For example, the 1998 sample mean is under-estimated by between 5% and 6% and the 1998 Gini coefficient is under-estimated by 3% to 4%.[11] Interestingly, the differences between predicted and sample values are much the same in proportionate terms as the corresponding ones reported by Dagum (1990), who referred to the 'exceptionally good' fit of his model (1990, p. 55).

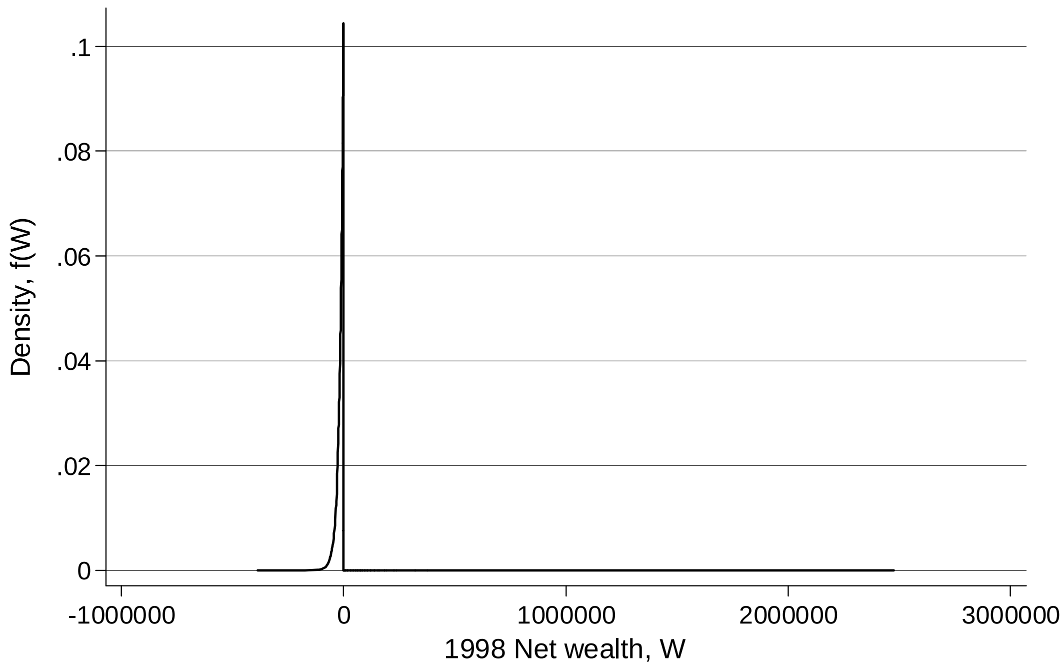
Figure 8
CDFs for net wealth in Finland, 1998: empirical versus fitted



Note. Fitted CDF derived from estimates of 'Dagum III' finite mixture model: see text.

[11] The predicted mean conditional on net wealth being negative was almost exactly the same as the sample conditional mean. On the other hand, the predicted mean conditional on wealth being positive under-estimated its sample counterpart.

Figure 9
Fitted PDF for net wealth in Finland, 1998



Note. Fitted PDF derived from estimates of 'Dagum III' finite mixture model: see text.

Finite mixture models such as the Dagum III one deserve further attention in future, combined with explorations of alternative distributions to characterize positive wealth values (e.g. Singh-Maddala or generalized Beta distributions). A feature of these models is that each of the parameters may be made a function of covariates summarizing household characteristics.[12] For example, we have estimated versions of the Dagum III model in which parameters were specified as functions of the age of the household head (see Appendix 2). Estimation of 'conditional' wealth distributions such as these provides a route to decomposition analysis of the sources of trends in wealth distributions over time or differences between countries, complementary to that based on kernel density estimates that

[12] The generalized linear regression model of Burbidge et al. (1988), applied to Canadian net wealth data, in effect allowed only the distribution mean to depend on covariates. I.e. there was heterogeneity in scale but not in shape.

was popularized by DiNardo et al. (1996). For an application of parametric models in this manner, but to income, see Biewen and Jenkins (2005).

7. Concluding remarks

This review has argued that many of the tools developed for summarizing income distributions can also be straightforwardly applied to wealth distributions, albeit with some care because of the distinctive features of wealth data. These features are the relatively high prevalence of zero values and also, for net wealth variables, the relatively high prevalence of negative values. With our empirical illustrations based on Finnish data, we hope to have shown how standard methods may be applied or adapted to summarize and compare distributions of wealth. There are a number of issues that we have not addressed, including the treatment of extreme values and statistical inference – issues that are of relevance to analysis of income as well as wealth.

References

Amiel Y., Cowell F.A., and Polovin A. (1996). 'Inequality among the kibbutzim', *Economica*, 63, S63–S85.

Atkinson A.B. (1970). 'On the measurement of inequality, *Journal of Economic Theory*, 2, 244–263.

Atkinson A.B. (1987). 'On the measurement of poverty', *Econometrica*, 55, 749–764.

Atkinson A.B. and Brandolini A. (2004). 'Global world inequality: absolute, relative or intermediate?', paper presented at the 28th General Conference of the International Association for Research on Income and Wealth, Cork, 22–28 August 2004. http://www.iariw.org/papers/2004/brand.pdf

Atkinson A.B. and Harrison A.J. (1978). *The Distribution of Wealth in Britain*, Cambridge University Press, Cambridge.

Biewen M. and Jenkins S.P. (2005) 'Accounting for differences in poverty between the USA, Britain and Germany', *Empirical Economics*, forthcoming.

Burbidge J.B., Magee L., and Robb A.L. (1988). 'Alternative transformations to handle extreme values of the dependent variable', *Journal of the American Statistical Association*, 83, 123–127.

Chen, C.-N., Tsaur, T.-W., and Rhai, T.-S. (1982) 'The Gini coefficient and negative income', *Oxford Economic Papers*, 34, 473–478.

Cowell F.A. (2000). 'Measurement of inequality', in *Handbook of Income Distribution Volume 1*, A.B. Atkinson and F. Bourguignon (eds), North-Holland, Amsterdam, 87–166.

Cowell F.A., Jenkins S.P., and Litchfield J.A. (1996). 'The changing shape of the UK income distribution: kernel density estimates", in J. Hills (ed.) *New Inequalities: The Changing Distribution of Income and Wealth in the United Kingdom*, Cambridge University Press, Cambridge, 49–75.

