Distributional comparisons for ordinal data

Stephen P. Jenkins
LSE
Email: s.jenkins@lse.ac.uk

Canazei Winter School, January 2020

Motivation / background

- 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 in New Zealand? How does the distribution of well-being in New Zealand compare with that in Australia? And so on
- Increasing weight is being put on subjective well-being measures by international agencies such as the OECD and national governments (including New Zealand's)
- Making distributional comparisons using subjective well-being measures raises analytical challenges because we have ordinal measures rather than cardinal ones such as income

LSE

Outline of talk

- A general review of distributional comparisons for ordinal data (e.g. personal subjective well-being, SWB):

1. Some preliminaries
2. Benchmark: methods for comparing distributions of *income* to assess whether we are 'better off'
   - Levels (and % with low, high values); inequality; 'welfare'
3. An analogous 'toolbox' for comparing SWB distributions (and other ordinal data)
4. Empirical illustrations using World Values Survey (WVS) data for AU, GB, NZ, US, ZA about life satisfaction
5. Lessons to be drawn
6. Calculations in Stata: illustrate my program *ineqord*

1. Some preliminaries

- The rise of “well-being” around the world, including in NZ
- Cardinal versus ordinal measures

Well-being to the fore in NZ

- Reflecting new emphasis on well-being around the world
- OECD, UK, NZ, and other countries

Well-being: a strategic priority in NZ

Indicators Aotearoa New Zealand

July 2018

Stats NZ
Tatauranga Aotearoa

A wellbeing approach to strategic decision-making

The Government wants to improve the wellbeing of current and future generations of New Zealanders. It wants a public sector that is focused on achieving this.

The diagram on the reverse shows how a wellbeing approach will improve strategic decision-making by drawing on better strategic advice, underpinned by indicators that track the wellbeing of New Zealanders.

The various strands of work and how they contribute to the wellbeing approach are described below.

Stats NZ

Indicators Aotearoa New Zealand (IANZ) will provide a clear view around how we are tracking as a nation, using a wellbeing and sustainable development lens.

It will provide information on current and future wellbeing, and the impact New Zealand is having on the rest of the world.

The Treasury

The Treasury is developing a Living Standards Framework.

An important next step is the development of a Living Standards Dashboard, which will be updated regularly. The dashboard is expected to largely draw its indicators from IANZ.

The Treasury also intends to publish a Living Standards Report that tells the story about the wellbeing of New Zealanders.

Sectors and agencies

Other sectors and agencies will inform the Government's wellbeing approach through:

- domain monitoring – using a subset of indicators from IANZ complemented by other indicators relevant to a specific domain (eg health)
- domain reporting – using commentary and visualisations to tell the story of progress within a domain (eg environmental reporting)
- sector and agency strategies – using specific strategies designed to achieve a long-term goal for a sector, agency, or population group (eg Child Wellbeing Strategy).

Cabinet and committees

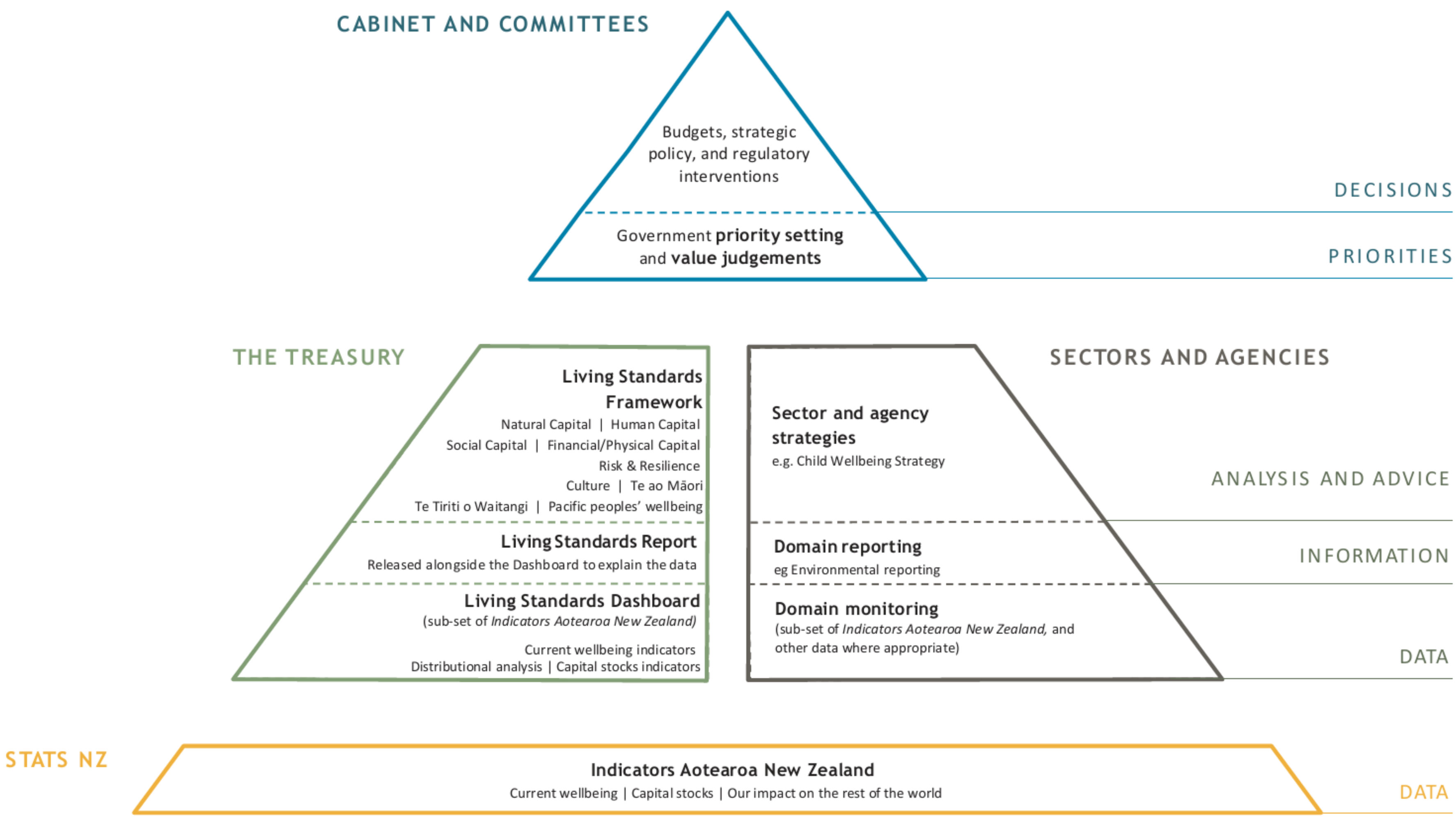
Cabinet and its committees will draw on the overarching advice they receive from the Treasury and other agencies to frame strategic wellbeing and investment decisions.

