Better off? Distributional comparisons for ordinal data about personal well-being

Stephen P. Jenkins

Published online: 07 Dec 2019.

Submit your article to this journal

Article views: 55

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

Better off? Distributional comparisons for ordinal data about personal well-being

Stephen P. Jenkins[a,b,c,d]

[a]London School of Economics and Political Science, London, UK; [b]ISER, University of Essex, Colchester, UK; [c]MIAESR, University of Melbourne, Melbourne, Australia; [d]IZA, Bonn, Germany

ABSTRACT

How to undertake distributional comparisons when personal well-being is measured using income is well-established. But what if personal well-being is measured using subjective well-being indicators such as life satisfaction or self-assessed health status? Has average well-being increased or well-being inequality decreased? How does the distribution of well-being in New Zealand compare with that in Australia, or between young and old people in New Zealand? This paper addresses questions such as these, stimulated by the increasing weight put on subjective well-being measures by international agencies such as the OECD and national governments including New Zealand's. The paper reviews the methods appropriate for distributional comparisons in the ordinal data context, comparing them with those routinely used for comparisons of income distributions. The methods are illustrated using data from the World Values Survey.

1. Introduction

How to undertake distributional comparisons when personal well-being is measured using income is well-established. But what if personal well-being is measured using subjective well-being indicators such as life satisfaction or self-assessed health status? Has average well-being increased, and what has happened to the inequality of well-being? How does the distribution of well-being in New Zealand compare with that in Australia, or between young and old people and men and women? This paper reviews and illustrates the methods for addressing questions such as these, stimulated by the increasing weight put on subjective well-being measures by international agencies such as the OECD and national governments including New Zealand's.

A catalyst for the new emphasis on subjective well-being was the *Report by the Commission on the Measurement of Economic Performance and Social Progress* (Stiglitz, Sen, & Fitoussi, 2009), which set out a comprehensive agenda for going 'Beyond GDP'. The Report's Quality of Life sections emphasize that 'well-being is multidimensional' (2009, p. 14), and that 'objective and subjective dimensions of well-being are both important' (2009, p. 16). The OECD has played an important role in implementing the Report's recommendations in this area, launching its Better Life Initiative (in 2011), regularly reporting on well-being outcomes (*How's Life*; see e.g. OECD, 2017), and developing the Better life Index and multiple online resources (see https://www.oecd.org/statistics/better-life-initiative.htm). In parallel, the national statistical agencies of OECD member countries have also introduced initiatives to address the Beyond GDP agenda, including a greater emphasis on collection of and reporting

---

CONTACT Stephen P. Jenkins ✉ s.jenkins@lse.ac.uk 🖃 Department of Social Policy, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

on subjective well-being data. For a recent review of initiatives to date, see the symposium on 'New measures of well-being: perspectives from statistical offices' in the March 2015 issue of the *Review of Income and Wealth*. Subjective well-being measures continue to be highlighted by high-level groups taking forward the Stiglitz-Sen-Fitoussi commission's work: see e.g. Stiglitz, Fitoussi, and Durand (2018, especially section 3.7).

New Zealand has embraced the Beyond GDP agenda too. New Zealand Treasury began developing its Living Standards Framework in 2011 and has continued to refine it, drawing heavily on the OECD's Better Life approach (New Zealand Treasury, 2018a). Again, the motivation is that there is 'more to wellbeing than just a healthy economy' (New Zealand Treasury, 2018b). New Zealand Treasury assesses current well-being across twelve domains of people's lives, ranging from their civic engagement and governance through to the quality and quantity of their leisure and recreation time, including income and consumption and also subjective well-being. A Living Standards Dashboard, consistent with the OECD's dashboard, summarizes the twelve domains using a large number of indicators, including life satisfaction and self-assessed health status. For a recent overview of New Zealand's situation relative to other OECD countries using this approach, see OECD (2017). More recently, Statistics New Zealand has been developing a new set of indicators for measuring social progress. The 'Indicators Aotearoa New Zealand' are intended to support various cross-government initiatives and international reporting, and to align with those in the Treasury's Living Standards Framework and the UN's Sustainable Development Goals (Statistics New Zealand, 2019).

These sets of indicators are not only for ex post monitoring of social progress but also for improving strategic decision-making in the realms of economic and social policy (Statistics New Zealand, 2018). The aim is to have an integrated and coordinated approach linking government (setting priorities and making policy decisions) and the Treasury and Statistics New Zealand (providing analysis and advice, information and data). This brings us to New Zealand's first 'Well-Being Budget'. According to the Minister of Finance,

> [this] is a new approach to how government works, . . . placing the wellbeing of New Zealanders at the heart of what we do . . . This approach represents a significant departure from the status quo. Budgets have traditionally focused on a limited set of economic data. Success has been declared on the basis of a narrow range of indicators, like GDP growth. . . . Many countries around the world have begun to look at different ways of measuring success to better reflect the wellbeing of their people. This Budget goes further and puts wellbeing at the heart of everything we do. To set the priorities for this Budget, we used evidence and expert advice to tell us where we could make the greatest difference to the wellbeing of New Zealanders. (Robertson, 2019, p. 3)

In a similar spirit but for the UK, Frijters and Krekel (2019) provide a detailed handbook on how to design policies to maximize subjective well-being.

The discussion so far demonstrates the substantial role that non-traditional indicators of personal well-being are now playing in assessments of social progress and the design of socio-economic policy. However, to date, the dominant approach to summarizing the distribution of well-being in a given year, and tracking changes over time or differences across groups, has examined distributions of income, rather than (say) distributions of subjective well-being (e.g. measures of life satisfaction or self-assessed health status). Leading examples of the income-based approach are the annual reports on *Household Incomes* from New Zealand's Ministry of Social Development (Perry, 2018), the UK's *Households Below Average Incomes* (Department for Work and Pensions, 2019), and the US's *Income and Poverty in the United States* (Semega, Kollar, Creamer, & Mohanty, 2019). Since its landmark report *Growing Unequal* (2008), the OECD has also provided frequent reports on the distributions of income within and between member countries, and an income distribution database (https://www.oecd.org/social/inequality-and-poverty.htm).

Against this background, this paper is based on two observations and a question. First, income is a *cardinal* variable, and many of the most commonly-used indicators of subjective well-being, including life satisfaction and self-assessed health status, are *ordinal* variables. Second, one cannot undertake

distributional comparisons for ordinal variables using the same methods as are routinely applied to cardinal variables. This raises the question: are there analogous methods for ordinal indicators?

I show that the answer is 'yes', broadly speaking, and illustrate the methods using data from the World Values Survey. Before proceeding to address the main question, I elaborate on the first two observations. Section 2 explains the distinction between cardinal and ordinal variables. Section 3 reviews the standard and well-known methods for undertaking distributional comparisons of income, a cardinal variable, considering both income levels and inequality. It is analogues of these that we require for distributional comparisons of ordinal data.

I provide a non-technical review of methods for distributional comparisons of ordinal variables in Section 4 and, in Section 5, I apply the methods to compare life satisfaction levels and life satisfaction inequality in New Zealand and four other countries (Australia, the United Kingdom, the USA, and South Africa) in the mid-2000s. I also consider how the distribution of life satisfaction in New Zealand has developed over time (I compare 1998, 2004, and 2011), and how life satisfaction distributions in 2004 differ between men and women, and by age group (Section 6). For brevity, I summarize the findings on these aspects in the main text with the charts summarizing the results available in Appendices A and B.

2. Cardinal versus ordinal variables (e.g. income versus life satisfaction)

Cardinal variables refer to quantities or amounts. Income is an example of such a variable; so too is financial wealth, or consumption measured using household expenditure. If we have a distribution of a cardinal variable for a set of individuals, we can order those individuals according to that variable from lowest to highest. In addition, differences between values, and ratios of values are well-defined (which also means that a value of zero is a true zero). It is meaningful to refer to a $100 difference in income or to say that John's income is twice as large as Janet's. Thus, income is measurable on both an interval and ratio scale.

Ordinal variables encapsulate people's choices from a set of categories with a well-defined ordering. Life satisfaction is a leading example. In the World Values Survey (WVS), respondents are asked 'All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means 'completely dissatisfied' and 10 means you are 'completely satisfied' where would you put your satisfaction with your life as a whole?'. The card looks like this:

Completely dissatisfied   Completely satisfied

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

With a distribution of life satisfaction responses, we can order the individuals in terms of their response from lowest to highest. However, differences in response values, and their ratios, are not well-defined. We cannot say that the difference in life satisfaction between 1 and 2 is the same as the difference between 8 and 9, or that a response of 8 corresponds to four times as great a life satisfaction as a response of 2, only that a value of 9 is greater than 8 and 2 is greater than 1.

The fundamental measurement issue for ordinal variables is that the labeling of the response categories is arbitrary. We do not know the underlying life satisfaction scale. The categories could be relabelled $(-5, -4, -3, -2, -1, 0, +1, +2, +3, +4)$ or $(1, 2, 4, 5, 7, 8, 10, 11, 13, 14)$ and there would be the same information about how a respondent places him- or herself across the 10 categories. There is no life satisfaction scale counterpart to the 'dollars per month' scale used for income; we only know whether a response is in the third, or fifth, or top category. The same issue arises for all other measures of subjective well-being in which survey respondents are presented with some form of Likert scale. These include self-assessed health, happiness, and so on.

As we shall see in Section 4, the arbitrariness of scale means that we cannot reliably compare distributions of life satisfaction (or other ordinal variables) in terms of their means or their inequalities

summarized using the Gini coefficient – commonly-used for comparisons of income distributions. Analogous but different tools are required.

An additional issue for distributional comparisons is cross-individual comparability of response values. That is, in the life satisfaction context and using the WVS scale above, we must assume that a response of '7', say, corresponds to the same life satisfaction value for all the individuals being compared, whether New Zealanders in 1998 and 2011, New Zealanders or South Africans, or men and women. This is the assumption of a 'common reporting function' and its plausibility may be questioned (Bond & Lang, 2019). For example, New Zealand and South African respondents may have different reporting functions because of the substantial differences between their two societies, but the assumption is more plausible if comparing respondents from a single country to surveys fielded not too many years apart, or if comparing groups of respondents to a survey in a given year. In this paper, I assume common reporting functions, following virtually all the literature on subjective well-being.

Readers should also be aware that analogous issues of comparability apply to 'income'. If we are interested in 'personal economic well-being', the appropriate measure of 'income' is not the raw response from the data source. Researchers and statistical agencies make assumptions about how to: (i) aggregate income sources across the individuals within an income unit (typically the household or the family); (ii) adjust observed money incomes for differences in household/family size and composition ($2000 per month for a single person living alone is worth a lot more in living standards terms than $2000 for a family of four); and (iii) adjust observed money incomes for differences in price levels across regions with a given country, across years (for temporal comparisons), or across countries (for cross-national comparisons). In other words, distributional comparisons are based on income variables that are transformed from the raw data using assumptions about aggregation, equivalence scales, within-country price differences, inflation rates, and market or purchasing power exchange rates, and these choices are not simply objective scientific ones but incorporate normative elements. There are extensive literatures about the sensitivity of income distribution comparisons to different choices under these headings. In sum, although cross-individual comparability is often cited as an issue for subjective well-being comparisons, the issue also applies to income comparisons. In the next section, I assume that appropriate adjustments to observed income values have already been made, and simply refer to 'income'.

3. Assessing better-offness: a toolbox of methods for income distributions

The two main distributional features relevant to assessing whether we are better off are well-being *levels* and well-being *inequality*. Better off means having higher well-being levels or less inequality, or potentially some combination of the two if we are willing to trade-off improvements in one dimension against deterioration in another. It is helpful to think of assessments being undertaken using a 'social welfare function' that aggregates the information in a distribution of well-being to a single number. That is, we employ a social welfare function $W = W(x_1, x_2, x_3, \ldots, x_n) = W(\boldsymbol{x})$ and distribution $\boldsymbol{x}$ is judged better than distribution $\boldsymbol{y}$ if $W(\boldsymbol{x}) > W(\boldsymbol{y})$. Individual-level outcomes $\boldsymbol{x}$ and $\boldsymbol{y}$ may refer to either income or to a subjective well-being measure.

The income distribution literature has developed two main approaches to assessing better-offness. The first consists of 'dominance' checks according to which we see whether we can *rank* two distributions A and B. The advantage of this strategy is that we may be able to say that A is better than B (A dominates B) without strong assumptions about the nature of the social welfare function. The weaker the assumptions, the less likely that there is room for disagreement about the overall assessment. The downside of relying on dominance checks is that we may not be able to derive a clear-cut ranking. However, regardless of whether dominance holds, the checks are useful because they involve graphical comparisons, and these pictures are powerful ways of summarizing the distributions – they 'show the data'.

The second approach is to summarize distributions using numerical indices. Indices encapsulate stronger and specific assumptions about the nature of the social welfare function – about which there

may be disagreement – but they have the powerful advantage that we can not only rank distributions (regardless of whether dominance holds), but also talk about magnitudes of difference. (For example, we are often interested not only in whether inequality increased or decreased but by how much.) Of course, the two approaches are complementary and analysts often employ both. In addition, it is also useful to look at density functions as well as cumulative density functions. Although there are no dominance results associated with the density functions, they are useful tools for 'showing the data' in a compact form.

