

DISCUSSION PAPER SERIES

IZA DP No. 13057

Comparing Distributions of Ordinal Data

Stephen P. Jenkins

MARCH 2020

DISCUSSION PAPER SERIES

IZA DP No. 13057

Comparing Distributions of Ordinal Data

Stephen P. Jenkins

London School of Economics and Political Science and IZA

MARCH 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Comparing Distributions of Ordinal Data

To compare distributions of ordinal data such as individuals' responses on Likert-type scale variables summarizing subjective well-being, we should not apply the toolbox of methods developed for cardinal variables such as income. Instead we should use an analogous toolbox which takes account of the ordinal nature of the responses. This paper reviews these methods and introduces a new Stata command `ineqord` for undertaking distributional comparisons. As the empirical illustrations demonstrate, `ineqord` can be used for dominance checks as well as for estimation of indices of polarization and inequality.

JEL Classification: D31, D63, I31

Keywords: inequality, ordinal data, subjective well-being, life satisfaction, Annual Population Survey

Corresponding author:

Stephen P. Jenkins
Department of Social Policy
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom
E-mail: s.jenkins@lse.ac.uk

1. Introduction

This is a paper about how to compare distributions of personal well-being where well-being is measured using an ordinal scale, and it introduces a new Stata command `ineqord` for undertaking these comparisons.

Leading examples of personal well-being indicators are self-assessed ('subjective') life satisfaction or health status for which individuals provide responses on a Likert-type scale. For instance, regarding life satisfaction, respondents may be presented with a linear integer scale running from 0 to 10 (11 levels) and asked to respond to the question "overall, how satisfied are you with your life nowadays where 0 is 'not at all satisfied' and 10 is 'completely satisfied'?" (Data based on this scale are used in the Examples section below.) Other life satisfaction scales use five, seven, or ten levels. Some subjective well-being (SWB) scales employ a mixture of negative and non-negative integers to label the levels. For example, people are asked to rate how satisfied they are with their life, choosing between 'completely dissatisfied' (scaled as -3), 'mostly dissatisfied' (-2), 'somewhat dissatisfied' (-1), 'neither satisfied nor dissatisfied' (0), 'somewhat satisfied' (1), 'mostly satisfied' (2), and 'completely satisfied' (3).

SWB measures are increasingly being used in tandem with the monetary measures of personal economic well-being such as income or wealth that national and international statistical agencies and most researchers have conventionally focused on. A catalyst for the new emphasis was the *Report by the Commission on the Measurement of Economic Performance and Social Progress* (Stiglitz, Sen, and 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: 14), and that 'objective and subjective dimensions of well-being are both important' (2009: 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 2020), 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 introduced initiatives to address the Beyond GDP agenda, including a greater emphasis on collection of and reporting on SWB data.

Income and wealth are cardinal variables and there are well-established methods for comparing distributions of them in terms of levels and inequality. There are also many community-contributed Stata commands for undertaking distributional comparisons of

cardinal variables, including my `ineqdeco`, `ineqdec0`, `sumdist`, `svylorenz`, and `svy_atk` and `svy_gei` (with Martin Biewen) and `glcurve` (with Philippe Van Kerm), all available from SSC.

In contrast, SWB measures are ordinal in nature, which raises the question of how to undertake distributional comparisons in this situation. How do we assess whether average well-being or well-being inequality has increased over time or differs between countries or social groups? A growing literature (cited below) has shown on the one hand that it is inappropriate to apply comparison methods developed for cardinal well-being measures to ordinal SWB measures, although many researchers continue to do this – the *World Happiness Report* (Helliwell et al. 2019) is a leading example. However, on the other hand, there is now a toolbox of methods for application to ordinal data that is analogous to the toolbox long applied to distributions of cardinal variables such as income. See Jenkins (2019a) for development of this argument and illustrations. `ineqord` provides the means to implement methods that are appropriate for comparisons of distributions of ordinal data.

`ineqord` produces estimates of inequality and polarization indices: the Allison-Foster index, the normalized Average Jump index, multiple Apouey indices, multiple Abul Naga-Yalcin indices, multiple Cowell-Flachaire indices, Jenkins indices. Optionally, `ineqord` also derives estimates of cumulative distribution functions and related objects that can be used to describe ordinal distributions and to undertake dominance checks of differences between distributions.

`ineqord` assumes the user has respondent-level data with responses referring to ordinal well-being scores. If the user has grouped data describing the distribution of the well-being variable, the user needs first to construct a dataset using this information. See the Examples section for illustrations.

2. Comparisons of distributions of ordinal data

This section provides a brief overview of methods used for undertaking comparisons of distributions of ordinal data and discusses `ineqord`'s functionality against this background.

Let us suppose that we have individual-level SWB data held in a variable called `swb`. The inequality and polarization indices calculated by `ineqord` summarize dispersion in the distribution of responses across the levels of `swb`. There are $K \geq 3$ levels of the ordinal

variable in principle though one or more levels might receive no responses in practice, a situation to which I return below. The levels have numerical labels c_1, c_2, \dots, c_K , where $-\infty < c_1 < c_2 \dots < c_K < \infty$. The ‘linear integer’ scale is the one with $c_k = k$, for each $k = 1, 2, \dots, K$. The empirical distribution of responses is described by the proportion of the individuals who report the k^{th} level, f_k , for each k . The cumulative distribution function (CDF) is described by the proportion of individuals reporting the k^{th} level or lower, $F_k = \sum_{j=1}^k f_j$, for each k . The survivor function is described by the proportion of individuals reporting the k^{th} level or higher, $S_k = \sum_{j=k}^K f_j$, for each k . This is a non-standard definition – usually the survivor function is defined as $1 - F_k$ – but it is what is used to characterize a class of Cowell-Flachaire inequality indices (see Section 2.2).

A commonly-used measure of inequality of such ordinal data, especially life satisfaction and happiness data, is the standard deviation. Use of this measure is inappropriate because it assumes that **swb** is measured on a ratio scale. Kalmijn and Veenhoven (2005) acknowledge this issue but claim that the standard deviation is an appropriate measure nonetheless.

Economists specialising in inequality measurement have long been critical of the application to ordinal data of the standard deviation and other inequality indices typically applied to variables measured on a ratio scale. These indices use the mean as the reference point for assessing spread but, with ordinal data, the value of the mean is contingent on the scale used. Orderings of distributions according to their means or standard deviations are not robust to changes in the scale used.

Critiques by economists include the papers by Allison and Foster (2004), Cowell and Flachaire (2017), and Dutta and Foster (2013). These authors and others propose measures that respect the ordinal nature of the data. In one tradition, indices characterize greater inequality as greater spread about the median. The other tradition characterizes greater inequality as greater spread away from a maximum value.

2.1 Polarization indices

The Allison-Foster index is the difference between the mean score for respondents with scores above the median minus the mean score for respondents with scores below the median. This index was first proposed by Allison and Foster (2004). Dutta and Foster (2013) provide

more extensive discussion of it, and the formulae used by **ineqord** are based on their equations 1 and 2 (p. 398).

The two-parameter indices proposed by Abdul Naga and Yalcin (2008), $ANY(a, b)$, with $a, b \geq 1$, are a form of weighted difference between the cumulative percentages of individuals in the lower half of the distribution and the cumulative percentages in the upper half of the distribution. The parameters tune the weights given to the two halves. $ANY(1, 1)$ weights the two halves equally. Broadly speaking, when $b > a$, $ANY(a, b)$ gives greater weight to the bottom half of the distribution; when $a > b$, it gives greater weight to the top half of the distribution. According to Abul Naga and Yalcin (2008: 1621), ‘For a given value of β , ..., as $\alpha \rightarrow \infty$, the inequality index abstracts from the dispersion below the median’. (Their α and β correspond to my a and b .) On the other hand, when $b > a$, $ANY(a, b)$ gives greater weight to the bottom half of the distribution. For a given value of a , choosing larger and larger values of b places less weight on the distribution in categories above the median. In the limiting case when $b \rightarrow \infty$, only below-median categories are relevant. Thus, for example, the indices $ANY(1, 1)$, $ANY(1, 2)$, and $ANY(1, 4)$ respectively give increasingly greater weight to the lower half of the distribution when assessing overall polarization.