Visit the website: www.stats.govt.nz
Email: indicators@stats.govt.nz

New Zealand Government

Government, NZT, and Stats NZ

Using a wellbeing approach to improve strategic decision-making



Well-being and NZT's Living Standards Framework

6

Current wellbeing outcomes

New Zealand's average level of current well-being: Comparative strengths and weaknesses

Note: This chart shows New Zealand's relative strengths and weaknesses in well-being when compared with other OECD countries. For both positive and negative indicators (such as homicides, marked with an "*"), longer bars always indicate better outcomes (i.e. higher well-being), whereas shorter bars always indicate worse outcomes (i.e. lower well-being). If data are missing for any given indicator, the relevant segment of the circle is shaded in white.

Additional information, including the data used in this country note, can be found at:

www.oecd.org/statistics/Better-Life-Initiative-2017-country-notes-data.xlsx

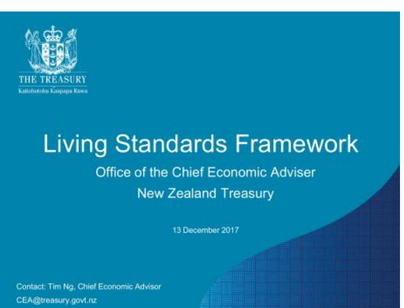
Current outcomes are assessed using indices of both quality of life and material conditions. It is a snapshot measure.

Each element has a distribution in the population and may vary across subpopulations (e.g. by gender or ethnicity). It is not just income or wealth that might be unequally distributed.

It cannot be measured as a single number without making significant implicit or explicit value judgements, for example, how important is health relative to income.

We are investigating supplements to OECD measures to capture New Zealand specific issues (see next slide).

Note: New Zealand household income data available but not included here because New Zealand is not part of the OECD Luxembourg Income Project. Follow OECD links on the web page for more information.



And in the UK …

- Treasury's *Green Book*
- And a new Handbook (Nov 2019) …

A handbook for

wellbeing policy-making in the UK: history, measurement, theory, implementation, and examples.

Paul Frijters and Christian Krekel

Green Book Central Government Guidance on Appraisal and Evaluation

4.15 Individual and society's wellbeing is influenced by a number of interrelated fact including health, relationships, security and purpose. At the long-list appraisal stage, evidence on the determinants of wellbeing can help describe Business As Usual and the purpose or scope of an intervention throug SMART objectives. It may help to identify interventions which have an impact on wellbeing or another outcome which is affected by wellbeing. This supports the development of a long-list of options or the most efficient way of implementing a proposed solution.

4.16 Where appropriate evaluations of previous or similar interventions, international and wellbeing evidence, shou be used to design options that build on wha works, to avoid repeating past mistakes. Th is particularly important when considering the scope of a proposal and the service solution (the technical means of delivering the intervention). When assessing the relevance of previous evaluation, allowance should be made for differences in context, circumstances and culture.

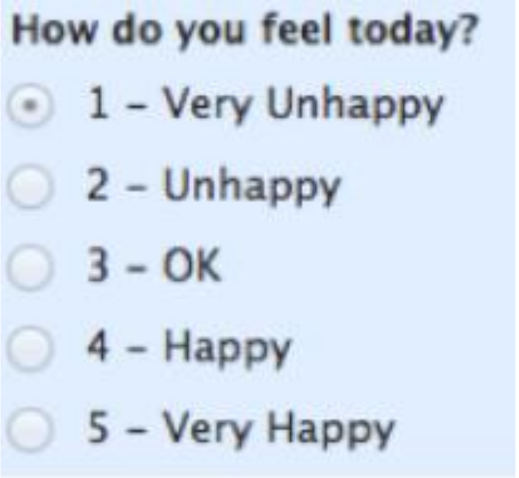
6.21 Subjective wellbeing evidence aims to capture the direct impact of a policy on wellbeing. The evidence can challenge decision makers to think carefully about the full range of an intervention's impacts and consider a wider range of interventions.

Caveat

- This talk is about assessing social progress – monitoring the (distribution of) well-being outcomes per se
  - Are we better off, in well-being terms?
- This talk is *not* about the relationships between these outcomes and policy interventions
  - How can and will the Living Standards Framework and well-being approach be used to guide policy design and analysis?
- NZ (and elsewhere) need to know about both aspects!
  - But you cannot do the second task properly without the first one sorted out

Cardinal versus ordinal measures

- **Cardinal** (e.g. income)
  - Order of values, and absolute differences between values are well-defined, and so too are ratios of values because a value of zero is well-defined
- **Ordinal** (e.g. life satisfaction and other subjective well-being measures)
  - Order of values is well-defined, but not differences in magnitude (or ratios)
  - We don't know if the difference in happiness between 'OK' and 'Happy' is the same as the difference between 'Happy' and 'Very Happy'
  - The labels [1,2,3,4,5] are arbitrary; could be [1,2,3,5,10]; or [–4, –2, 0, 2, 4]; or …



LSE

WVS Life Satisfaction question

10 point scale:

(*Show Card C*)

V22. All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means you are "completely dissatisfied" and 10 means you are "completely satisfied" where would you put your satisfaction with your life as a whole? (*Code one number*):

Completely dissatisfied									Completely satisfied
1	2	3	4	5	6	7	8	9	10

What would you respond?

- Remember your answer and compare it with the estimates shown later

Common 'reporting function'

- Is there comparability of responses?
- Does my "8" correspond to your "8", or is a "7" today comparable with a "7" from a decade ago?
- Bond and Lang (*JPolEcon* 2019), among others, express scepticism regarding existence of a common reporting function for SWB questions
- SPJ: their case is over-stated because the common reporting function issue also applies to 'income' – the most commonly used cardinal measure of personal economic well-being
  - Raw survey data are aggregated across individuals within households; adjustments for differences in household size and composition (equivalence scales); adjustments for difference in prices levels across time and/or geographies
- I assume common reporting functions!

2. Assessing 'better off': distributional comparisons for income data

A benchmark 'toolbox'
– are there corresponding methods
for ordinal data?

Standard methods for income data

Two main **outcomes of 'social welfare' interest**: income *levels* (including mean; poverty), and *inequality*, and potential trade-off between them (overall *social welfare*)

Two main **approaches in the assessment toolbox**

- **Pictures**: to show the data and to conduct 'dominance' checks
  - Dominance: can we derive unanimous rankings of a pair of distributions *A* and *B* using only minimal assumptions about the social welfare function?
- **Indices**: to summarise numerically
  - Ranking distributions when dominance does not hold (so stronger assumptions required about social welfare function)
  - Providing magnitudes, not simply orderings

The power of pictures

- Density functions: to show the data
- Levels and Cumulative Distribution Functions (CDFs)
  - CDF: $F(x) = p, p \in [0,1]$
    - Alternatively: compare Pen's Parades (quantile functions): $x = F^{-1}(p)$
  - To show the data and undertake first-order dominance checks
- Inequality and Lorenz curves
  - To show the data and undertake dominance checks
- 'Social welfare' and Generalized Lorenz curves
  - To show the data and undertake second-order dominance checks

