Pareto Models, Top Incomes, and Recent Trends in UK Income Inequality

Stephen P. Jenkins (London School of Economics)

Paper prepared for the 34th IARIW General Conference

Dresden, Germany, August 21-27, 2016

Session 7E: Household Income and Wealth: Distribution and Trends

Time: Friday, August 26, 2016 [Morning]

Pareto models, top incomes, and recent trends in UK income inequality

Stephen P. Jenkins
(LSE, ISER (University of Essex), and IZA)

Revised version, 31 July 2016

Abstract

I determine UK income inequality levels and trends by combining inequality estimates from tax return data (for the 'rich') and household survey data (for the 'non-rich'), taking advantage of the better coverage of top incomes in tax return data (which I demonstrate) and creating income variables in the survey data with the same definitions as in the tax data to enhance comparability. For top income recipients, I estimate inequality and mean income by fitting Pareto models to the tax data, examining specification issues in depth, notably whether to use Pareto I or Pareto II (generalised Pareto) models, and the choice of income threshold above which the Pareto models apply. The preferred specification is a Pareto II model with a threshold set at the 99th or 95th percentile (depending on year). Conclusions about aggregate UK inequality trends since the mid-1990s are robust to the way in which tax data are employed. The Gini coefficient for gross individual income rose by around 7% or 8% between 1996/97 and 2007/08, with most of the increase occurring after 2003/04. The corresponding estimate based wholly on the survey data is around –5%.

Keywords: inequality, top incomes, Pareto distribution, generalized Pareto distribution, survey under-coverage, HBAI, SPI

JEL Classifications: C46, C81, D31

Acknowledgements

This paper is inspired by and draws on Frank Cowell's contributions to the analysis using Pareto models, income data problems, and inequality decomposition (see, inter alia, Cowell 1977, 1980, 1989, 2011, 2013; Cowell et al. 1998; Cowell and Flachaire 2007; Cowell and Kuga 1980; Cowell and Victoria-Feser 1996, 2007). Frank: saluto tu! My research is part supported by an Australian Research Council Discovery Grant (award DP150102409, with Richard Burkhauser, Nicolas Hérault, and Roger Wilkins) and core funding of the Research Centre on Micro-Social Change at the Institute for Social and Economic Research by the University of Essex and the UK Economic and Social Research Council (award ES/L009153/1). I wish to thank Nicolas Hérault for preparing the individual-level income data; Tony Atkinson, Facundo Alvaredo, and Christian Schluter for helpful discussions about Pareto distributions and top incomes; David Roodman and Philippe Van Kerm for sharing their Stata programs; and the handling editor, two anonymous referees, and audiences at Aix-Marseilles University, KU Leuven, and Ca' Foscari University of Venice for comments on a first draft.

Correspondence: Stephen P. Jenkins, Department of Social Policy, London School of Economics, Houghton Street, London WC2A 2AE, UK. Email: s.jenkins@lse.ac.uk.

1. Introduction

There is a bifurcation in the literature on income inequality levels and trends. On the one hand, most official statistics and academic analysis utilise data from household surveys and report estimates of the inequality of family or household disposable income summarised using Gini coefficients and other inequality indices calculated using all incomes from poorest to richest. (See e.g. OECD 2008, 2011, 2015, on cross-national comparisons, and Department of Work and Pensions 2015 on UK trends.) On the other hand, there is the 'top incomes' literature that uses administrative record data on personal income tax returns, reporting estimates of top income shares – the share of total income received by the richest 1% or richest 10%, and so on. (See e.g. Alvaredo et al. 2013, Atkinson and Piketty 2007 on cross-national comparisons, and Atkinson 2005 on UK trends.)

The two literatures differ in their findings about recent inequality trends: estimates from tax return data show a substantial rise in inequality over the last two decades in both the UK and USA, for instance, whereas survey-based estimates of inequality show much less change. For the UK, for example, the share of total income held by the richest 1% increased by 29% between fiscal years 1996/97 and 2007/08 whereas the Gini coefficient increased by 7%. For the USA, the corresponding increases over the same period are 30% and 2%. (See Burkhauser et al. (2016: Figures 1 and A1), for further details about estimates and sources.)

The divergent findings about inequality trends from the two data sources arise partly because different inequality indices and income definitions are employed (more on this later). However, another important explanation is that household surveys do not capture top incomes very well, whereas tax data do a much better job of this.

In this paper, I determine UK income inequality levels and trends since the mid-1990s by combining estimates from tax return data (for the 'rich') and household survey data (for the 'non-rich'), taking advantage of the better coverage of top incomes in tax return data (which I demonstrate) and creating income variables in the survey data with the same definitions as in the tax data to enhance comparability. I also analyse how estimates of inequality trends differ by inequality index.

There are multiple sources of under-coverage of top incomes in survey data. The first is under-reporting among high-income respondents or top-coding of their responses by survey administrators. In these cases, survey data are right-censored. A second source of under-coverage is the sampling of high-income respondents per se. Respondents may provide sparse coverage of the top income ranges and, in addition, there may be no respondents at all from the extreme right-hand tail, because the survey organisation does not target potential high income respondents by design, or it is unable to contact them, or there is contact but refusal to participate. In this case, the observed income data are a right-truncated sample of the 'true' distribution. Both types of under-coverage contribute downward bias to survey estimates of inequality for a given year because there is not enough income observed in the very top income ranges. A by-product of sparse coverage of the top income ranges is that the high-income observations present in the survey data have the characteristics of outliers (even if they are genuine rather than an error) and have substantial influence on the conventional non-parametric estimate of an inequality measure: see Cowell and Victoria-Feser (1996, 2007) and Cowell and Flachaire (2007). This sensitivity can introduce spurious volatility in time series of inequality estimates.

There are three approaches to estimating inequality measures that address these under-coverage problems: see Figure 1 for a schematic summary. Approach A is based entirely on survey data. It derives an inequality estimate for the poorest *p*% using non-parametric methods applied to survey unit-record data, and derives an inequality estimate from the richest (1–*p*)% by fitting a Pareto Type I distribution to the top income observations from the same source. The estimate of total inequality, mostly summarised using the Gini coefficient, is calculated by adding together three components: inequality within the top group, inequality within the non-top group, and between-group inequality.

<Figure 1 near here>

Cowell and Flachaire (2007) provide a thorough examination of the properties of Approach A motivated by, and focusing on, the problem of sparse coverage of top income ranges. Their headline conclusion is that such 'use of appropriate semiparametric methods for modelling the upper tail can greatly improve the performance of those inequality indices that are normally considered particularly sensitive to extreme values' (2007: 1044). Alfons et al. (2013) also motivate their application of Approach A, using EU-SILC survey data for Austria and Belgium, with reference to sensitivity issues. Neither article refers to under-coverage per se. By contrast, Ruiz and Woloszko motivate their application to survey data for OECD countries in terms of 'correcting household survey data for underreporting in the upper-tail of income distributions' (2015: 6). Burkhauser et al. (2012) use Approach A to adjust for the systematic under-coverage of high incomes in public use Current Population Survey datasets introduced by US Census Bureau top-coding. In both applications, the idea is that the upper tail to the income distribution implied by the parametric model estimates will capture more income than non-parametric estimates.

There is evidence that Approach A's ability to address survey under-coverage at the top is limited. For example, survey-based estimates of the share of total income held by the top 1% are several percentage points less than the estimates from tax return data according to the analysis of Atkinson et al. (2011) and Burkhauser (2012) for the USA. Put differently, fitting a parametric upper tail may obviate the sparsity problem (there is density mass at all points of the distribution's support, by assumption), but the estimate of the 'true' upper tail based on model-based extrapolation from the observed survey observations may not be reliable. This motivates the use of tax return data, as they have better coverage of the upper tail.

Approaches B and C both use tax return data but take different routes to addressing under-coverage issues. Approach B replaces the highest incomes in the survey with cell-mean imputations based on the corresponding observations in the tax return data. The 'SPI adjustment' to Family Resource Survey income data – used to derive the UK's official income distribution statistics since the early 1990s – is an example of this approach (see e.g. Department for Work and Pensions 2015). Burkhauser et al. (2016) apply Approach B in a more extensive and comprehensive manner and use World Top Incomes Database (Alvaredo et al. 2015) estimates of top income shares as a benchmark. Bach et al. (2009) is an application to Germany.

Approach C, used in this paper, *combines estimates* from the two types of data source rather than *combining data* per se as Approach B does. It is thus identical to Approach A except that it uses both survey and tax data rather than only the former; it is this feature that addresses the under-coverage problem. Approach C was developed by Atkinson (2007: 19–20) with an application to the USA by Atkinson et al. (2011), and extended by Alvaredo (2011) who also included applications to Argentina and the USA. Subsequent applications include those by Alvaredo and Londoño Vélez (2015) and Diaz-Bazan (2015) to Colombia, and by Lakner and Milanovic (2016) and Anand and Segal (2016) to global income inequality. Each of the applications cited uses a Pareto I model to describe the upper tail of the income distribution. In principle, researchers could employ non-parametric estimates of inequality indices for the top incomes in the tax data, but there is then the issue of whether these would be subject to the sensitivity problems mentioned earlier. The issue has not been studied using tax data before: I do so in this paper.

To perform well, Approaches B and C both rely on the researcher using the same 'income' definition in both data sources and ensuring that calculations refer to the same population. Otherwise, there is an 'apples + bananas' problem: non-comparability introduces bias. To avoid this, we may exploit a comparative advantage of survey data. The ability to change income definitions in tax return data is limited but, with access to unit record survey data, we can do a cross-walk from survey to tax data definitions. That is what I do in this paper, employing the same harmonized income variables for the survey and tax data as Burkhauser et al. (2016). For more details, see below.

This paper makes several contributions. First, there is the substantive application to UK inequality trends since the mid-1990s. How much income inequality has been growing is of much public interest. Second, related, there is question of whether Approaches C and B tell the same story about trends when applied to the same data sources. I contrast my Approach C estimates with the Approach B estimates provided by the official statistics (Department for Work and Pensions 2015; see also Belfield et al. 2015) and Burkhauser et al. (2016). Third, I provide new evidence about the extent to which there is under-coverage by survey data of the UK income distribution, using comparable tax data as the benchmark.

Fourth, I provide new analysis of issues that arise when fitting a Pareto model to the upper tail of the income distribution, and hence of direct relevance to researchers applying the semiparametric Approaches A and C. My findings are relevant to analysis of other heavy-tailed distributions such as wealth (Shorrocks et al. 2015, Vermeulen 2014), and city and firm size (Eeckhout 2004; Gabaix 2009, 2016). I use unit record tax return data rather than grouped (bracketed) data and so have flexibility to explore a number of econometric issues. (On estimation issues that arise with grouped tax return data – the only source available for deriving very long historical series – see Atkinson 2005, 2007 and references therein.) For instance, for the Pareto Type I model, I compare the performance of ordinary least squares, maximum likelihood, and maximum likelihood-robust estimators. I also address two implementation questions.

The first question is: what model should be fitted to top incomes? To date, researchers have invariably used the Pareto Type I model. This has a single shape parameter and there are simple formulae for calculating mean income and inequality indices from parameter estimates. There is also a widespread view that Pareto Type I models fit top income data well (Atkinson et al. 2014: 14). However, many of the goodness of fit checks that researchers have employed do not reliably distinguish Pareto distributions from other heavy-tailed distributions. In addition, most of the goodness of fit approaches used can only check whether data are consistent with a distribution in the Pareto family, i.e. not with the Pareto Type I specifically (Cirillo 2013). I provide the first systematic comparison of the goodness of fit of Pareto Type I and Pareto Type II (‘generalised Pareto’) models to top income data, and show that the latter outperforms the former except at extremely high thresholds – thresholds that are well above those typically employed.

The second and related implementation question is: if we assume that incomes are described by a Pareto model above some threshold, what should that threshold be? In particular, when implementing Approaches C or A, what is the cut-off to use to distinguish between top incomes and non-top incomes? Is the top income group the top 10% (Ruiz and Woloszko 2015), or the top 5% (Atkinson 2016), or the top 1% (Alvaredo 2011)? There is some evidence that a higher cut-off decreases the estimate of the Pareto Type I shape parameter, i.e. increases inequality among top incomes, other things being equal (see e.g. Burkhauser 2012: Appendix A). However, the impact on total inequality estimated using Approach C of changing the threshold is unclear, because inequality and the mean among non-top incomes and between-group inequality also change.

Several criteria have been proposed for choosing Pareto thresholds (see e.g. Clauset et al. 2009, Coles 2001) and I employ them. However, I also argue that there is an additional issue to be taken into account when applying Approach C. That is, because non-coverage issues motivate the approach, it is important to ascertain precisely where along the top income range it is that survey non-coverage occurs. There is little evidence about this for the UK. I show that survey non-coverage is apparent from around the 99th percentile upwards in the mid- to late-1990s or from around the 95th percentile in the 2000s. I use the 99th and 95th percentiles as the Pareto threshold when deriving my inequality estimates, as well as the 90th percentile as a robustness check.