Apouey’s (2007) $P2(e)$ indices each aggregate the ‘distances’ between F_k and 0.5 (the value of F_k at the median) across the levels of **swb**. $P2(0.5)$ uses the square root of the absolute differences to summarize distance and $P2(1)$ uses a ‘city block’ (linear) distance function. $P2(2)$ uses a Euclidean distance metric and is the same as the the ‘1 – l-squared’ index of Blair and Lacy (2000). (The Blair-Lacy index may also be calculated using community-contributed command **ordvar** on SSC.) In general, the value of parameter e determines how concentration within the groups below the median and within the groups above the median contributes to overall polarization.

The Average Jump index is the average across respondents of the absolute difference between each observed value of **swb** and the median value, normalized by the maximum value for the index. For a linear integer scale, the Average Jump index equals the Allison-Foster index divided by the total number of levels of **swb** minus one (Allison and Foster 2004, p. 514). In this case, the index summarizes the (normalized) average number of category ‘jumps’ required to change from the observed level to the median level. For a linear integer scale, the Average Jump index is the same as the $ANY(1, 1)$ index and the $P2(1)$ index.

2.2 Inequality indices

Cowell and Flachaire (2017) build inequality measures from axiomatic first principles, providing two families of one-parameter indices based on downward-looking and upward-looking measures of individual ‘status’, respectively. **ineqord** uses the ‘peer-inclusive’ (rather than ‘peer-exclusive’) definitions of these, reflecting the focus of Cowell and Flachaire (2017) and other authors. For an individual reporting a response corresponding to the k^{th} level of the scale, peer-inclusive downward-looking status is given by F_k and peer-inclusive upward-looking status is given by S_k . The inequality indices aggregate ‘distances’ between each individual’s status and the maximum possible status value (which is one, given their definition of status).

Members of the two Cowell-Flachaire inequality index families $I(\alpha)$, are distinguished by parameter α which encapsulates the sensitivity of overall inequality to the dispersion of individual status in different ranges of the status distribution, with $0 \leq \alpha < 1$. The smaller that α is, the more sensitive is the overall index to differences in status at the bottom of the status distribution rather than at the top. If the distribution of responses on **swb** is symmetric across the levels, $F_k = S_k$ and each downward-looking Cowell-Flachaire index has the same value as its upward-looking counterpart with the same α .

Jenkins’s (2019b) J_d index is defined for Cowell and Flachaire’s peer-inclusive downward-looking status measure and his J_u index for their peer-inclusive upward-looking status measure. Each index is equal to the area between the Generalized Lorenz curve for the relevant status distribution and the Generalized Lorenz curve for the distribution with no status inequality (in which case the Generalized Lorenz curve is a straight line between the origin and point (1,1)), divided by the total area beneath the perfect equality curve (= 0.5). Equivalently, each index is equal to one minus twice the area beneath the Generalized Lorenz curve for status. The Generalized Lorenz curve for status, $GL(p)$, plots cumulative status per capita against cumulative population share, $0 \leq p \leq 1$, of individuals ranked in ascending order of status. $GL(0) = 0$ and $GL(1)$ is the arithmetic mean of status. See Jenkins (2019b) for details.

2.3 Index properties

All of the polarization and inequality indices calculated by **ineqord** equal their minimum value, zero, if all respondents report the same value for **swb**. The Allison-Foster, Average Jump, Apouey, and Abul Naga and Yalcin indices each summarize polarization of responses relative to the median. These indices reach their maximum value when half the responses on **swb** refer to the minimum value of the scale and half the responses refer to the maximum value, i.e. the distribution of responses is totally polarized. In this case, the maximum value equals one – except for the Allison-Foster index, for which the maximum value depends on the number of categories.

Cowell-Flachaire $I(\alpha)$ and J indices need not reach a maximum value with this distribution of responses: this is because the indices summarize inequality as spread rather than as polarization. For example, for any given K , $I(\alpha)$ and J indices record greater inequality for a uniform distribution than for a totally polarized distribution (Jenkins 2019b).

$I(\alpha)$ and J indices are invariant to order-preserving transformations of the ordinal scale variable, i.e. scale independent. The Allison-Foster index is not scale independent and hence Dutta and Foster (2013) provide estimates based on linear, convex, and concave scales in their empirical application. Abul Naga-Yalcin and Apouey indices are scale independent (but also see the remarks in Section 2.5).

2.4 Dominance checks for unanimous orderings by classes of indices

ineqord also provides users with the ability to undertake dominance checks. In general, dominance means that finding an appropriately defined graph for one distribution lies everywhere on or above the corresponding graph for another distribution is equivalent to a unanimous ranking of the two distribution by all measures satisfying a specific set of properties. There are several different types of dominance in this context.

Allison and Foster (2004) provide results for ‘ F -dominance’ and ‘ S -dominance’. The former refers to comparisons of CDFs and rankings by average well-being levels (first-order dominance): if the CDF for distribution A lies everywhere on or above the CDF below that for distribution B , then A has higher average well-being than B , regardless of scale. S -dominance (spread dominance) refers to comparisons of S -curves, which are derived from CDFs and so the criterion can also be expressed in terms of these. That is, if A and B have the

same median, and the CDF for A lies above that for B at scale values below the median but above that for B at scale values at the median and above, all polarization indices respecting the property that greater spread about the median corresponds to greater polarization will show A as having greater polarization than B . S -dominance can only arise if the pair of distributions have a common median and if there is no F -dominance.

Jenkins (2019b) shows that, for each of the two Cowell-Flachaire definitions of status, if the Generalized Lorenz curve for status distribution A lies nowhere above the Generalized Lorenz curve for status distribution B , all Cowell-Flachaire $I(\alpha)$ indices and the J index will record A as having more inequality than B . These Generalized Lorenz curve comparisons can be applied if the distributions have different medians.

ineqord can also be used to undertake the H -dominance checks proposed by Gravel, Magdalou, and Moyes (2020). These authors start from the principle that the inequality of an ordinal variable increases if there is a shift in density mass away from a specific level (one person moving up a level and one moving down). This is the concept of a disequalizing ‘Hammond transfer’ (cf. the concept of a disequalizing Pigou-Dalton transfer for a cardinal variable such as income). Gravel et al. (2020) define H^+ and H^- curves (called H and \bar{H} curves in their paper) which are specifically-defined recursive cumulations of CDFs (just as Generalized Lorenz curves are, but differently defined). The authors prove a dual dominance result: distribution A being more equal than distribution B according to the Hammond transfer concept is equivalent to finding (i) the H^+ curve for A lying nowhere above the H^+ curve for B and (ii) the H^- curve for A lying nowhere above the H^- curve for B . They also show that, if there is F -dominance, there is also H^+ dominance. The dual dominance check can be applied if the distributions have different medians.

Gravel et al. (2020) do not refer to any existing indices when discussing their dual dominance criteria. The relationships between the dual dominance and Generalized Lorenz dominance criteria are a topic of current research.

2.5 Some computational and conceptual issues

For correct calculation of the Abul Naga-Yalcin, Average Jump, Apouey, and Jenkins indices, **ineqord** must know the total number of possible levels of the ordinal response variable. This number may be greater than the maximum observed in the data, for example, if there are no responses on some scale values or if there is total polarization. By default,

ineqord assumes that the total number of possible levels of the ordinal response is the number of levels observed to contain responses. If this assumption is incorrect, it is the user's responsibility to specify the maximum number of levels of response using the **nlevels** option, described below. See also the discussion of scale dependence below.

Apouey $P2(e)$ indices refer to the case in which the ordered response categories are labelled with positive integers (1 for the lowest level, 2 for the second lowest, etc.), which is a linear integer scale. For correct calculation of these indices, it is the user's responsibility to check that the scale underlying **swb** is appropriate. Optionally, **ineqord** relabels the observed responses to calculate the Apouey indices using a linear transformation:

response_new = response - minlevel + 1, where **minlevel** is the value specified by the **minlevel** option and **response** in this case would be **swb**. For example, using the option, the life satisfaction scale cited above (0, 1, ..., 10) is converted to (1, 2, ..., 11) by setting **minlevel** = 0. Scale (-1, 0, 1) is converted to (1, 2, 3) by setting **minlevel** = -1. Be aware that if the response scale values were instead (2, 4, 6), say, and the user sets **minlevel** = 2, **ineqord**'s calculation would be based on transformed responses (1, 3, 5), not (1, 2, 3), and correct calculation of the Apouey (and J) indices would also require setting the maximum number of levels to 5 using the **nlevels** option. Calculation of the indices would assume that scale values 2 and 4 are possible (and this is relevant to the assessment of how polarized **swb** is), but there are no responses observed for them. On related issues, see the discussion of the 'mergers principle' by Cowell and Flachaire (2017).