CDFs and distributional comparisons

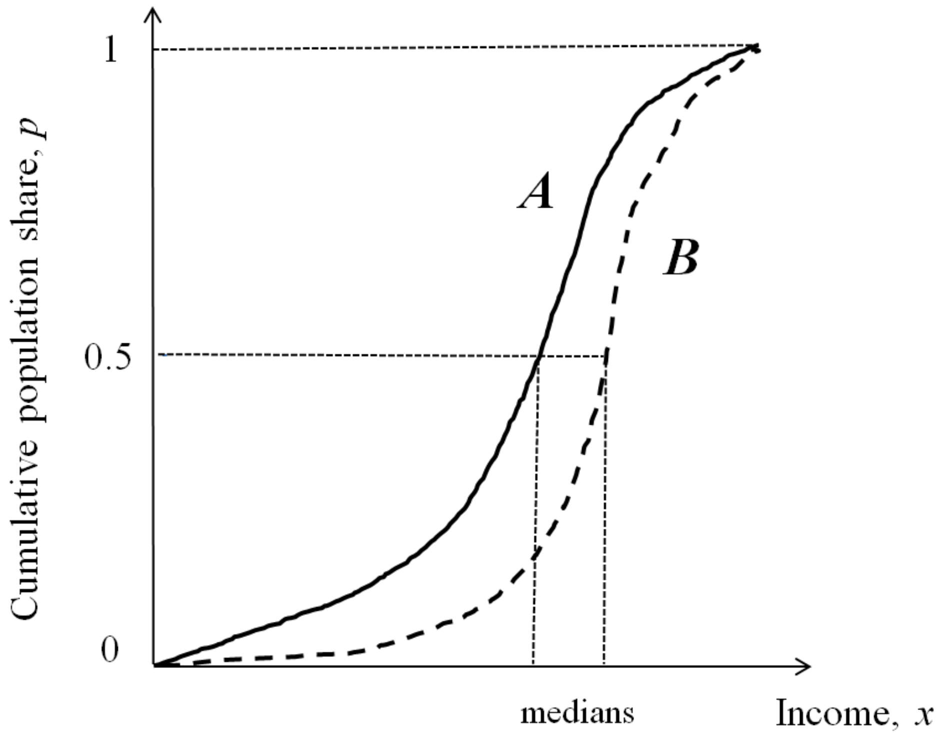
- Link configurations of CDFs with first-order Welfare dominance and with Poverty dominance

Suppose that the social welfare function is $W = W(x_1, x_2, \ldots, x_n)$

- Class $\mathcal{W}_1$ characterized by all W that satisfy properties:
  - More is better, other things equal ('Pareto' principle, a.k.a. monotonicity)
    - increasing ($\partial W/\partial x_i > 0$, all i), or at least non-decreasing
  - All that matters is the value of your 'income' (so 'anonymity'); no relative income effects
    - Symmetric, i.e. invariant to permutations of the income vector
  - We can compare distributions from populations that are of different sizes
    - Invariant to replications of the population
- NB $\mathcal{W}_1$ is essentially concerned with *real income levels*

CDF and dominance (a)

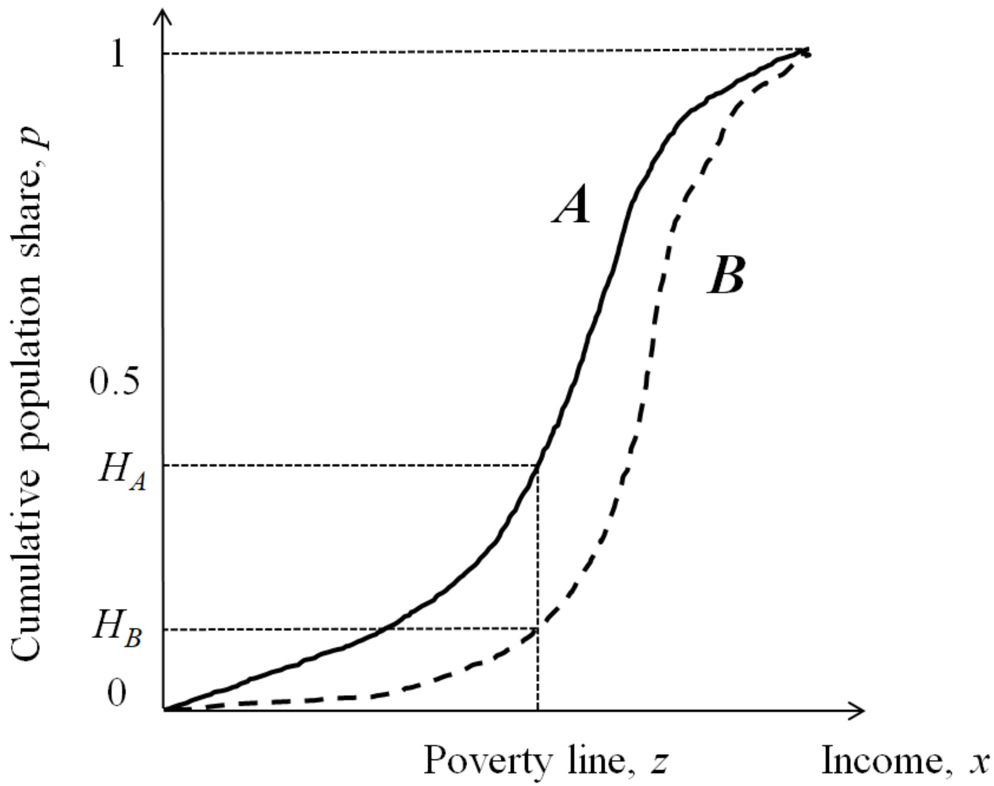
- First order welfare dominance result (Saposnik):
  - CDF(B) lies everywhere below CDF(A) $\Leftrightarrow$ $W(B) > W(A)$ for all $W \in \mathcal{W}_1$, i.e. symmetric replication-invariant social welfare functions increasing in each income
  - Distribution B is better than A in a social welfare sense
  - At any specific p, quantile $x = F^{-1}(p)$ is larger in B than A
  - In particular, Mean(B) > Mean(A)



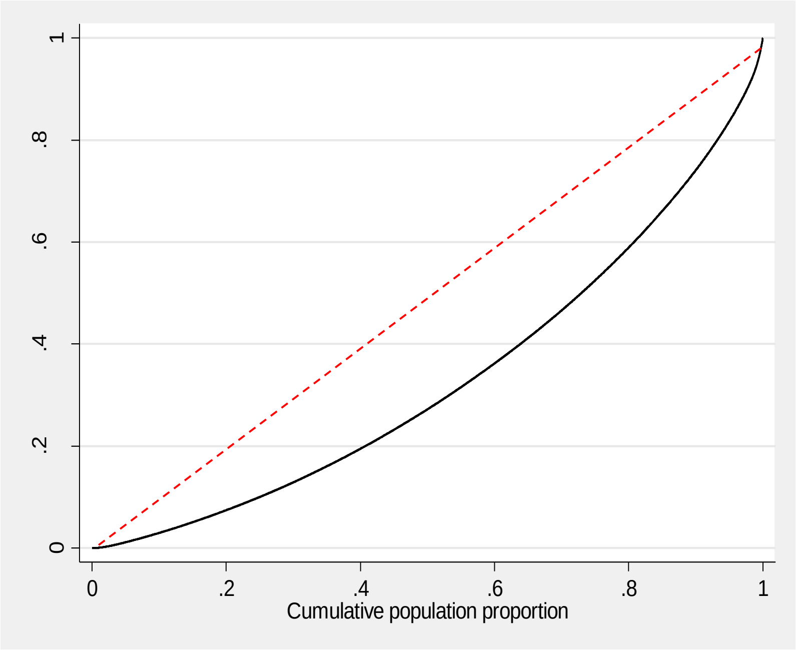
LSE

CDFs and dominance (b)