Cowell, F.A. and Victoria-Feser M.-P. (1996). 'Robustness properties of inequality measures', *Econometrica*, 64, 77–101.

Dagum C. (1990). 'A model of net wealth distribution specified for negative, null and positive wealth. A case study: Italy', in C. Dagum and M. Zenga (eds), *Income and Wealth Distribution, Inequality and Poverty*, Springer, Berlin and Heidelberg, 42–56.

Davies J.B. and Shorrocks A.F. (2000). 'The distribution of wealth', in *Handbook of Income Distribution Volume 1*, in A.B. Atkinson and F. Bourguignon (eds), North-Holland, Amsterdam, 605–675.

DiNardo J., Fortin N.M., and Lemieux T. (1996). 'Labour market institutions and the distribution of wages, 1973–1992: a semi-parametric approach', *Econometrica*, 64, 1001–1046.

Donaldson D. and Weymark J.A. (1980). 'A single parameter extension of the Gini indices of inequality', *Journal of Economic Theory*, 22, 67–86.

Foster J.E. (1985). 'Inequality measurement', in H.P. Young (ed.), *Fair Allocation*, Proceeding of Symposia in Applied Mathematics, volume 33, American Mathematical Society, Providence RI, 31–68.

Inland Revenue (2003). 'Distribution of personal wealth', *Inland Revenue Statistics*, Inland Revenue, London. http://www.inlandrevenue.gov.uk/stats/personal_wealth/menu.htm

Jäntti M. (2004). 'Trends in the distribution of income and wealth – Finland 1987–1998', unpublished paper, Åbo Akademi University, Turku, Finland. http://www.abo.fi/~mjantti/TrendsinWealthandIncome.pdf

Kleiber C. and Kotz S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley & Sons, Hoboken NJ.

Klevmarken N.A. (2004). ‘On the wealth dynamics of Swedish families, 1984–98’, *Review of Income and Wealth*, 50, 469–491.

Kolm S.C. (1976). ‘Unequal inequalities I and II’, *Journal of Economic Theory*, 12, 416–442 and 13, 82–111.

Mosler K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach*, Springer, Berlin and New York.

Moyes P. (1987). ‘A new concept of Lorenz dominance’, *Economics Letters*, 23, 203–207.

Paap R. and van Dijk H.K. (1998). ‘Distribution and mobility of wealth of nations’, *European Economic Review*, 42, 1269–1294.

Pen J. (1972). *Income Distribution*, Penguin, Harmondsworth.

Shorrocks A.F. (1983). ‘Ranking income distributions’, *Economica*, 50, 3–17.

Shorrocks A.F. and Foster J.E. (1987). ‘Transfer sensitive inequality measures’, *Review of Economic Studies*, 54, 485–497.

Sierminska, E. and Smeeding T.M. (2005) ‘Measurement issues: equivalence scales, accounting framework, and reference Unit’, paper presented at the Luxembourg Wealth Study Workshop, Perugia, 27–29 January 2005.

Yitzhaki S. (1983). ‘On an extension of the Gini index’, *International Economic Review*, 24, 617–628.

Appendix 1. Wealth survey data for Finland

Statistics Finland collected broadly comparable wealth data using surveys in 1987, 1988, 1994 and 1998. The first two surveys were a panel, but used a somewhat different definition of wealth than the two later surveys, which is why we use data for 1994 and 1998 in this paper. The wealth surveys were administered in conjunction with the Income Distribution Surveys of the relevant year, so they contain high-quality data on income and other variables as well. The net sample sizes were 5210 and 3893 in 1994 and 1998, reflecting response rates of 75.2% and 65.6%. The public use file that Statistics Finland provides for researchers contains sampling weights that have been calibrated to correct for non-response with respect to a few marginal distributions, including that of taxable income. We use these calibrated weights in all our empirical illustrations.

The wealth data were collected by interviewing the household head and asking detailed questions about the holding of different categories wealth and debts. For instance, for publicly listed companies, respondents were asked to provide a list with the number of shares, rather then the value of their portfolios on the day of the interview. This list was then evaluated using the stock exchange prices at the end of the year. Similarly, respondents were asked to list the automobiles and other vehicles that they owned, rather than estimate their value. Market prices for used vehicles of the type and age owned were used to value the vehicles.

The definition of wealth used by Statistics Finland is fairly standard, but lacks one substantial component, namely pension wealth (except for voluntary pension insurance). This is explained, in part, by the fact that Finland has a defined benefit, rather than a defined contribution pension system, which means that pension income becomes known only at retirement. In their survey of wealth inequality, Davies and Shorrocks (2000) defined pension wealth to belong to 'augmented wealth'. While inclusion of some estimate of future pension wealth would lead to higher estimated average wealth and lower wealth inequality, this omission is unimportant for the purposes of this paper.

For ease of comparison with other sources, we have converted survey values expressed in 1994 and 1998 Finnish Markkas to Euros using the cost-of-living index (http://www.stat.fi) and the official Markka-Euro conversion rate of 5.946. To convert Euros to 'international dollars', we then used the OECD's PPP exchange rate for Finland in 2000, i.e. 0.979 (http://www.oecd.org/dataoecd/61/54/18598754.pdf).

Appendix 2. Practicing what we preach: Stata code to derive the estimates and draw the figures

The scripts that follow were used to derive the estimates and graphs that are reported in the paper. We circulate the files in order to maximize the dissemination of the methods that we have discussed, observing that Stata (http://www.stata.com) is widely used by social scientists of many disciplines at both introductory and advanced levels. In other words, much distributional analysis can be undertaken using readily-available statistical software.

The scripts consist of five Stata do files (*an01.do*, *an02.do*, *an03.do*, *an03.do*, *an04.do*, *an05.do*) paralleling the sections of the paper. An up-to-date version of Stata 8.2 is required to run the scripts without modification. All programs used are either built in to Stata or are in the public domain. Use the Stata *findit* command to find and install the public domain programs (with the exception of *dagumfit* which is available from SPJ).