The precise definition of the median is fundamental to the estimation of polarization indices. I use Stata's definition of the median, as set out in the Methods and Formula section of the entry for **summarize** in the [R] Base Reference manual, with one rarely-used modification.

There are other possible definitions of the median. For example, Abul Naga and Yalcin's definition is that level ' m is the median ... if $P_{m-1} \leq 0.5$ and $P_m \geq 0.5$ ' (2008: 1616), where P_k is the fraction of individuals reporting level k or less, i.e. what I have referred to as F_k . The definition means that the median is undefined if the fraction reporting the lowest level $k = 1$ is greater than one half ($P_1 > 0.5$) though, of course, this case is likely to be rare in practice. Cowell and Flachaire (2017: 300) discuss other potential issues and refer to them when motivating their non-median-based approach.

Although use of Stata's definition of the median almost invariably works well in real-world situations, there is one tricky special case to deal with – the situation in which Stata

reports a non-integer median (having taken the average value of the scale in two adjacent categories – see the [R] Base Reference manual again). This is most likely to occur if there are scale levels in the middle of the range that do not receive any responses. Using a non-integer median ‘as is’ leads to an error when calculating *ANY* indices using Abul Naga and Yalcin’s (2008) formulae. Thus the code for **ineqord** uses Stata’s definition by default except that, in the (rarely experienced) non-integer median case, it applies the **ceil()** function to the non-integer median and then proceeds using the revised (integer-valued) median. If **ceil()** changes the median, **r(newmedian)** differs from **r(median)** in the stored results. With this adjustment, **ineqord** generates the estimates expected.

Bootstrapped standard errors for the indices can be derived using **bootstrap** or, for example, **rhsbbsample** (Van Kerm 2013) implementing Saigo et al.’s (2001) repeated-half sample bootstrap approach. See the Examples section below. Analytical formulae for variance estimates exist for some of the indices and curve ordinates but not for all of them, and the formulae that are provided do not account for sample design features such as weights, clustering, or stratification.

Finally, note that the indices and the dominance results cited earlier refer to levels and dispersion of a categorical well-being variable with an arbitrary scale. They do not refer to levels and dispersion of some underlying unobserved SWB variable. This is an important distinction because it is often assumed that discrete categorical responses on a Likert-type scale are manifestations of a latent continuous variable. For example, Delhey and Kohler’s (2011) adjustment to the standard deviation measure to account for the bounded nature of a Likert-type scale, implemented in **sdlim** on SSC, refers to a latent SWB variable. Stevenson and Wolfers (2008) suppose that the ordinal data responses are realisations of a latent continuous well-being variable which is assumed to be normally distributed within a population, with moments of the latent variable estimated using ordinal regression techniques. Bond and Lang (2019) emphasize the distinction between manifest categorical and latent continuous SWB variables and they highlight the strong assumptions required to identify distributions of the latter from the former. More positively, Kaplan and Zhuo (2019) provide some results about what can be learnt about latent SWB distributions when manifest categorical distributions are available.

3. Syntax

This section describes the syntax of the **ineqord** command. The command works with Stata version 14 or later.

```
ineqord varname [weight] [if] [in]  
    [, alpha(real) nlevels(integer) minlevel(integer)  
    ustatusvar(string) dstatusvar(string)  
    catvals(string) catprops(string) catcprops(string) gldvar(string)  
    catsprops(string) gluvar(string) hplus(string) hminus(string) ]
```

3.1 Options

alpha(#) Use this to calculate an additional Cowell-Flachaire index with parameter value α . The value must be between 0 and 1.

nlevels(#) Use this to specify the total number of possible levels of the ordinal response variable. Required for correct calculation of the Apouey indices if the observed number of levels is less than the maximum possible.

minlevel(#) Use this to specify the minimum level of the ordinal response variable. Required for correct calculation of the Blair-Lacy and Apouey indices if the observed minimum is not equal to 1.

ustatusvar(*string*) Use this to save the Cowell-Flachaire ‘upward-looking’ status variable after calculation.

dstatusvar(*string*) Use this to save the Cowell-Flachaire ‘downward-looking’ status variable after calculation.

catvals(*string*) Use this to save the distinct values of the response variable after calculation. There is one value per level.

catprops(*string*) Use this to save the sample proportions for each level after calculation.

catcprops(*string*) Use this to save the sample cumulative proportions after calculation.

catsprops(*string*) Use this to save the sample cumulative survivor proportions after calculation.

gldvar(*string*) Use this to save the Generalized Lorenz ordinates of the Cowell-Flachaire ‘downward-looking’ status variable after calculation.

gluvar(*string*) Use this to save the Generalized Lorenz ordinates of the Cowell-Flachaire ‘upward-looking’ status variable after calculation.

hplus(*string*) Use this to save the ordinates of H^+ curve after calculation.

hminus(*string*) Use this to save the ordinates of H^- curve after calculation.

by and **statsby** are allowed; so too are **aweight**s, **fweight**s, **pweight**s, and **iweight**s.

ustatusvar and **ustatusvar** are observation-level variables. The **cat*** variables as well as the Generalized Lorenz and H curve ordinates are category-level variables (with K values). See the examples in Section 4.

3.2 Stored results

ineqord stores the following in **r()**:

Scalars

r(n_distinct_cats)	Number of distinct levels in <i>varname</i>
r(mean)	Mean of <i>varname</i>
r(mean_rescaled)	Mean of <i>varname</i> if minlevel() option applied
r(median)	Median of <i>varname</i>
r(median_rescaled)	Median of <i>varname</i> if minlevel() option applied
r(newmedian)	ceil(Median of <i>varname</i>); see discussion in text
r(Var)	Variance of <i>varname</i>
r(sd)	Standard deviation of <i>varname</i>
r(sumw)	Sum of weights
r(N)	Number of observations on <i>varname</i>
r(min)	Minimum value of <i>varname</i>
r(max)	Maximum value of <i>varname</i>
r(min_rescaled)	Minimum value of <i>varname</i> if minlevel() option applied
r(max_rescaled)	Maximum value of <i>varname</i> if minlevel() option applied
r(nlevels)	Number of levels specified in nlevels() option

r(minlevel)	Value of minimum level specified in minlevel() option
r(dfmeanabove)	Mean of <i>varname</i> for obs above the median (Allison-Foster definition)
r(s_H)	Mean of <i>varname</i> for obs above the median minus the median of <i>varname</i>
r(dfmeanbelow)	Mean of <i>varname</i> for obs below the median (Allison-Foster definition)
r(s_L)	Median of <i>varname</i> minus mean of <i>varname</i> for obs above the median
r(allisonfoster)	Allison-Foster index
r(avjump)	Average Jump index
r(apoueypt5)	Apouey $P2(0.5)$ index
r(apouey1)	Apouey $P2(1)$ index
r(apouey2)	Apouey $P2(2)$ index
r(blairlacy)	Blair-Lacy index ('1 - l-squared') = $P2(2)$
r(any11)	Abul Naga-Yalcin (1, 1) index
r(any21)	Abul Naga-Yalcin (2, 1) index
r(any12)	Abul Naga-Yalcin (1, 2) index
r(any41)	Abul Naga-Yalcin (4, 1) index
r(any14)	Abul Naga-Yalcin (1, 4) index
r(i0d)	Cowell-Flachaire downward-looking index ($\alpha = 0$)
r(ioneqd)	Cowell-Flachaire downward-looking index ($\alpha = 0.25$)
r(ihalfd)	Cowell-Flachaire downward-looking index ($\alpha = 0.5$)
r(ithreeqd)	Cowell-Flachaire downward-looking index ($\alpha = 0.75$)
r(ixd)	Optional Cowell-Flachaire downward-looking index ($\alpha = \#$)
r(i0u)	Cowell-Flachaire upward-looking index ($\alpha = 0$)
r(ionequ)	Cowell-Flachaire upward-looking index ($\alpha = 0.25$)
r(ihalfu)	Cowell-Flachaire upward-looking index ($\alpha = 0.5$)
r(ithreequ)	Cowell-Flachaire upward-looking index ($\alpha = 0.75$)

<code>r(ixu)</code>	Optional Cowell-Flachaire upward-looking index ($\alpha = \#$)
<code>r(Jd)</code>	Jenkins downward-looking index
<code>r(Ju)</code>	Jenkins upward-looking index

Locals

<code>r(cats_list)</code>	List of levels of <i>varname</i>
---------------------------	----------------------------------

4. Examples

This section illustrates **ineqord** in action. For further examples of distributional comparisons based on numerical indices and dominance checks applied to World Values Survey (WVS) data on life satisfaction, see Jenkins (2019a, 2019b). These WVS data were also used by Cowell and Flachaire (2017), and **ineqord** produces the same estimates that they report – they focus on $I(0)$ using both peer-inclusive downward- and upward-looking definitions of status.