- Poverty dominance according to the Headcount Ratio measure, *H*, a.k.a. the proportion poor (Foster & Shorrocks)
  - If CDF(*A*) lies everywhere above CDF(*B*) at every income in interval [0, *z*], then $H(B) < H(A)$ for all common poverty lines between 0 and *z*
- Analogous results for proportions *above* a threshold



Lorenz curves and inequality



- A Lorenz curve plots cumulative income shares against cumulative population shares (*p*), with obs in ascending order by income
- Complete equality: Lorenz curve coincides with 45° ray through origin
- Inequality is greater, the further the Lorenz curve from the 45° ray
- Gini index of inequality equals the area between the Lorenz curve and the 45° ray, divided by total area under ray (= 1 – 2*area under L curve)

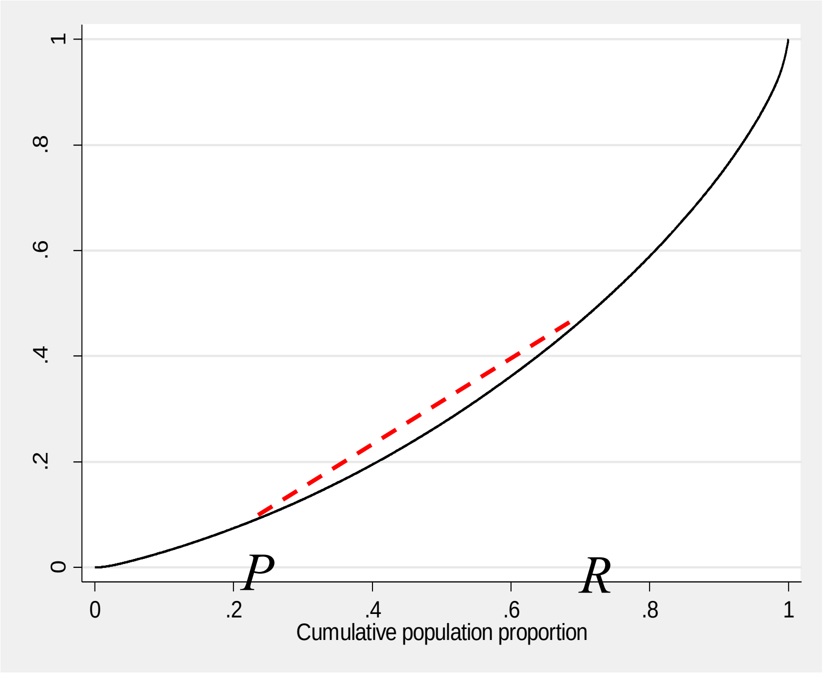
Lorenz curves and inequality comparisons

- Class of social welfare functions, $\mathcal{W}_2$ with $W \in \mathcal{W}_2$ if increasing in each income (i.e. respects *more is better*), symmetric, replication-invariant and *concave* (i.e. respects *equality preference*)
  - This means that a mean-preserving spread of income lowers social welfare ≈ ‘inequality aversion’ ≈ ‘equality preference’
    - E.g. a transfer £10 from a poor person to a millionaire reduces overall societal welfare
  - And vice-versa: a mean-preserving transfer from a rich person to a poor person raises overall societal welfare

  - Example of function satisfying $\mathcal{W}_2$: $w = \sum_i U(y_i)$, with $U(y_i) = \log(y_i)$
  - NB common assessment function $U(.)$

A mean-preserving progressive transfer shifts the Lorenz curve inwards (inequality reduction)

- The *Poorer* person has a bit more income; *Richer* person has a little less, all other incomes and total income unchanged
- So cumulative income shares above cumulative population shares corresponding those between *Poorer* and *Richer* must be greater , i.e. Lorenz curve moves inwards over range between *Poorer* and *Richer*