I introduce in Section 2 the UK tax return and survey data that I use, and explain the creation of income variables using harmonized definitions and hence on a comparable basis. Section 3 provides evidence about under-coverage of the survey data using the tax data as the benchmark. I analyse the fitting of Pareto models to top incomes in tax return data in Section 4, and present estimates of overall inequality levels and trends since the mid-1990s in Section 5. Section 6 provides a summary and conclusions. Applying Approach C, I show that choosing different Pareto models and different thresholds has noticeable impacts on estimates of inequality among the rich. However, my conclusions about overall inequality trends are broadly robust to the choice of Pareto model and percentile threshold, and there are similar results if upper tail inequality and mean income are estimated non-parametrically. The estimated inequality trends from Approach C are also similar to those derived using Approach B (Burkhauser et al. 2016). For example, the Gini coefficient for gross individual income rose by around 7% or 8% between 1996/97 and 2007/08, with most of the increase occurring after 2003/04. The corresponding estimate based wholly on the survey data is around –5%.

2. Survey and tax data, and the definition of income

The income tax return data are from the public-release files of the *Survey of Personal Incomes* (SPI) for each year 1995/96 through 2010/11, with the exception of 2008/09 for which no data have been released. Atkinson (2005) uses these data, as well as published tabulations from the SPI and from supertax and surtax returns for earlier years, in his pioneering analysis of trends in UK top income shares since 1908. (See also Atkinson 2016 for Pareto I parameter estimates back to 1799.) The SPI data underlie the UK top income share estimates in the World Top Incomes Database (WTID) (Alvaredo et al. 2015). Each year's SPI is a stratified sample of the universe of tax returns. The number of individuals in the data has increased from around 57,000 individuals in 1995/96 to nearly 677,500 in 2010/11, corresponding to around 32 million taxpayers. For further details, see HM Revenue and Customs KAI Data, Policy and Co-ordination (2014) and Burkhauser et al. (2016). The data are comparable over time, except for a small discontinuity between 1995/96 and later years (the effect of which I show later). Self-assessment was introduced that year and there were changes to the SPI methodology (personal communication with HMRC). Hence, I use 1996/97 as the base year for analysis of inequality trends rather than 1995/96.

Throughout the period of my analysis (and since 1990), the unit of assessment in the UK income tax system has been the individual. For this reason, the SPI income variables are all individual-level variables, rather than referring to the incomes of families or households (as in the survey data and official income distribution statistics). The SPI income variable I use is *individual gross income* (total taxable income from the market plus taxable government transfers, and before the deduction of income tax), i.e. the same variable that the WTID and the top income shares literature focuses on.

In addition, and to further align my research with the WTID and top income shares literature, I restrict analysis to the population of individuals aged 15 years or more. Because the SPI does not cover all individuals in the UK population or all of their income, the WTID uses external population and income control totals for each year, i.e. estimates of the total number of individuals aged 15 or more, and of the total income held by them. I use the WTID control totals throughout. In practice, I accomplish this by introducing some observations with zero income into each year’s unit record data and adjusting the grossing-up weights supplied with the data.

The unit record survey data I employ come from the *Family Resources Survey* (FRS), and the accompanying subfiles of derived income variables called the *Households Below Average Income* (HBAI) dataset (Department for Work and Pensions 2013, Department for Work and Pensions et al. 2014). I use data for the same period as the SPI data, 1995/96–2010/11. The FRS is a large continuous cross-sectional survey with data released annually for around 20,000 respondent households and the individuals within them. The Department for Work and Pensions (DWP) administers the FRS, and DWP staff produce the HBAI subfiles that they use to derive the UK’s official income distribution statistics published annually using a variant of Approach B, i.e. the ‘SPI adjustment’. (Despite its label, the HBAI provides information about the income distribution as a whole.) In essence, the HBAI subfiles contain a set of FRS income variables that DWP statisticians have cleaned.

Because the DWP’s focus is on family and household post-tax post-transfer income variables (reflecting the needs of official statistics), there is a definitional mismatch between the income variables in the HBAI and the SPI. As it happens, the DWP’s public-use files do contain an individual-level gross income variable but only from 2005/06 onwards. Burkhauser et al. (2016) create a complete time series for the period 1995/96–2010/11 (as for the SPI data) from FRS variables and show that their derived individual-level gross income variable is virtually identical to the DWP’s for the years for which they can make comparisons. I use Burkhauser et al.’s individual gross income variables derived from the HBAI in this paper. (None of these variables are SPI-adjusted in the sense described earlier.) Burkhauser et al. (2016) go on to create a second set of individual-level income variables when implementing Approach B. These data reflect a more extensive ‘SPI adjustment’ procedure than employed by the DWP for the official statistics, and Burkhauser et al. (2016) label it ‘SPI2’ accordingly.

In sum, there are two main individual-level gross income data series employed in the paper to implement Approach C: that from the tax data (‘SPI’) and from the DWP’s cleaned-up survey data (‘HBAI’). In Section 5, I contrast my results for overall inequality based on the SPI and HBAI series (combining estimates) with those derived using Approach B (combining data). I refer to the DWP’s (2015) inequality series as ‘HBAI-SPI’ and the Burkhauser et al. (2016) series as ‘HBAI-SPI2’.

To fully align the survey data with the tax return data, I restrict attention to individuals aged 15 years or more. I use the FRS weights in all calculations with the survey data and SPI weights with the tax data ones. All income variables (from tax and survey data) are expressed in pounds per year in 2012/13 prices.

3. Under-coverage of top incomes by household survey data

Ascertaining the point on the income range at which survey under-coverage of top incomes begins is an integral part of implementing Approaches A and C and of independent interest as well.

Table 1 shows estimates of percentiles *p*90, *p*95, *p*99, *p*99.5 and *p*99.9 derived from the survey and tax data as well as the ratio of each corresponding survey and tax data estimate (in %), by year. (For brevity, henceforth I refer to tax years 1995/96 as '1995', 1996/97 as '1996', and so on.) Real incomes at the top of the distribution generally rose over the period according to either source (look down each column of Table 1), except that there is fall in the uppermost percentiles after 2007, especially in the tax data estimates.

<Table 1 near here>

There are two explanations for the post-2007 fall in the uppermost percentiles. One is the recession at that time. The second, particularly relevant here, is the incentive for high income taxpayers to declare income in tax year 2009/10 rather than 2010/11 in order to avoid the increase in top marginal tax rate from 45% to 50% with effect from April 2010. The subsequent reintroduction of the 45% top marginal rate with effect from April 2013 provided an incentive to defer declaration of income. On these issues of 'forestalling' and 'reverse forestalling', see HM Revenue and Customs (2012). Because of these issues (and having no SPI data for 2008), although I provide annual estimates for the full period between 1995 and 2010, I mostly focus discussion on inequality trends through to 2007.

Table 1 provides clear evidence of under-coverage in top incomes and that its nature changed over the period. Survey estimates of the very top percentiles are more volatile over time than are the tax data estimates, which is indicative of the sparsity aspect of under-coverage. Regarding under-coverage per se, look at the 'ratio' columns: values less than 100% suggest under-coverage. Throughout the period, there is a broad correspondence between survey and tax incomes up to around *p*99. In the mid- to late-1990s, one might refer to 'over-coverage' of the survey up to *p*95. However, in the 2000s, there is a substantial uplift in the very highest incomes shown by the tax data. This is not picked up by the survey data. Between 2000 and 2007, the ratio of survey *p*99 to the tax data *p*99 fell from around 100% to

82%. There is a similar decline in the corresponding ratio for *p*99.5 starting from around 1997 (when it was 100%), down to 78% in 2007. These changes in under-coverage over time suggest that it may be inappropriate to use the same percentile cut-off to define the top income group for all years. I return to this issue. This aside, the table also suggests that the optimal threshold for application of Approach C (or A) should not be lower than *p*95, because survey coverage is adequate up to this point.

Figure 2 provides a complementary perspective on the nature of survey under-coverage. It focuses on 1996 and 2007; the full series for all years is shown in Appendix A. I show densities derived from a histogram for the full distribution of log(income) in the survey data and for the tax data for each year. (I use the logarithmic scale in order to focus on the upper tail.) There are three plots for each year. The leftmost one shows densities plotted for log(income) greater than 10 (i.e. income > £22,026), and the vertical dotted lines mark *p*90, *p*95, *p*99, and *p*99.5 for the relevant year. The other two graphs provide more detailed views on the upper tail by plotting the same densities by plotted only for log(income) greater than 11 (i.e. income > £59,874; middle graph) and log(income) greater than 12 (i.e. income > £162,755; rightmost graph). Histogram areas reflect survivor function proportions, and so comparisons of areas provide information about under-coverage in the sense of how much of top income being captured by the survey data. The histograms also provide information about sparsity and 'outlierness' in top income ranges. Sensitivity issues are likely to be more important, the more that the histograms do not approximate a continuous function and show clumping of density mass.

<Figure 2 near here>

The leftmost plots suggest that the concentration of incomes in the tax and survey data is quite similar for most of the income range if one focuses on the top 5% to 10% of the distribution as a whole. Coverage, summarised by differences in histogram areas, is not so different – though it is clearly worse in 2007 than 1996. Both survey and tax densities appear quite smooth and continuous, though the tax data distribution has a long tail that is not present in the survey data, especially in 2007.

However, differences in income concentration across data sources are much more apparent if one focuses on the extreme top: look at the middle and rightmost plots. In 1996, both densities are discontinuous: extreme incomes are spread sparsely across the top income range, and this range is much greater for tax data. There are greater proportions at the very top in the tax data than in the survey data (the total area of the dark bars is greater than the total area of the light bars). By 2007, and with the secular growth of incomes over the previous decade (Table 1), the survey data are even more clumpy and the proportion with extremely high incomes is more markedly less than for the tax data. In the tax data, the density is relatively continuous up to extremely high incomes.

Overall, Figure 2 suggests that, from the point of view of survey undercoverage of top incomes, the cut-off used to implement Approach C (or A) should lie at around *p*95 or higher, depending on the year. In addition, the sparse spread of incomes along the very top income range means that there are potentially ‘high leverage’ outliers (Cowell and Flachaire 2007) even in the tax data, which could bias estimation of Pareto model parameters. I address this issue below.

4. Fitting Pareto models to top incomes

An integral part of inequality estimation using Approaches A and C is to fit a parametric model to top income data, but there are implementation issues concerning the choice of model and the top income range over which they are fitted. There is also a prior question of whether top incomes are described better by a model other than a Pareto one. This issue has rarely been addressed though one exception is Harrison (1979, 1981) who compares the fit to UK men’s top earnings data of Pareto I, lognormal, and sech^2 distributions. Addressing all these issues is complicated by a chicken and egg problem: most methods for choosing the appropriate model are conditional on a given threshold; and most methods for choosing the threshold have been applied to a single model. One can use multiple models and thresholds but there can be an information overload, and this is and potentially worsened by having 15 years of data covering a period when the income distribution changed. What is appropriate for one year’s data may not be appropriate for another. To address the implementation issues, I have had to make some judicious choices regarding empirical strategy and what I report. A full set of estimates is provided in appendices.

My analysis focuses on comparisons of Pareto I and Pareto II models fitted to SPI tax data. In this section, first I explain the model properties and different parameter estimation methods. (Important references on Pareto distributions include Arnold 2008, 2015, Coles 2001, Cowell 2011, 2015, and Kleiber and Kotz 2003.) Next, I report on tests checking whether Paretianity is an appropriate assumption, and whether answers depend on the income threshold used. Then I consider the relative goodness of fit of Pareto I and II models using two methods and multiple thresholds. Finally, I address the choice of threshold issue using both rule-of-thumb and more formal statistical methods. Overall, I demonstrate that the choice of model and threshold is not as clear cut as typical practice might suggest.

Pareto Type I and Type II models

If income x is characterised by a Pareto Type I model, the survivor function showing the fraction of the population with incomes greater than x, $S(x)$, i.e. one minus the cumulative distribution function, $F(x)$, is:

$$S(x) = 1 - F(x) = \left(\frac{x}{x_m}\right)^{-\alpha} \tag{1}$$

where $x \geq x_m > 0$, and $x_m > 0$ is the lower bound on incomes. Parameter α is the shape parameter ('tail index') describing the heaviness of the right tail of the distribution, with smaller values corresponding to greater tail heaviness. The k^{th} moment exists only if $k < \alpha$.

The survivor function for the Pareto Type II model is:

$$S(x) = \left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \xi > 0 \tag{2}$$

where $x > \mu$ (a location parameter), and $\sigma > 0$ is a scale parameter. Parameter ξ is the shape parameter. In principle, ξ can take on any real value (including the limiting case of $\xi = 0$, which implies an exponential distribution), but the restriction $\xi > 0$ yields heavy-tailed distributions of the 'Pareto' kind. The k^{th} moment exists only if $k < 1/\xi$. The Pareto Type II model is equivalent to a Pareto Type I model when $\xi = 1/\alpha$, $\mu = x_m$, and $\sigma = x_m/\alpha$. With one additional parameter, the Pareto Type II model has the potential to fit real-world top incomes better. But the improvement in goodness-of-fit may be negligible and this has be balanced against the greater simplicity of the Pareto I model.