---

an01.do (means; proportions of negative, zero, positive; extreme values, etc.)

```
version 8
clear
set more off
capture log close
cd d:/home/stephenj/myprojects/lws/finland
log using an01.log, replace
******************  an01.do ***************************************
*
* Analysis for paper: Finland, 1994, 1998
*
* Section 1: means, proportions of negative, zero, positive, strange obs etc
*
********************************************************************************
use finland
******* summary stats for each wealth vble, by year **********************

* mean and extreme values

bysort year: su gross debt net [aw=wgt], de

* proportions with neg, zero, positive etc

bysort year: tab gcat [aw = wgt]
bysort year: tab dcat [aw = wgt]
bysort year: tab ncat [aw = wgt]

* conditional means

bysort year: su gross [aw = wgt] if gcat == 3
bysort year: su debt [aw = wgt] if dcat == 3
bysort year: su net [aw = wgt] if ncat == 3
bysort year: su net [aw = wgt] if ncat == 1

* more about extreme values from boxplots

graph box gross, over(year) saving(boxplot_gross, replace)
graph export boxplot_gross.eps, replace
graph box debt, over(year) saving(boxplot_debt, replace)
graph export boxplot_debt.eps, replace
graph box net, over(year) saving(boxplot_net, replace)
graph export boxplot_net.eps, replace

* Means if trim bottom and top 1% of observations

_pctile gross [aw = wgt] if year == 1, p(1 99)
local gp194 = r(r1)
local gp9994 = r(r2)
```

```
_pctile debt [aw = wgt] if year == 1, p(1 99)
local dp194 = r(r1)
local dp9994 = r(r2)

_pctile net [aw = wgt] if year == 1, p(1 99)
local np194 = r(r1)
local np9994 = r(r2)

_pctile gross [aw = wgt] if year == 2, p(1 99)
local gp12 = r(r1)
local gp992 = r(r2)

_pctile debt [aw = wgt] if year == 2, p(1 99)
local dp12 = r(r1)
local dp992 = r(r2)

_pctile net [aw = wgt] if year == 2, p(1 99)
local np12 = r(r1)
local np992 = r(r2)

su gross [aw = wgt] if gross >= `gp194' & gross <= `gp9994' & year == 1
su gross [aw = wgt] if gross >= `gp12' & gross <= `gp992' & year == 2
su debt [aw = wgt] if debt >= `dp194' & debt <= `dp9994' & year == 1
su debt [aw = wgt] if debt >= `dp194' & debt <= `dp992' & year == 2
su net [aw = wgt] if net >= `np194' & gross <= `np9994' & year == 1
su net [aw = wgt] if net >= `np12' & gross <= `np992' & year == 2

log close
```

an02.do (CDF/parades, quantiles, density estimation)

```
version 8
clear
set more off
capture log close
cd d:/home/stephenj/myprojects/lws/finland
log using an02.log, replace
*******************  an02.do ****************************************
*
* Analysis for paper, Finland 1994, 1998
*
* Section 2: CDF/parades, tables of quantiles, density estimations
*
*******************************************************************************
use finland

****** quantiles **************

bys year: su gross debt net [aw=wgt], de

sumdist gross [aw=wgt] if year == 1, n(10)
sumdist gross [aw=wgt] if year == 2, n(10)
sumdist debt [aw=wgt] if year == 1, n(10)
sumdist debt [aw=wgt] if year == 2, n(10)
sumdist net [aw=wgt] if year == 1, n(10)
sumdist net [aw=wgt] if year == 2, n(10)

********** CDFs **************

bys year: cumul gross [aw = wgt], gen(cdf_gross)
lab var cdf_gross “Cumulative population share, p”
sort year cdf_gross
graph twoway (line gross cdf_gross  if year ==1)            ///
        (line gross cdf_gross if year ==2)                  ///
        , legend(label (1 "1994") label(2 "1998")    ///
        region(lstyle(none)) ) saving(cdf_gross, replace)
graph export cdf_gross.eps, replace

bys year: cumul debt [aw = wgt], gen(cdf_debt)
lab var cdf_debt “Cumulative population share, p”
sort year cdf_debt
graph twoway (line debt cdf_debt  if year ==1)              ///
        (line debt cdf_debt if year ==2)                    ///
        , legend(label (1 "1994") label(2 "1998")    ///
        region(lstyle(none)) ) saving(cdf_debt, replace)
graph export cdf_debt.eps, replace

bys year: cumul net [aw = wgt], gen(cdf_net)
lab var cdf_net “Cumulative population share, p”
```

```
sort year cdf_net
graph twoway (line net cdf_net  if year ==1)                ///
        (line net cdf_net if year ==2)                      ///
        , legend(label (1 "1994") label(2 "1998")    ///
        region(lstyle(none)) ) saving(cdf_net, replace)
graph export cdf_net.eps, replace

bys year: su cdf* [aw=wgt]

********* histograms, 100 bins *********************

hist gross, by(year) bin(100) saving(hist_gross, replace)
graph export hist_gross.eps, replace
hist debt, by(year) bin(100) saving(hist_debt, replace)
graph export hist_debt.eps, replace
hist net, by(year) bin(100) saving(hist_net, replace)
graph export hist_net.eps, replace

********* kernel density estimates *************
*
* all use Epanechnikov kernel,
*       bandwidth = 0.9*m/(nobs)^.2 where m = min( sqrt(var), (interquartile_range)/1.349 )
*
graph twoway (kdensity gross [aw=wgt] if year == 1, n(1000) )              ///
               (kdensity gross [aw=wgt] if year == 2, n(1000))      ///
               , ytitle(Density) xtitle("Gross wealth")     ///
               legend(label (1 "1994") label(2 "1998")      ///
               region(lstyle(none)) ) saving(kdensity_gross, replace)
graph export kdensity_gross.eps, replace

graph twoway (kdensity debt [aw=wgt] if year == 1, n(1000))         ///
               (kdensity debt [aw=wgt] if year == 2, n(1000))       ///
               , ytitle(Density) xtitle("Debt")     ///
               legend(label (1 "1994") label(2 "1998")      ///
               region(lstyle(none)) ) saving(kdensity_debt, replace)
graph export kdensity_debt.eps, replace

graph twoway (kdensity net [aw=wgt] if year == 1, n(1000))          ///
               (kdensity net [aw=wgt] if year == 2, n(1000))        ///
               , ytitle(Density) xtitle("Net wealth")       ///
               legend(label (1 "1994") label(2 "1998")      ///
               region(lstyle(none)) ) saving(kdensity_net, replace)
graph export kdensity_net.eps, replace

********* adaptive kernel density estimates *************
*
* all use Epanechnikov kernel
*

akdensity gross [aw=wgt] if year == 1, n(1000) gen(g1 fg1) nograph
akdensity gross [aw=wgt] if year == 2, n(1000) gen(g2 fg2) nograph
graph twoway (line fg1 g1) (line fg2 g2)                    ///
               , legend(label (1 "1994") label(2 "1998")    ///
               region(lstyle(none)) ) saving(akdensity_gross, replace)
graph export akdensity_gross.eps, replace

akdensity debt [aw=wgt] if year == 1, n(1000) gen(d1 fd1) nograph
akdensity debt [aw=wgt] if year == 2, n(1000) gen(d2 fd2) nograph
graph twoway (line fd1 d1) (line fd2 d2)                    ///
               , legend(label (1 "1994") label(2 "1998")    ///
               region(lstyle(none)) ) saving(akdensity_debt, replace)
graph export akdensity_debt.eps, replace

akdensity net [aw=wgt] if year == 1, n(1000) gen(n1 fn1) nograph
akdensity net [aw=wgt] if year == 2, n(1000) gen(n2 fn2) nograph
graph twoway (line fn1 n1) (line fn2 n2)                    ///
               , legend(label (1 "1994") label(2 "1998")    ///
               region(lstyle(none)) ) saving(akdensity_net, replace)
graph export akdensity_net.eps, replace

log close
```