For a more detailed background to the results that I review in the rest of this section, as well as references to the original literature and additional results, see inter alia Foster, Seth, Lokshin, and Sajara (2013). For an earlier survey of the same material, see Jenkins (1991).

3.1. *Three dominance results and many indices*

There are three main dominance results employed in income distribution analysis. The first is first-order welfare dominance based on comparisons of cumulative distribution functions (CDFs). We order the population from poorest to richest in terms of income and calculate each person's normalized rank (equivalently, cumulative population share), p, and plot p against income. Finding the CDF for distribution B lying everywhere below the CDF for distribution A is equivalent to B being ranked better than A according to all social welfare functions that are increasing in each person's income. (These functions are individualistic – there are no relative income effects.) The assumption is that 'more is better' – there is a social preference for increases in income levels, other things being equal. (One could go from A to B with a finite sequence of small increments to incomes.) This is a relatively weak assumption, likely to have widespread support. Observe that nothing has been said about *how much* social welfare increases if someone's income increases; and the result holds regards regardless of attitudes to inequality.

If this first order dominance holds, then at any specific value of p, quantile $x_p = F^{-1}(p)$ is larger in B than in A. It is also the case that the mean of distribution A is greater than the mean of distribution B.

The second result is Lorenz dominance and relates to inequality comparisons. A Lorenz curve plots cumulative income shares against cumulative population shares (p). The perfect equality reference point is when the Lorenz curve is a $45°$ ray through the origin: everyone has the same share of total income; each person's income is the mean income. With inequality, the Lorenz curve hangs below the $45°$ line: the poorest 20% have less than one fifth of total income and so on. Intuitively, inequality is greater, the further the Lorenz curve is from the $45°$ ray. There is also a fundamental link between mean-preserving progressive transfers of income and inequality reduction. If a small amount of income is transferred from person R to person P, where R's income is larger than P's, and no other incomes are changed, then the Lorenz curve must move inwards towards the perfect equality reference point over the range of p between P and R.

Atkinson's (1970) Lorenz dominance result formalizes these ideas: finding the Lorenz curve for distribution B lying everywhere below the Lorenz curve for distribution A is equivalent to finding that A has less inequality than B according to all social welfare functions that reflect 'equality preference' (and 'more is better'). Equality preference is the idea that a mean-preserving progressive income transfer as described above improves social welfare – though the magnitude of the change is not stated, thus allowing a wide range of attitudes to be incorporated from almost totally inegalitarian to strongly egalitarian views. In fact, we can go from the Lorenz dominated distribution to the dominating one by a finite sequence of mean-preserving progressive transfers. The assumption that 'more is better' is maintained but differences in mean income are controlled for: Lorenz curves are based on income shares rather than income levels.

The third result is Generalized Lorenz dominance and relates to 'social welfare' by taking account of both income levels and income inequality. A Generalized Lorenz curve is derived from a Lorenz curve by scaling each Lorenz ordinate vertically by overall mean income. Equivalently, the vertical axis of the Generalized Curve plots against cumulative population share p, the mean income among

the poorest $100p\%$ multiplied by p. The curvature inherited from the Lorenz curve reflects inequality; the height reflects income levels (note that the Generalized Lorenz curve value at $p = 1$ equals mean income).

Shorrocks' (1983) Generalized Lorenz dominance result states that finding the Generalized Lorenz curve for distribution B lying everywhere below the Generalized Lorenz curve for distribution A is equivalent to finding that B has less social welfare than A according to all social welfare functions that reflect 'more is better' and 'equality preference' – we are concerned about both the overall size of the economic pie as well as the size of each of the slices going to individuals. The Generalized Lorenz criterion encapsulates the idea that a society may be better off if a rise in inequality is accompanied by a sufficiently large increase in income levels. In cases where two Generalized Lorenz curves do not cross, we can go from the dominated distribution to the dominating one by a sequence of either mean-preserving progressive transfers or increments to individuals' income levels or a combination of both.

If we do not find dominance in a distributional comparison, we cannot order the distributions without additional assumptions about nature of the social welfare function. A numerical index reflects these specifically. Even if two CDFs cross, we can compare mean incomes (the arithmetic mean is one index of well-being levels; the geometric mean is another). If two Lorenz curves cross, we can assess inequality by comparing inequality indices. A leading example is the Gini coefficient which is equal to twice the area between the Lorenz curve and the line of perfect inequality.

The choice of index is a normative issue since different indices aggregate income differences in different ranges of the distribution in different ways. For this reason, economists are particularly fond of using families of inequality indices in which this sensitivity to differences in income shares in different ranges is varied in a systematic way. For example, members of the Atkinson family of inequality indices are characterized by an 'inequality aversion' parameter $\varepsilon > 0$, The larger is ε, the greater the inequality aversion – the greater is equality preference – and the greater the weight that is put by the index on income differences at the bottom of the income distribution. Generalized Entropy indices are the other leading family of inequality indices and are characterized by an income sensitivity parameter α. With $\alpha = 0$, we have the Mean Logarithmic Deviation; with $\alpha = 1$, we have the Theil index, and setting $\alpha = 2$ yields half the squared coefficient of variation. The smaller (and more negative) α is, the more 'bottom sensitive' the index (corresponding to larger values of ε for Atkinson indices). The larger α is, the more sensitive is the index to income differences at the top of the income distribution.

In sum, the income distribution toolbox consists of three sorts of dominance check and associated graphical device, and many indices incorporating different social judgements about how to aggregate incomes and income gaps. These basic tools are used routinely by income distribution researchers. Not all of the tools are used by statistical agencies but virtually all of the agencies (including the New Zealand, UK and US ones cited earlier) report estimates or provide access to raw data in a form that allow other researchers to apply the tools.

4. Assessing better-offness: a toolbox of methods for distributional comparisons of ordinal variables (e.g. life satisfaction)

The vast majority of users of subjective well-being data treat their well-being variables as cardinal rather than ordinal. For distributional comparisons, this has a great advantage: one can simply apply the toolbox described in the previous section. A leading example of current practice is the *World Happiness Report* which considers trends over time within countries and cross-national differences in subjective well-being measured variously in terms of life evaluation (scored using a Cantril ladder) and positive and negative affect. Helliwell, Huang, and Wang (2019) summarize well-being levels using means and well-being inequalities using standard deviations. Similarly, a major study of subjective well-being trends in the UK summarizes trends in four subjective well-being measures in terms of their means (Bangham, 2019: Figure 11). The OECD (2017) *How's Life* report examines cross-national

differences in life satisfaction by comparing country means and country inequalities summarized by $S80/S20$ indices – the average score of the top 20% of scores divided by the average score of those in the bottom 20% of scores.

The fundamental problems with using the mean to summarize the distribution of an ordinal variable are that the mean is not a stable reference point and the ranking of a pair of distributions can change if one changes the scale. The arguments are set out clearly by Allison and Foster (2004). See also Schröder and Yitzhaki (2017) and Bond and Lang (2019).

For a cardinal variable like income, the mean is a stable reference point but, for ordinal variables, any transformation of the numerical category labels that preserves their ordering is legitimate. With these changes in scale, the mean's location relative to the distribution as a whole or a subset of observations can vary a lot. Allison and Foster (2004, p. 510) illustrate this problem with reference to a distribution of 17 responses on a five-category variable represented by (1,3,3,6,4), meaning 1 observation reporting the lowest category through to 4 observations reporting the fifth (top) category. If the scale is linear with labels taking integer values from 1 to 5, the distribution has a mean of 3.53, located between 7th lowest and 8th lowest of the 17 observations. But if the scale is changed so that the label of the top category is 8 rather than 5, the mean is 4.24, lying between 13th and 14th observations. Allison and Foster comment that '[r]ecalling the categorical nature of the original data, it is difficult to accept the notion of a central reference value that jumps across observations and values with such ease' (2004, p. 510). Moreover, there is no robust answer to questions such as what percentage of population has above-average SWB?

To illustrate how distributional rankings according to the mean may change if the scale is changed, Allison and Foster consider distributions $x = (2,2,2,2,2)$ and $y = (3,2,1,1,3)$. With the same integer linear scale as before, running from 1 to 5, the means of x and y are 3 and 2.9 respectively. But if the numerical label attached to the top category is changed to 10, the means of x and y are 4 and 4.4. In other words, the distributional ordering is reversed. This illustration is not a theoretical curiosum: Allison and Foster (2004) provide a real-world example using self-assessed health data.

Rankings of distributions of ordinal variables based on comparisons of standard inequality indices – the commonly-used ones cited in the previous section – are also not robust to changes of scale. This is because all the indices are based on ratios of each observation's well-being score to the mean score (this is why they are sometimes called 'relative' indices). The OECD's $S80/S20$ index is also not robust for the same reason. The same problem rules out comparisons based on Lorenz and Generalized Lorenz curves.

The standard deviation of subjective well-being scores is often used as an inequality measure (as by the *World Happiness Report* op. cit.). Inequality orderings based on this measure, and other 'absolute' indices such as the absolute Gini coefficient or inter-quartile range, are robust to absolute changes in scale, e.g. if the five-category linear scale cited earlier with labels running from 1 to 5 were changed to a scale running from 3 to 7. However, there is no reason to support the legitimacy of this sort of scale change and rule out other transformations – an issue not considered by Kalmijn and Veenhoven (2005) in their defence of the standard deviation as an inequality measure for ordinal variables.

The arguments made so far imply that distributional comparisons of ordinal variables need new tools; the toolbox for cardinal variables should not be applied without adaption or modification.

4.1. Dominance results and indices

For dominance results relating to subjective well-being levels, comparisons of CDFs continue to play a role, though they now refer to CDFs for discrete distributions (and hence step functions) rather than continuous distributions. Although comparisons of mean subjective well-being are not robust to scale changes, comparisons of quantiles are. The category that is the median, for instance, remains the median if the scale is changed. One can also compare proportions with responses in a number of categories, for instance in the lowest category or the top two categories (this corresponds to looking

at the inverse of the CDF). For example, the UK's Office for National Statistics (2018) report on subjective well-being summarizes trends in distributions over time with reference to the proportion of persons with low well-being (scores of 0–4 on an 11-point integer scale running from 0 to 10), and the proportion with very high well-being (scores of 9 or 10).

The robustness of a distributional ordering by the mean holds only in the special case in which two well-being CDFs do not cross. Allison and Foster (2004, Theorem 1) show that finding the CDF for distribution B lying everywhere below the CDF for distribution A is equivalent to B having higher mean well-being than A for all possible scales applied to the categories. In the language of Dutta and Foster (2013) distribution B 'F-dominates' distribution A. This result does not tell us which scale of the infinite possible scales should be used to calculate and compare mean well-being, only that robustness to scale exists. For numerical comparisons of well-being levels, comparisons should be based on quantiles or category proportions.

Note that the Allison-Foster theorem, and other results cited in this section, refer to dominance for a categorical well-being variable (with an arbitrary scale). They do not refer to dominance for an underlying unobserved subjective well-being variable – it is often assumed that the discrete categorical responses are manifestations of a latent continuous variable. Bond and Lang (2019) emphasize this point, highlighting the strong assumptions required to identify the distribution of the latent variable.

In the conventional Lorenz-based approach to inequality, 'more unequal' is associated with a greater spread around the mean. This concept is inappropriate for ordinal data because the mean is scale-dependent. In contrast, the median is a natural reference point because its relative position is invariant to the choice of scale. Hence most research by economists has characterized 'more unequal' for ordinal variables as meaning greater dispersion around the median. The perfect equality benchmark is when all responses fall in the same category (which need not be the median). The perfect inequality benchmark from this perspective is when half the responses fall in the lowest category and half the responses fall in the top category. Strictly speaking this is a situation of maximum polarization rather than maximum inequality in the sense of maximal spread across categories, a distinction to which I return below.

Allison and Foster (2004) and Dutta and Foster (2013) discuss how a concept of 'S-dominance' ('S' for spread) encapsulates the idea of greater spread around the median. They show that distribution A has greater spread away from the median than distribution B (B S-dominates A) if and only if the following three conditions hold: (i) A and B have the same median; (ii) for every category k below the median category, the cumulative population share in the lowest k categories of A is no smaller than the cumulative population share in the lowest k categories of B; and (iii) for all categories at and above the median category, the cumulative population share in the lowest k categories of A is no greater than the cumulative population share in the lowest k categories of B. In terms of CDFs, the CDF for A is above (or nowhere below) the CDF for B up to the common median, and then is below (or nowhere above) the CDF for B above the median through to the top category. Allison and Foster (2004) show how one can construct an 'S-curve' from the CDF in order to implement dominance checks but, in practice, it is straightforward to compare CDFs to check S-dominance. If there is S-dominance, the CDFs for A and B must cross each other (once), and so, if there is F-dominance, there cannot be S-dominance as well.