To implement Approaches A and C, we need formulae for the mean and inequality for the top income group (those with incomes greater than x_m or, equivalently, μ) expressed in terms of the model parameters. I display the formulae for these statistics in Table 2, and clearly they are simpler for the Pareto I model.

<Table 2 near here>

Estimation

Estimation of the two Pareto models proceeds by assuming x_m or μ is a threshold pre-specified by the researcher (not estimated) with its choice determined by a simple rule-of-thumb (such as the 95$^{\text{th}}$ or 99$^{\text{th}}$ percentile) or other means. I return to this issue below.

There are two methods commonly used to estimate the Pareto I shape parameter α. The first is an Ordinary Least Squares (OLS) regression of the log of empirical survivor function on the log of income and a constant term. The idea is that, if eq. (1) holds, then the Zipf plot – a plot of the log of the survivor function against logarithms of income (for incomes in ascending order and greater than x_m) – is a straight line with slope equal to $-\alpha$. Atkinson (2016) explains that α may be estimated by OLS in two other ways. (The Zipf approach uses data on income and the survivor function; the other two approaches utilize information about the total income received by income units.) I have estimated α using all three methods, but find that the Zipf method performed best, and so report only estimates from this in the main text. For the full set of estimates for all years, see Appendix B.

The OLS estimate of α is consistent (Quandt 1966) but the standard error is incorrect because no account is taken of the positive autocorrelation in the residuals introduced by the ranking of incomes. In contrast, the Maximum Likelihood (ML) estimator of α and its standard error is consistent, efficient, and asymptotically normal (Hill 1975, Quandt 1966). I implement the ML estimator using software by Jenkins and Van Kerm (2015). Both OLS and ML estimators are potentially biased in small samples, but the sample sizes in the tax return data employed in this paper are never 'small' – an advantage of using this source.

The ML estimator of α is susceptible to bias when there are a few high outlier incomes, the values of which may be potentially genuine or may reflect error and data contamination in the sense of Cowell and Victoria-Feser (1996, 2007) and Cowell and Flachaire (2007). The influence function for the ML estimator is unbounded in this situation. Figure 2 (and Appendix A) suggest that this issue may be relevant, even for tax data. I address this potential problem by using the ML 'Optimal b-robust estimator' (ML-OBRE) of Ronchetti and Victoria-Feser (1994). (The software implementation is by Van Kerm 2007.) The idea is to use the ML score function for most of the data (and exploit the efficiency of the ML estimator) but to place an upper limit c on the score function for high values in the interests of robustness. Ronchetti and Victoria-Feser (1994) show that, with 95% efficiency, the optimal value in the Pareto case is $c = 3$, and this is what I use. I use both ML and ML-OBRE estimators because only the former can be used for likelihood ratio tests of Pareto I versus Pareto II models. Differences between their estimates are indicative of the empirical importance of the robustness problem.

There are several estimators of the Pareto Type II model: see e.g. Singh and Guo (2009) for a review. However, ML is the most commonly used and provides consistent, efficient and asymptotically normal estimates. The software implementation is by Roodman (2015); software for an ML-OBRE estimator is not available.

Are top incomes Pareto distributed?

Researchers commonly check for Pareto properties by inspecting whether Zipf plots are linear above some income threshold (while perhaps discounting apparent non-linearity in the very highest income range given the sparsity of observations there). However, Cirillo (2013) argues persuasively that we should not check Paretianity in this way: our eyes are unreliable detectors of linearity, and what we see as linearity is also consistent with non-Pareto distributions including lognormal distributions that do not have a heavy tail. As it happens, Zipf plots for each year of SPI data do appear roughly linear above a threshold (with the exception of 1995 – see below). However, given Cirillo's critique, I relegate these plots to Appendix C.

Mean excess plots are another tool used for checking Pareto properties. They plot mean income above a threshold against a series of thresholds. For Pareto distributions, the graph is a positively-sloped straight line above some minimum income; deviations from linearity are evidence of non-Paretianity. I show mean excess plots for selected years in Figure 3, using thresholds ranging from £10,000 per year to £600,000 per year. The graphs also show pointwise 95% confidence bands. The estimates for all years are shown in Appendix D.

<Figure 3 near here>

It is difficult to draw definitive conclusions from the mean excess plots. On the one hand, the plots are roughly linear at thresholds above approximately £50,000 per year though perhaps accompanied by some small decrease in slope at extremely high thresholds. On the other hand, in every plot, confidence intervals (CIs) become very wide as the income threshold increases (there are few observations at extremely high incomes), and so it is difficult to cite non-linearities with confidence. The plot for 1995 is an exception because non-linearity is much clearer. However, this is no doubt due to the SPI discontinuities cited in the previous section. The non-linearity in the 1996 plot arises at thresholds of £300,000 or more and hence relates to a tiny number of incomes.

Cirillo (2013: 5983) also points out that mean excess plots provide a reliable means of differentiating between Pareto distributions and lognormal distributions only if the number of observations is very large (he mentions 10,000). The most reliable conclusion that we can draw from the mean excess plots (and Zipf plots) is that there is no decisive rejection of Paretianity.

Zenga curves provide a much better means for discriminating between different types of model (Cirillo 2013). A Zenga curve, $Z(u)$, is a transformation of the Lorenz curve:

$$Z(u) = \frac{u - L(u)}{u[1 - L(u)]}\ , 0 < u < 1, \qquad (3)$$

where $L(u)$ is the Lorenz curve for the distribution of incomes above a pre-specified threshold. For Pareto distributions, the Zenga curve is positively-sloped and rises as $u \to 1$ and, the higher the curve, the more heavy-tailed the distribution is. By contrast, for a lognormal distribution, the Zenga curve is horizontal. Figure 4 shows plots for 1996 and 2007 for thresholds of £60,000 and £120,000 (the higher threshold provides greater resolution over the top income range). See Appendix E for other years and thresholds.

The Zenga plots provide strong evidence in favour of Paretianity for all years (with the exception of 1995 for the reasons cited earlier.) At the same time, the location and precise shape of the curves changes over time and with threshold. This suggests that not only do Pareto tail indexes vary from year to year but also with the threshold chosen. I return to these issues below.

<Figure 4 near here>

Which distributional model for top incomes? Pareto I or Pareto II?

We cannot reliably differentiate between Pareto Type I and Type II models with these graphical checks. To do this, I use two approaches. The first is a straightforward likelihood ratio test. The second is comparisons of probability plots, specifically 'PP' plots graphing values of $p = F(x)$ predicted from each model against the values of p in the data, with a different plot for each threshold. Plots that lie wholly along the 45° line from the origin indicate perfect goodness of fit. The better fitting model is the one with less deviation from the 45° line.

Figure 5 summarizes likelihood ratio test statistics – equal to twice the difference in estimated log-likelihoods of ML-estimated Pareto I and II models – for thresholds up to £300,000 for 1996, 2001, 2007, and 2010. I cap the test statistics at 100 for plotting purposes. The dotted lines show critical values of the $\chi^2(1)$ distribution at significance levels 0.05, 0.01, and 0.001. (Plots for other years are in Appendix F.) Regardless of the critical value chosen, the findings are clear. Using a likelihood criterion, we should choose the Pareto I model over Pareto II only if the threshold used to fit the models is extremely high. For 1996, the balance in favour of Pareto II is at all thresholds below around £100,000, which lies between *p*99 and *p*99.5. For the other three years shown in Figure 5, the cut-off threshold is at the same high level or even higher, and hence above the income level at which survey non-coverage starts (Table 1, Figure 2). The plots for other years confirm this general finding.

<Figure 5 near here>

The PP plots shown in Figure 6 compare model goodness of fit over the full range of incomes above the pre-specified threshold. Plots for the Pareto I model are on the left and for the Pareto II model on the right. For brevity, I show results only for 2007 and thresholds of £60,000 and £80,000 (between p95 and *p*99 in 2007), with plots for other years and thresholds in Appendix G. The fit of each model is good: the curves shown are closer to the 45° line than most textbook illustrations of PP plots. However, there is evidence that the Pareto II model fits better than Pareto I at the lower of the two thresholds (consistent with the likelihood ratio test findings). Below the median of the left-truncated distribution, Pareto I under-predicts empirical probabilities. More evidence in favour of Pareto II is apparent for other years and thresholds (see Appendix G). Overall, probability plots provide evidence in favour of the Pareto II model over the Pareto I model, but the differences in goodness of fit are generally not large.

<Figure 6 near here>

The results from the two types of goodness of fit check suggest that the choice between Pareto models is threshold-contingent. What, then, is the optimal threshold?

What is the optimal high income threshold?

Clauset et al. (2009) and Coles (2001) review methods for determining the threshold. The most commonly-used approaches are reviews of Zipf plots or minimum excess plots, as discussed above. Another intuitively attractive approach is to plot estimated parameters against thresholds and to choose as optimal threshold, the minimum income above which the plot is horizontal. For the Pareto I model the plot is of fitted α against threshold t; for the Pareto II model, the plots are of fitted ξ and 'modified scale parameter' $\sigma^* = \sigma\xi - t$ against t (Coles 2001: 83).

Clauset et al. argue against these 'subjective' approaches and in favour of a 'more objective and principled approach based on minimizing the "distance" between the power-law model and the empirical data' (2009: 670). After reviewing alternatives, they favour measuring distance between fitted and empirical distributions using the Kolmogorov-

Smirnov (KS) statistic, i.e. the maximum distance between their cumulative distribution functions, D:

$$D = \max_{x \geq x_m}[F(x) - P(x)] \quad (3)$$

where $F(x)$ is the empirical CDF for incomes at the threshold x_m or above and $P(x)$ is the model-predicted CDF over the same range. (D is thus a numerical summary of information shown in a PP plot.) The optimal threshold is the value of x_m that minimizes D.

Figure 7 displays plots of estimated parameters against thresholds for both models, for 1996 and 2007. (Plots for other years are in Appendix H.) The vertical dashed lines show, from left to right, the percentiles $p90$, $p95$, $p99$, and $p99.5$ in the SPI data.

The figure shows that the choice of estimator matters when fitting a Pareto I model. On the one hand, the OLS estimator produces estimates of α that are distinctly smaller than those derived from ML and ML-OBRE estimators, except at extremely high thresholds. On the other hand, the ML and ML-OBRE estimates are remarkably similar.

<Figure 7 near here>

Regardless of estimator, the choice of threshold for the Pareto I model is not clear cut if the information in Figure 7 and Appendix H is used as the guide. The graphs are relatively flat only at extremely high thresholds, though the flattening out occurs at thresholds that are lower in later years – but they are very high nonetheless. The pattern for 2007 is also apparent from the start of the 2000s (Appendix H). Put differently, if we restrict the range of thresholds to between $p95$ and $p99.5$, i.e. in the range commonly used, then in 1996 the estimate of α varies between around 2.5 and 2. This is a wide range: it corresponds to Gini coefficients between 0.25 and 0.33 (according to the formula in Table 1). For 2007 and over the same range, the α estimates vary between 2.2 and 1.8, and hence Gini coefficients between 0.29 and 0.38.

In contrast, this sensitivity of parameter estimates is not apparent for the Pareto II model for thresholds in the range of $p95$ and $p99.5$. The curves are relatively flat and there is evidence for an optimal threshold lying between $p95$ and $p99$, with the precise range depending on the year.

Figure 8 displays optimal thresholds derived using the KS minimum distance criterion for both Pareto models. For the Pareto I case, the optimal thresholds are very similar for each year for ML and ML-OBRE estimators, with the exception of 1996 and 2004. It is striking that the optimal thresholds for the Pareto I model are typically much larger than those for the Pareto II model (except in 2007). For the Pareto I model, the optima are at around $p99.5$ or higher; for the Pareto II model, they are at about £50,000 which corresponds to around *p*99 in the mid- to late-1990s or *p*95 in 2000. Although there is variation in the estimated optimal threshold from year to year, there is much less variability in the optima derived for the Pareto II model than for those derived for the Pareto I model.

<Figure 8 near here>

The general lesson of this analysis is that Pareto I model estimates from top income data are sensitive to the choice of threshold, and perhaps more so than has been appreciated by researchers to date. Put differently, the range of thresholds for which the Pareto I model estimates are stable is well above the thresholds commonly used. Pareto II model estimates are more robust to the choice of threshold.

The specific lesson for applications of Approach C to determining total inequality is that estimates may be sensitive to choice of both the model of top incomes and the threshold. The criterion regarding threshold choice discussed earlier – that it should be in the income range at which survey under-coverage becomes apparent – further complicates matters. For the period considered here, this criterion implies a threshold somewhere between *p*95 and *p*99, with the former more appropriate in later years, the latter more appropriate in earlier years. This income range is broadly consistent with optimal thresholds derived for the Pareto II model but not those for the Pareto I model. In the light of these results, and in order to check the robustness of findings about overall inequality, my implementation of Approach C uses both Pareto models and multiple thresholds.