4.1 Life satisfaction data from the UK Annual Population Survey (APS)

Most of my examples are based on data about life satisfaction drawn from the ‘Annual Population Survey Three-Year Pooled Dataset January 2015 - December 2017’ (Office for National Statistics, Social Survey Division 2019), a nationally-representative survey of UK adults. The data and documentation are downloadable from the UK Data Service by researchers who register with them. For brevity I refer to the data as ‘the APS’. Data drawn from the APS are used by the UK’s Office for National Statistics to provide annual reports on personal well-being: see for example, Office for National Statistics (2019).

The dataset contains 530,300 (unweighted) observations of which 275,336 provide a non-missing response to the life satisfaction question set out in Section 1. Responses are held in the variable named **SATIS**. Sample weights are provided in variable **PWTA17C**. Missing values are recorded as values of -8 and -9 and all variables names are in upper case. Before any use of **ineqord**, I convert the missing values to Stata missing values and, for convenience, put all variable names in lower case.

```
. use aps_3yr_jan15dec17_eul, clear
```

```
. rename _all, lower
```

```
. ta satis
```

Satisfied with your life	Freq.	Percent	Cum.
Does not apply	252,574	47.63	47.63
No answer	2,390	0.45	48.08
not at all satisfied	1,777	0.34	48.41
1	886	0.17	48.58
2	2,124	0.40	48.98
3	3,165	0.60	49.58
4	5,246	0.99	50.57
5	19,531	3.68	54.25
6	18,684	3.52	57.77
7	48,054	9.06	66.84
8	90,366	17.04	83.88
9	42,993	8.11	91.98
completely satisfied	42,510	8.02	100.00
Total	530,300	100.00	

```
. replace satis = .a if satis == -8  
(2,390 real changes made, 2,390 to missing)
```

```
. replace satis = .b if satis == -9  
(252,574 real changes made, 252,574 to missing)
```

```
. lab define SATIS .b "Does not apply" .a "No answer", modify
```

```
. ta satis
```

Satisfied with your life	Freq.	Percent	Cum.
not at all satisfied	1,777	0.65	0.65
1	886	0.32	0.97
2	2,124	0.77	1.74
3	3,165	1.15	2.89
4	5,246	1.91	4.79
5	19,531	7.09	11.89
6	18,684	6.79	18.67
7	48,054	17.45	36.13
8	90,366	32.82	68.95
9	42,993	15.61	84.56
completely satisfied	42,510	15.44	100.00
Total	275,336	100.00	

The weighted distribution of valid responses is as follows:

```
. ta satis [aw = pwtal7c]
```

Satisfied with your life	Freq.	Percent	Cum.
not at all satisfied	1,707.0425	0.62	0.62
1	846.473904	0.31	0.93
2	2,051.4235	0.75	1.67
3	3,103.0005	1.13	2.80
4	5,037.9697	1.83	4.63
5	19,173.955	6.97	11.61
6	19,141.885	6.96	18.57
7	49,671.19	18.06	36.63
8	90,855.553	33.03	69.66
9	42,814.464	15.57	85.23
completely satisfied	40,628.044	14.77	100.00
Total	275,031	100.00	

Only around 12% of respondents report a value of 5 or lower on the 0 to 10 scale. Almost 7% report that they are completely satisfied with their life (scale point 10), with the modal value equal to 8.

To proceed further, we have to address the fact that the linear integer scale runs from 0 to 10. It does not start at 1. If `ineqord` were applied ignoring this, it would provide incorrect estimates for some of the indices. There are two ways to proceed: either (i) create a new variable to ensure the scale goes from 1 to 11 and then run `ineqord` using this variable; or (ii) run `ineqord` using its `minlevel(0)` option and the variable `satis`. To implement strategy (i), I create a new variable named `ls`:

```
. ge ls = satis + 1
(254,964 missing values generated)

. lab var ls "= satis + 1"

. lab def ls 1 "Not at all satisfied: 1" 11 "Completely satisfied: 11" ///
>          .b "Does not apply" .a "No answer"

. lab val ls ls
```

Applying strategy (ii), we derive the following estimates for the UK adult population. It is easily verified that the code `ineqord ls [aw = pwta17c]` gives exactly the same estimates as those shown, whereas `ineqord satis [aw = pwta17c]` gives incorrect estimates (output not shown).

```
. ineqord satis [aw = pwta17c], minlevel(0)
Note: satis rescaled for calculation of Apouey indices (see help file)
Warning: summary statistics for rescaled responses differ from those for observed responses
```

Summary statistics for observed levels

All obs	min	max	# levels	median
	0	10	11	8

Mean, variance, and standard deviation of observed levels

All obs	mean	variance	sd
	7.67653	3.12008	1.76638

Polarization indices: Allison-Foster; Average Jump; Apouey P2(2); Apouey P2(1); Apouey P2(.5)

All obs	A-F	Av. Jump	P2(2)	P2(1)	P2(.5)
	2.45139	0.24514	0.37056	0.24514	0.14555

Polarization indices: Abul Naga-Yalcin(a,b)

All obs	ANY(1,1)	ANY(2,1)	ANY(1,2)	ANY(4,1)	ANY(1,4)
	0.24514	0.21218	0.28417	0.31366	0.34238

Inequality indices: Cowell-Flachaire, downward-looking status

All obs	I(0)	I(.25)	I(.5)	I(.75)
	0.76610	0.82952	1.04890	1.81626

Inequality indices: Cowell-Flachaire, upward-looking status

All obs	I(0)	I(.25)	I(.5)	I(.75)
	0.66864	0.76828	1.00823	1.78788

Inequality indices: J_d (downward-looking status); J_u (upward-looking status)

All obs	Jd	Ju
	0.56744	0.55128

The first components of the output provide descriptive statistics. For example, we see that the median response is 8 on the 0–10 scale. The Average Jump index estimate is 0.24514.

Because we have a linear integer scale, the average number of category ‘jumps’ required to change from the observed level to the median level (normalised by the total number of levels minus one = 10) is 0.24514, and is also equal to the estimates of *ANY*(1, 1) and *P2*(1) in this case. The Allison-Foster index value, 2.45139, is ten times the Average Jump index.

The earlier tabulation of **satis** shows dispersion in life satisfaction responses, and this is reflected in estimates of the indices of polarization and inequality that are greater than zero. The specific values of the estimates are otherwise hard to interpret; they become more valuable when there are estimates from multiple distributions that can be compared.