The 'greater spread around the median' idea leads naturally to numerical indices summarizing the degree of inequality (polarization). Allison and Foster (2004), and later Dutta and Foster (2013), propose a measure based on areas under an S-curve. Specifically, the AF index is the difference between the mean response for above-median categories minus the mean response for below-median categories. Although it has an intuitive interpretation, AF is scale-dependent.

Scale-independent indices have been developed by Abdul Naga and Yalcin (2008), based on the weighted difference between the proportion of individuals in above-median categories and the proportion in below-median categories. Every member of the $ANY(\alpha, \beta)$ family of indices takes on values between 0 (the case of equality) and 1 (total polarization), where $\alpha, \beta \geq 1$. The parameters determine

the weights that are given to the top and bottom halves of the distribution in the aggregation of differences from the median. Relatively more weight is given to the bottom the more that β is greater than α, and relatively more weight is given to the top the more that α is greater than β. Thus, by analogy with the Generalized Entropy family of inequality indices, the $ANY(\alpha, \beta)$ family may have bottom- and top-sensitive members. For the case of a linear integer scale $(1, 2, 3, \ldots)$, the $ANY(1, 1)$ index can be interpreted as an Average Jump index – the average number of category 'jumps' required to change from the observed level to the median level, normalized by the total number of categories. For this scale, $ANY(1, 1)$ is also equal to the Allison-Foster index divided by the total number of categories minus one (Allison & Foster, 2004, p. 514).

Other families of inequality/polarization indices have been proposed by, for example, Apouey (2007), whose measures summarize 'distances' between the CDF and one half (the CDF value at the median) across the categories of the well-being variable. The greater the aggregate distance, the greater the spread around the median, with different distance concepts leading to different indices. Apouey's $P2(0.5)$ and $P2(2)$ indices incorporate 'square root' and Euclidean distance, respectively. $P2(2)$ is the $1 - l^2$ index of Blair and Lacy (2000). $P2(1)$ is the Average Jump index. In the empirical analysis that follows, I estimate $P2(0.5)$ and $P2(2)$ in addition to $ANY(1,1)$, $ANY(1,2)$, and $ANY(1,4)$ indices.

A totally different approach to measuring inequality in ordinal variables is taken by Cowell and Flachaire (2017). They argue in favor of a multi-step approach to the derivation of an inequality index. First, one decides how to summarize each individual's 'status' given the responses on the ordinal variable. Second, overall inequality is an aggregate summary of the 'distances' between each person's status and an appropriate reference status value. One of the advantages of their general approach is that it incorporates the conventional approach to inequality measurement for cardinal variables as well. Cowell and Flachaire consider several status measures but focus on the definition labeled 'peer-inclusive downward-looking' status: each person responding with category k is attributed with a status equal to the value of the empirical CDF for category k, with all persons in the same category having the same status. An important property of this status measure (and Cowell's and Flachaire's other peer-exclusive and upward-looking status measures) is that they are scale-independent. The next definitional step is to specify the reference status value. Cowell and Flachaire argue persuasively that, for the case of a peer-inclusive status measure, this reference point should be the maximum status value (which is one, given that status is defined in terms of CDF values). The final step in building the inequality index is to aggregate the individual differences between observed and maximum status.

The result is a one-parameter family of inequality indices, $CF(\gamma)$, with $0 \leq \gamma < 1$. The smaller that γ is – the closer to zero – the greater the weight that is put on small status values relative to high status values in the aggregation or, put differently, greater weight is put on status values that are a large distance from the reference point value of 1 rather than a small distance. In common with the median-based indices, the minimum value of $CF(\gamma)$ indices, zero, arises when all individuals are in the same category. The maximum value of a $CF(\gamma)$ index cannot be determined a priori, but one can show that the index value is greater for a uniform distribution (respondents spread evenly across all of the categories) than for a totally polarized distribution (Jenkins, 2019).

Jenkins (2019) reports a dominance check relating to $CF(\gamma)$ inequality indices based on comparisons of Generalized Lorenz curves for the distributions of status (as defined above). If all individuals report the same category and so have the same status (one), the Generalized Lorenz curve is a 45° ray through the origin and $(1, 1)$. If there is inequality, the Generalized Lorenz curve hangs below this reference line with a vertical intercept at $p = 1$ that is less than one. Drawing on Shorrocks (1983), Jenkins shows that finding the Generalized Lorenz curve for status distribution A everywhere below the Generalized Lorenz curve for status distribution B is equivalent to $CF_A(\gamma) > CF_B(\gamma)$ for all possible values of γ.

A fresh look at dominance results for ordinal variables is taken by Gravel, Magdalou, and Moyes (2015) using the concept of a Hammond transfer. A progressive Hammond transfer is one in which there is a small shift in the frequency density mass downwards from some higher category combined with a small shift in frequency density mass upwards from some lower category, where the 'recipient'

category in the former is less than or the same as the 'recipient' category in the latter. This is analogous to the mean-preserving progressive transfer concept cited earlier for income except that it does not specify the 'amount' transferred – this is inappropriate where ordinal variables are concerned. Gravel *et al.* (2015) derive a social welfare dominance result analogous to Generalized Lorenz dominance for income: they show that (i) a ranking of distribution A above distribution B according to all social welfare functions that are a scale-weighted sum of category proportions where the weights are strongly concave is equivalent to (ii) being able to go from B to A with a combination of progressive Hammond transfers or increments in frequency density mass (a shift of a person from one category to the next highest category). And (i) and (ii) are equivalent to (iii) 'H^+ dominance' of A over B, which is a set of restrictions on the configurations of the CDFs for the two distributions (see their paper for details). Although the conditions defining H^+ dominance are complicated, they are relatively easy to check in practice. Gravel *et al.* (2015) also prove a second dominance result related to an 'H^- dominance' condition on CDFs which is equivalent to a sequence of Hammond transfers and/or decrements.

By analogy with results showing that the Lorenz dominance criterion for income is equivalent to the simultaneous satisfaction of Generalized Lorenz dominance (known as 'supermajorization' in the mathematics literature) and a related concept ('submajorization'), Gravel *et al.* argue that the 'ranking of distributions generated by the intersection of the two domination criteria H^+ and H^- could serve as a plausible instance of a clear inequality reduction in an ordinal setting. Indeed, ..., any finite sequence of Hammond transfers would be recorded as an improvement by this intersection ranking' (2015, p. 6). A formal proof of the equivalence between pure inequality reduction (relating to sequences of Hammond transfers alone) and the intersection ranking is provided in the November 2019 revision of their paper. I therefore report H^- dominance checks below in addition to H^+ dominance checks.

An advantage of the Gravel *et al.* (2015) approach to ordinal distributional comparisons is that the dominance checks are not restricted to distributions with a common median as is the case for Allison and Foster's (2004) S-dominance criterion. This is also true for a $CF(\gamma)$ dominance check but the relationship between this and Gravel *et al.*'s (2015) dual-H dominance is yet to be ascertained. Observe, finally, that H^+ dominance is a form of second-order stochastic dominance and so, if there is F-dominance, there must also be H^+ dominance.

In sum, I have shown that there is a toolbox for distributional comparisons of ordinal variables that is analogous to the more widely-known and commonly-used toolbox for income distribution comparisons. Both toolboxes consist of dominance checks and associated graphical devices, and also indices incorporating different social judgements about to weight low values relative to high values of the well-being variable. The next Section illustrates the toolbox in action.

5. Comparisons of life satisfaction distributions: New Zealand and Australia, the UK, the USA, and South Africa

My empirical analysis is based on data from the *World Values Survey* (WVS), documented by Inglehart *et al.* (2014). I wish to focus on New Zealand in this paper and the WVS is the only accessible source with unit record data on subjective well-being variables for New Zealand. Unfortunately – and unnecessarily – access to confidential unit record files from other New Zealand data sources on subjective well-being is currently restricted to residents with access to a Statistics New Zealand-approved datalab.

The WVS consists of nationally representative surveys in almost 100 countries using a common questionnaire that includes questions on a range of subjective well-being variables. The WVS collects data from one adult (individual aged 18+ years) per household. Six WVS survey rounds have been conducted to date but New Zealand participated only in 1998 (wave 3), 2004 (wave 5), and 2011 (wave 6).

The focus here is on comparisons of WVS wave 5 life satisfaction distributions for New Zealand (NZ; 954 respondents), Australia (AU; 1,421), Great Britain (GB; 1,041), the United States of America

(US; 1,249), and South Africa (ZA; 2,988). The distributions refer to responses to the 10-category life satisfaction question set out in the Introduction. All the estimates I report are based on weighted data using the weights supplied with the WVS.

I restrict attention to only four countries besides New Zealand for the sake of brevity and focus. The countries chosen are all English-speaking; four are former British colonies; and four are close rivals in many sports. South Africa is a middle-income country in the Global South, whereas the other four countries are relatively rich (OECD member countries; in the Global North).

Country differences in income distribution provide a reference point for comparisons of life satisfaction distributions. It is clear from Table 1 that South Africa stands out from the other four countries, with an average income measured using real GDP per capita very much lower than the other four countries (about one third that of New Zealand's) and with substantially greater income inequality (about twice as large as New Zealand's according to the Gini coefficient). The USA also stands out, though to a lesser degree, because of its relatively high average income and second-ranked inequality level. The rest of this section considers whether these cross-national differences in income distribution are mirrored in the cross-national differences in life satisfaction. For example, are New Zealanders better off in life satisfaction terms than South Africans or Australians?

Figure 1 shows the proportions of individuals in each of the ten life satisfaction categories, country by country. Australia, Britain, and the USA appear to have similarly-shaped distributions with relative frequencies rising toward the modal value (8) and then decreasing thereafter. The mode is also 8 for New Zealand and South Africa, but these two countries stand out for the relatively high fractions of individuals reporting that they are 'completely satisfied' (score of 10): 22.0% and 17.5% respectively, by contrast with 10.9% for Britain, 8.2% for Australia, and only 6.7% for the USA. South Africa is also distinctive for having relatively high fractions of individuals reporting very low life satisfaction, with almost 6% in the lowest two categories, whereas the fraction is less than half that for the other four countries (highest for Australia at 2.6%). The country means are, given the 1–10 scale, 7.9 (NZ), 7.3 (AU), 7.6 (GB), 7.3 (US), and 7.0 (ZA). All the pairwise differences in means are statistically significant with the exception of the AU-US comparison. The median is 8 for all countries except South Africa for which it is 7. Thus, a first look suggests that life satisfaction levels and inequality rankings across the countries broadly correspond with those reported for income in Table 1. But to what extent are the rankings robust to scale and the use of methods appropriate for ordinal data?

Figure 2 presents pair-wise comparisons of the life satisfaction CDF for New Zealand with each of the other four countries. The horizontal dashed lines at 0.5 identify the median category. New Zealand's CDF lies everywhere on or below the CDF for Australia, the USA, and South Africa, i.e. the life satisfaction distribution for New Zealand F-dominates that for these three countries. (The largest differences are in the top life satisfaction categories, as expected from Figure 1.) Hence, we can say that average life satisfaction is higher for New Zealand than for any of those three countries, regardless of the life satisfaction scale. There is only one other F-dominance result: Britain F-dominates Australia (graph not shown). In every other pair-wise comparison, the CDFs cross.

Table 1. Income distribution summary: AU, GB, NZ, US, ZA.