5. UK income inequality: estimates from combining estimates and combining trends

To implement Approaches C and A, we exploit the properties of inequality indices that are additively decomposable by population subgroup. For all such indices, we may write:

$$\begin{aligned}\text{Total inequality} = \ & \text{inequality among the top incomes group} \\ & + \text{inequality among the non-top incomes group} \\ & + \text{between-group inequality}\end{aligned} \qquad (4)$$

where between-group inequality is the inequality that would arise if each individual is attributed the mean of his or her income group. Additively decomposable indices include all members of the generalized entropy class I_a, including the mean logarithmic deviation (I_0 or '*L*'), the Theil index (I_1, '*T*'), and half the squared coefficient of variation (I_2, HSCV), that were cited in Table 2. The larger that a is, the more sensitive is I_a to income differences at the top of the distribution compared to the bottom. HSCV is particularly top-sensitive. Because the incomes of the top income group and the non-top income group do not overlap (by construction), the Gini coefficient is also additively decomposable in this context. For further discussion of decomposable inequality indices, see inter alia, Cowell (1980) and Cowell and Kuga (1981).

The decomposition formula for the Gini coefficient, G, derived by Atkinson (2007) and Alvaredo (2011), is also set out clearly by Cowell (2013: 43):

$$G = P_R\, S_R\, G_R \;+\; P_N\, S_N\, G_N \;+\; G_B. \tag{5}$$

P_R is the proportion of the population in the top income group ('Rich') in a given year; $P_N = 1 – P_R$ is the proportion of the population in the non-rich group; $S_R = P_R\, \mu_R\, /\mu$ and $S_N = P_N\, \mu_N\, /\mu$ are the shares in total income of each group; μ_R and μ_N are the group mean incomes; and $\mu = P_R\, \mu_R + P_N\, \mu_N$ is the overall mean. Between-group inequality $G_B = S_R – P_R$.

Pareto I and Pareto II models fitted using the same threshold and data provide different estimates of total inequality G in a given year because they imply different estimates of G_R and μ_R. (G_R and μ_R may also be estimated non-parametrically: see below.) A higher estimate of μ_R from one model implies larger S_R and G_B. That model's estimate of G will be greater as well unless the higher μ_R coincides with a sufficiently lower value of G_R. For either model, what happens to estimates of G when one changes the threshold (and thence P_R) is less clear cut because there are changes in G_N and μ_N as well as in G_R and μ_R.

The researcher has to choose the value of P_R. In the light of the analysis in previous sections, I use three thresholds for each year, *p*99, *p*95, and *p*90, estimating them non-parametrically from the *survey* data. (Although *p*90 is substantially below the thresholds discussed earlier, I include it as a robustness check; it has been used by Ruiz and Woloszko 2015.) Because the survey estimates differ from their tax data counterparts (Table 1), P_R in the *tax* data is close to but not exactly equal to 1%, 5%, or 10% respectively (see Appendix I for the values for each year). I also estimate μ_N and G_N non-parametrically from the survey data, and μ_R and G_R, L_R, and T_R from the estimates of the two Pareto models using the formulae shown in Table 1. (I report estimates for Pareto I derived using the ML-OBRE estimator.)

I calculate the combined estimate G using the formula in (5) and employ analogous steps to calculate estimates of L and T. I could not derive T for the Pareto II model (there were numerical integration problems) and I did not calculate HSCV because of its strong top-sensitivity and because the requisite moments of the fitted Pareto distribution do not always exist (Figure 7, Appendix H). Appendix I contains the estimates derived from the SPI data of the Pareto model parameters and their standard errors; μ_R, G_R, L_R, T_R and their standard errors (derived from the Pareto parameters using Table 2 formulae; and also calculated non-parametrically), plus μ_N and G_N, L_N, and T_N (derived non-parametrically from HBAI data). Appendix I also contains the combined estimates G, L and T, for all years, and for each of the three sets of estimates of mean income and inequality among the Rich. I focus discussion initially on the Pareto model-based estimates for the Gini coefficient, and later compare them with the fully non-parametric estimates, together with corresponding estimates for L and T.

Figure 9 charts the Pareto-based estimates of mean income among the Rich (μ_R), the share of total income held by the Rich (S_R), inequality among the Rich (G_R), and the overall combined estimate (G), for each of the three percentile thresholds. The Pareto I estimates are on the left; the Pareto II estimates are on the right.

The headline finding is that income inequality summarized by the Gini coefficient distinctly increased between the mid-1990s and 2007: see panel (a). It then fell back to late-1990s levels by 2010, though assessment of the fall is complicated by the forestalling issues mentioned earlier. Most of the inequality increase occurred between 2004 and 2007. These conclusions hold regardless of which Pareto model and threshold is used.

<Figure 9 near here>

Using a higher threshold leads to higher estimates of μ_R, S_R, and G_R in each year, for both Pareto models, and especially going from P_R = 5% to P_R = 1%. The S_R estimates closely track those shown in the World Top Incomes Database for the UK (based exclusively on SPI data), though there are some differences in levels (the S_R depend also on survey data).

However, when looking at overall inequality summarised by G, the Pareto II estimates are less sensitive to the choice of threshold than are the Pareto I estimates: see panel (a). Each yearly Pareto II estimate of G differs by at most one percentage point across series for the three thresholds (in the mid-1990s), and the series for P_R = 5% and P_R = 10% are virtually identical up to 2006. For the Pareto I model, the corresponding range is larger, reaching a maximum of around 2.5 percentage points (2009). Otherwise, again, the largest differences are between the series for P_R = 1% on the one hand, and P_R = 5% or 10% on the other hand. The variation in estimates relates back to the earlier findings regarding choice of the optimal threshold. The thresholds used in this section correspond to range of optimal thresholds for the Pareto II model, but well below those for the Pareto I model.

I now contrast my estimates of inequality trends derived using Approach C with estimates derived using other approaches. For brevity, I show only the results for P_R = 5%: see Figure 10. Conclusions are largely insensitive to choice of threshold in any case: for the corresponding graphs for the other two thresholds, see Appendix I.

The three series of Approach C estimates ('HBAI & SPI' variants) differ according to whether top incomes are summarised using the Pareto I or Pareto II models or non-parametrically. There are two HBAI series showing trends in inequality for the poorest 95% and the poorest 100% of the survey data, i.e. not including any information from the tax data. The HBAI-SPI series uses the estimates in the UK official income statistics derived using a variant of Approach B (the DWP's SPI adjustment, cited earlier). It is important to note that the HBAI-SPI series uses a different income definition and refer to a different population than all of the other series shown in the figure. It refers to inequality of equivalized household net income among all individuals (adults and children) rather than to individual gross income among adults. Estimates of inequality *levels* based on HBAI-SPI definitions are substantially smaller than estimates based on the tax data definition, but the differences in definitions have little effect on estimates of inequality *trends* (Burkhauser et al. 2016). The HBAI-SPI2 series is from Burkhauser et al. (2016) and uses a different variant of Approach B (see earlier). I summarise inequality not only using the Gini coefficient (panel a), but also using *L* and *T* (panels b and c). The DWP does not publish estimates for *L* or *T*: I derived them non-parametrically from public use HBAI unit record data.

<Figure 10 near here>

Figure 10 shows that if one restricts attention to the poorest 95% in the survey data each year, all three indices show a marked decline in inequality over the period as a whole, with the greatest fall being between 1998 and 1999. (These estimates are unlikely to be contaminated by the 'forestalling' issue.) Inequality also appears to be falling according to the series that uses 100% of the survey data observations (the Gini fell by around 5% between 1996 and 2007), but another distinctive feature of the series is its volatility. This is particularly acute for the Theil index, which is unsurprising because it is the most top-sensitive of the three indices. Thus Figure 10 illustrates well the sensitivity problems analysed by Cowell and Flachaire (2007) and also their conclusion that in terms of performance in finite samples there is little to choose between the Gini coefficient and the mean logarithmic deviation (*L*). But what happens if one utilizes information about top incomes from tax data?

According to all three Approach C variants, and all three inequality indices, inequality increased between 1996 and 2007. The Gini coefficient increased by around 5% according to the Pareto I estimates, by around 8% according to the Pareto II estimates, and by around 7% according to the non-parametric ones. For *L*, the corresponding increases are 1%, 5%, and 4%. For *T*, the increase in the Pareto I estimate is 24% and 33% for the non-parametric estimate. These results indicate that using a Pareto II model for top incomes leads to larger estimates of the rise in inequality over this period than does a Pareto I model, not only according to the Gini coefficient (Figure 9) but also according to *L*. In addition, the Pareto II estimates of trends in *G* and *L* are quite similar to the non-parametric ones. This is reassuring evidence for analysts that the Pareto II model provides a parsimonious but good description of distributions of top incomes. Using a more top-sensitive index (*T* rather than *L*) leads to a greater estimated increase in inequality, reflecting the marked increase is top income shares over the period.

Figure 10 also shows how the three Approach C estimates of inequality trends compare with the two Approach B series. The Burkhauser et al. (2016) HBAI-SPI2 series is very similar to the Pareto I-based Approach C series for all three inequality indices.

By contrast, trends in the DWP's official statistics series (HBAI-SPI) appear at first sight to differ markedly from those of all three Approach C estimates and for all three inequality indices. However, closer inspection of the figure reveals that the differences in trends arise almost entirely between 1996 and 1998. The official series shows a sharp increase in inequality over those two years; trends are much more similar across series in subsequent years. It is difficult to explain the sub-period inconsistency across series. One possible source is the way in which the DWP's SPI adjustment derives the cell mean estimates for top income groups. According to Department for Work and Pensions (2015*a*: 11), values to be used in year *t* of the HBAI data are derived by HMRC statisticians by 'projections' from SPI data for year *t*–1. (because year *t*'s SPI data are not yet available). No further details of the projection method are given. By contrast, all my Approach C estimates and Burkhauser et al.'s (2016) Approach B estimates combine information HBAI data for year *t* and SPI data for year *t*. I conjecture that a larger than usual difference between projected cell-means and actual out-turns for 1996–1998 are the source of the inconsistency observed for that period. The public-use SPI data do not contain the variables that would allow me to check if this is the case.

In sum, apart from the exceptional and short sub-period just discussed, there is substantial consistency across the different methods for combining survey data and tax data about top incomes. Compared to the estimates that are wholly survey-based, all show a rise in income inequality over the decade prior to the onset of the Great Recession, whereas the estimates that are wholly survey-based show no increase in inequality.

6. Summary and conclusions

Statistical agencies and other researchers typically estimate income inequality levels and trends from either survey data or tax return data, but rarely combine the information in the two types of data source. The result is that very different impressions about how inequality is changing over time may arise, as the examples for the UK and the USA in the Introduction show. Research users may reasonably ask what the 'true' picture of inequality trends is. There is a good case for providing them with answers using methods that combine information from survey and tax data in order to take advantage of the strengths of each source. In particular, tax return data provide better coverage of top incomes than do survey data; and survey data provide the ability to create income variables with the same definitions, so that combination is done on a like-for-like basis.

I have analysed income inequality levels and trends for the UK by combining inequality estimates from survey and tax data (Approach C), contrasting these estimates with those derived by combining data per se (Approach B). As part of this analysis, I have also provided new findings about survey under-coverage of top incomes in UK survey data (Section 3). The problem becomes apparent at around the 99th percentile in the 1990s but at around the 95th percentile in the 2000s.

I have found that that conclusions about aggregate UK inequality trends since the mid-1990s are broadly robust to the way in which tax data are employed in Approach C. One may conclude for example that the Gini coefficient for gross individual income rose by around 7% to 8% between 1996/97 and 2007/08, with most of the increase occurring after 2003/04. When I use only survey data, with tax data not exploited at all, the Gini is estimated to decrease by around 5% over the same period.

The result that combining information about top incomes from tax data with information about the rest of the distribution in survey data leads to an estimated increase in inequality is unsurprising given knowledge of survey under-coverage of top incomes and the marked rise in top income shares in the UK over the last two decades. But I have shown how we may go beyond the qualitative conjecture and provide specific quantitative estimates of inequality trends, and for a range of inequality indices.

Indeed, the analysis highlights the continuing importance of normative judgements for inequality analysis. Different inequality indices incorporate different assumptions about how to evaluate income differences in different parts of the income distribution (Cowell 1977, 2011). A focus on the income share of the top 1% measure means that zero weight is placed on income differences among the poorest 99% or on differences between the rich and non-rich groups. This paper has considered inequality indices that give a non-zero weight to everyone. It is because of this, and because the rich are assumed to form such small proportion of the population, that I estimate the increase in UK income inequality over the last two decades to be substantially smaller than the rise in the income share of the top 1% whether shown by the tax data alone or by the combined estimate (Figure 9).

The portfolio of inequality indices is also constrained by practical considerations: the indices used cannot be too top-sensitive. Application of the semiparametric Approach C is problematic if the distribution of top incomes is particularly heavy-tailed. Various moments of the fitted Pareto distributions do not exist in this case, and hence nor do many top-sensitive inequality indices (Figure 7; Appendix H). Cowell and Flachaire (2007) make this argument in the context of Approach A; I have shown that it also applies even if one uses tax data to describe top incomes. One might instead consider implementing Approach C using non-parametric estimates of top-sensitive inequality measures for the top income group, but I have found that such estimates are non-robust and volatile in the sense described in the Introduction, even using SPI tax data rather than HBAI survey data. (See the nonparametric estimates of the HSCV for the rich that are reported in Appendix I.)