Let us therefore proceed to some distributional comparisons, considering how life satisfaction distributions differ between UK adults according to their marital status. I create a new variable **mstat** collapsing the information held in the **masta** variable. I treat individuals in a cohabiting relationship as married.

```
. ge mstat = .
(530,300 missing values generated)

. lab var mstat "Marital status"

. replace mstat = 1 if marsta == 1
(234,840 real changes made)

. replace mstat = 2 if marsta == 2
(219,411 real changes made)

. replace mstat = 3 if inlist(marsta, 3, 4, 5)
```

```
(75,081 real changes made)

. replace mstat = 4 if marsta == 6
(968 real changes made)

. lab def mstat 1 "Single, never married" ///
>      2 "Married, living with spouse" ///
>      3 "Separated, divorced, or widowed" ///
>      4 "Other (current/prev civil partnership)"

. lab val mstat mstat

. ta mstat [aw = pwtal7c]
```

Marital status	Freq.	Percent	Cum.
Single, never married	248,085.17	46.82	46.82
Married, living with spouse	210,891.57	39.80	86.62
Separated, divorced, or widowed	69,898.505	13.19	99.81
Other (current/prev civil partnership)	996.753415	0.19	100.00
Total	529,872	100.00	

A first look at the distributions of life satisfaction broken down by marital status indicates that married individuals are more satisfied than single, never-married (SNM) or separated, divorced, or widowed (SDW) individuals. (In what follows, I ignore the ‘other’ group given their small size.) For convenience I use the rescaled variable **ls** henceforth (rather than **satis**).

```
. ta ls mstat [aw = pwtal7c], col nofreq
```

= satis + 1	Marital status				Total
	Single, n	Married,	Separated	Other (cu	
Not at all satisfied:	0.66	0.29	1.31	0.00	0.62
2	0.36	0.16	0.57	0.85	0.31
3	0.92	0.38	1.33	0.87	0.75
4	1.51	0.61	1.78	0.46	1.13
5	2.20	1.09	3.02	1.53	1.83
6	7.87	4.76	10.77	4.44	6.97
7	8.72	5.19	8.60	5.83	6.96
8	22.13	15.84	17.61	19.69	18.06
9	32.24	35.21	29.20	33.35	33.03
10	12.45	18.81	12.39	17.86	15.57
Completely satisfied:	10.93	17.65	13.41	15.12	14.77
Total	100.00	100.00	100.00	100.00	100.00

4.2 Dominance checks

I begin by reporting dominance checks rather than indices, for two reasons. First, from a robustness point of view, it is useful to know if a pair of distributions can be unanimously ranked by all indices of a given family sharing key common characteristics. Even if you and I disagree about which is the best index within the family, but there is dominance, you and I will agree about how to rank a pair of distributions – though of course we may disagree about

the magnitudes of differences. Second, since dominance checks are usually implemented using graphs, using them is also a way of ‘showing the data’.

All the raw materials for the various dominance checks can be created by **ineqord** using the ‘cat*’, ‘gl*’, and ‘h*’ options shown in the syntax diagram. When **ineqord** runs using these options, it creates new variables that can be **listed** or displayed graphically. To compare distributions of life satisfaction by marital status group, I run the following code. The output for the indices is not shown here but is summarized later.

```
.           // single, never married
. ineqord ls [aw = pwtal7c] if mstat == 1, alpha(.9) ///
>         catv(v_snm) catpr(f_snm) catcpr(F_snm) catspr(S_snm) ///
>         gldvar(gld_snm) gluvar(glu_snm) hplus(hp_snm) hminus(hm_snm)

<output not shown>

.           // married, living with spouse
. ineqord ls [aw = pwtal7c] if mstat == 2, alpha(.9) ///
>         catv(v_m) catpr(f_m) catcpr(F_m) catspr(S_m) ///
>         gldvar(gld_m) gluvar(glu_m) hplus(hp_m) hminus(hm_m)

<output not shown>

.           // Separated, divorced, or widowed
. ineqord ls [aw = pwtal7c] if mstat == 3, alpha(.9) ///
>         catv(v_sdw) catpr(f_sdw) catcpr(F_sdw) catspr(S_sdw) ///
>         gldvar(gld_sdw) gluvar(glu_sdw) hplus(hp_sdw) hminus(hm_sdw)

<output not shown>
```

There is a listing of the values of new variables for married individuals below. Going from left to right, there are scale labels followed by the estimates of the density function, the CDF (with estimates corresponding to those shown by the earlier **tabulate** command), the survivor function, Generalized Lorenz ordinates for (peer-inclusive) downward-looking status, and H^+ and H^- ordinates, respectively. **ineqord** creates the zeros in the bottom row by default in order to facilitate drawing of Generalized Lorenz and H curves.

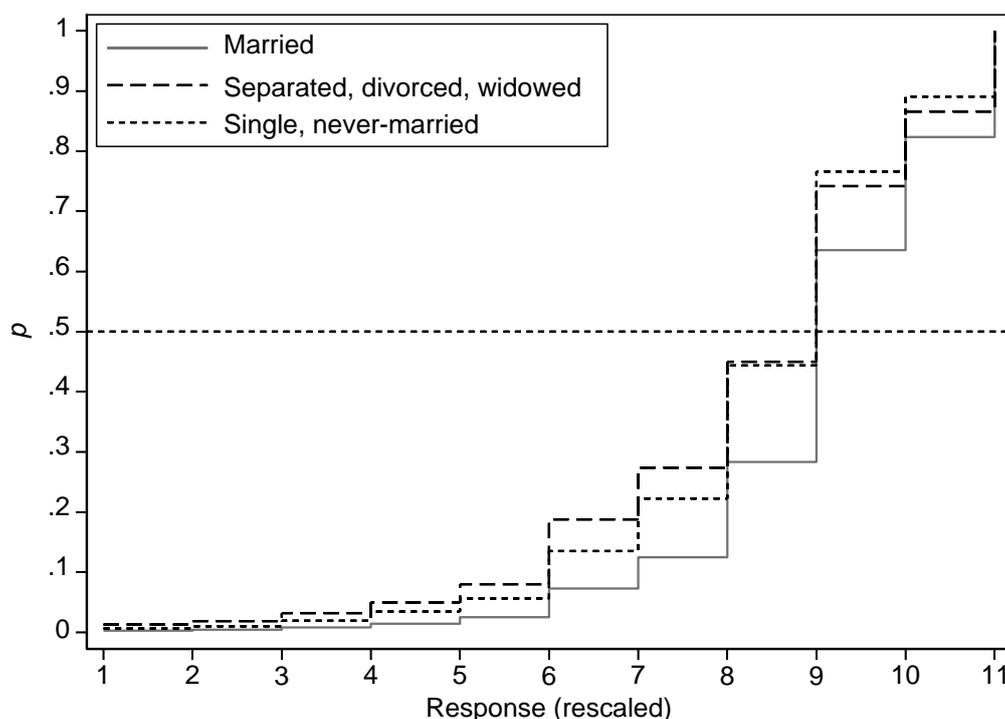
```
. sort v_m
. list v_m f_m F_m S_m gld_m hp_m hm_m if !missing(F_m), noobs
```

v_m	f_m	F_m	S_m	gld_m	hp_m	hm_m
1	.0029368	.0029368	1	8.62e-06	.0029368	292.5553
2	.0015666	.0045034	.9970632	.0000157	.0074403	146.2769
3	.0038127	.0083161	.9954966	.0000474	.0186932	73.13654
4	.0061448	.014461	.9916838	.0001362	.0435313	36.5652
5	.0108831	.0253441	.985539	.0004121	.0979457	18.27716
6	.0476428	.0729869	.9746559	.0038894	.2435341	9.114757
7	.0519418	.1249287	.9270132	.0103784	.5390101	4.531407
8	.1584475	.2833762	.8750713	.0552786	1.236468	2.18648
9	.352071	.6354473	.7166238	.2790012	2.825006	.9172045
10	.188099	.8235463	.3645527	.4339095	5.838112	.3645527
11	.1764537	1	.1764537	.6103632	11.85268	0
.	.	0	0	0	0	0

Figure 1 shows the CDFs for the three marital status groups. The code used to produce the graph follows below. (Stata 14 users should omit the ‘%55’, which refers to a transparency option introduced in Stata 15.)

```
. tw (line F_m v_m, sort c(stairstep) lcolor(black%55) ) ///
> (line F_sdw v_sdw, sort c(stairstep) lcolor(black) lpatt(dash) ) ///
> (line F_snm v_snm, sort c(stairstep) lcolor(black) lpatt(shortdash) ) ///
> , xlab(1(1)11) yline(0.5, lpatt(shortdash) lcol(black)) ///
> ylab(0(.1)1, angle(0)) ytitle("{it:p}") xtitle("Response (rescaled)") ///
> legend(label(1 "Married") label(2 "Separated, divorced, widowed") ///
> label(3 "Single, never-married") col(1) ///
> ring(0) position(11) ) ///
> scheme(slc) graphregion(color(white)) ///
> saving(aps01_Fdom_m-sdw-snm.gph, replace)
(file aps01_Fdom_m-sdw-snm.gph saved)
```

Figure 1. Cumulative distribution functions (CDFs) for life satisfaction, by marital status group