Country	(Acronym)	Relative GDP per capita (%)	Gini coefficient (%)
Australia	(AU)	122	31.5
Great Britain	(GB)	116	34.6
New Zealand	(NZ)	100	33.5
USA	(US)	155	38.5
South Africa	(ZA)	34	65.7

Notes: GDP per capita is in 2011 PPP international dollars in 2005, expressed relative to NZ's value (source: World Bank Development Indicators online database). Gini coefficient refers to distributions of equivalized net household income among persons for AU in 2004, GB in 2005, NZ in 2005, US in 2006, and ZA in 2008 (source: UNU-WIDER's WIID4 database).
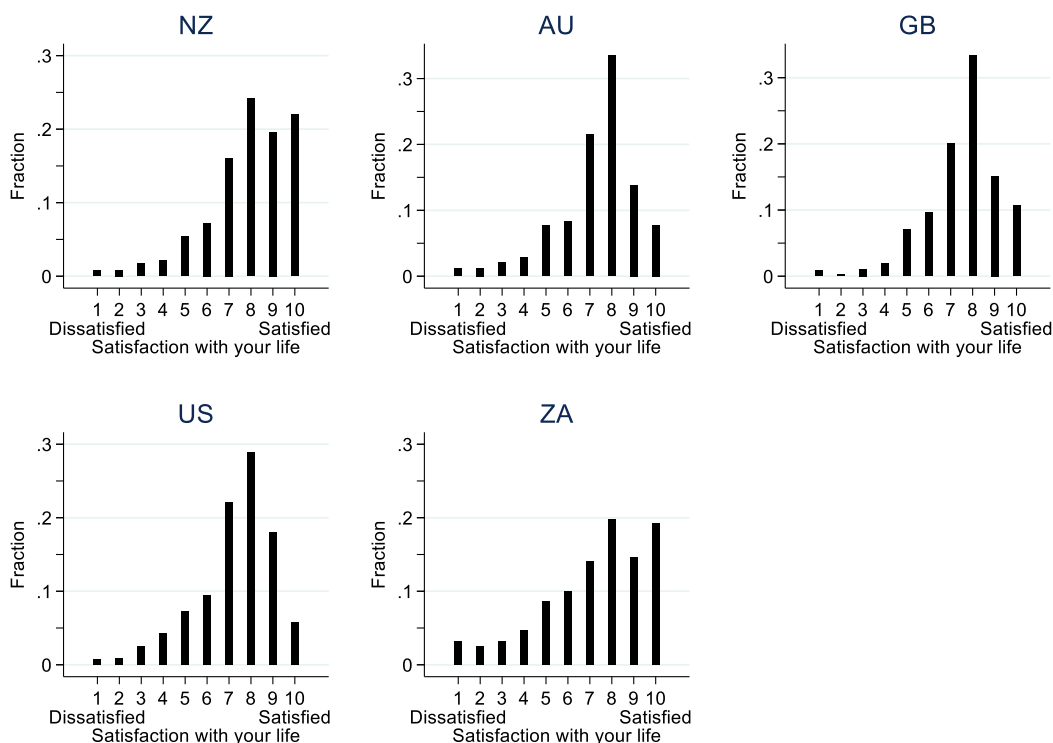
Figure 1. The distribution of life satisfaction in 5 countries, mid-2000s.
Notes: Weighted estimates from WVS data. The country means are: 7.9 (NZ), 7.3 (AU), 7.6 (GB), 7.3 (US), and 7.0 (ZA). The median is 8 for all countries except ZA (7).
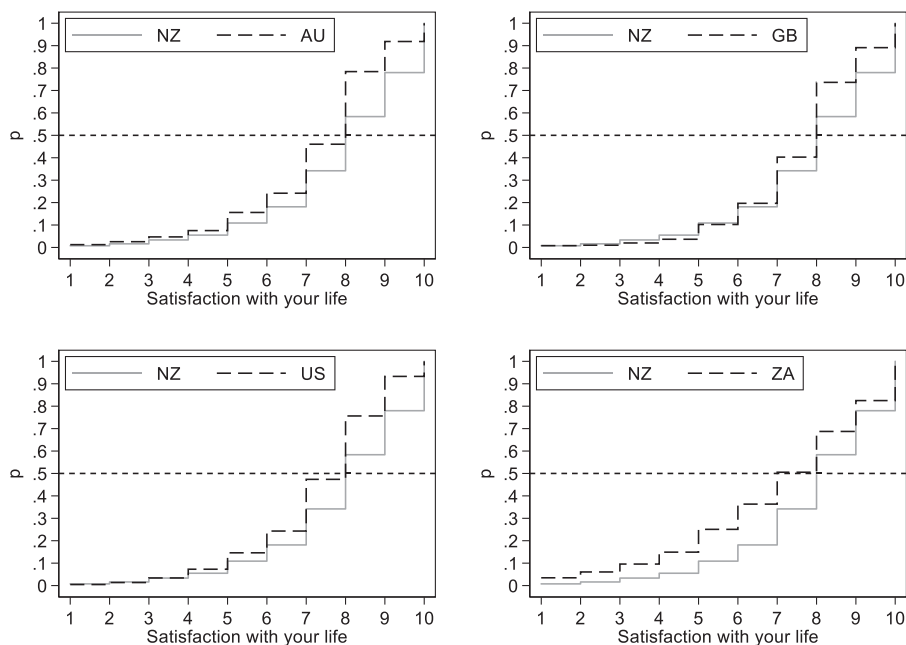

Figure 2. Cumulative distribution functions (CDFs) for life satisfaction.
Notes: Weighted estimates from WVS data. Country *A* *F*-dominates country *B* if the CDF for *A* lies everywhere on or below that for *B* (see main text). *p* is the cumulative proportion of individuals ordered from lowest to highest life satisfaction.

What about life satisfaction polarization and inequality dominance? There is no pair-wise country comparison that yields an *S*-dominance result. There are three reasons for this. First, the criterion is inapplicable when the two countries have different medians, as with any comparison involving South Africa. Second, where there is *F*-dominance, as in four pair-wise comparisons, there cannot also be *S*-dominance. Third, in the remaining comparisons, the configurations of CDFs are inconsistent with *S*-dominance. Thus, the minimal assumptions about the social welfare function that underlie the *S*-dominance criterion are insufficient to order the countries in terms of median-referenced inequality/polarisation. But what if the common-median requirement is removed?

Figure 3 shows the results of $CF(\gamma)$ inequality index dominance checks for New Zealand with each of the other four countries. New Zealand's Generalized Lorenz curve for life satisfaction status lies everywhere on or above that for South Africa, so New Zealand's life satisfaction distribution is more equally distributed than South Africa's according to all members of the $CF(\gamma)$ family of indices. However, Britain's Generalized Lorenz curve lies everywhere on or above New Zealand's, so New Zealand's distribution is more unequal than Britain's according to this set of indices. However, the curves for New Zealand and Australia, and New Zealand and US cross, and so there is no dominance in these cases: whether New Zealand has greater or lower inequality according to a $CF(\gamma)$ index
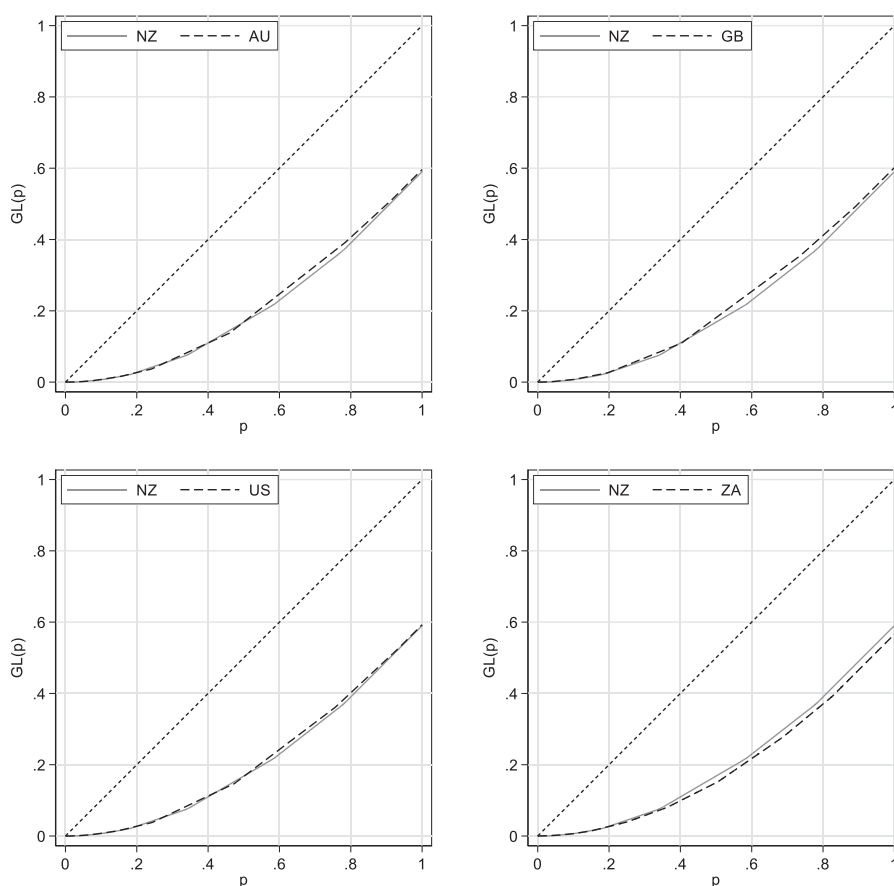
Figure 3. Checks for dominance according to all $CF(\gamma)$ inequality indices using Generalized Lorenz curve comparisons of life satisfaction 'status' distributions.
Notes: Weighted estimates from WVS data. The figure shows Generalized Lorenz curves for life satisfaction 'status', where status is defined using Cowell and Flachaire's peer-inclusive downward-looking definition (see main text). If the Generalized Curve for Country *A* lies everywhere on or below that for *B*, life satisfaction inequality is greater in *A* than *B*, for all members of the $CF(\gamma)$ family of inequality indices (see main text). *p* is the cumulative proportion of individuals ordered from lowest to highest life satisfaction status.

can depend on the value of γ that is used. Generalized Lorenz curves for comparisons of the USA with Britain, and the USA with Australia also cross (graph not shown). However, checks show that Britain is more equal than Australia (graph not shown). Moreover, and consistent with the discussion of frequency distributions earlier, the checks show that South Africa's life satisfaction distribution is more unequal than that for each of the other four countries (graphs not shown). The differences between the curves are quite small, however, suggesting that some cross-national differences may not be statistically significant – an issue I return to below.

Figure 4 shows pairwise comparisons of H^+ curves (left) and H^- curves (right) for New Zealand and each of the other four countries. Recall that New Zealand F-dominates each of Australia, the USA, and South Africa (Figure 2) and therefore we also see that New Zealand H^+ dominates each of these countries. New Zealand's curve lies everywhere on or below that for these three. However, New Zealand's H^+ curve lies everywhere on or above that for Britain. Although there was no first-order dominance (F-dominance), Britain's distribution second-order dominates New Zealand's. Again the relatively small differences between the pairs of curves compares warns us that differences may not be statistically significant.

For 'pure' inequality dominance according to Gravel *et al.*'s (2015) results, we require the dual dominance criterion to hold, i.e. to have both H^- and H^+ dominance. In Figure 3, the only comparison for which this occurs is for New Zealand and Britain. Thus, the distribution of life satisfaction in Britain is more equal than the distribution in New Zealand according to the dual dominance criterion as well as according to the $CF(\gamma)$ dominance criterion. South Africa's distribution is also more unequal than the distributions for each of Australia, Britain, and the USA, according to both
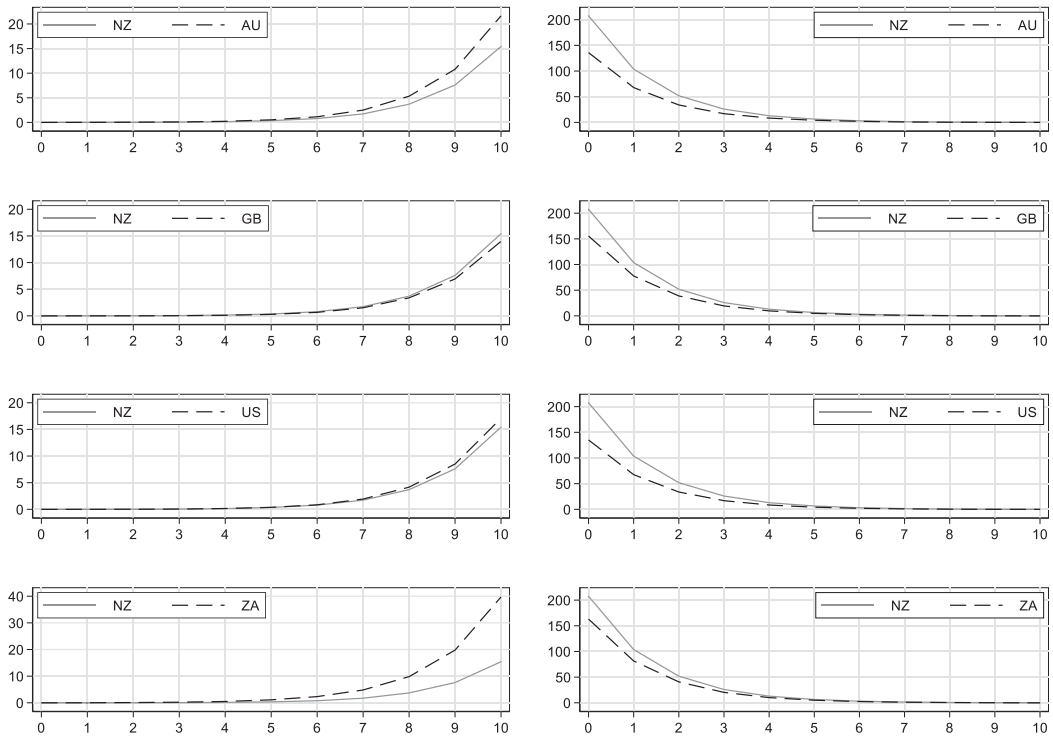
Figure 4. Checks for life satisfaction H^+ dominance (left) and H^- dominance (right).
Notes: Weighted estimates from WVS data. The figure shows H^+ curves (left) and H^- curves (right) If the H^+ curve for Country *A* lies everywhere on or below that for *B*, *A* H^+-dominates *B* meaning social welfare is higher; and analogously for non-crossing H^- curves (see main text). Gravel *et al.*'s (2015) result is that if both H^+- and H^--dominance hold (dual dominance), there is also equality dominance (see main text).

criteria (H curves not shown). However, the dual dominance criterion is not satisfied for the NZ-ZA comparison (or, indeed, for any other pairwise comparison of countries not mentioned so far). New Zealand's distribution clearly H^+ dominates (and F-dominates) South Africa's, but there is no H^- dominance, a result essentially due to the relatively high proportions of individuals in the very top life satisfaction categories in both these countries (Figure 2).