an03.do (Lorenz curves of various types)

```
version 8
clear
set more off
capture log close
cd d:/home/stephenj/myprojects/lws/finland
log using an03.log, replace
******************  an03.do ***************************************
```

```
*
* Analysis for paper, Finland 1994, 1998
*
* Section 3: Lorenz curves of various types
*
********************************************************************************
use finland

************ Relative Lorenz curves ***********************

glcurve gross [aw=wgt] , by(year) split pvar(pgross) glvar(rlcg) lorenz nograph
lab var rlcg_1 "1994"
lab var rlcg_2 "1998"
lab var pgross "Cumulative population share, p"

sort pgross
graph twoway (line rlcg_1 pgross, yaxis(1 2) )                      ///
       (line rlcg_2 pgross, yaxis(1 2) )                    ///
       (function y = x, range(0 1) yaxis(1 2) )             ///
       , aspect(1) xtitle("Cumulative population share, p")        ///
       ytitle("Lorenz ordinate (gross wealth)", axis(1)) ytitle(" ", axis(2)) ///
       legend(label (1 "1994") label(2 "1998") label(3 "Equality")        ///
       region(lstyle(none)) ) saving(rlc_gross, replace)
graph export rlc_gross.eps, replace
su rlcg* [aw=wgt], de

glcurve debt [aw=wgt] , by(year) split pvar(pdebt) glvar(rlcd) lorenz nograph
lab var rlcd_1 "1994"
lab var rlcd_2 "1998"
lab var pdebt "Cumulative population share"
sort pdebt
graph twoway (line rlcd_1 pdebt, yaxis(1 2) )                       ///
       (line rlcd_2 pdebt, yaxis(1 2) )                     ///
       (function y = x, range(0 1) yaxis(1 2) )             ///
       , aspect(1) xtitle("Cumulative population share, p")        ///
       ytitle("Lorenz ordinate (debt)", axis(1)) ytitle(" ", axis(2)) ///
       legend(label (1 "1994") label(2 "1998") label(3 "Equality")        ///
       region(lstyle(none)) ) saving(rlc_debt, replace)
graph export rlc_debt.eps, replace
su rlcd* [aw=wgt], de

glcurve net [aw=wgt] , by(year) split pvar(pnet) glvar(rlcn) lorenz nograph
lab var rlcn_1 "1994"
lab var rlcn_2 "1998"
lab var pnet "Cumulative population share"
sort pnet
graph twoway (line rlcn_1 pnet, yaxis(1 2) )                        ///
       (line rlcn_2 pnet, yaxis(1 2) )                      ///
       (function y = x, range(0 1) yaxis(1 2) )             ///
       , aspect(1) xtitle("Cumulative population share, p")        ///
       ylabel(-.2(.2)1, axis(1)) ylabel(-.2(.2)1, axis(2))  ///
       ytitle("Lorenz ordinate (net wealth)", axis(1)) ytitle(" ", axis(2)) ///
       legend(label (1 "1994") label(2 "1998") label(3 "Equality")        ///
       region(lstyle(none)) ) saving(rlc_net, replace)
graph export rlc_net.eps, replace
su rlcn* [aw=wgt], de

capture drop pgross pdebt pnet

************ Generalized Lorenz curves ***********************

glcurve gross [aw=wgt] , by(year) split pvar(pgross) glvar(glcg) nograph
lab var glcg_1 "1994"
lab var glcg_2 "1998"
lab var pgross "Cumulative population share"
sort pgross
graph twoway (line glcg_1 pgross, yaxis(1 2) )                      ///
       (line glcg_2 pgross, yaxis(1 2) )                    ///
       , aspect(1) xtitle("Cumulative population share, p")        ///
       ytitle("Generalized Lorenz ordinate (gross wealth)", axis(1)) ytitle(" ", axis(2)) ///
       legend(label (1 "1994") label(2 "1998")      ///
       region(lstyle(none)) ) saving(glc_gross, replace)
graph export glc_gross.eps, replace
su glcg* [aw=wgt], de

glcurve debt [aw=wgt] , by(year) split pvar(pdebt) glvar(glcd) nograph
lab var glcd_1 "1994"
lab var glcd_2 "1998"
lab var pdebt "Cumulative population share"
sort pdebt
graph twoway (line glcd_1 pdebt, yaxis(1 2) )                       ///
       (line glcd_2 pdebt, yaxis(1 2) )                     ///
       , aspect(1) xtitle("Cumulative population share, p")        ///
```

```
        ytitle("Generalized Lorenz ordinate (debt)", axis(1)) ytitle(" ", axis(2)) ///
        legend(label (1 "1994") label(2 "1998")      ///
        region(lstyle(none)) ) saving(glc_debt, replace)
graph export glc_debt.eps, replace
su glcd* [aw=wgt], de

glcurve net [aw=wgt] , by(year) split pvar(pnet) glvar(glcn) nograph
lab var glcn_1 "1994"