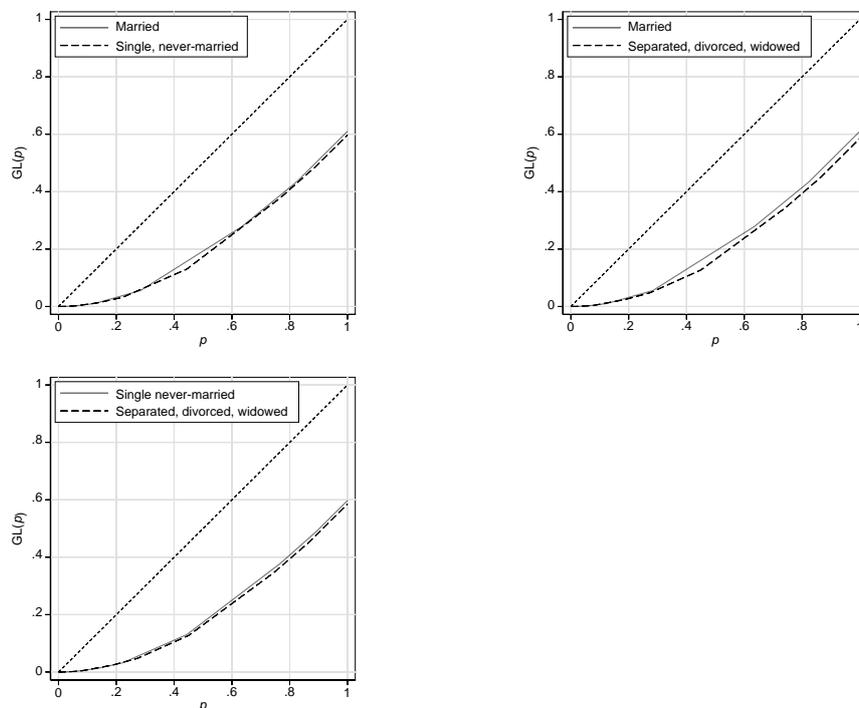
We can see immediately that 9 is the median value of $\mathbf{1s}$ for all three groups (the value where cumulative population share, $p = 0.5$). The CDF for married adults lies everywhere on or below the CDFs for the other two groups (F -dominance) and so we can say that married adults have higher average life satisfaction than the other two groups, regardless of the scale used. The CDFs for the SNM and SDW groups cross and so there is no F -dominance result. However, there is S -dominance. Below the median, the CDF for the SDW

group is further from the median than the CDF for the SNM group and the reverse is the case above the median. Thus, there is greater polarization in the distribution of life satisfaction among the SDW group than among the SNM group according to all standard polarization indices, and so including all members of the $ANY(a, b)$ and $P2(e)$ families of indices.

To check for unanimous rankings by Cowell-Flachaire and J indices, I focus on the peer-inclusive downward-looking definition of status for brevity. Figure 2 shows the results of the three pair-wise comparisons between groups. Below I show the code used for the Married and SDW groups' comparison. Analogous code for the other two pair-wise comparisons followed by a `graph combine` produced Figure 2.

```
. tw (function y = x, lpatt(shortdash) lcol(black)) ///
> (line gld_m F_m, sort lcolor(black%55) ) ///
> (line gld_sdw F_sdw, sort lcolor(black) lpatt(dash) ) ///
> , xtitle("{it:p}") ytitle("GL({it:p})") ///
> xlab(0(.2)1, grid) ylab(0(.2)1, grid angle(0)) ///
> legend( label(2 "Married") label(3 "Separated, divorced, widowed") ///
>         ring(0) position(11) order(2 3) col(1) ) ///
> aspect(1) scheme(sicolor) ///
> saving(aps01_gld_sdw-m.gph, replace)
(file aps01_gld_sdw-m.gph saved)
```

Figure 2. Generalized Lorenz curve comparisons for life satisfaction, by marital status group



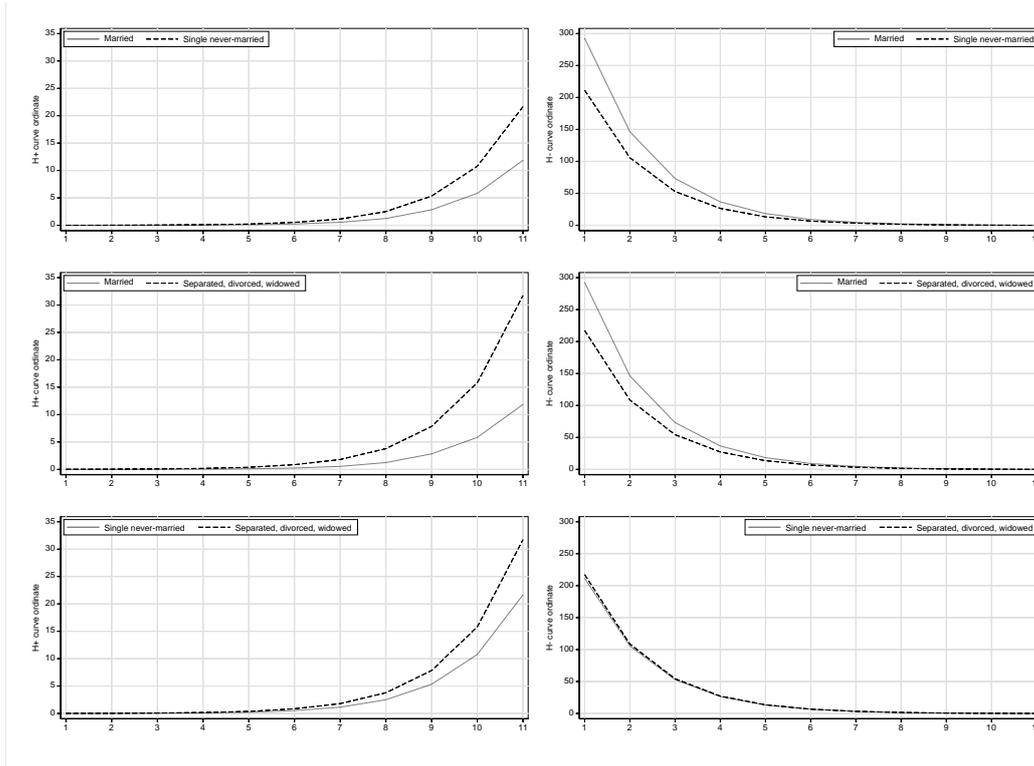
Note. Graphs drawn using Cowell and Flachaire's (2017) peer-inclusive downward-looking definition of status.

All three pair-wise comparisons reveal dominance. The clearest result, in the sense that the gap between the Generalized Lorenz curves for status is greatest, is in the top-right picture: life satisfaction is more unequal for the SDW group than the Married group according to all Cowell-Flachaire indices and the J index. The other two charts show that inequality is greater among the SNM group than the Married group, and among the SDW group compared to the SNM group. Thus, there is an unambiguous ranking from highest to lowest inequality according to all C-F indices and J , with the SNM group the most unequal, the Married group the least unequal, and the SDW group in between.

Figure 3 summarizes checks of Gravel et al.'s (2020) dual dominance criteria based on H^+ and H^- curve comparisons. The code used for the comparison of H^+ curves for married and single-married groups is shown below. Analogous code for the other two pair-wise H^+ comparisons, and for the three H^- comparisons, followed by a `graph combine` produced Figure 3.

```
. * H-plus dominance single never-married and married
.
. tw (line hp_m v_m, sort lcol(black%50) ) ///
>     (line hp_snm v_snm, sort lcolor(black) lpatt(dash) ) ///
>     , legend(label(1 "Married") label(2 "Single never-married") ) ///
>         ring(0) position(11) ) ///
>     xlab(1(1)11, grid) xtitle(" ") ///
>     ylab( 0(5)35, angle(0) grid) ///
>     scheme(sicolor) graphregion(color(white)) ///
>     saving(aps01_hplus_snm-m.gph, replace)
(file aps01_hplus_snm-m.gph saved)
```

Figure 3. H^+ and H^- curve comparisons for life satisfaction, by marital status group



Note. H^+ curve comparisons are shown on the left and H^- curve comparisons on the right.

Recall that for Gravel et al.'s dual dominance criteria to be satisfied, we need to find the H^+ and H^- curves for one group nowhere above the corresponding curves for another group. For these data, the orderings of the groups according to the H^+ criterion are the same as the orderings by the F -dominance criterion (because F -dominance implies H^+ dominance). See the charts on the left-hand side of Figure 3. However, there is dual dominance in only one case. The distribution of life satisfaction among the single never-married group is more equal than the distribution among the separated, divorced, widowed group: the H^+ and H^- curves for the former group are nowhere below those for the latter group. This ordering of the two groups, based on Hammond transfer principles, is the same as their ordering according to the S -dominance criterion (referring to greater polarization about the median), and is also the same as their Generalized Lorenz dominance ordering (see Figure 2).