In sum, the two types of inequality dominance check yield only one consistent result – that Britain's life satisfaction distribution is more equal than South Africa's. According to the CF dominance checks, Britain's distribution is more equal than New Zealand's and Australia's and the distribution for each of these two countries is in turn more equal than South Africa's. Thus, in order to completely order the five countries in terms of life satisfaction inequality, we have to use indices, in the knowledge that different indices may provide different orderings.

Figure 5 shows estimates of inequality indices for each of the five countries together with their associated 95% confidence intervals based on standard errors based on a repeated half-sample boot-strapping method that handles sampling weights correctly (see Van Kerm 2013 for details of the implementation). There are nine indices: three $CF(\gamma)$ with γ chosen to cover the range of possible values; three $ANY(\alpha, \beta)$ indices with the parameters chosen so there is variation in the extent to which the indices are 'bottom sensitive'; two Apouey indices incorporating different concepts of 'distance' of the CDF from the median; plus, for reference, the commonly-used standard deviation measure which treats the data as if they are cardinal rather than ordinal.

Figure 3 shows that life satisfaction inequality is greater in South Africa than in each of the other four countries, regardless of which index is chosen. Also, Britain is the country with the most equal distribution. However, the magnitudes of the differences depend on the index chosen. For example, according to $CF(0)$, inequality is 9% higher in South Africa than in Britain and 4% higher than in New Zealand. However, if the $ANY(1,1)$ index is used, the corresponding differences are 67%
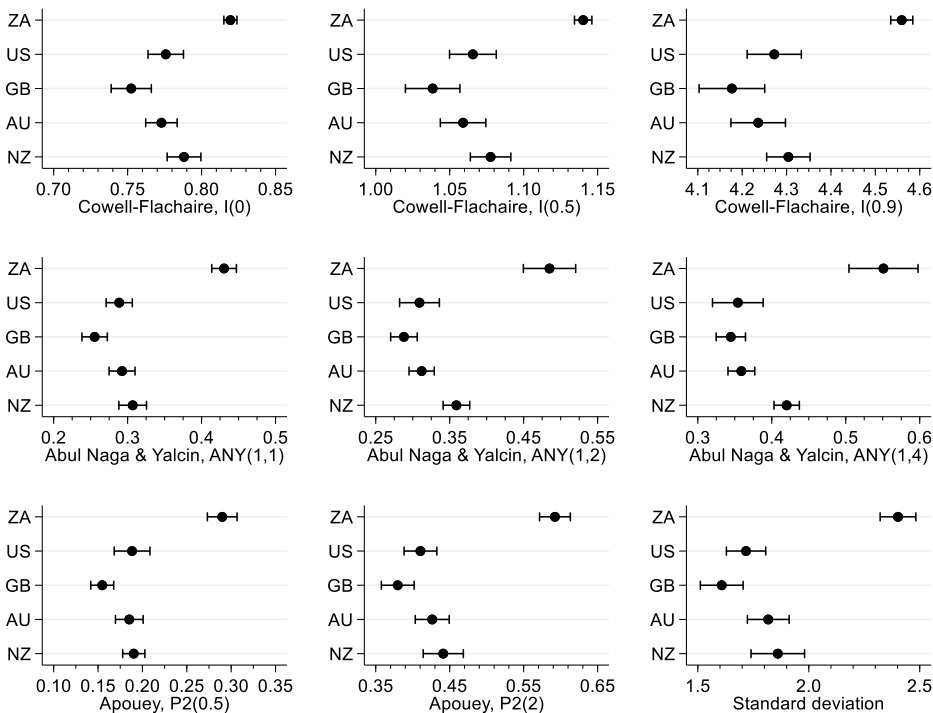

Figure 5. Indices of life satisfaction inequality and polarization.
Notes: Weighted estimates from WVS data. The figure shows point estimates of inequality and polarization indices and their associated 95% confidence intervals (derived using bootstrap standard errors; 500 replications using appropriate bootstrap weights).

and 40%. For the standard deviation (scale dependent), the differences are in between: 49% and 29% respectively. In part, these differences in magnitude arise because the three indices cited incorporate different conceptualizations of 'inequality', i.e. deviation of values from different reference points – the maximum, the median, and the mean. Differences also arise because of the different ways that indices weight distances from the relevant reference point. For example, $ANY(1,4)$ is more sensitive to differences below the median than $ANY(1,1)$. According to the former, life satisfaction inequality is 41% higher in South Africa than in New Zealand and 35% higher than in Britain. These cross-country differentials are slightly smaller than the corresponding ones for $ANY(1,1)$. Going from $ANY(1,1)$ to $ANY(1,4)$ makes more of a difference when summarizing how much greater New Zealand's inequality is than Australia's: for the former index, the ratio of country estimates is 1.06; for the latter, it is 1.17. (Look at Figure 5 to see this too.)

Changing the index may also change the ordering of the countries in terms of life satisfaction inequality. Figure 5 shows, and as we would expect from the dominance analysis, the differences in country rankings largely refer to the ordering of Australia and the USA relative to each other. For seven of the indices, inequality is very slightly greater in the USA; for $P2(2)$ and the standard deviation, US inequality is slightly lower than Australia's.

However, the confidence intervals shown in Figure 5 remind us to be cautious about making inferences about cross-national differences. For example, the null hypothesis that inequality is the same in Australia and the USA cannot be rejected for $CF(0)$ (test t-statistic 0.37, which is very much less than a critical value of 2), $P2(2)$ (t-statistic 0.98), or the standard deviation (t-statistic 1.55). On the other hand, the differences between South Africa and each of the other four countries are statistically significant. This is not only because estimated inequality is so large for South Africa, but also because the sample size for that country is much larger than for the others and so estimates are more precise (confidence intervals narrower). For example, the null hypothesis that life satisfaction inequality is the same in New Zealand and South Africa is decisively rejected. The test t-statistics are: 5.1 for $CF(0)$, 9.9 for $ANY(1,1)$, 4.9 for $ANY(1,4)$, and 9.2 for $P2(2)$, for instance.

In sum, there is one clear similarity between the differences in the distributions of life satisfaction and of income across the five countries: for both measures of well-being, South Africa has the lowest average well-being and the highest inequality. Otherwise, however, rankings of the other four countries are not the same. For example, by contrast with the rankings shown in Table 1, it is not the USA that has the highest average life satisfaction or the second-highest level of life satisfaction inequality. And, compared with the USA, New Zealand has higher levels of average life satisfaction and higher life satisfaction inequality.

6. Life satisfaction in New Zealand: trends over time and differences across groups

The toolbox of methods for ordinal data illustrated in Section 5 can also be applied to other types of comparisons. One can address questions such as whether the distribution of life satisfaction has improved over time and how life satisfaction distributions differ between groups within a given population. Differences by sex and by age are common indicators of stratification in a society. Differences by ethnic background are also important but they cannot be examined using these WVS data.

Using WVS data, I have compared life satisfaction distributions in New Zealand for 1998, 2004, and 2011 (unweighted sample sizes: 1,201, 954, and 841, respectively). In addition, for 2004, I have compared life satisfaction between men and women (unweighted sample sizes 513, 420), and between three age groups defined as individuals aged 18–29 years, 30–59 years, and 60+ years (unweighted sample sizes 111, 560, and 257).

For brevity, all the charts summarizing the comparisons undertaken – in the same formats as those discussed in Section 5 – are reported in two Appendices (A for trends over time; B for differences across groups). In this section, I simply highlight the main findings without citing the charts on which they are based.

6.1. Trends over time: life satisfaction in 1998, 2004, and 2011

The main issue of interest here is whether New Zealanders have got better off in successive periods. In terms of the metrics discussed so far, we are interested in what happened to life satisfaction levels, inequality, and their combination ('social welfare'). The frequency distribution changed a little over time: although the median response is 8 for 1998, 2004, 2011, the means are 7.7, 7.9, and 7.6, respectively. The higher mean for 2004 largely reflects a small increase in relative frequencies in the very top category. Pairwise comparisons of CDFs reveal that 2004 F-dominates 1998 and also F-dominates 2011, but the CDFs for 1998 and 2011 cross. It seems that, regardless of the scale used, average life satisfaction levels rose and then fell back.

What about life satisfaction inequality trends? There are no S-dominance results and so no dominance results are available for median-based inequality indices. According to comparisons of Generalized Lorenz curves for life satisfaction 'status', 2011 is more equal than 1998, and 2004 is also more equal than 1998 according to all $CF(\gamma)$ indices. However, the differences between the curves are very small and hence likely to be not statistically significant.

Given the F-dominance results cited above, it must also be the case that 2004 H^+-dominates 1998 and also that 2004 H^+-dominates 2011. But Gravel *et al.*'s (2015) inequality dominance result also requires H^- dominance to hold and it does not for these two pairs of years. The only dual-dominance result is that 2011 is more equal than 1998. This is consistent with the Generalized Lorenz dominance result for $CF(\gamma)$ indices for that pair of years. However, the differences between the relevant H curves are very small once again.

Estimates of the nine equality indices used earlier confirm that inequality is lower in 2011 than 1998, though the estimated decrease over time depends on which specific index is used. For example, according to $CF(0)$ the decrease is less than 1% (from 0.794–0.792). The decrease registered by $P2(2)$ is one of the largest but, at around 5%, is still small. Inequality in 2004 is estimated to be lower than in 1998 according to all nine indices, but whether it is larger or smaller than 2011 depends on which index is employed. For six indices, 2004 has lower inequality but, for the two bottom-sensitive indices $ANY(1,2)$ and $ANY(1,4)$ and also $P2(0.5)$, inequality is higher. Again, the changes over time are small and none of the pairwise differences between years is statistically significant.

Overall, the data suggest that there was both a fall in average life satisfaction and in inequality between 1998 and 2011. These changes contrast with the rise in mean and median equivalized household income over the same period and relatively unchanged income inequality for the same period that are reported by Perry (2018). However, one cannot state with confidence that the changes in the distribution of life satisfaction are statistically significant.

6.2. Differences across groups: life satisfaction in 2004, by sex and by age group

Although differences by sex and by age are of great interest, it is difficult to draw robust conclusions about life satisfaction differences using WVS data. As before, the issue is one of statistical rather than substantive significance. This is particularly pertinent when looking at subgroup differences because WVS sample sizes for subgroups are small, especially for the age breakdowns. Indeed, breakdowns by sex and age combined, or other interesting combinations of characteristics, all of which lead to even smaller sample sizes, are largely worthless.

The frequency distributions for men and women look different. Although the medians are the same (8) and the means very similar (7.92 and 7.87 respectively), the spread across the top four categories differs: in particular, relatively more women than men report 10 rather than 9. Comparisons of CDFs reveal neither F- nor S-dominance. The Generalized Lorenz curve for women's life satisfaction status lies everywhere on or below that for men, so inequality is greater for women according to all $CF(\gamma)$ indices, but the differences between the curves are negligible. The higher inequality among women conclusion is also supported by the comparisons of H curves: there is dual dominance, though it should be noted that the H^- curves largely coincide. The estimates of all nine inequality indices

are larger for women than men but differences are not statistically significant from zero. Among the largest differences are those registered by $P2(2)$ and $ANY(1,4)$, 10% and 8% respectively, but the test t-statistics are only 1.7 and 1.9 respectively.

Overall, one cannot state with confidence that the life satisfaction distributions of New Zealand men and women differ in 2004, which is a positive conclusion if the social goal is to eliminate gender divides in well-being. Note also that life satisfaction questions provide an individual-based perspective that conventional income distribution analysis is unable to do because of the standard assumption that income is equally distributed within families or households.

There appear to be some remarkable differences across age groups in life satisfaction but robust conclusions about these differences are bedevilled by small sample sizes. The outstanding feature of the subgroup frequency distributions is the extremely large fraction of individuals aged 60+ reporting that they are completely satisfied (a score of 10), 36%, which is around twice the fraction for the other two groups – 16% for individuals aged 30–59 and 17% for individuals aged 18–29. Related, the median life satisfaction score for the 60+ group is 9, but 8 for the other two groups, and the means are 8.3, 7.4, and 7.7, respectively. (The differences in means for the 60+ group and each of the other two age groups are statistically significant; the difference in means for the two younger groups is not.) The relatively high average for the 60+ group is robust to changes in scale: their distribution F-dominates the distributions for each of the other two groups. The high rates of individuals aged 60+ group at the top of the life satisfaction distribution is in contrast with their position in the distribution of household income (on this, see e.g. Perry, 2018: Figures B1 and B2).

There are no S-dominance results, nor Generalized Lorenz or H dual dominance results. Estimates of the nine inequality indices do not provide conclusive results either. On the one hand, this is because the inequality ranking of the groups depends on the index: the 60+ group has the lowest inequality according to the three $CF(\gamma)$ indices but the highest inequality according to $ANY(1,1)$ and $P2(0.5)$. The various indices account for the very high number of completely satisfied individuals aged 60+ in different ways. On the other hand, the 95% confidence intervals for all the estimates are extremely wide (especially for the 18–29 group with a sample size of only 111), and the intervals overlap substantially. No cross-group inequality difference differs significantly from zero.