lab var glcn_2 "1998"
lab var pnet "Cumulative population share"
sort pnet
graph twoway (line glcn_1 pnet, yaxis(1 2) )                        ///
        (line glcn_2 pnet, yaxis(1 2) )                      ///
        , aspect(1) xtitle("Cumulative population share, p")        ///
        ytitle("Generalized Lorenz ordinate (net wealth)", axis(1)) ytitle(" ", axis(2)) ///
        legend(label (1 "1994") label(2 "1998")      ///
        region(lstyle(none)) ) saving(glc_net, replace)
graph export glc_net.eps, replace
su glcn* [aw=wgt], de

capture drop pgross pdebt pnet

************ Absolute Lorenz curves ***********************

glcurve gross [aw=wgt] , by(year) split pvar(pgross) glvar(alcg) nograph
su gross [aw=wgt] if year == 1, meanonly
replace alcg_1 = alcg_1 - pgross*r(mean)
su gross [aw=wgt] if year == 2, meanonly
replace alcg_2 = alcg_2 - pgross*r(mean)
lab var alcg_1 "1994"
lab var alcg_2 "1998"
lab var pgross "Cumulative population share"

sort pgross
graph twoway (line alcg_1 pgross, yaxis(1 2) )                      ///
        (line alcg_2 pgross, yaxis(1 2) )                    ///
        , aspect(1) xtitle("Cumulative population share, p")        ///
        ytitle("Absolute Lorenz ordinate (gross wealth)", axis(1)) ytitle(" ", axis(2)) ///
        legend(label (1 "1994") label(2 "1998")      ///
        region(lstyle(none)) ) saving(alc_gross, replace)
graph export alc_gross.eps, replace
su alcg* [aw=wgt], de

glcurve debt [aw=wgt] , by(year) split pvar(pdebt) glvar(alcd) nograph
su debt [aw=wgt] if year == 1, meanonly
replace alcd_1 = alcd_1 - pdebt*r(mean)
su debt [aw=wgt] if year == 2, meanonly
replace alcd_2 = alcd_2 - pdebt*r(mean)
lab var alcd_1 "1994"
lab var alcd_2 "1998"
lab var pdebt "Cumulative population share"
sort pdebt
graph twoway (line alcd_1 pdebt, yaxis(1 2) )                       ///
        (line alcd_2 pdebt, yaxis(1 2) )                     ///
        , aspect(1) xtitle("Cumulative population share, p")        ///
        ytitle("Absolute Lorenz ordinate (debt)", axis(1)) ytitle(" ", axis(2)) ///
        legend(label (1 "1994") label(2 "1998")      ///
        region(lstyle(none)) ) saving(alc_debt, replace)
graph export alc_debt.eps, replace
su alcd* [aw=wgt], de

glcurve net [aw=wgt] , by(year) split pvar(pnet) glvar(alcn) nograph
su net [aw=wgt] if year == 1, meanonly
replace alcn_1 = alcn_1 - pnet*r(mean)
su net [aw=wgt] if year == 2, meanonly
replace alcn_2 = alcn_2 - pnet*r(mean)
lab var alcn_1 "1994"
lab var alcn_2 "1998"
lab var pnet "Cumulative population share"
sort pnet
graph twoway (line alcn_1 pnet, yaxis(1 2) )                        ///
        (line alcn_2 pnet, yaxis(1 2) )                      ///
        , aspect(1) xtitle("Cumulative population share, p")        ///
        ytitle("Absolute Lorenz ordinate (net wealth)", axis(1)) ytitle(" ", axis(2)) ///
        legend(label (1 "1994") label(2 "1998")      ///
        region(lstyle(none)) ) saving(alc_net, replace)
graph export alc_net.eps, replace
su alcn* [aw=wgt], de

* What if trim bottom and top 1% of observations of net wealth

capture drop pgross pdebt pnet alcn*
_pctile net [aw = wgt] if year == 1, p(1 99)
```

```
replace net = . if year == 1 & ( net < r(r1) | net > r(r2) )
_pctile net [aw = wgt] if year == 2, p(1 99)
replace net = . if year == 2 & ( net < r(r1) | net > r(r2) )
glcurve net [aw=wgt] , by(year) split pvar(pnet) glvar(alcn) nograph
su net [aw=wgt] if year == 1, meanonly
replace alcn_1 = alcn_1 - pnet*r(mean)
su net [aw=wgt] if year == 2, meanonly
replace alcn_2 = alcn_2 - pnet*r(mean)
lab var alcn_1 "1994"
lab var alcn_2 "1998"
lab var pnet "Cumulative population share"
sort pnet
graph twoway (line alcn_1 pnet, yaxis(1 2) )                        ///
        (line alcn_2 pnet, yaxis(1 2) )                     ///
        , aspect(1) xtitle("Cumulative population share, p")        ///
        ytitle("Absolute Lorenz ordinate (net wealth)", axis(1)) ytitle(" ", axis(2)) ///
        legend(label (1 "1994") label(2 "1998")      ///
        region(lstyle(none)) ) saving(alc_net_trimmed, replace)
graph export alc_net_trimmed.eps, replace
su alcn* [aw=wgt], de

log close
```