In this paper, I have focused on inequality estimation issues related to data quality and ignored issues of statistical significance (as has virtually all previous work using Approaches B and C). It is relatively straightforward to estimate standard errors or the various elements in the inequality decomposition equation (1) using standard asymptotic formulae (these estimates are provided in Appendix I). However, there are non-trivial challenges to overcome in providing reliable inference for the overall inequality estimate. Cowell and Flachaire (2007, especially Section 3.3) discuss these issues with reference to generalised entropy inequality indices. The case of the Gini coefficient appears not to have been discussed in the literature to date.

As part of deriving the substantive results about inequality levels and trends, this paper has also provided new evidence about which Pareto model to fit to the upper tail of a heavy-tailed distribution, and which threshold to use when doing this. Although the application has been to income, the analysis should be of broad interest because Pareto distributions are commonly used in many other contexts,. In his recent review of power laws in economics and finance, Gabaix argued that 'the Pareto law has survived the test of time: It fits still quite well. The extra degree of freedom allowed by a lognormal might be a distraction from the essence of the phenomenon' (2009: 285). He might have substituted 'Pareto II' for 'lognormal'. My analysis has shown that there is a good case for exploiting the extra degree of freedom provided by the Pareto II model, especially given the top income thresholds that are typically used (*p*99 or less). Put differently, the Pareto I model is as good as Pareto II only at extremely high incomes, beyond the range of thresholds usually considered. My conclusions refer to income rather than other variables such as wealth or city and firm size, and to the UK rather than to other countries, so checking the robustness of my findings in other contexts would be a useful topic for future research.

References

Alfons, A., Templ, M. and Filzmoser, P. (2013). Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Journal of the Royal Statistical Society, Series C*, 62 (2), 271–286.

Alvaredo, F. (2011). A note on the relationship between top income shares and the Gini coefficient. *Economic Letters*, 110 (3), 274–277.

Alvaredo, F., Atkinson, A. B., Piketty, T., and Saez, E. (2013). The top 1 percent in international and historical perspective. *Journal of Economic Perspectives*, 27 (3), 3–20.

Alvaredo, F., Atkinson, A. B., Piketty, T., Saez, E. (2015). The World Top Incomes Database. http://g-mond.parisschoolofeconomics.eu/topincomes/.

Alvaredo, F. and Londoño Vélez, J. (2013). High incomes and personal taxation in a developing economy: Colombia 1993–2010. CEQ Working Paper 12. New Orleans, LA: CEQ Institute, Tulane University. http://www.commitmentoequity.org/publications_files/CEQWPNo12%20HighTaxationDevEconColombia1993-2010_19March2013.pdf

Anand, S. and Segal, P. (2016). Who are the global top 1%? International Development Institute Working Paper 2016-02. London: Kings College London. http://www.kcl.ac.uk/sspp/sga/idi/Research/IDI-Working-Papers.aspx

Arnold, B. C. (2008). Pareto and generalized Pareto distributions. In: D. Chotikapanich (ed.), *Modeling Income Distributions and Lorenz Curves*. New York: Springer Science+Business Media.

Arnold, B. C. (2015). *Pareto Distributions*, second edition. Boca Raton, FL: Chapman and Hall/CRC.

Atkinson, A. B. (2005). Top incomes in the UK over the twentieth century. *Journal of the Royal Statistical Society, Series A*, 168 (2), 325–343.

Atkinson, A. B. (2007). Measuring top incomes: methodological issues. In: A. B. Atkinson and T. Piketty (eds), *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*. Oxford: Oxford University Press.

Atkinson, A. B. (2016). Pareto and the upper tail of the income distribution in the UK: 1799 to the present. Unpublished paper, Nuffield College, Oxford. Forthcoming in *Economica* special issue in honour of Frank Cowell.

Atkinson, A. B., and Piketty, T. (eds). (2007). *Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*. Oxford: Oxford University Press.

Atkinson, A. B., Piketty, T., and Saez, E. (2011). Top incomes in the long run of history. *Journal of Economic Literature*, 49 (1), 3–71.

Bach, S., Corneo, G., and Steiner, V. (2009). From bottom to top. The entire income distribution in Germany, 1992–2003. *Review of Income and Wealth*, 55 (2), 303–330.

Belfield, C., Cribb, J., Hood, A., and Joyce, R. (2015). *Living Standards, Poverty and Inequality in the UK: 2015*. Report 107. London: Institute for Fiscal Studies. http://www.ifs.org.uk/publications/7878

Burkhauser, R.V., Feng, S., Jenkins, S. P., and Larrimore, J. (2012). Recent trends in top income shares in the USA: reconciling estimates from March CPS and IRS tax return data, *Review of Economics and Statistics*, 94 (2), 371–388.

Burkhauser, R. V., Hérault, N., Jenkins, S. P., and Wilkins, R. (2016). What has been happening to UK income inequality since the mid-1990s? Answers from reconciled and combined household survey and tax return data. ISER Working Paper 2016-03.

Colchester: Institute for Social and Economic Research.
https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2016-03

Cirillo, P. (2013). Are your data really Pareto distributed? *Physica A*, 392, 5947–5962.

Clauset, A., Shaliza, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51 (4), 661–703.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: London.

Cowell, F. A. (1977). *Measuring Inequality*. Oxford: Philip Allan.

Cowell, F. A. (1980). On the structure of additive inequality measures. *Review of Economic Studies*, 47, 521–531.

Cowell, F. A. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics*, 42, 27–41.

Cowell, F. A. (2011). *Measuring Inequality*, third edition. Oxford: Oxford University Press.

Cowell, F. A. (2013). UK wealth inequality in international context. In: J. Hills, F. Bastagli, F. Cowell, H. Glennerster, E. Karagiannaki, and A. McKnight, *Wealth in the UK, Distribution, Accumulation, and Policy*. Oxford: Oxford University Press.

Cowell, F. A., Ferreira, F. H. G., and Litchfield, J. A. (1998). Income distribution in Brazil 1981–1990. Parametric and non-parametric approaches. *Journal of Income Distribution*, 8, 63–76.

Cowell, F. A. and Flachaire, E. (2007). Income distribution and inequality measurement: the problem of extreme values. *Journal of Econometrics*, 141, 1044–1072.

Cowell, F. A. and Kuga, K. (1981). Inequality measurement: an axiomatic approach. *European Economic Review*, 15, 287–305.

Cowell, F. A. and Victoria-Feser, M.-P. (1996). Robustness properties of inequality measures. *Econometrica*, 64 (1), 77–101.

Cowell, F. A. and Victoria-Feser, M.-P. (2007). Robust stochastic dominance: A semi-parametric approach. *Journal of Economic Inequality*, 5, 21–37.

Department for Work and Pensions (2013). *Households Below Average Income, 1994/95-2012/13* [computer file]. *7th Edition.* Colchester, Essex: UK Data Archive [distributor], October 2013. SN: 5828.

Department for Work and Pensions (2015). *Households Below Average Income An Analysis of the Income Distribution 1994/95 – 2013/14*. London: Department for Work and Pensions.

Department for Work and Pensions (2015*a*). *Households Below Average Income. Quality and Methodology Information Report – 2013/14*. London: Department for Work and Pensions.

Department for Work and Pensions, National Centre for Social Research, and Office for National Statistics Social and Vital Statistics Division (2014), *Family Resources Survey, 2010-2011* [computer file]. *2nd Edition.* Colchester, Essex: UK Data Archive [distributor], October 2014. SN: 7085, http://dx.doi.org/10.5255/UKDA-SN-7085-2

Diaz-Bazan, T. (2015). Measuring inequality from top to bottom. World Bank Policy Research Paper 7237. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/2015/04/24353317/measuring-inequality-top-bottom

Eeckhout, J. (2004). Gibrat's law for (all) cities. *American Economic Review*, 94 (5), 1429–1451.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1, 255–293.

Gabaix, X. (2016). Power laws in Economics: an introduction. *Journal of Economic Perspectives*, 30 (1), 185–206.

Harrison, A. (1979). The upper tail of the earnings distribution. Pareto or lognormal? *Economic Letters*, 2, 191–195.

Harrison, A. (1981). Earnings by size: a tail of two distributions. *Review of Economic Studies*, 48 (4), 621–631.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3 (5), 1163–1174.

HM Revenue and Customs (2012). The Exchequer effect of the 50 per cent additional rate of income tax. London: HMRC.

HM Revenue and Customs KAI Data, Policy and Co-ordination (2014), *Survey of Personal Incomes, 2010-2011: Public Use Tape* [computer file]. Colchester, Essex: UK Data Archive [distributor], November 2014. SN: 7569, http://dx.doi.org/10.5255/UKDA-SN-7569-1.

Jenkins, S. P. and Van Kerm, P. (2015). PARETOFIT: Stata module to fit a Type 1 Pareto distribution. Statistical Software Components archive. Boston, MA: Boston College. https://ideas.repec.org/c/boc/bocode/s456832.html

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

Lakner, C. and Milanovic, B. (2016). Global income distribution. From the fall of the Berlin Wall to the Great Recession. *World Bank Economic Review*, 30 (2), 203–232.

OECD (2008). *Growing Unequal? Income Distribution and Poverty in OECD Countries*. Paris: OECD Publishing.

OECD (2011). *Divided We Stand: Why Inequality Keeps Rising*. Paris: OECD Publishing.

OECD (2015). *In It Together: Why Less Inequality Benefits All*. Paris: OECD Publishing.

Roodman, D. (2016). EXTREME: Stata module to fit models used in univariate extreme value theory. Statistical Software Components archive. Boston, MA: Boston College. https://ideas.repec.org/c/boc/bocode/s457953.html

Ruiz, N. and Woloszko, N. (2015). What do household surveys suggest about the top 1% incomes and inequality in OECD countries? Economics Department Working Paper No. 1265. Paris: OECD. http://dx.doi.org/10.1787/5jrs556f36zt-en

Shorrocks, A. F., Davies, J. and Lluberas, R. (2015). *Credit Suisse Global Wealth Databook 2015*. Zurich: Credit Suisse Institute.

Singh, V. P. and Guo, H. (1995). Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME). *Hydrological Sciences Journal*, 40 (2), 165–181.

Van Kerm, P. (2007). Extreme values and robust distribution analysis. Presentation at 13th UK Stata User Group, London. http://www.stata.com/meeting/13uk/vankerm-uksug_slides.pdf

Vermeulen, P. (2014). How fat is the top tail of the wealth distribution? ECB Working Paper 1962. Frankfurt am Main: European Central Bank. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1692.pdf

Victoria-Feser, M.-P. and Ronchetti, E. (1994). Robust methods for personal income distribution models. *Canadian Journal of Statistics*, 22 (2), 247–258.

Table 1. Percentiles of individual gross income (£ p.a., 2012/13 prices), survey and tax data estimates

Year	*p*90			*p*95			*p*99			*p*99.5		
	HBAI	SPI	Ratio	HBAI	SPI	Ratio	HBAI	SPI	Ratio	HBAI	SPI	Ratio
1995	35,551	30,964	115	45,602	40,056	114	83,362	78,340	106	107,080	105,630	101
1996	37,028	31,209	119	47,623	41,043	116	85,141	80,869	105	109,948	113,469	97
1997	37,317	32,381	115	48,084	42,133	114	89,995	86,330	104	122,952	122,466	100
1998	38,285	33,595	114	50,485	44,541	113	95,853	92,694	103	127,830	131,499	97
1999	39,075	35,333	111	51,065	46,776	109	96,053	98,250	98	126,214	138,335	91
2000	40,371	37,273	108	53,048	49,786	107	106,413	105,722	101	142,586	150,871	95
2001	41,854	38,370	109	54,721	51,117	107	105,741	108,620	97	140,193	153,585	91
2002	41,360	38,498	107	54,401	51,370	106	103,532	108,742	95	136,152	153,379	89
2003	41,932	38,174	110	54,644	50,860	107	100,354	108,587	92	135,495	153,474	88
2004	42,460	40,575	105	55,847	53,533	104	107,367	114,855	93	144,244	165,249	87
2005	42,691	42,191	101	56,158	56,172	100	108,183	125,574	86	140,916	183,723	77
2006	43,335	42,885	101	56,101	57,369	98	109,910	131,012	84	157,894	193,230	82
2007	42,735	43,994	97	56,007	59,149	95	111,282	136,392	82	156,536	201,972	78
2008	43,312			56,703			109,812			143,624		
2009	42,564	41,917	102	56,839	55,782	102	113,420	124,753	91	156,384	183,777	85
2010	41,036	40,690	101	54,377	53,777	101	106,812	114,368	93	147,956	163,524	90

Notes. Author’s estimates from HBAI (survey) and SPI (tax) data. Years refer to fiscal years (‘1995’ means 1995/96, and so on). No SPI unit record data have been released for 2008. Ratio: ratio of HBAI estimate to SPI estimate, in percent.