Overall, it appears that the prevalence of extremely high life satisfaction levels in New Zealand (in 2004) is much greater for individuals aged 60+ than for younger people, but robust conclusions about inequality differences across age groups cannot be made because of small sample sizes.

7. Summary and conclusions

I have argued that we should compare distributions of ordinal data using methods appropriate to such data, rather than applying methods developed for cardinal data as is commonly done. That the application of cardinal methods sometimes yields the same substantive conclusion is not a strong argument against my injunction because there can also be situations in which different methods lead to different findings and, in any case, there is now a well-developed toolbox for undertaking distributional comparisons of subjective well-being data. As I have shown, easily-applicable tools are available to 'show the data' and to compare well-being levels and inequality using dominance checks and indices.

Areas for further research on aspects of the subjective well-being toolbox remain of course. One example is the relationship between H curve dual dominance and rankings by Cowell-Flachaire (2017) inequality indices. Another example is the development of appropriate methods for statistical inference, accompanied by software, enabling practitioners to implement the methods routinely.

Better methods also need to be accompanied by better data. My empirical illustrations based on WVS data have thrown up some intriguing results about how life satisfaction in New Zealand compares with other countries, how it has changed over time, and differs by sex and age. These perspectives on the nation's well-being differ in several ways from those provided when well-being is measured using income. However, it is also the case that the small sample sizes available in the WVS

substantially constrain our ability to conclude whether a finding is statistically significant in addition to being substantively significant. This is particularly so for subgroup breakdowns.

Fortunately, there are larger sample sizes in New Zealand's General Social Survey (GSS) – around 8,000 which is around eight times larger than New Zealand's WVS sample sizes – with surveys carried out in 2008, 2010, 2012, 2014, and 2018. The GSS enables some additional important breakdowns for groups such as Pakeha, Māori, and Pasifika, and by region of residence. Te Kupenga, a specialist survey of Māori well-being undertaken in 2013, also enables detailed breakdowns but comparisons with GSS data are constrained because the GSS used a 5-point life satisfaction scale in its 2008, 2010, and 2012 rounds. Another promising development is the recent substantial expansion in the Household Economic Survey's sample size to around 20,000 households combined with the introduction of additional well-being questions.

Access to the unit record data from these data sources should be improved in order to maximize the returns to the large investments that they represent. At present, researchers based outside New Zealand cannot access the confidential unit record files and New Zealand-based researchers can only do so from within Statistics New Zealand-approved datalabs. I support Statistics New Zealand's goals of preserving respondent confidentiality and privacy. Yet, at the same time, experience elsewhere shows that it is possible to provide better data access while still meeting these goals. Given the priority that New Zealand is placing on information about subjective well-being nowadays, it is important to increase the amount of data analysis undertaken on the topic. Facilitating an increase in the pool of analysts is one important way to do this.

Acknowledgements

Disclosure statement

Funding

References

Abdul Naga, R., & Yalcin, T. (2008). Inequality measurement for ordered response health data. *Journal of Health Economics*, *27*, 1614–1625.

Allison, R.A., & Foster, J.E. (2004). Measuring health inequality using qualitative data. *Journal of Health Economics*, *23*, 505–524.

Apouey, B. (2007). Measuring health polarization with self-assessed health data. *Health Economics*, *16*, 875–894.

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

Bangham, G. (2019). *Happy now? Lessons for economic policy makers from a focus on subjective well-being*. London: Resolution Foundation.

Blair, J., & Lacy, M.G. (2000). Statistics of ordinal variation. *Sociological Methods and Research*, *28*(3), 251–280.

Bond, T.N., & Lang, K. (2019). The sad truth about happiness scales. *Journal of Political Economy*, *127*(4), 1629–1640.

Cowell, F.A., & Flachaire, E. (2017). Inequality with ordinal data. *Economica*, *84*, 290–321.

Department for Work and Pensions. (2019). *Households below average incomes 1994/5–2017/8*. London: Department for Work and Pensions. Retrieved from https://www.gov.uk/government/collections/households-below-average-income-hbai–2

Dutta, I., & Foster, J.E. (2013). Inequality of happiness in the US: 1972–2010. *Review of Income and Wealth*, *59*, 393–415.

Foster, J., Seth, S., Lokshin, M., & Sajara, Z. (2013). *A unified approach to measuring poverty and inequality*. Washington, DC: World Bank. Retrieved from http://documents.worldbank.org/curated/en/281001468323965733/A-unified-approach-to-measuring-poverty-and-inequality-theory-and-practice

Frijters, P., & Krekel, C. (2019). *A handbook for wellbeing policy-making in the UK: History, measurement, theory, implementation, and examples*. Working Paper. London: LSE.

Gravel, N., Magdalou, B., & Moyes, P. (2015). Ranking distributions of an ordinal attribute. Working paper. Retrieved from https://halshs.archives-ouvertes.fr/halshs-01082996v2

Helliwell, J.F., Huang, H., & Wang, S. (2019). Changing world happiness', Chapter 2. In J.F. Helliwell, R. Layard, & J.D. Sachs (Eds.), *World Happiness report 2019* (pp. 11–45). New York, NY: Sustainable Development Solutions Network. Retrieved from https://worldhappiness.report/ed/2019/

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., . . . Puranen, B. (Eds.). (2014). *World values survey: All rounds - country-pooled datafile version*. Madrid: JD Systems Institute. Retrieved from http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp

Jenkins, S.P. (1991). The measurement of income inequality', chapter 1. In L. Osberg (Ed.), *Economic inequality and poverty: International perspectives* (pp. 3–38). Armonk, NY: M. E. Sharpe.

Jenkins, S.P. (2019). Inequality comparisons with ordinal data. Unpublished paper.

Kalmijn, W., & Veenhoven, R. (2005). Measuring inequality of happiness in nations. In search of proper statistics. *Journal of Happiness Studies*, 6, 357–396.

New Zealand Treasury. (2018a). *Our people our country our future. Living standards framework: Background and future work*. Wellington: New Zealand Treasury. Retrieved from https://treasury.govt.nz/sites/default/files/2018-12/lsf-background-future-work.pdf

New Zealand Treasury. (2018b). Our living standards framework. Retrieved from https://treasury.govt.nz/information-and-services/nz-economy/living-standards/our-living-standards-framework

OECD. (2008). *Growing unequal? Income distribution and poverty in OECD countries*. Paris: OECD.

OECD. (2017). *How's life? 2017: Measuring well-being*. Paris: OECD Publishing.

Office for National Statistics. (2018). *Personal well-being in the UK, July 2017 to June 2018*. London: ONS. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/measuringnationalwellbeing/july2017tojune2018

Perry, B. (2018). *Household incomes in New Zealand: Trends in indicators of inequality and Hardship 1982 to 2017*. Wellington: Ministry of Social Development. Retrieved from https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/monitoring/household-incomes/

Robertson, G. (2019). From the Minister of Finance. *The wellbeing budget*, 30 May 2019. Wellington: New Zealand Treasury. Retrieved from https://treasury.govt.nz/publications/wellbeing-budget/wellbeing-budget-2019

Schröder, C., & Yitzhaki, S. (2017). Revisiting the evidence for cardinal treatment of ordinal variables. *European Economic Review*, 92, 337–358.

Semega, J., Kollar, M., Creamer, J., & Mohanty, A. (2019). *Income and poverty in the United States: 2018*. Washington, DC: US Census Bureau. Retrieved from https://www.census.gov/library/publications/2019/demo/p60-266.html

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

Statistics New Zealand. (2018). A well-being approach to strategic decisionmaking. Retrieved from https://www.stats.govt.nz/assets/Consultations/indicators-aotearoa-new-zealand-nga-tutohu-aotearoa-consultation/a-well-being-approach-to-strategic-decision-making.pdf

Statistics New Zealand. (2019). *Indicators Aotearoa New Zealand – Ngā Tūtohu Aotearoa*, 9 July 2019. Retrieved from https://www.stats.govt.nz/indicators-and-snapshots/indicators-aotearoa-new-zealand-nga-tutohu-aotearoa/

Stiglitz, J.E., Fitoussi, J.-P., & Durand, M. (2018). *Beyond GDP: Measuring what counts for economic and social performance*. Paris: OECD Publishing.

Stiglitz, J.E., Sen, A.K., & Fitoussi, J.-P. (2009). *Report by the commission on the measurement of economic performance and social progress*. Retrieved from https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report

Van Kerm, P. (2013). RHSBSAMPLE: Stata module for repeated half-sample bootstrap sampling. Statistical Software Components S457697, Boston College Department of Economics, revised 17 Nov 2013.

APPENDICES

A. **Life satisfaction in New Zealand: comparisons of 1998, 2004, and 2011**
B. **Life satisfaction in New Zealand: comparisons of men and men, and of three age groups (18–29, 30–59, 60+ years)**

The 5 figures in each appendix correspond with the 5 figures presented in the main text. All estimates are derived from weighted WVS data.
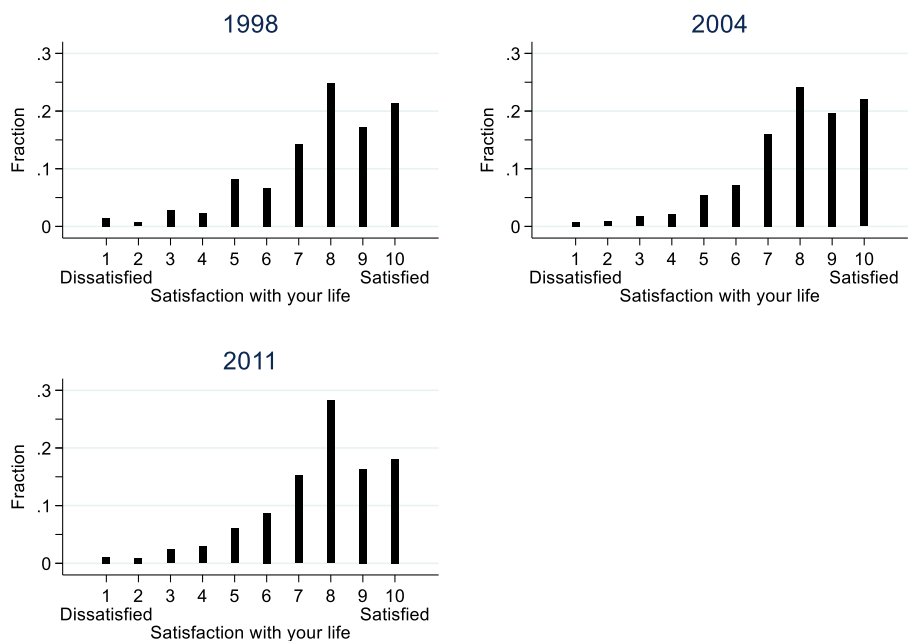
Figure A1. The distribution of life satisfaction in NZ: 1998, 2004, and 2011.
Notes: The median (and mode) is 8 for all 3 years. The means are 7.7 for 1998, 7.9 for 2004, and 7.6 for 2011.
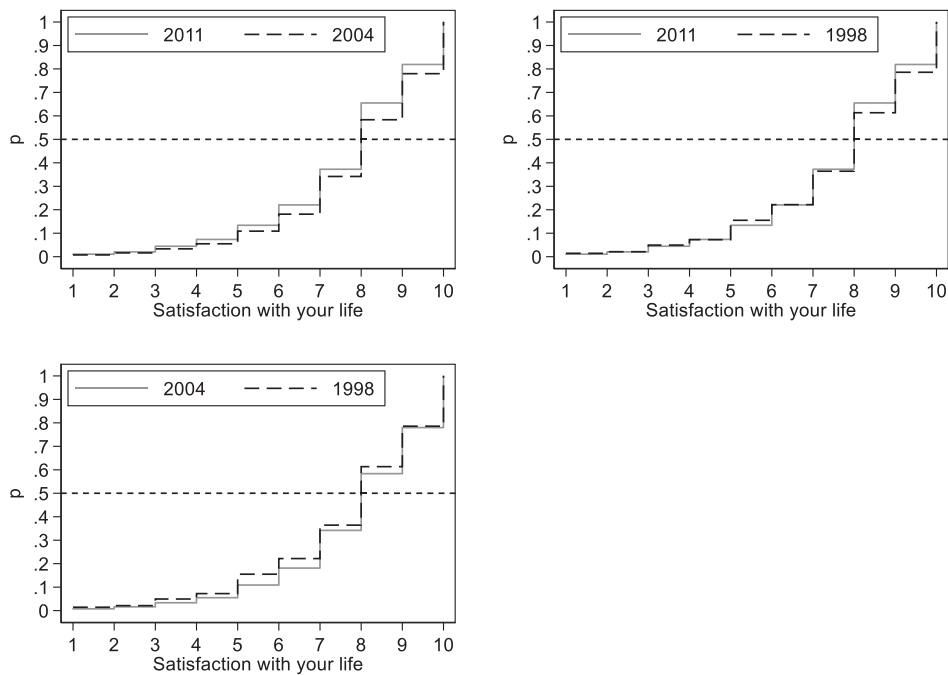