an04.do (Inequality indices, relative and absolute)

```
version 8
clear
set more off
capture log close
cd d:/home/stephenj/myprojects/lws/finland
log using an04.log, replace
*******************  an04.do ***************************************
*
* Analysis for paper, Finland, 1994, 1998
*
* Section 4: Summary indices (allowing for zeros and negatives!)
*
******************************************************************************
use finland

* relative mean deviation = .5*mean* SUM f_i * abs(w_i - mean)
cap program drop rmd
program define rmd
        quietly {
                tempvar fi
                su `1' [aw=wgt] if year == `2'
                local sumw = r(sum_w)
                local mean = r(mean)
                ge `fi' =  wgt/r(sum_w) if year == `2'
                egen rmd_`1'`2' = sum(`fi'*abs(`1'-`mean')) if year == `2'
                replace rmd_`1'`2' = .5*rmd_`1'`2'/`mean' if year == `2'
                sum rmd_`1'`2', meanonly
                local rmd = r(max)
                noi di "RMD for " "`1'" " in year " `2' " = " `rmd'
        }
end

* Kolm indices: (1/k)*ln[ SUM fi * exp(k*(mean - w_i)) ], k > 0
* Three values of k:
*       k1: inverse of 1994 mean
*       k2: inverse of 1998 mean
*       k3: simple average of k1 and k2
quietly {
        su gross [aw=wgt] if year == 1
        global mg1 = r(mean)
        su gross [aw=wgt] if year == 2
        global mg2 = r(mean)

        global grossk1 = 1/$mg1
        global grossk2 = 1/$mg2
        global grossk3 = .5*($grossk1 + $grossk2)

        su debt [aw=wgt] if year == 1
        global md1 = r(mean)
        su debt [aw=wgt] if year == 2
        global md2 = r(mean)

        global debtk1 = 1/$md1
        global debtk2 = 1/$md2
        global debtk3 = .5*($debtk1 + $debtk2)
```

```
        su net [aw=wgt] if year == 1
        global mn1 = r(mean)
        su net [aw=wgt] if year == 2
        global mn2 = r(mean)

        global netk1 = 1/$mn1
        global netk2 = 1/$mn2
        global netk3 = .5*($netk1 + $netk2)
}
cap program drop kolm
program define kolm
        quietly {
                tempvar fi
                su `1' [aw=wgt] if year == `2'
                local sumw = r(sum_w)
                local mean = r(mean)
                ge `fi' =  wgt/r(sum_w) if year == `2'

                egen kolm1_`1'`2' =   sum(`fi'*exp(${`1'k1}*(`mean'-`1'))) if year == `2'
                replace kolm1_`1'`2' = (1/${`1'k1})*ln(kolm1_`1'`2') if year == `2'
                sum kolm1_`1'`2', meanonly
                local kolm = r(max)
                noi di "Kolm1 for " "`1'" " in year " `2' " = " `kolm' " ; k1 = " ${`1'k1'}

                egen kolm2_`1'`2' =   sum(`fi'*exp(${`1'k2'}*(`mean'-`1'))) if year == `2'
                replace kolm2_`1'`2' = (1/${`1'k2'})*ln(kolm2_`1'`2') if year == `2'
                sum kolm2_`1'`2', meanonly
                local kolm = r(max)
                noi di "Kolm2 for " "`1'" " in year " `2' " = " `kolm' " ; k2 = " ${`1'k2'}

                egen kolm3_`1'`2' =   sum(`fi'*exp(${`1'k3'}*(`mean'-`1'))) if year == `2'
                replace kolm3_`1'`2' = (1/${`1'k3'})*ln(kolm3_`1'`2') if year == `2'
                sum kolm3_`1'`2', meanonly
                local kolm = r(max)
                noi di "Kolm3 for " "`1'" " in year " `2' " = " `kolm' " ; k3 = " ${`1'k3'}
        }
end

********** Relative indices: Gini, GE(2), RMD ****************

ineqdec0 gross [aw=wgt] if year == 1
local gini_g1 = $S_gini
local i2_g1 = $S_i2
rmd gross 1

ineqdec0 gross [aw=wgt] if year == 2
local gini_g2 = $S_gini
local i2_g2 = $S_i2
rmd gross 2

ineqdec0 debt [aw=wgt] if year == 1
local gini_d1 = $S_gini
local i2_d1 = $S_i2
rmd debt 1

ineqdec0 debt [aw=wgt] if year == 2
local gini_d2 = $S_gini
local i2_d2 = $S_i2
rmd debt 2

ineqdec0 net [aw=wgt] if year == 1
local gini_n1 = $S_gini
local i2_n1 = $S_i2
rmd net 1

ineqdec0 net [aw=wgt] if year == 2
local gini_n2 = $S_gini
local i2_n2 = $S_i2
rmd net 2

* % change in inequality indices

di "% change in Gini (gross) = " 100*(`gini_g2' - `gini_g1')/`gini_g1'
di "% change in GE(2) (gross) = " 100*(`i2_g2' - `i2_g1')/`gini_g1'

di "% change in Gini (debt) = " 100*(`gini_d2' - `gini_d1')/`gini_d1'
di "% change in GE(2) (debt) = " 100*(`i2_d2' - `i2_d1')/`i2_d1'

di "% change in Gini (net) = " 100*(`gini_n2' - `gini_n1')/`gini_n1'
di "% change in GE(2) (net) = " 100*(`i2_n2' - `i2_n1')/`i2_n1'

********** Absolute indices: absolute Gini, Kolm x 3 **********
```

```
di "Absolute Gini for gross in year 1 = " `gini_g1'*$mg1
di "(mean for gross in year 1 = " $mg1  ")"
kolm gross 1

di "Absolute Gini for gross in year 2 = " `gini_g2'*$mg2
di "(mean for gross in year 2 = " $mg2  ")"
kolm gross 2

di "Absolute Gini for debt in year 1 = " `gini_d1'*$md1
di "(mean for debt in year 1 = " $md1  ")"
kolm debt 1

di "Absolute Gini for debt in year 2 = " `gini_d2'*$md2
di "(mean for debt in year 2 = " $md2  ")"
kolm debt 2

di "Absolute Gini for net in year 1 = " `gini_n1'*$mn1
di "(mean for net in year 1 = " $mn1  ")"
kolm net 1

di "Absolute Gini for net in year 2 = " `gini_n2'*$mn2
di "(mean for net in year 2 = " $mn2  ")"
kolm net 2

****** Effect of extreme values on Gini, GE(2) for net ****************

qui sum net [aw=wgt] if year == 1, de
local max = r(max)
local p1 = r(p1)
local p99 = r(p99)

* drop max
ineqdec0 net [aw=wgt] if year == 1 & net < `max'
local i2_1m = $S_i2
local gini_1m = $S_gini
* trim top and bottom 1%
ineqdec0 net [aw=wgt] if year == 1 & net >= `p1' & net <= `p99'
local i2_1t = $S_i2
local gini_1t = $S_gini

qui sum net [aw=wgt] if year == 2, de
local max = r(max)
local p1 = r(p1)
local p99 = r(p99)

* drop max
ineqdec0 net [aw=wgt] if year == 2 & net < `max'
local i2_2m = $S_i2
local gini_2m = $S_gini
* trim top and bottom 1%
ineqdec0 net [aw=wgt] if year == 2 & net >= `p1' & net <= `p99'
local i2_2t = $S_i2
local gini_2t = $S_gini

di "% change in Gini (excluding max obs) = " 100*(`gini_2m' - `gini_1m')/`gini_1m'
di "% change in Gini (trimming top and bottom 1%) = " 100*(`gini_2t' - `gini_1t')/`gini_1t'
di "% change in GE(2) (excluding max obs) = " 100*(`i2_2m' - `i2_1m')/`i2_2m'
di "% change in GE(2) (trimming top and bottom 1%) = " 100*(`i2_2t' - `i2_1t')/`i2_1t'

log close
```