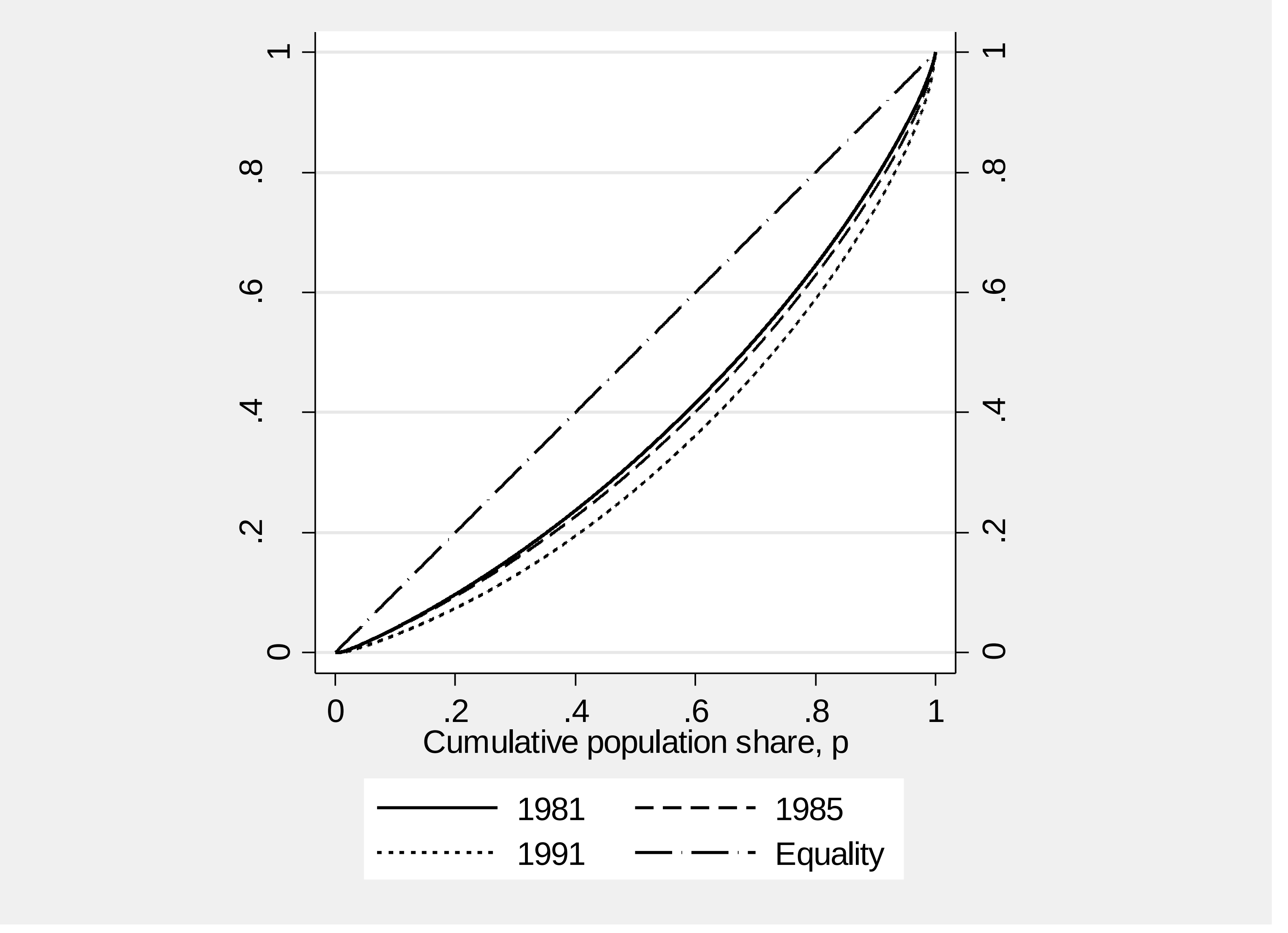


Lorenz curves and inequality dominance

Atkinson (1970) theorem [also Kolm]:

- For 2 distributions with the same mean, if the Lorenz curve for distribution $\boldsymbol{x}$ lies above the Lorenz curve for $\boldsymbol{y} \Leftrightarrow W(\boldsymbol{x}) > W(\boldsymbol{y})$ for all $W \in \mathcal{W}_2$, i.e. symmetric replication-invariant increasing and *concave* social welfare functions
- This is a result about inequality comparisons for distributions in which one 'controls' for differences in mean income
  - by deflating every income by the mean of the distribution to which it belongs, the two distributions being compared have the same mean (equal to one)

Inequality comparisons, UK incomes: 1981, 1985, 1991



Inequality indices often used with income data

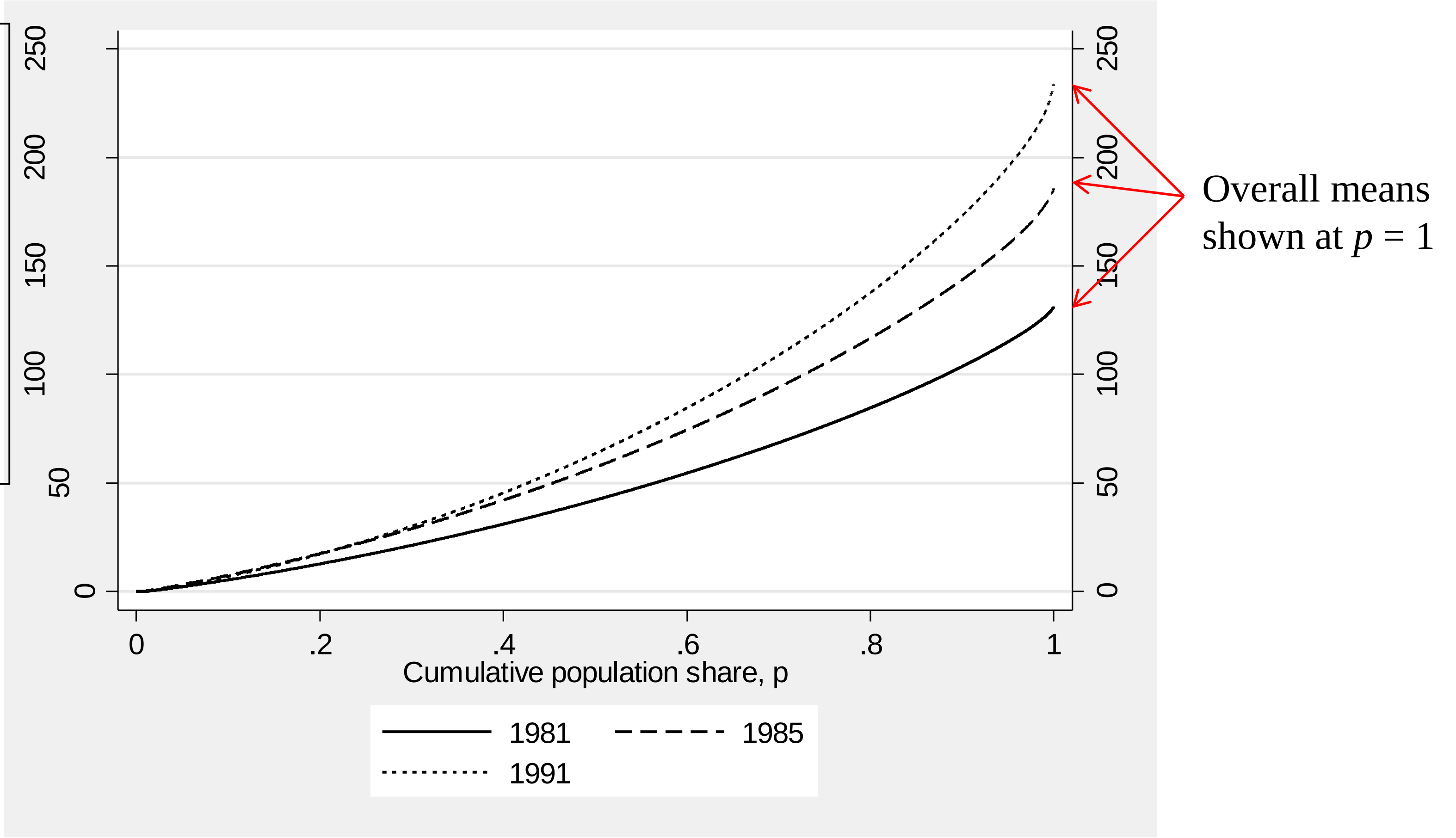
- Gini index, *G*
- Generalised Entropy indices, GE(a), including
  - GE(0): mean logarithmic deviation
  - GE(1): Theil
  - GE(2): half the squared coefficient of variation
  - Larger a ~ more 'top sensitive'; smaller a ~ more 'bottom sensitive'
- Atkinson indices, A(e)
  - Parameter e: degree of inequality aversion
  - Larger e: more inequality-averse ~ more 'bottom sensitive'
- All the indices are forms of social welfare functions defined over a distribution of relative incomes, where relative income = income ÷ mean
- All indices incorporate normative assumptions

Generalized Lorenz curves and social welfare

- ‘Social welfare’ assessments incorporate the idea of a trade-off between increases in inequality and increases in real income levels (sufficient rise in latter offset former)
- Generalized Lorenz curve: Lorenz curve scaled up at each p by population mean income, i.e. a plot of $p\mu_p$ (‘cumulative mean’) against p, where units are ordered in ascending order of income
- Class of social welfare functions, $\mathcal{W}_2$ with $W \in \mathcal{W}_2$ if increasing in each income, symmetric, replication-invariant and *concave*
  - Same SWF as underlies the Atkinson theorem
- Second Order Welfare Dominance result (Shorrocks): GLC($\boldsymbol{x}$) above GLC($\boldsymbol{y}$) at every $p \Leftrightarrow W(\boldsymbol{x}) > W(\boldsymbol{y})$ for all $W \in \mathcal{W}_2$
  - Also implies poverty dominance by poverty gap measures