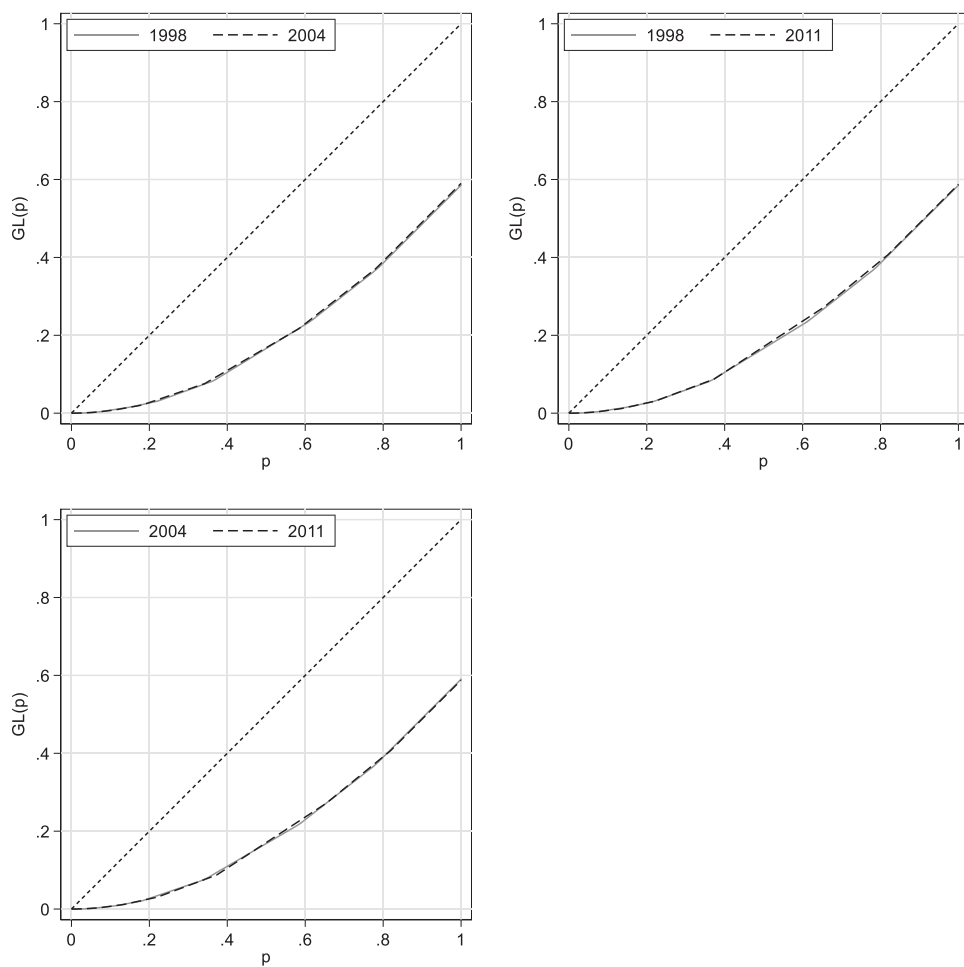**Figure A2.** Cumulative distribution functions (CDFs) for life satisfaction, 1998, 2004, and 2004.

Figure A3. Checks for dominance according to all $CF(\gamma)$ inequality indices using Generalized Lorenz curve comparisons of life satisfaction 'status' distributions.
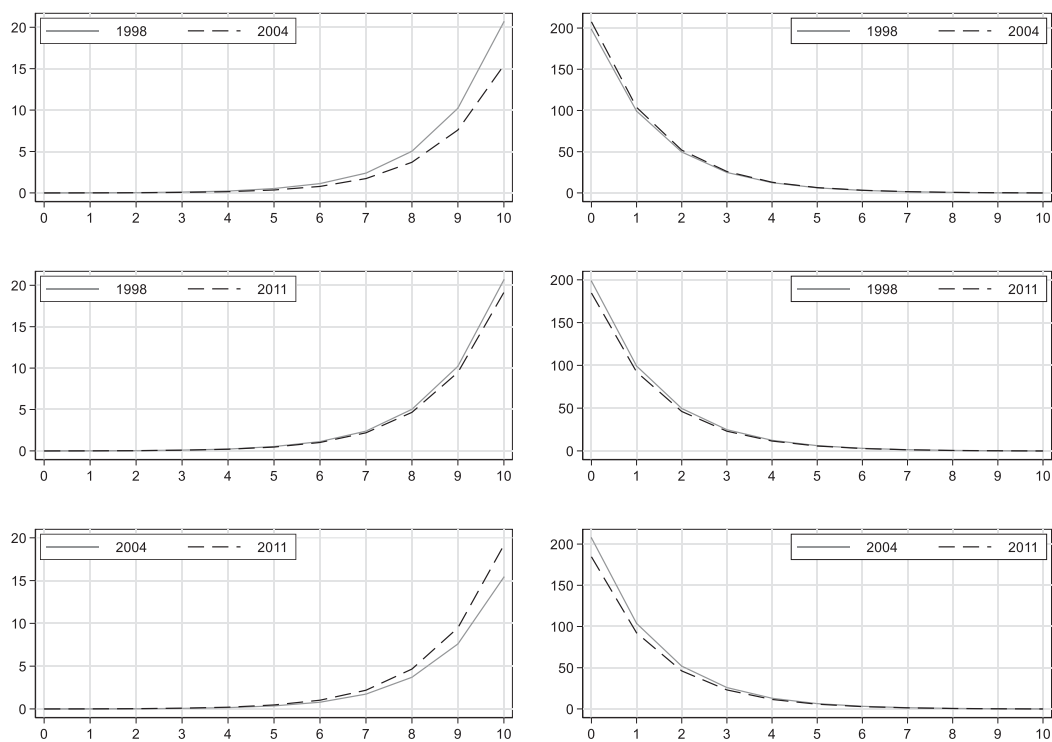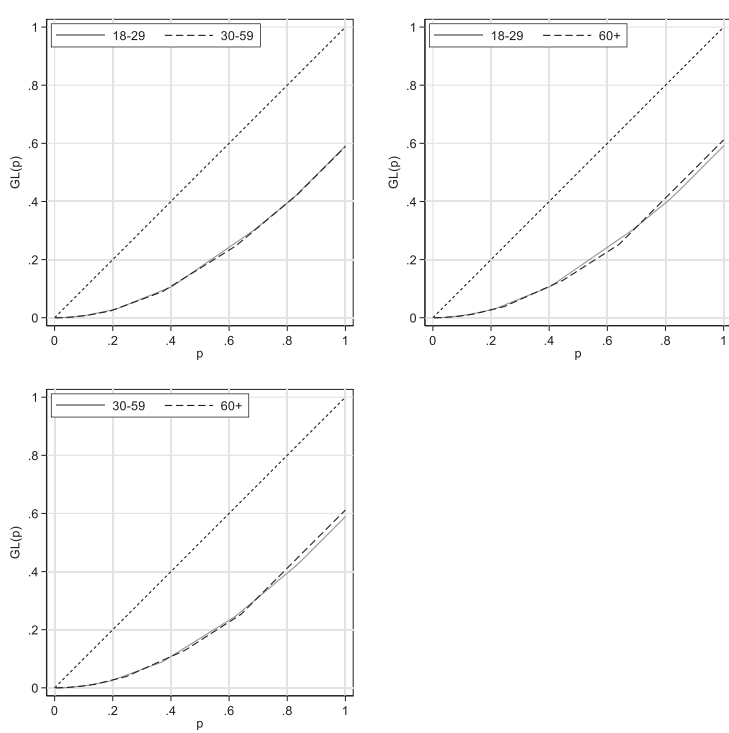
Figure A4. Checks for life satisfaction H^+ dominance (left) and H^- dominance (right).
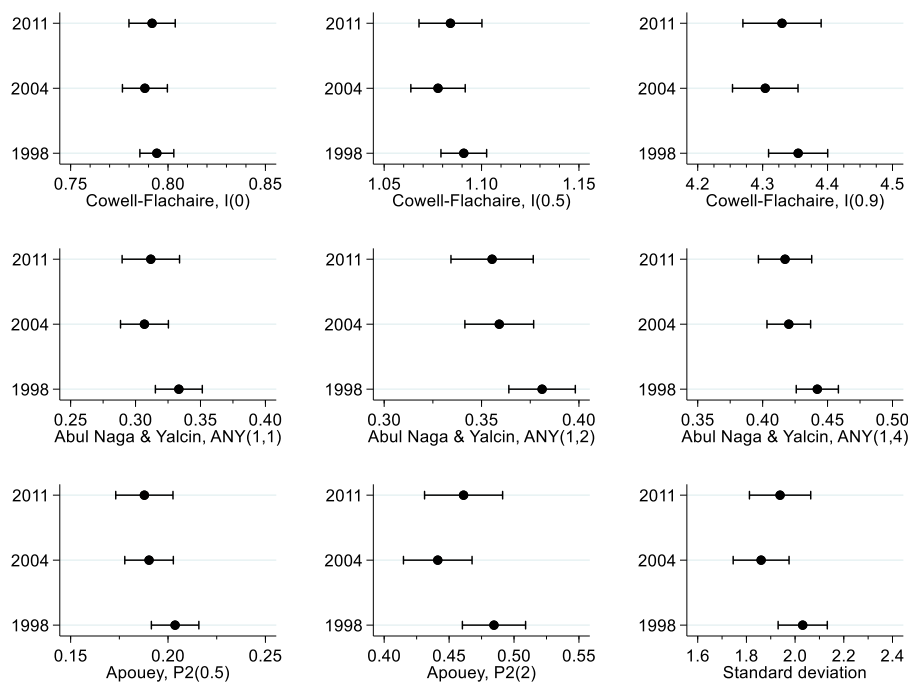
Figure A5. Indices of life satisfaction inequality and polarization.

(a) Men and women
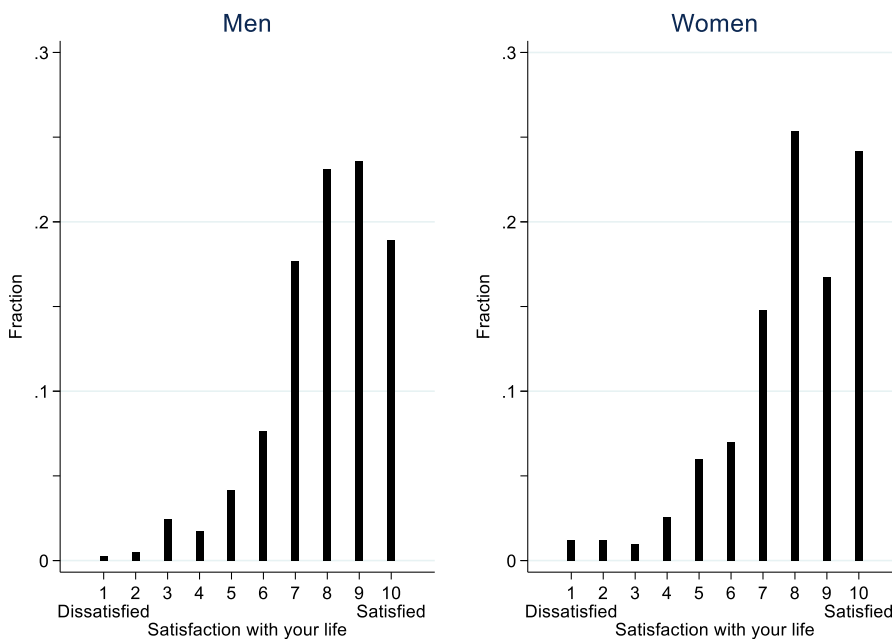
(b) Age group
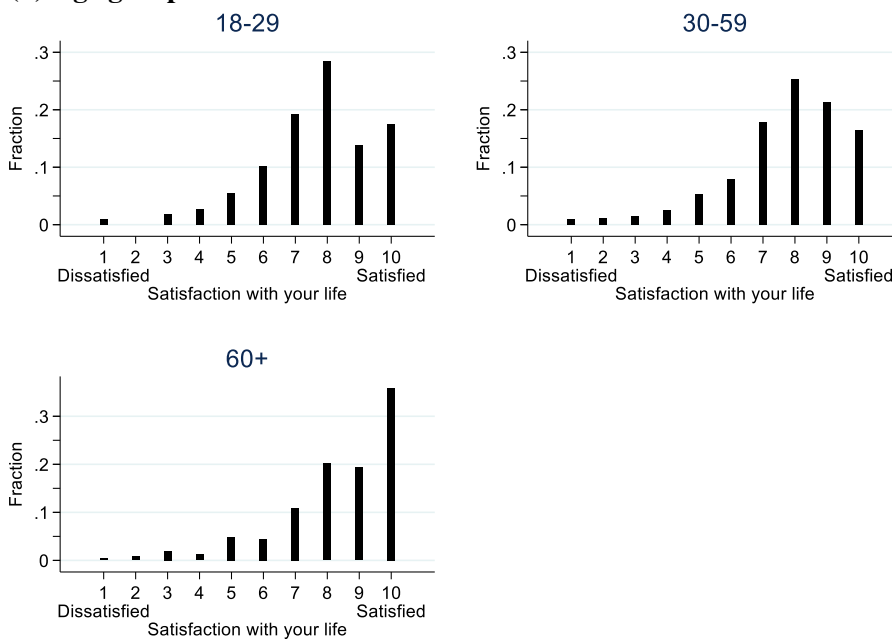
Figure B1. The distribution of life satisfaction in NZ, by sex and age group. (a) Men and women. Notes: The median is 8 and the mean is 7.9 for both men and women. (b) Age group. Notes: The median is 8 for individuals aged 18–29 or aged 30–59 years, and 9 for individuals aged 60+. The means are 7.7, 7.4, and 8.3, respectively.