an05.do (Dagum III finite mixture model)

```
version 8
clear
set more off
capture log close
cd d:/home/stephenj/myprojects/lws/finland
log using an05.log, replace
*******************  an05.do ****************************************
* Dagum3 model fitting for net wealth only
*  Finland 1994, 1998
use finland, clear

********************** 1994 *************************************
keep if year == 1

// net wealth variable into global macro
global w "net"

* summarize net wealth vble
```

```
su $w [aw=wgt], de
ineqdec0 $w  [aw=wgt]
sumdist $w  [aw=wgt]
cumul  $w [aw=wgt], gen(ecdf)

******** preliminary fitting of separate models to get starting values ***********

* Proportions with neg, zero, pos (trickier to get since need weighted proportion)
sum net [aw=wgt], meanonly
local sumw = r(sum_w)
sum $w [aw=wgt] if $w <0, meanonly
local b1 = r(sum_w)/`sumw'
di "Proportion with value < 0 = " `b1'
sum $w [aw=wgt] if $w == 0, meanonly
local b2 = r(sum_w)/`sumw'
di "Proportion with value = 0 = " `b2'
local b3 = 1 - `b1' - `b2'
di "Proportion with value > 0 = " `b3'

* fit Dagum model to positive values
dagumfit $w [aw=wgt] , stats
* e(ba) e(bb) e(bp) contain the estimates
local a = e(ba)
di "`a'"
local b = e(bb)
di "`b'"
local p = e(bp)
di "`p'"
di "Predicted mean among positives: "  `e(mean)'
su $w [aw=wgt] if $w > 0

* fit exponential model to negative values
cap program drop dagumm_ll
program define dagumm_ll
        version 8.2
        args lnf t
        quietly replace `lnf' = ln(`t') + (`t' * $w) if $w < 0
end
ml model lf dagumm_ll (t: `tvar')  [aw=wgt] if $w < 0
* ml check
ml search
ml maximize
mat b = e(b)
local t = b[1,1]
di "`t'"
di "Predicted mean among negatives: "  -1/`t'
su $w  [aw=wgt] if $w < 0

*********** fit general (mixture) model ************************

capture program drop dagum3_ll
program define dagum3_ll
        version 8.2
        args lnf a b p b1 b2 t
        quietly        {
                replace `lnf' = ln(`b1') + ln(`t') + (`t' * $w) if $w < 0
                replace `lnf' = ln(`b2') if $w == 0
                replace `lnf' = ln( 1 - `b1' - `b2' )                ///
                              + ln(`a') + ln(`p') + `a'*ln(`b')     ///
                              - (`a'+1)*ln($w)              ///
                              - (`p'+1)*ln(1+(`b'/$w)^(`a'))        ///
                               if $w > 0
        }
end

ml model lf dagum3_ll  (a: `avar') (b: `bvar')  (p: `pvar') ///
               (b1: `b1var') (b2: `b2var') (t: `tvar')  [aw=wgt]

ml init  a:_cons = `a' b:_cons = `b' p:_cons = `p' b1:_cons = `b1' ///
                b2:_cons = `b2' t:_cons = `t'
ml search
ml maximize
mat list e(b)
mat b = e(b)
local a = b[1,1]
local b = b[1,2]
local p = b[1,3]
local b1 = b[1,4]
local b2 = b[1,5]
local t = b[1,6]
nlcom b3: 1 - [b1]_cons - [b2]_cons, post
mat c = e(b)
local b3 = c[1,1]
```

```
ge cdf = `b1' * exp(`t'*$w) if $w <=0
replace cdf = `b1' + `b2' if $w == 0
replace cdf = `b1' + `b2'                ///
        + `b3' * ( (1 + (`b'/$w)^`a')^-(`p') ) if $w > 0
su cdf ecdf, de

su $w  [aw=wgt]
su $w  [aw=wgt] if $w < 0
su $w  [aw=wgt] if $w > 0

local m =  -`b1'/`t' + `b3'*( `b'*exp(lngamma(1-
1/`a'))*exp(lngamma(`p'+1/`a'))/exp(lngamma(`p')) )

di "Mean = " `m'

di "Mean among negatives: " -1/`t'
di "Mean among positives: "  `b'*exp(lngamma(1-
1/`a'))*exp(lngamma(`p'+1/`a'))/exp(lngamma(`p'))

local gini = 1 + 3*(`b1')^2/(2*`t'*`m')                      ///
        - (2*(`b3')^2*`p'*`b'              ///
        *exp(lngamma(1-1/`a'))*exp(lngamma(`p'+1/`a'))/exp(lngamma(1+`p'))) ///
        * (1 - .5*( exp(lngamma(`p')) ///
        * exp(lngamma(2*`p'+1/`a'))/(exp(lngamma(2*`p'))*exp(lngamma(`p'+1/`a'))) ))/`m'
di "Gini = " `gini'

ge pdf1 = `b1'*exp(`t'*$w)*($w < 0)
ge pdf2 = `b2' * ($w == 0)
ge pdf3 =  `b3' * (`a'*`p' * (`b'/$w)^`a'/$w) * (1 + (`b'/$w)^`a')^ -(`p'+1) if $w > 0
replace pdf3 = 0 if $w <= 0
ge pdf = pdf1 + pdf2 + pdf3
su pdf*  [aw=wgt]

* predicted and sample median

list $w ecdf if abs(ecdf - .5) < 0.001
list $w cdf if abs(cdf - .5) < 0.001

lab var cdf "CDF (fitted)"
lab var ecdf "CDF (empirical)"
lab var pdf "PDF (fitted)"

graph twoway (line ecdf $w) (line cdf $w)                    ///
        , xtitle("1994 Net wealth, W") ytitle("p = F(W)")    ///
        legend( region(lstyle(none)) ) saving(cdf_net_fitted_94, replace)
graph export cdf_net_fitted_94.eps, replace

graph twoway (line pdf $w)                  ///
        , xtitle("1994 Net wealth, W") ytitle("Density, f(W)")      ///
        legend( region(lstyle(none)) ) saving(pdf_net_fitted_94, replace)
graph export pdf_net_fitted_94.eps, replace

* example of general (mixture) model with covariates *

ml model lf dagum3_ll  (a: age agesq) (b: age agesq)  (p: age agesq) ///
         (b1: age agesq) (b2: age agesq) (t: age agesq)  [aw=wgt] , tech(bhhh nr)