Generalized Lorenz curves: UK example

$p\mu_p$ = $\Sigma y_i/N$ = Total income among poorest p, divided by N (i.e **'cumulative income per capita'**)

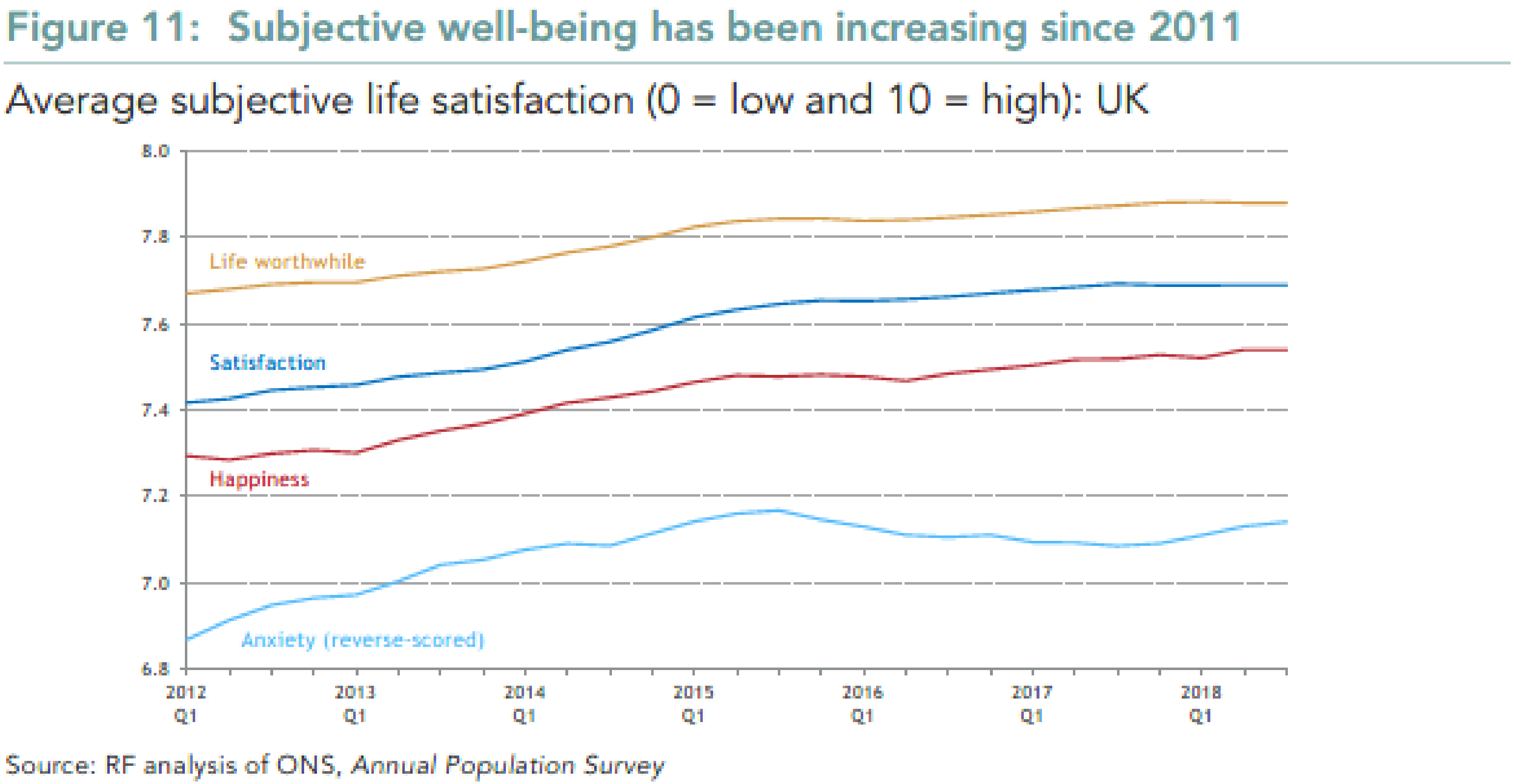


- Since differences in means are typically relatively large compared to inequality differences (Lorenz curve), GL comparisons are 'driven' by differences in income levels

3. Well-being comparisons for ordinal variables

Most SWB aficionados treat well-being measures as if they are cardinal rather than ordinal

Tracking well-being: averages over time (UK)



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

The mean is not order-preserving under scale changes (Allison-Foster, *JHE* 2004)

- Cardinal data: the mean is a stable reference point of the distribution
- Ordinal data: any re-scaling is possible, and the mean's location relative to the distribution (or subset of observations) can vary widely with changes in scale
  - E.g. distribution $x = (1,3,3,6,4)$ with scale $c = [1,2,3,4,5]$ has mean $\mu(x, c) = 3.53$, located between 7th lowest and 8th lowest of 17 obs
  - But if instead scale $c' = [1,2,3,4,8]$, $\mu(x, c') = 4.24$, lying between 13th and 14th of the 17 obs
    - "Recalling 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" (Allinson & Foster 2004: 510)
    - Non-robust answers to "What percentage of population has above-average SWB?"
- Ordinal data: ranking of 2 distributions by the mean may change when scale changed
  - E.g. distribution $x = (2,2,2,2,2)$ and $y = (3,2,1,1,3)$: with scale $c = [1,2,3,4,5]$, mean $\mu(x, c) = 3 > \mu(y, c) = 2.9$. But if $c' = [1,2,3,4,\mathbf{10}]$, $\mu(x, c') = 4 < \mu(y, c') = 4.4$, i.e. reversed ranking
  - Allison & Foster (2004) have a real world example using self-assessed health data

Tracking well-being: proportions with low and very high SWB levels (UK)

Figure 2a: Proportion of people reporting "poor" ratings of personal well-being across the UK, years ending June 2012, June 2017 and June 2018