(a) Men and women
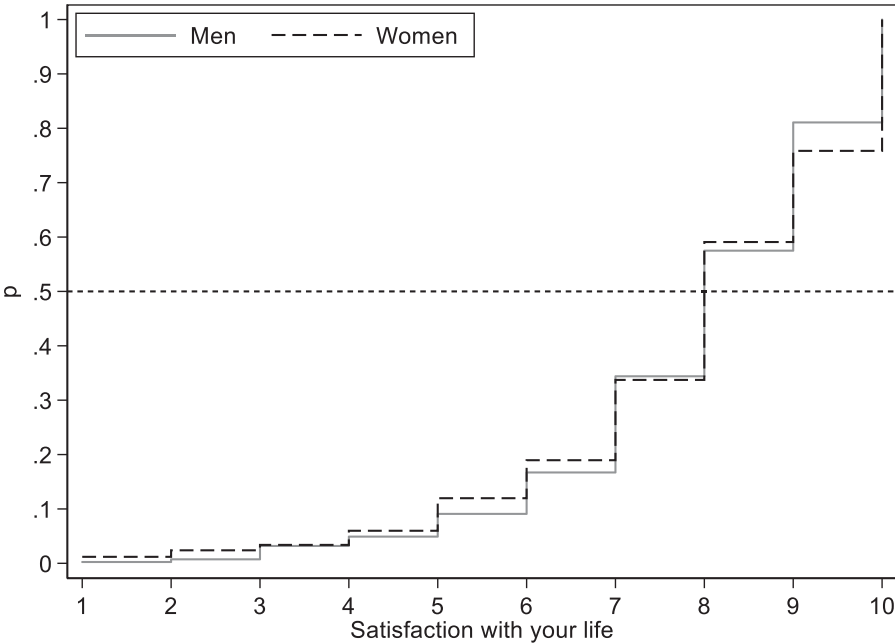
(b) Age group
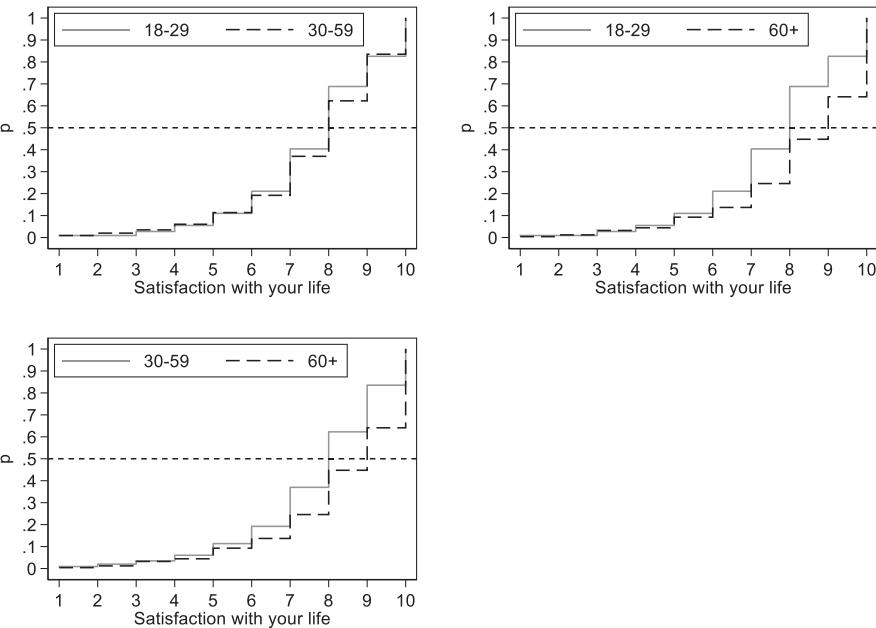
Figure B2. Cumulative distribution functions (CDFs) for life satisfaction, by subgroup. (a) Men and women. (b) Age group.
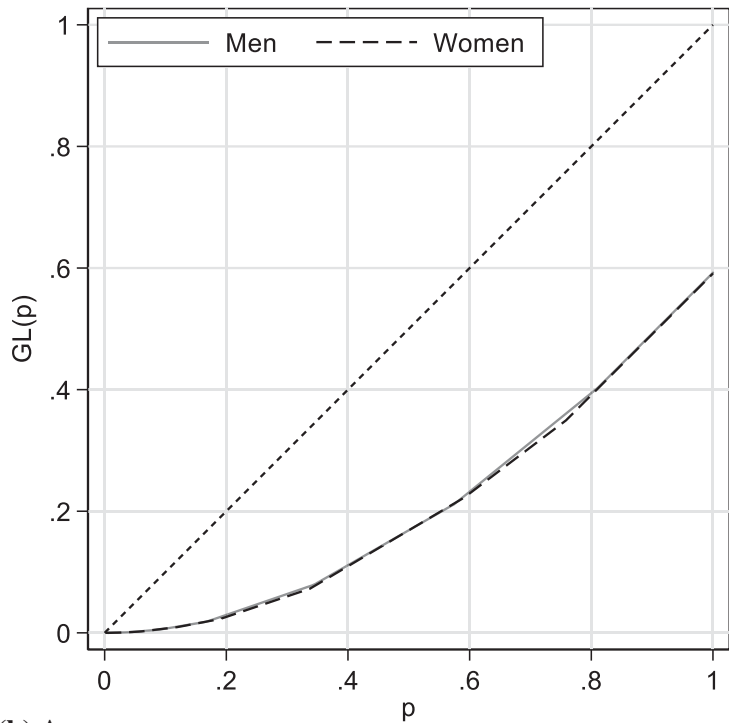
(a) Men and women
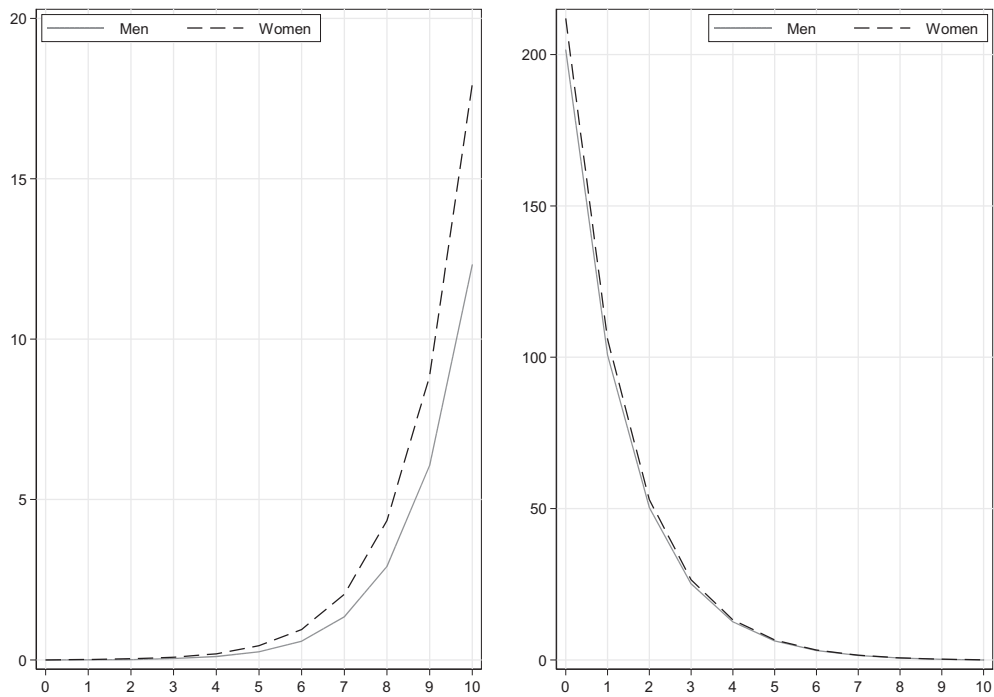

(b) Age group

Figure B3. Checks for dominance according to all $CF(\gamma)$ inequality indices using Generalized Lorenz curve comparisons of life satisfaction 'status' distributions. (a) Men and women. (b) Age group.
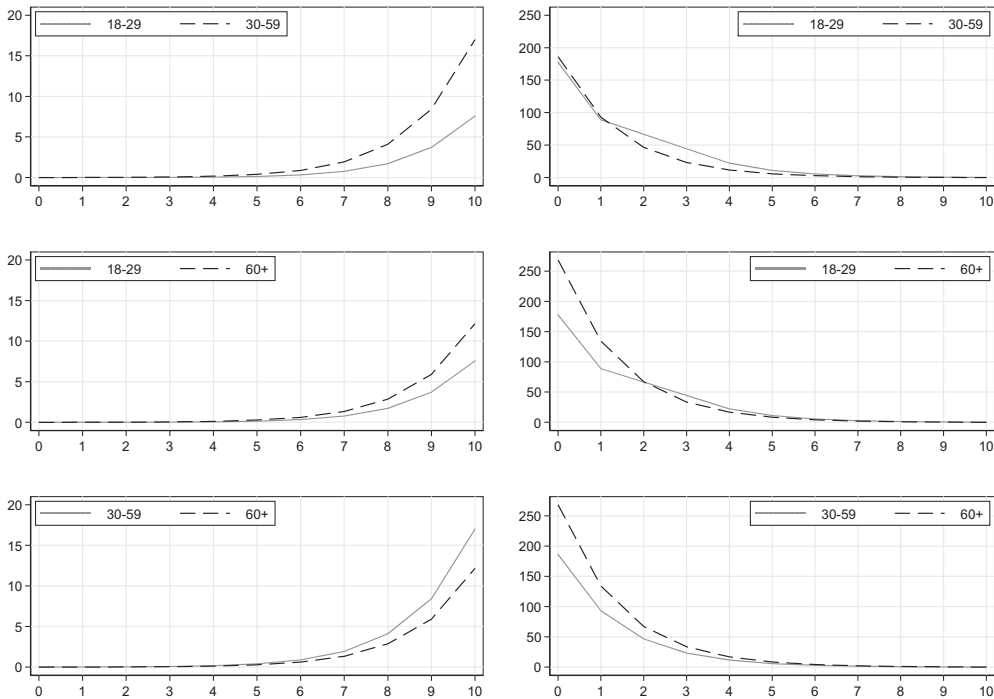
(a) Men and women

(b) Age group

Figure B4. Checks for life satisfaction H^+ dominance (left) and H^- dominance (right). (a) Men and women. (b) Age group.
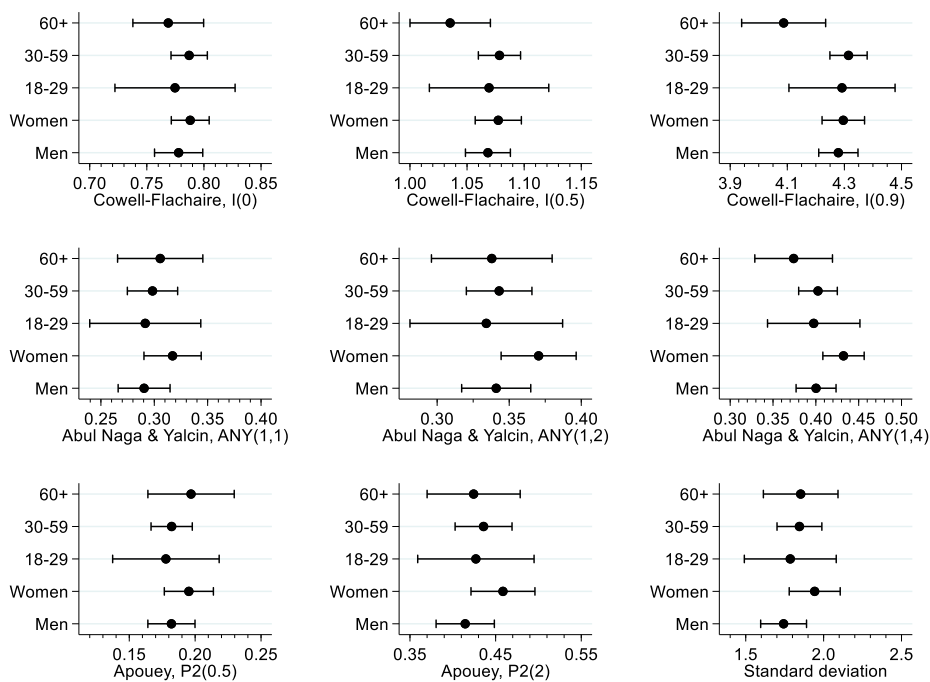
Figure B5. Indices of life satisfaction inequality and polarization, by subgroup.