Table 2. Pareto Type I and Type II models: means and inequality indices

Statistic	Pareto Type I	Pareto Type II
Mean	$\frac{\alpha x_m}{\alpha - 1}, \alpha > 1$	$\mu + \left(\frac{\sigma}{1-\xi}\right), \xi < 1$
Gini coefficient (G)	$\frac{1}{2\alpha - 1}, \alpha > 1$	$1 - \left[\frac{\mu + 2\left(\frac{\sigma}{\xi^2}\right) B\left(\frac{2-\xi}{\xi}, 2\right)}{\mu + \left(\frac{\sigma}{\xi^2}\right) B\left(\frac{1-\xi}{\xi}, 2\right)}\right], \xi < 1$
Mean logarithmic deviation (L)	$\log\left(\frac{\alpha}{\alpha - 1}\right) - \left(\frac{1}{\alpha}\right), \alpha > 1$	No closed form expression
Theil index (T)	$\left(\frac{1}{\alpha - 1}\right) - \log\left(\frac{\alpha}{\alpha - 1}\right), \alpha > 1$	No closed form expression
Half the coefficient of variation squared (HSCV)	$\frac{1}{2\alpha(\alpha - 2)}, \alpha > 2$	$\frac{\sigma^2}{2(1-2\xi)[\mu(1-\xi)+\sigma]^2}, \xi < \frac{1}{2}$

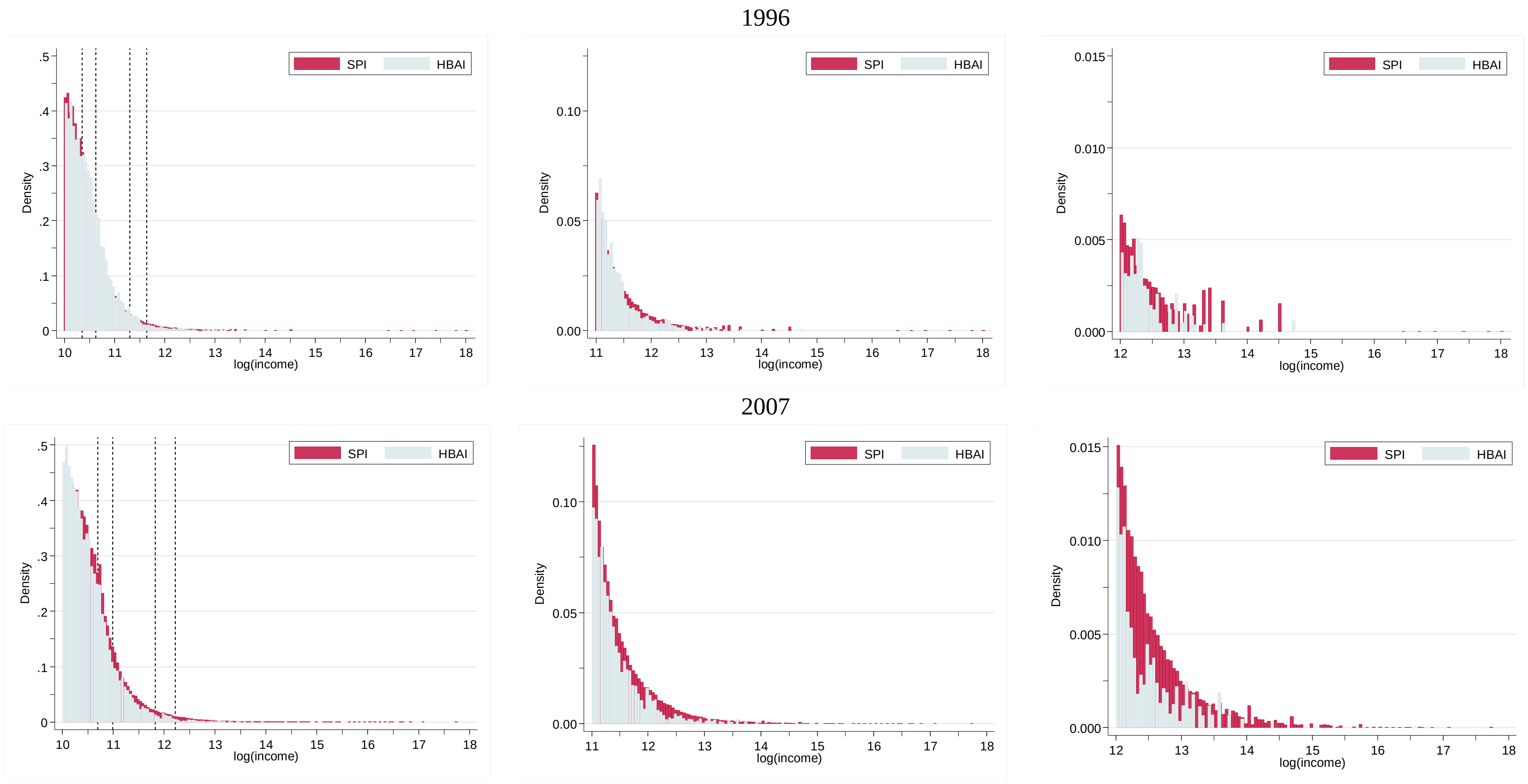
Notes. For the formulae for the survivor functions of the Pareto I and II models, see the main text. $B(.)$ is the Beta distribution. Sources for formulae: Arnold (2008), Cowell (2007), Kleiber and Kotz (2003), and Singh and Guo (1995). L, T, and HSCV are members of the generalized entropy family of inequality indices, $I(a)$, with a = 0, 1, and 2 respectively. Values of the L and T for the Pareto II distribution may be derived by numerical integration using the formulae for generalized moments in Cowell (1989), if the relevant moments exist.

Figure 1. Estimating inequality: approaches to addressing top income issues in survey data

Approach		Survey data		Tax data	Examples
A.	Semi-parametric: combine summary measures derived from within the survey data	Gini (& other measures) for poorest (1–p)% + Pareto-estimated Gini (& other measures) for richest p% ↓ Combined Gini & other measures		Not used	Alfons et al. (2013), Burkhauser et al. (2012), Cowell and Flachaire (2007), Ruiz and Woloszko (2015)
B.	Non-parametric: tax data cell-means replace top incomes in survey data	Survey ↓ Gini & other measures	←	Tax data	Bach et al. (2009), Burkhauser et al. (2016), Department for Work and Pensions (2015)
C.	Semi-parametric: combine inequality indices derived from survey and tax data	Survey data ↓ Gini (& other measures) for poorest (1–p)% ↓		Tax data ↓ Pareto-estimated Gini (& other measures) for richest p% ↓	Atkinson (2007), Atkinson et al. (2011), Alvaredo (2011), Alvaredo and Londoño Vélez (2015), Diaz-Bazan (2015), Anand and Segal (2016), Lakner and Milanovic (2016), this paper
		Combined Gini (& other measures)			

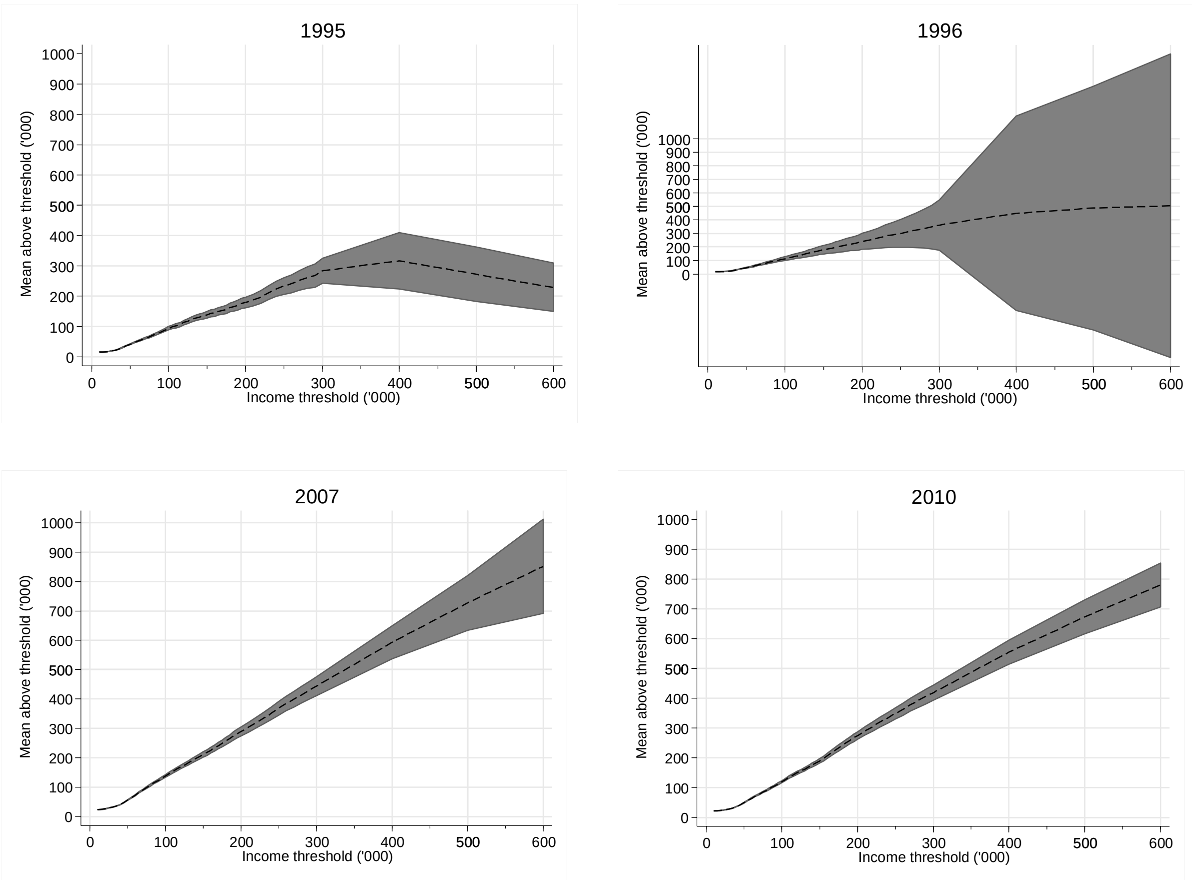
Note. Approach C may also be implemented using non-parametric estimates from tax data: see main text.

Figure 2. The concentration of income in high and extremely high income ranges: survey and tax return data compared, 1996 and 2007



Notes. Author's estimates from SPI (tax) and HBAI (survey) data. Income is in £ per year, 2012/13 prices. Vertical dashed lines show (from left to right) *p*90, *p*95, *p*99, *p*99.5. For plots for other years, see Appendix A.

Figure 3. Mean excess plots for top incomes, tax return data, by year

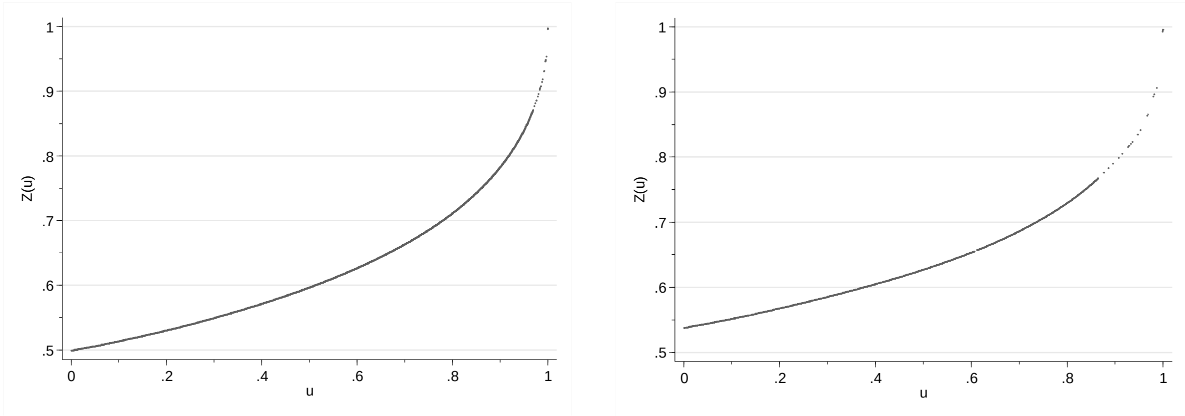


Notes. Author's estimates from SPI data. For plots for other years, see Appendix C. The shaded areas represent pointwise 95% confidence bands. Thresholds are in £ per year, 2012/13 prices. Plots estimated at intervals of £5,000 for thresholds between £10,000 and £200,000, £10,000 between £210,000 and £300,000, and £100,000 thereafter.

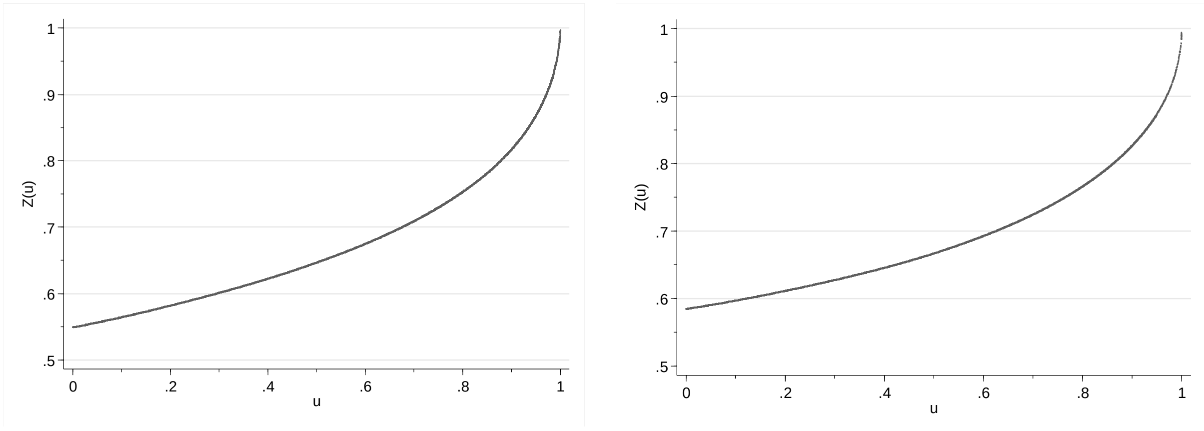
Figure 4. Zenga plots for top incomes, tax return data, by threshold and year

Threshold = £60,000 p.a. | Threshold = £120,000 p.a.