4.3 Indices of polarization and inequality

Estimates of specific polarization and inequality indices are consistent with this dominance result and also the S -dominance result cited earlier (the SDW group is more polarized about the median than the SNM group). Specific indices are also useful for deriving inequality and polarization orderings when there is no dominance result and, of course, can also be used to place a number on the magnitude of differences. To illustrate these points, I present estimates of a selection of inequality indices ($I(\alpha)$ for $\alpha = 0, 0.25, 0.5, 0.75, 0.9$; and J ; all using a peer-inclusive downward-looking status definition) and three polarization indices, $ANY(1, 1)$, the top-sensitive $ANY(4, 1)$, and the bottom-sensitive $ANY(1, 4)$. In addition, I show how one can derive standard errors for the indices using Saigo et al.'s (2001) repeated half-sample bootstrap using Philippe Van Kerm's (2013) `rhsbsample` (on SSC), with 500 bootstrap replications in this case. With the APS's very large sample size, the indices are going to be precisely estimated and confidence intervals narrow, even for subgroup calculations, but this is not generally the case with survey data. (See Jenkins 2019a for examples.) Hence this code may be usefully applied in other contexts.

The code below shows the derivations for the Married group. The first component drops observations with missing values and the second uses `rhsbsample` to create the bootstrap sample weights. Third, I `svyset` the data. If survey design variables other than weights – PSU and strata variables – had been available, this is where they would have been cited. The fourth step is to call `ineqord` using the `svy bootstrap` prefix command. Observe that I use the `alpha(0.9)` option in order to derive estimates of Cowell-Flachaire indices, $I(\alpha)$, for values of α spanning its range. (I also derived estimates for more polarization indices than I cited earlier, just in case I needed them.) The ‘d’ suffix on the estimates’ names reminds us that I am using Cowell and Flachaire’s peer-inclusive downward-looking status definition. The final step is to save the estimates of indices, standard errors and confidence intervals to a dataset using Roger Newson’s (2003) `parmest` utility command (latest version available from SSC).

```
. ** Married **
. use aps_3yr_jan15dec17_eul if MARSTA == 2, clear

. rename _all, lower

. ge ls = .
(219,411 missing values generated)

. replace ls = satis + 1 if satis >= 0 & !missing(satis)
(140,679 real changes made)
```

```

. lab var ls "Response (scaled)"
. lab def ls 1 "Not at all satisfied: 1" 11 "Completely satisfied: 11"
. lab val ls ls
. drop if missing(ls)
(164,851 observations deleted)
. rename pwtal7c wgt // for convenience
. drop if missing(wgt)
(0 observations deleted)
. loc R = 500
. forvalues i = 1/\`R' {
2.   qui gen rhsbrw`i' = .
3.   qui rhsbssample, weight(rhsbrw`i')
4.   qui replace rhsbrw`i' = rhsbrw`i' * wgt
5. }
. svyset [pw = wgt], vce(bootstrap) bsrweight(rhsbrw*) mse
    pweight: wgt
      VCE: bootstrap
      MSE: on
    bsrweight: rhsbrw1 .. rhsbrw500
Single unit: missing
Strata 1: <one>
  SU 1: <observations>
  FPC 1: <zero>
. svy bootstrap ithreequ = (r(ithreequ)) ithreeqd = (r(ithreeqd)) ///
>   ihalfu = (r(ihalfu)) ihalfd = (r(ihalfd)) ioneqd = (r(ioneqd)) ///
>   i0u = (r(i0u)) i0d = (r(i0d)) ipt9u = (r(ixu)) ipt9d = (r(ixd)) ///
>   any11 = (r(any11)) any21 = (r(any21)) any12 = (r(any12)) ///
>   any41 = (r(any41)) any14 = (r(any14)) ///
>   apoueypt5 = (r(apoueypt5)) blairlacy = (r(blairlacy)) ///
>   jd = (r(Jd)) ///
>   N = (r(N)) sumw = (r(sumw)) ///
>   median = (r(median)) mean = (r(mean)) sd = (r(sd)) ///
>   , dots ///
>   : ineqord ls, alpha(0.9)
(running ineqord on estimation sample)
< output omitted >
. parmest , idn(2) saving(aps02_bstrap-m_parmest.dta, replace)
(note: file aps02_bstrap-m_parmest.dta not found)
file aps02_bstrap-m_parmest.dta saved

```

I repeated this code for the SNM and SDW groups as well, specifying different arguments for **parmest**'s **idn(.)** option in each case in order to separately identify the estimates for the three marital status groups when I combined the three datasets using **append**. Using the combined-estimates dataset, it is straightforward to summarize differences across groups, by index, in graphical form. See Figure 4, created by first creating graphs for each index and then using **graph combine**. Here is the code used to display the estimates for $I(0)$:

```

. tw (rcap min95 max95 idnum if parm == "i0d", lcol(black) horizontal) ///
>   (scatter idnum estimate if parm == "i0d", mc(black)) ///
>   , ylab(1(1)3, angle(0) value label) ytitle(" ") ///
>   xlabel(0.7(.05).8, format(%3.2f)) xmtick(.7(.01).8) ///

```

```

>         xtitle("Cowell-Flachaire, I(0)") legend(off) ///
>         graphregion(color(white)) ///
>         saving(aps03_bstrap-all_parmest_i0d.gph, replace)
(note: file aps03_bstrap-all_parmest_i0d.gph not found)
(file aps03_bstrap-all_parmest_i0d.gph saved)

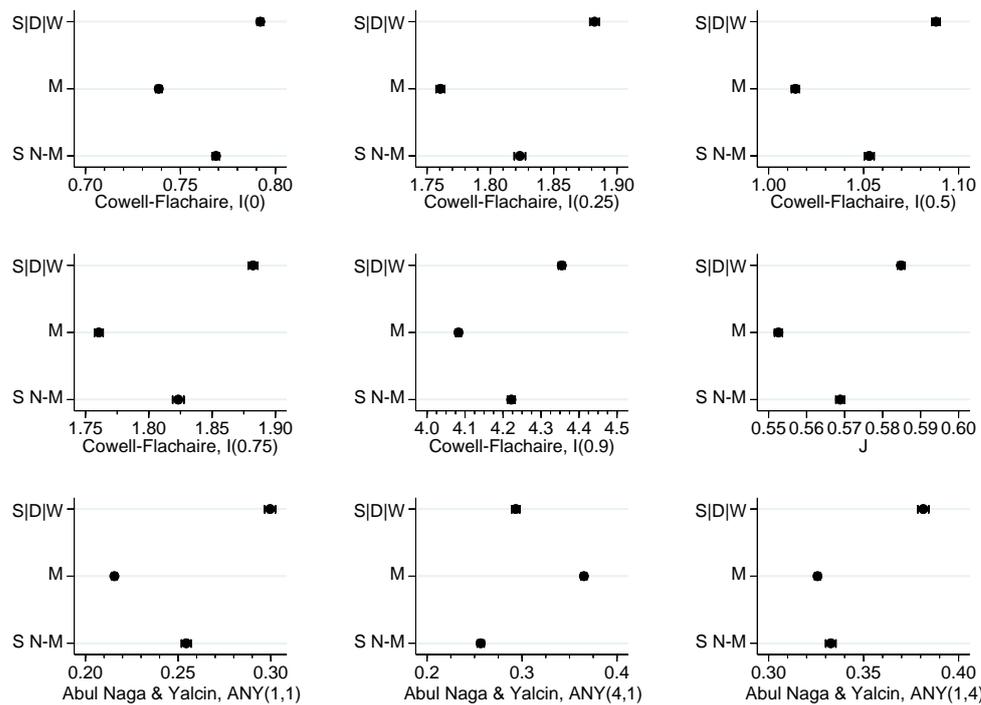
```

All indices are very precisely estimated, and all between-group differences are statistically significantly different from zero.