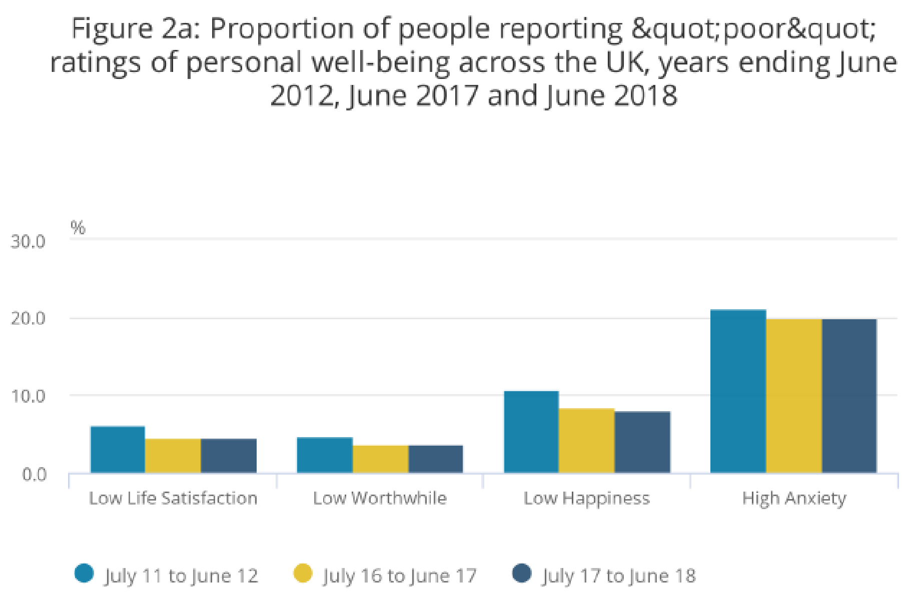
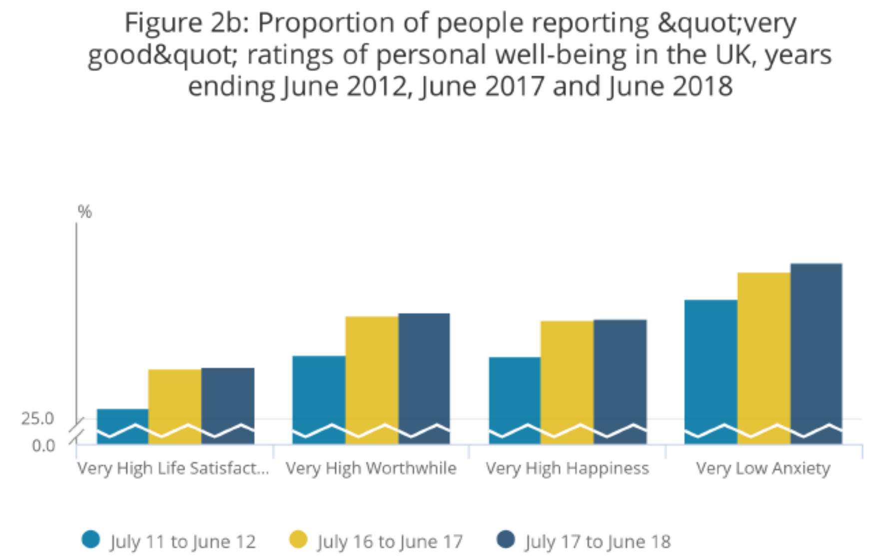


Figure 2b: Proportion of people reporting "very good" ratings of personal well-being in the UK, years ending June 2012, June 2017 and June 2018



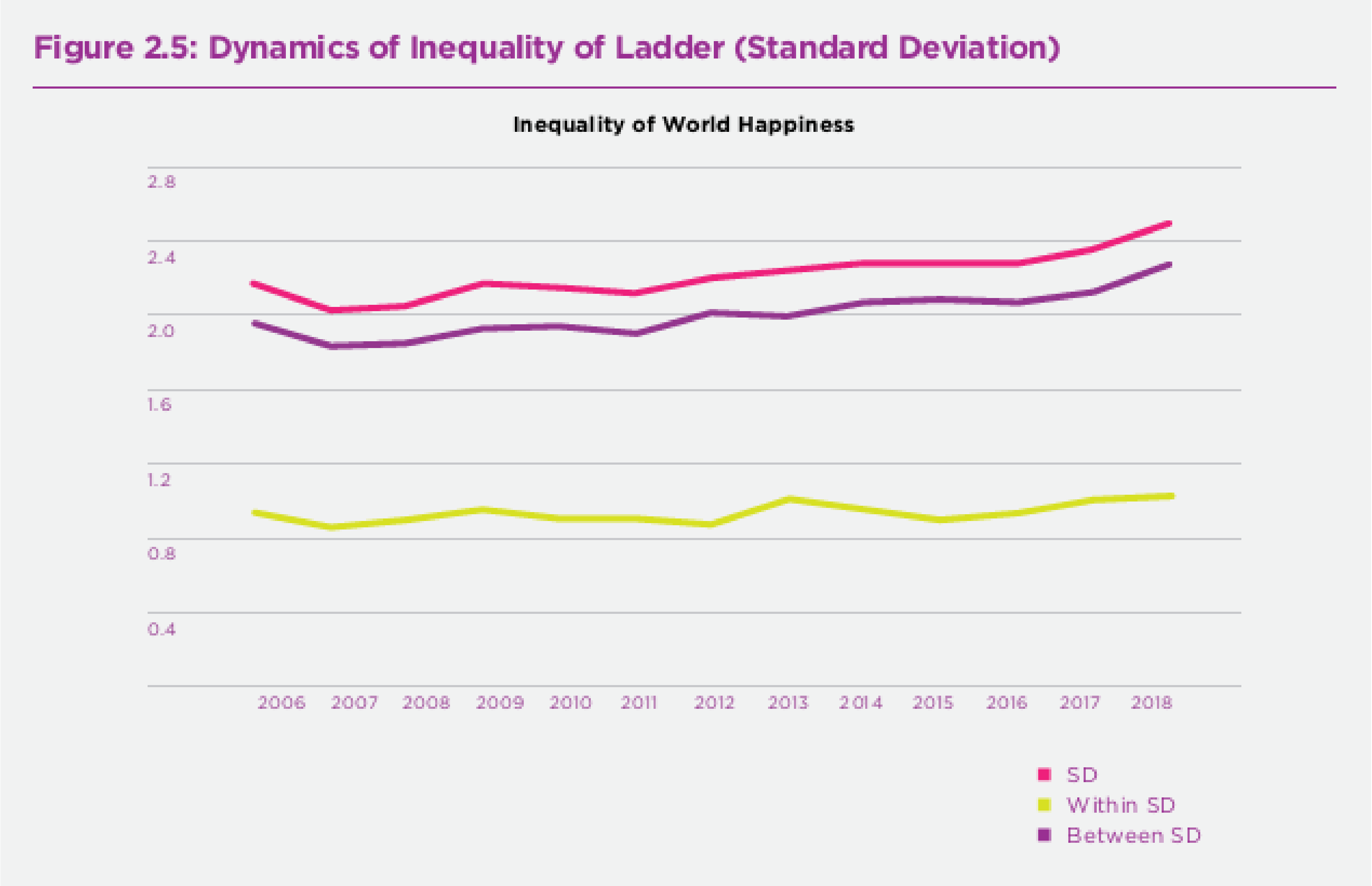
Note: “Poor” well-being refers to those providing life satisfaction, worthwhile, and happiness ratings of 0 to 4 on an 11-point scale, and anxiety ratings of 6 to 10. 'Very good' well-being refers to those providing a rating of 0-1 for anxiety and 9-10 for happiness, life satisfaction and worthwhile.