1996

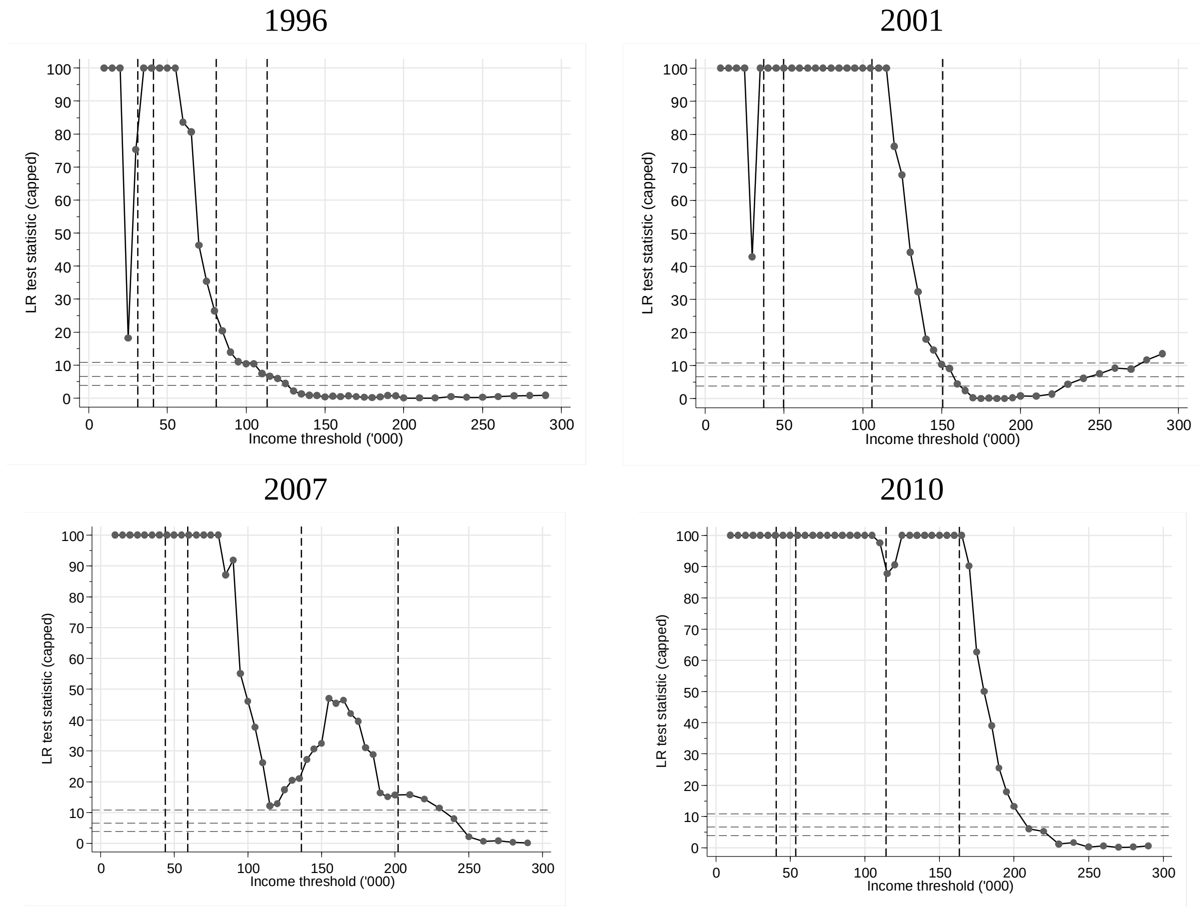


2007



Notes. Author's estimates from SPI data. For plots for other years and thresholds, see Appendix D. On the Zenga plot, see the main text and Cirillo (2013).

Figure 5. Likelihood ratio test statistics (Pareto I versus Pareto II), by threshold: tax return data for 1996, 2001, 2007, and 2010



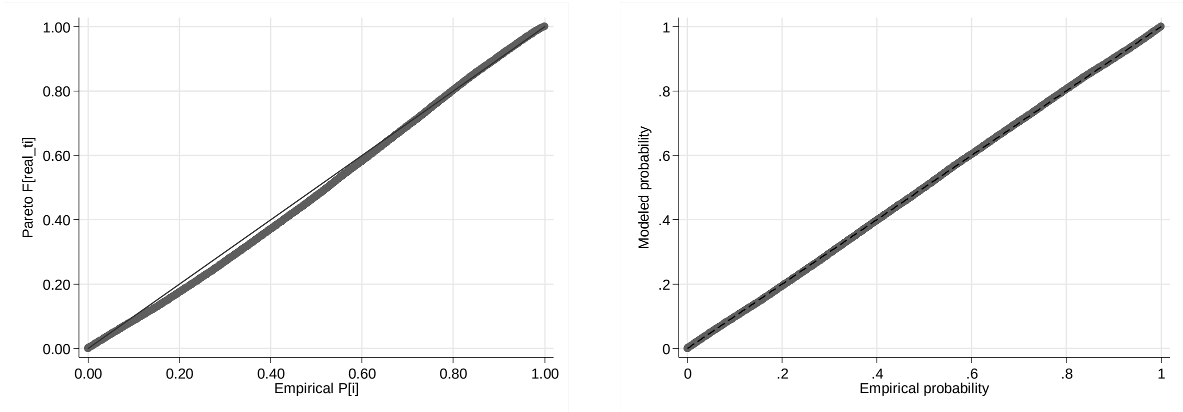
Notes. Author’s estimates from SPI data. The figures plot twice the difference in log-likelihood for Pareto I and II models (each fitted using ML). Test statistics are capped at 100 for plotting purposes. Dotted horizontal lines show critical values of the $\chi^2(1)$ distribution at significance levels 0.05, 0.01, and 0.001. Vertical dashed lines show (from left to right) *p*90, *p*95, *p*99, *p*99.5. For plots for other years, see Appendix E.

Figure 6. PP plots for top incomes, by threshold: tax return data for 2007

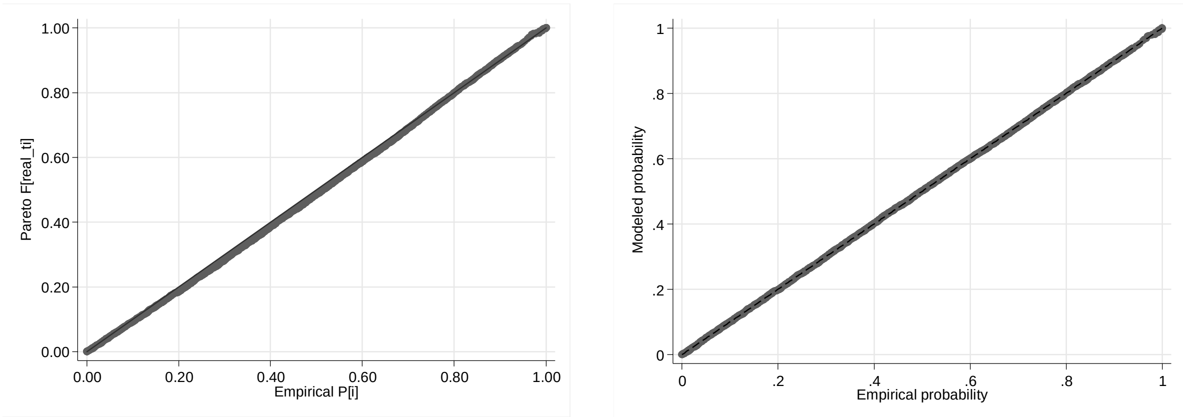
Pareto I

Pareto II

Threshold = £60,000 p.a.



Threshold = £80,000 p.a.

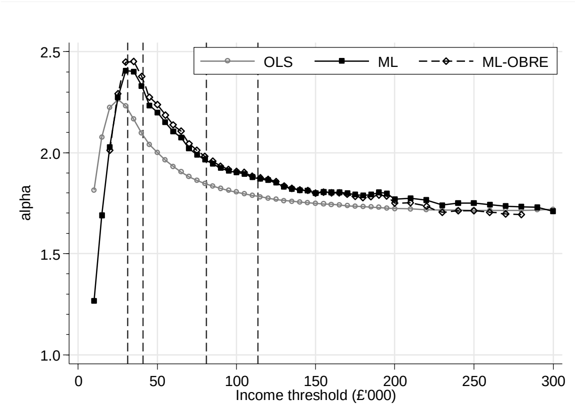


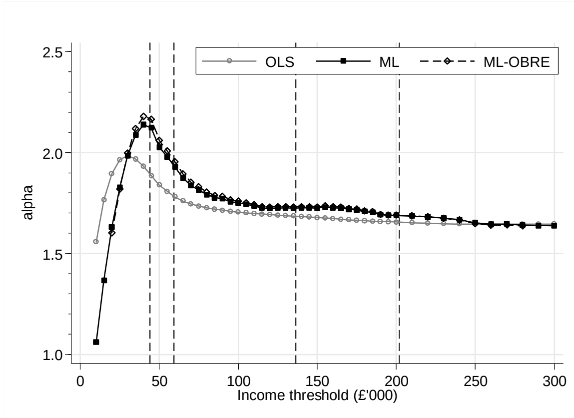
Notes. Author's estimates from SPI data. The charts plot modelled (cumulative) probabilities against empirical probabilities: see text. For plots for other years and thresholds, see the Appendix F. ML estimator used for both models.

Figure 7. Pareto I and II parameter estimates, by threshold, tax return data for 1996 and 2007