The rankings of marital status subgroups in Figure 3 are of course consistent with the dominance results discussed earlier. However, there was no *S*-dominance result for the polarization comparisons between the Married group and each of the other two groups and so index values are valuable for providing a polarization ranking. Interestingly, Figure 4 shows that this depends on the index chosen. For *ANY*(1, 1) and *ANY*(1, 4), the ranking is the same as for the inequality indices. However, for top-sensitive index, *ANY*(4, 1), the Married group shows the greatest polarization rather than the lowest. What is driving this result is that the Married group has relatively large fractions of responses in the top two life satisfaction scale categories by contrast with the other two groups (see the tabulation of life satisfaction by marital status shown earlier).

Indices also tell us about the magnitudes of differences across groups. As it happens, the $I(\alpha)$ and J indices provide similar estimates. For example, all of them indicate that the difference in life satisfaction inequality between the most unequal group (SDW) and the least unequal group (Married) is around 7%. More marked differences are apparent for the *ANY* polarization indices. For example, for *ANY*(1, 1) which is also the Average Jump index, the difference in polarization between the SDW and Married groups is around 39%, whereas for *ANY*(1, 4) it is around 17%. For *ANY*(4, 1) it is -20%.

Figure 4. Indices of life satisfaction inequality and polarization, by marital status group



Notes Each component 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). ‘S|D|W’: separated, widowed, or divorced. ‘M’: married. ‘S N-M’: single never-married.

4.4 Using grouped data with *ineqord*

ineqord is designed for use with datasets containing individual-level responses but it is straightforward to also use it if only grouped response data are available, specifically if one has information on the number of individuals reporting each response level (or fraction of individuals) or the empirical CDF.

For example, Abul Naga and Yalcin (2008, Table 2) report the empirical CDF for self-reported health status recorded on a 5-level scale (‘very bad’, ‘bad’, ‘so so’, ‘good’, ‘very good’) for each of seven statistical areas in Switzerland. The empirical CDF for the Central region, can be reproduced using the following code to characterize the distribution of responses:

```
. set obs 100
number of observations (_N) was 0, now 100

.
. ge central = .
(100 missing values generated)
```

```

. replace central = 2 if _n <= 2
(2 real changes made)

. replace central = 3 if _n > 2 & _n <= 13
(11 real changes made)

. replace central = 4 if _n > 13 & _n <= 76
(63 real changes made)

. replace central = 5 if _n > 76
(24 real changes made)

.
. ta central

```

central	Freq.	Percent	Cum.
2	2	2.00	2.00
3	11	11.00	13.00
4	63	63.00	76.00
5	24	24.00	100.00
Total	100	100.00	

No individuals in Central region reported ‘very bad’ health status and so simply typing `ineqord central` will produce incorrect results, for the reasons discussed earlier. However, typing `ineqord central, nlevels(5)` produces estimates that are the same as those reported by Abul Naga and Yalcin (2008: Table 4). Application of analogous code using information about the empirical CDFs for the other six regions reproduces the estimates for the regions reported in Abul Naga and Yalcin’s Table 4, as well as providing estimates for other indices that they did not consider. With appropriate use of `ineqord` options, it is easy to also derive the outputs required to undertake dominance checks.

Using the same grouped-data approach, I have verified that `ineqord` produces the same estimates of *ANY* polarization indices as reported by Madden (2010) for self-reported health status in Ireland in each year 2003–2006. I can also reproduce the estimates of the Blair-Lacy $1-l^2$ polarization index shown by Blair and Lacy (2000, Table 2), once account is taken of some typographical errors (estimates of l^2 are reported in the wrong table rows).

5. Summary and conclusions

The personal well-being of individuals is increasingly being measured using questions requiring responses on a Likert-type scale. Life satisfaction and self-assessed health status are leading examples of these measures and they yield distributions of ordinal data. To compare such distributions across groups of individuals or over time, we should not apply the toolbox of methods developed for cardinal variables such as income. These methods rely on the mean

as a reference point, but changing the scale in the ordinal data case can change the orderings of distributions according to their means or other measures based on the mean, including conventional inequality indices. Thus, we should use an analogous toolbox which takes account of the ordinal nature of the responses. This paper reviews these methods and introduces a new command **ineqord** for undertaking distributional comparisons. As the empirical illustrations demonstrate, **ineqord** can be used for dominance checks as well as for estimation of indices of polarization and inequality.

Acknowledgements

I developed **ineqord** as part of a project undertaken with Arthur Grimes and Florencia Tranquilli, both of Motu Research (Wellington, New Zealand). The research was part-supported by 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 gratefully acknowledge the hospitality of the School of Economics, University of Queensland, where I was based when I began drafting this paper, and the Stone Center on Socio-Economic Inequality, City University of New York Graduate Center, where I revised it. My thanks go to Benoît-Paul Hébert and an anonymous referee for their helpful comments on an earlier version.

References

- Abdul Naga, R. and Yalcin, T. 2008. Inequality measurement for ordered response health data. *Journal of Health Economics*, 27: 1614–1625.
- Allison, R.A. and 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.
- Blair, J., and Lacy, M. 2000. Statistics of ordinal variation. *Sociological Methods and Research*, 28: 251–280.

- Bond, T. N. and Lang, K. 2019. The sad truth about happiness scales. *Journal of Political Economy*, 127: 1629–1640.
- Cowell, F.A. and Flachaire, E. 2017. Inequality with ordinal data. *Economica*, 84: 290–321.
- Delhey, J. and Kohler, U. 2013. Is happiness inequality immune to income inequality? New evidence through instrument-effect-corrected standard deviations. *Social Science Research*, 40: 742–756.
- Dutta, I. and Foster, J.E. 2013. Inequality of happiness in the US: 1972–2010. *Review of Income and Wealth*, 59: 393–415.
- Gravel, N., Magdalou, B., and Moyes, P. 2020. Ranking distributions of an ordinal variable, *Economic Theory*, online ahead of print.
<https://link.springer.com/article/10.1007/s00199-019-01241-4>
- Helliwell, J. F., Huang, H., and Wang, S. 2019. Changing world happiness, chapter 2 in: J. F. Helliwell, R. Layard, and J. D. Sachs (eds), *World Happiness Report 2019*. New York: Sustainable Development Solutions Network, 11–45.
<https://worldhappiness.report/ed/2019/>
- Jenkins, S.P. 2019a. ‘Better off? Distributional comparisons for ordinal data about personal well-being’, *New Zealand Economic Papers*, online ahead of print.
<https://www.tandfonline.com/doi/full/10.1080/00779954.2019.1697729>
- Jenkins, S.P. 2019b. Inequality comparisons with ordinal data. IZA Discussion Paper 12811. Bonn: IZA. <http://ftp.iza.org/dp12811>
- Kalmijn, W. and Veenhoven, R. 2005. Measuring inequality of happiness in nations. In search of proper statistics. *Journal of Happiness Studies*, 6: 357–396.
- Kaplan, D.M. and Zhuo, L. 2019. Comparing latent inequality with ordinal data. Unpublished paper available at <https://faculty.missouri.edu/~kaplandm>.
- Madden, D. 2010. Ordinal and cardinal measures of health inequality: an empirical comparison. *Health Economics*, 19: 243–250.
- Newson, R. 2003. Confidence intervals and p-values for delivery to the end user. *The Stata Journal*, 3: 245–269.
- OECD (2020). *How’s Life? 2020. Measuring Well-being*. Paris: OECD Publishing.
- Office for National Statistics 2019. *Personal Well-being in the UK, April 2018 to March 2019*. London: ONS.
<https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/bulletins/measuringnationalwellbeing/april2018tomarch2019>

- Office for National Statistics, Social Survey Division. 2019. Annual Population Survey Three-Year Pooled Dataset, January 2015 - December 2017. [data collection]. 2nd Edition. UK Data Service. SN: 8370. <http://doi.org/10.5255/UKDA-SN-8370-2>
- Saigo, H., Shao, J., and Sitter, R. R. 2001. A repeated half-sample bootstrap and balanced repeated replications for randomly imputed data. *Survey Methodology*, 27: 189–196.
- Stiglitz, J. E., Sen, A. K., and Fitoussi, J.-P. 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress. <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.

About the author

Stephen P. Jenkins is Professor of Economic and Social Policy at the London School of Economics and Political Science (LSE). He is also a Research Fellow at IZA, Bonn. Much of his substantive research is about inequality and poverty and related topics. He also has interests in applied micro-econometrics including survival analysis and statistical graphics, and the use of survey and administrative record data. Stephen has contributed many commands to SSC and written articles for the *Stata Journal*.