ml init  a:_cons = `a' b:_cons = `b' p:_cons = `p' b1:_cons = `b1' b2:_cons = `b2' t:_cons =
`t'
ml search
ml maximize, difficult

*********************** 1998 *************************************
use finland, clear
keep if year == 2

// net wealth variable into global macro
global w "net"

* summarize net wealth vble
su $w [aw=wgt], de
ineqdec0 $w  [aw=wgt]
sumdist $w  [aw=wgt]
cumul  $w [aw=wgt], gen(ecdf)

******** preliminary fitting of separate models to get starting values ***********

* Proportions with neg, zero, pos (trickier to get since need weighted proportion)
sum net [aw=wgt], meanonly
local sumw = r(sum_w)
sum $w [aw=wgt] if $w <0, meanonly
local b1 = r(sum_w)/`sumw'
di "Proportion with value < 0 = " `b1'
```

```
sum $w [aw=wgt] if $w == 0, meanonly
local b2 = r(sum_w)/`sumw'
di "Proportion with value = 0 = " `b2'
local b3 = 1 - `b1' - `b2'
di "Proportion with value > 0 = " `b3'

* fit Dagum model to positive values
dagumfit $w [aw=wgt], stats
* e(ba) e(bb) e(bp) contain the estimates
local a = e(ba)
di "`a'"
local b = e(bb)
di "`b'"
local p = e(bp)
di "`p'"
di "Predicted mean among positives: "  `e(mean)'
su $w  [aw=wgt] if $w > 0

* fit exponential model to negative values
cap program drop dagumm_ll
program define dagumm_ll
        version 8.2
        args lnf t
        quietly replace `lnf' = ln(`t') + (`t' * $w) if $w < 0
end
ml model lf dagumm_ll (t: `tvar') [aw=wgt] if $w < 0
ml search
ml maximize
mat b = e(b)
local t = b[1,1]
di "`t'"

di "Predicted mean among negatives: "  -1/`t'
su $w [aw=wgt] if $w < 0

*********** fit general (mixture) model ************************

capture program drop dagum3_ll
program define dagum3_ll
        version 8.2
        args lnf a b p b1 b2 t
        quietly        {
                replace `lnf' = ln(`b1') + ln(`t') + (`t' * $w) if $w < 0
                replace `lnf' = ln(`b2') if $w == 0
                replace `lnf' = ln( 1 - `b1' - `b2' )               ///
                               + ln(`a') + ln(`p') + `a'*ln(`b')     ///
                               - (`a'+1)*ln($w)              ///
                               - (`p'+1)*ln(1+(`b'/$w)^(`a'))        ///
                                if $w > 0
        }
end

ml model lf dagum3_ll  (a: `avar') (b: `bvar')  (p: `pvar')  (b1: `b1var') ///
        (b2: `b2var') (t: `tvar')  [aw=wgt], tech(bhhh)
ml init  a:_cons = `a' b:_cons = `b' p:_cons = `p' b1:_cons = `b1' b2:_cons = `b2' t:_cons =
`t'
ml search
ml maximize
mat list e(b)
mat b = e(b)
local a = b[1,1]
local b = b[1,2]
local p = b[1,3]
local b1 = b[1,4]
local b2 = b[1,5]
local t = b[1,6]
nlcom b3: 1 - [b1]_cons - [b2]_cons, post
mat c = e(b)
local b3 = c[1,1]

ge cdf = `b1' * exp(`t'*$w) if $w <=0
replace cdf = `b1' + `b2' if $w == 0
replace cdf = `b1' + `b2'              ///
        + `b3' * ( (1 + (`b'/$w)^`a')^-(`p') ) if $w > 0
su cdf ecdf, de
su $w [aw=wgt]
su $w [aw=wgt] if $w < 0
su $w [aw=wgt] if $w > 0

local m =  -`b1'/`t' + `b3'*( `b'*exp(lngamma(1-
1/`a'))*exp(lngamma(`p'+1/`a'))/exp(lngamma(`p')) )
di "Mean = " `m'
di "Mean among negatives: " -1/`t'
```

```
di "Mean among positives: "  `b'*exp(lngamma(1-
1/`a'))*exp(lngamma(`p'+1/`a'))/exp(lngamma(`p'))

local gini = 1 + 3*(`b1')^2/(2*`t'*`m') - (2*(`b3')^2*`p'*`b'              ///
        *exp(lngamma(1-1/`a'))*exp(lngamma(`p'+1/`a'))/exp(lngamma(1+`p'))) ///
        * (1 - .5*( exp(lngamma(`p'))*                                     ///
        exp(lngamma(2*`p'+1/`a'))/(exp(lngamma(2*`p'))*exp(lngamma(`p'+1/`a'))) ))/`m'

di "Gini = " `gini'
ge pdf1 = `b1'*exp(`t'*$w)*($w < 0)
ge pdf2 = `b2' * ($w == 0)
ge pdf3 =  `b3' * (`a'*`p' * (`b'/$w)^`a'/$w) * (1 + (`b'/$w)^`a')^ -(`p'+1) if $w > 0
replace pdf3 = 0 if $w <= 0
ge pdf = pdf1 + pdf2 + pdf3
su pdf*  [aw=wgt]

* predicted and sample median
list $w ecdf if abs(ecdf - .5) < 0.001
list $w cdf if abs(cdf - .5) < 0.001

lab var cdf "CDF (fitted)"
lab var ecdf "CDF (empirical)"
lab var pdf "PDF (fitted)"
graph twoway (line ecdf $w) (line cdf $w)                   ///
        , xtitle("1998 Net wealth, W") ytitle("p = F(W)")    ///
        legend( region(lstyle(none)) ) saving(cdf_net_fitted_98, replace)
graph export cdf_net_fitted_98.eps, replace

graph twoway (line pdf $w)                   ///
        , xtitle("1998 Net wealth, W") ytitle("Density, f(W)")      ///
        legend( region(lstyle(none)) ) saving(pdf_net_fitted_98, replace)
graph export pdf_net_fitted_98.eps, replace

* example of general (mixture) model with covariates *

ml model lf dagum3_ll  (a: age agesq) (b: age agesq)  (p: age agesq) ///
         (b1: age agesq) (b2: age agesq) (t: age agesq)  [aw=wgt], tech(bhhh nr)
ml init  a:_cons = `a' b:_cons = `b' p:_cons = `p' b1:_cons = `b1' b2:_cons = `b2' t:_cons =
`t'
ml search
ml maximize, difficult

log close
```