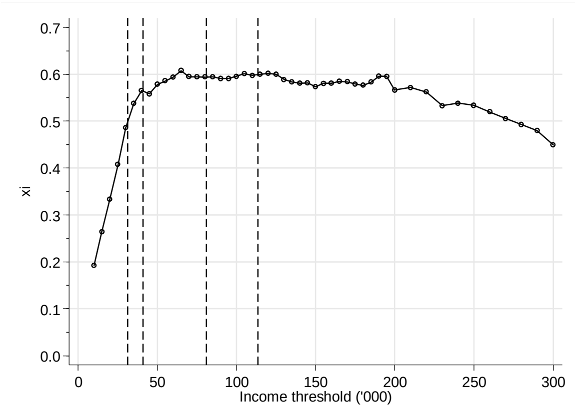
1996 2007

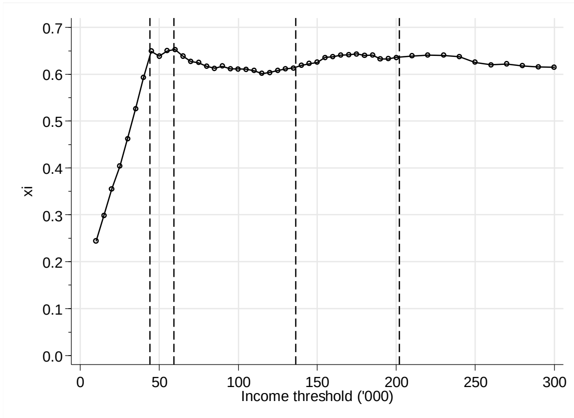
Pareto I model, shape parameter α



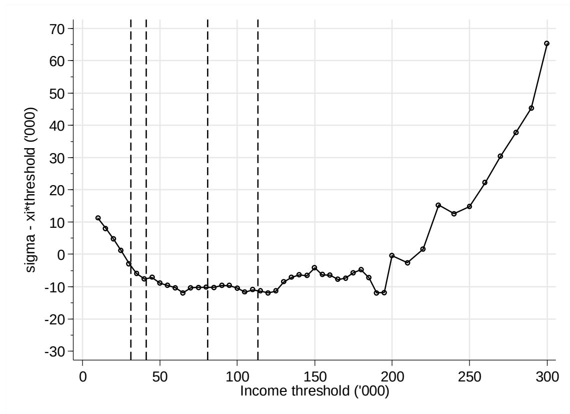


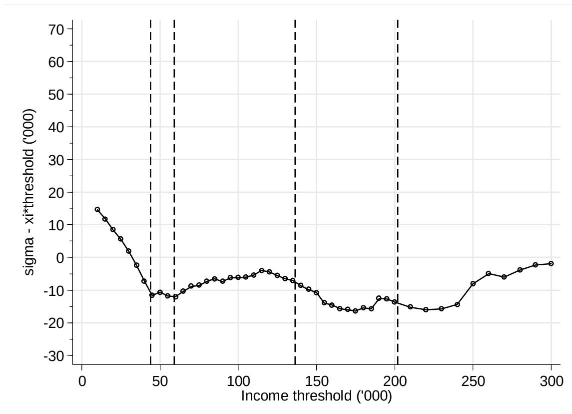
Pareto II model, shape parameter ξ





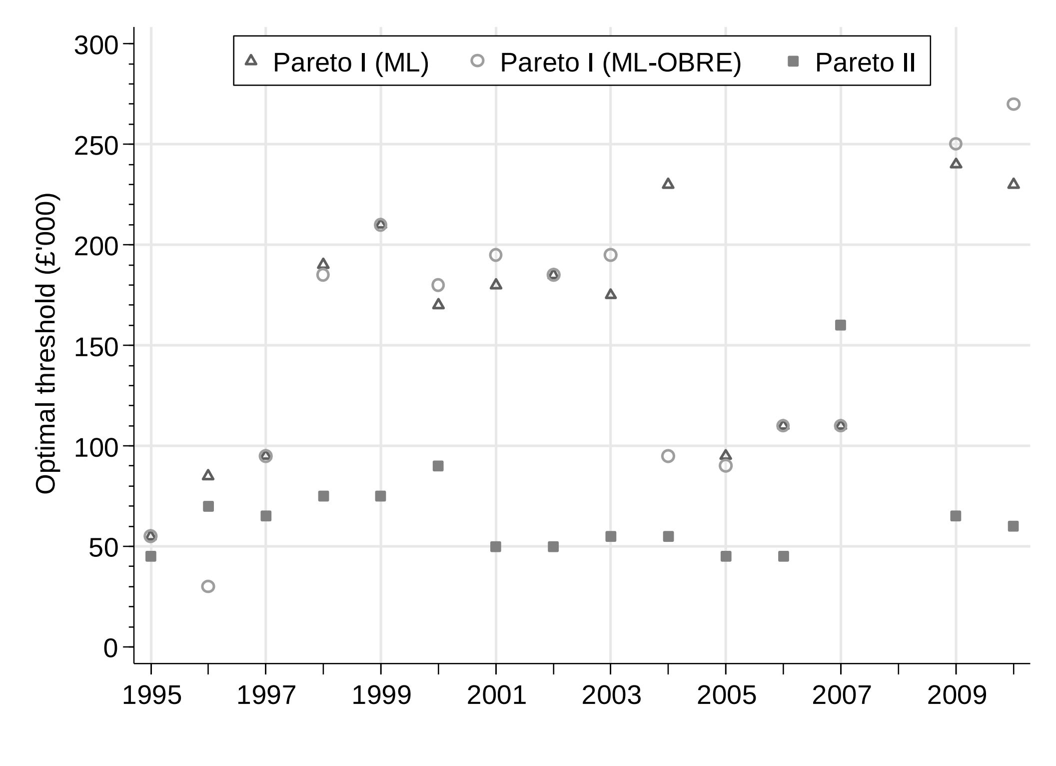
Pareto II model, modified scale parameter $\sigma^* = \sigma - \xi t$





Notes. Author's estimates from SPI data. Vertical dashed lines show (from left to right) *p*90, *p*95, *p*99, *p*99.5. For plots for other years, see Appendix G.

Figure 8. Optimal Pareto threshold (KS criterion), tax return data, by estimator and year

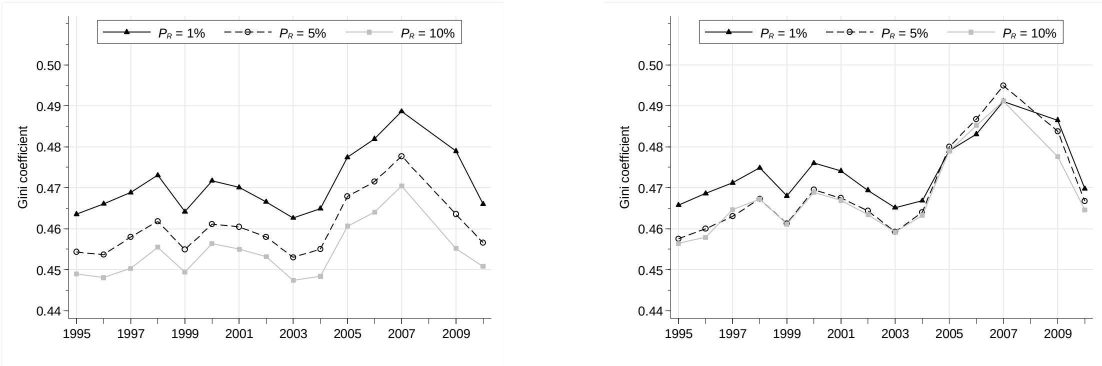


Notes. Author's estimates from SPI data. The figure plots the thresholds selected using the Kolmogorov-Smirnov criterion described in eq. (3) and main text.

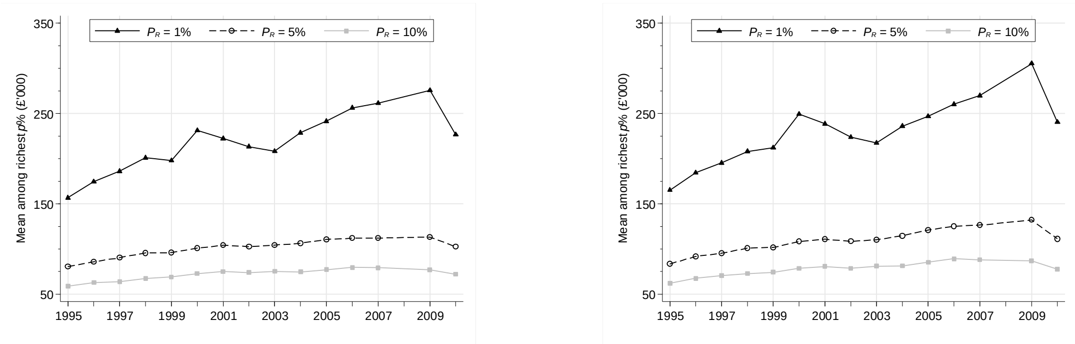
Figure 9. Combined data estimates (Approach C): Gini coefficient overall, mean income, income share, and Gini coefficient of the Rich, by Pareto model and high-income threshold

Pareto I estimates | Pareto II estimates

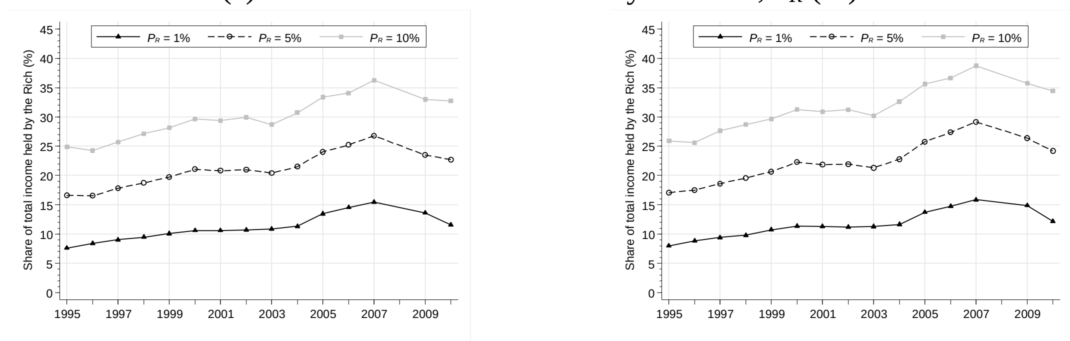
(a) Gini coefficient (all adults, Rich and Non-rich), *G*



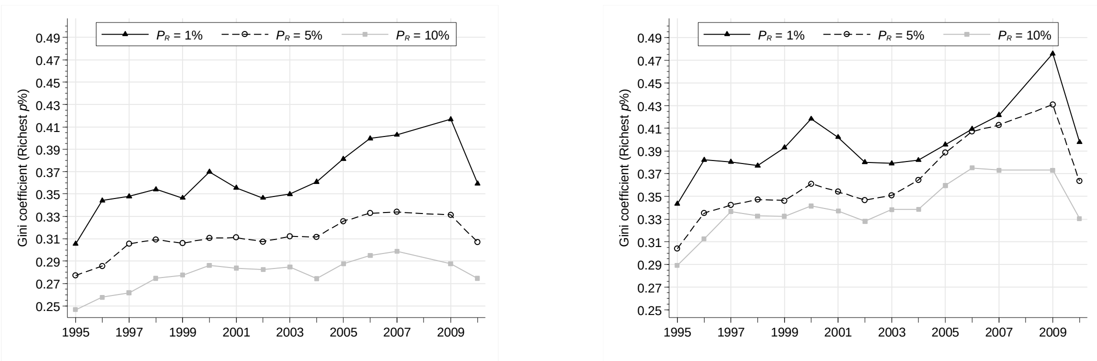
(b) Mean income of the Rich, μ_R



(c) Share of total income held by the Rich, S_R (%)



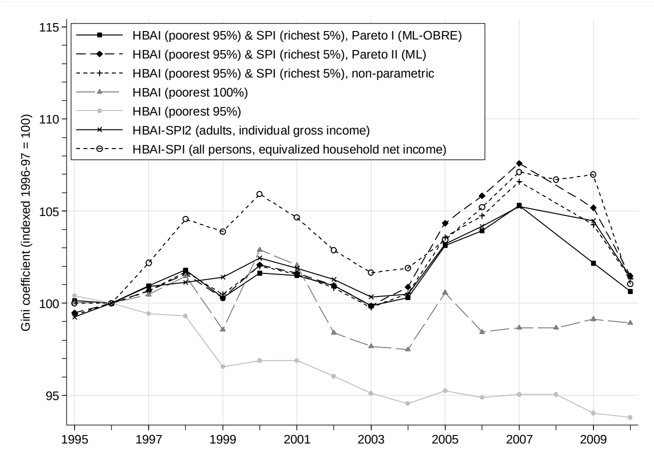
(d) Gini coefficient among the Rich, G_R



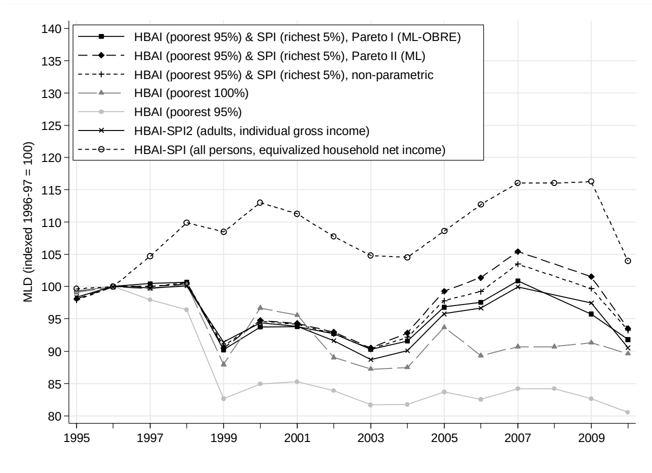
Notes. Author's derivations from SPI and HBAI data using eqn. (5). Pareto I (ML-OBRE) and II models fitted using each of three thresholds to define the Rich group: P_R = 1%, 5%, and 10% (cut-offs derived from survey data – see text for further explanation). All series refer to distributions of individual gross income among adults. Estimates of μ_R, μ_N, S_R, S_N, G_R, and G_N are listed in Appendix H.

Figure 10. UK inequality (indexed 1996 = 100), by series and inequality index

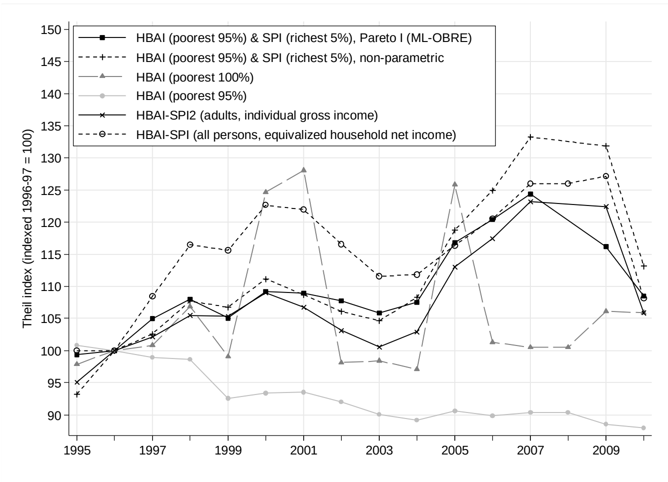
(a) Gini coefficient (G)



(b) Mean logarithmic deviation (L)



(c) Theil index (T)



Notes. Author's derivations from SPI, HBAI, HBAI-SPI (Department for Work and Pensions 2015), and HBAI-SPI2 data (Burkhauser et al. 2016). All series shown are based on the distribution of individual gross income among adults, with the exception of the HBAI-SPI series which refers to equivalized household net income among all individuals (see main text). There are no Pareto II estimates for the Theil index. Threshold: p95 in the HBAI data (see main text). The corresponding graphs for the other two percentile thresholds are shown in Appendix I.