Source: ONS, Personal Well-being in the UK, July 2017 to June 2018, 2018

Distributional orderings based on comparisons of CDFs are robust to changes in the scale

- The ONS's charts of trends in the fraction of people with 'poor' (or with 'very good') SWB are based on comparisons of points on the CDFs
- If the distribution for Year 1 has a smaller (larger) fraction of people with 'poor' SWB than does the distribution for Year 2, changing the scale does not change the result
  - Scale [0,1,2,3,**4**,5,6,7,8,9,10] with 'poor threshold '4' and scale [-5,-4,-3,-2,**-1**,0,1,2,3,4,5] with 'poor threshold '-1' would provide same fractions for the distributions in Years 1 and 2
- Hence strong case for placing comparisons of CDFs at heart of analysis of ordinal data

Tracking well-being: inequality over time



Note: "inequality" is measured using the standard deviation of the SWB score

Source: Helliwell, Layard, and Sachs (eds), World Happiness Report 2019

Standard inequality indices should not be applied to ordinal data

- Inequality orderings based on the SD and other ‘absolute’ measures (e.g. absolute Gini) don’t change if each scale value is changed by the same amount, e.g. if $c = [1,2,3] \rightarrow c' = [3,4,5]$.
  - But why this transformation only – why rule out others?
- Inequality orderings based on standard relative measures (as earlier) are not robust because the indices depend on ratios of incomes to the mean, and the mean is not order-preserving with changes to the scale

Inequality and ordinal data

Fundamental questions to be addressed:

1. What is the 'perfect equality' benchmark?
   - Everyone gives the same response (all in same category)
2. Dispersion in 'location' relative to what reference point?
   - Not the mean!
3. What is the 'perfect inequality' benchmark?
4. 'More unequal' is characterised how?
   - Cf. mean-preserving spread in traditional (Lorenz-based) approach

- Dominant approach among economists to date is to: take the median as the reference point; interpret 'more unequal' as greater dispersion around the median; and thence interpret 'perfect inequality' benchmark as being when half the responses are in lowest category and half in the highest
  - NB strictly speaking, given the perfect inequality benchmark, this approach summarises *polarisation*, rather than inequality
- There is a second (newer) tradition defining 'greater spread' differently

Median-based approaches to inequality of ordinal data (Allison & Foster, *JHE* 2004)

- Relative position of the median doesn't change if the scale changes
- ***F*-dominance** (First-order dominance): when CDFs don't intersect (and ordering of means robust to scale)
- ***S*-dominance**: Inequality as 'greater spread away from the median' (Allison & Foster
    - Cf. Lorenz-dominance: 'greater spread away from the mean'
  - Given two distributions *x* and *y*, *x* has greater spread than *y* if:
    a. Distributions *x* and *y* have the same median *m*; and
    b. for all categories $k < m$, cumulative population share of lowest *k* categories of $x \geq$ cumulative population share of lowest *k* categories of *y*; and
    c. for all categories $k \geq m$, cumulative population share of lowest *k* categories of $x \leq$ cumulative population share of lowest *k* categories of *y*
  - See over for graph contrasting *S*-dominance with *F*-dominance

Median-based approaches to inequality of ordinal data (Allison & Foster, *JHE* 2004)

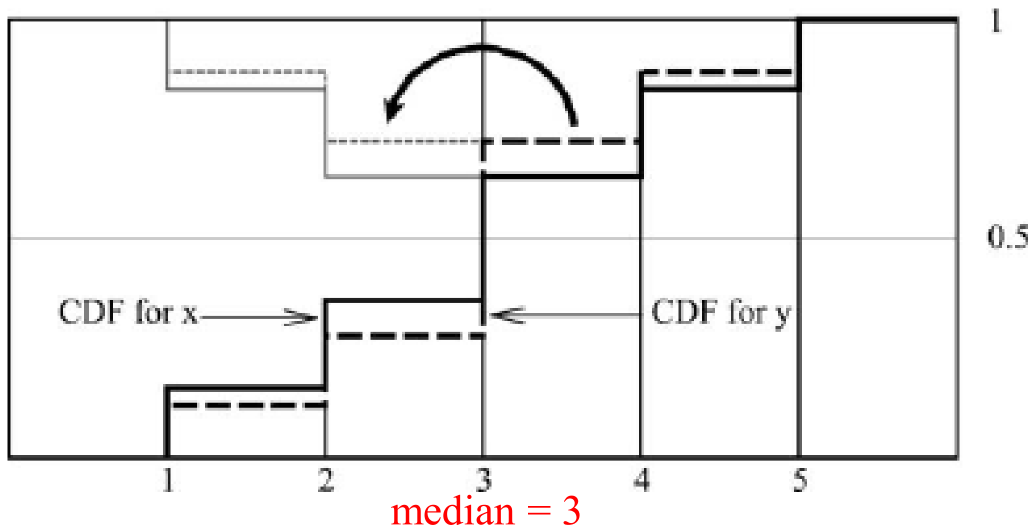


Fig. 2. Ranking of CDFs and the construction of *S*-curves. The CDFs for $x = (4, 5, 7, 5, 4)$ and $y = (3, 4, 11, 4, 32)$, depicted above with thicker lines, cross at a common median level of health, 3. The portion of the CDFs that lie above the median are 'flipped' across the median to form the upper halves of the *S*-curves for x and y.

- **Distribution *x* has greater spread about median than *y***: CDF(x) ≥ CDF(y) below the median category, and CDF(x) ≤ CDF(y) at median category and above (and common median)
  - more of density mass is concentrated in the extremes, unambiguously further away from the median

Inequality (polarisation) indices consistent with the median-based approach

- **Allison-Foster** (2004) and **Dutta-Foster** (*RIW*, 2013): mean level of SWB above the median *minus* mean level of SWB below the median
  - Scale-dependent index but 'gap' idea is intuitive and easily grasped
- **Abul Naga & Yalcin** (*JHE*, 2008): weighted difference between the proportion of people in the lower and upper halves of the distribution
  - Scale-independent, and ranges between 0 (equality) and 1 (complete polarisation)
  - Choice of parameters (α, β) affects weight given to top and bottom halves (e.g. more to bottom if $\beta \geq \alpha \geq 1$)
    - I use $\alpha = 1$, $\beta = 1, 2, 4$ in examples shown later
- And many other indices, including by **Apouey** (*HE*, 2007)
  - Apouey $P2(e)$ class … includes Blair-Lacy index = $P2(2)$

LSE

A different approach: Cowell and Flachaire (2017)

- An individual's 'status' is summarised by the fraction of persons in the same or lower categories as him/her
  - This is C-F's peer-inclusive 'downward-looking' definition
    - C-F's 'upward-looking' definition not considered here (but see later – time permitting)
  - I.e. status distribution is described by the CDF and is independent of scale
- Inequality summarised by people's 'distances' from a reference value, taken to be the **maximum** value of the status variable (one)
  - C-F argue that: (i) reference point has to be maximum if using peer-inclusive status definition; and, anyway, (ii) the median is not always well-defined for categorical data
  - Inequality = 0 if all have the same status (all in same category)
  - Perfectly polarised distribution is not necessarily the most unequal distribution; a uniform distribution has greater inequality (Jenkins)
- C-F family of inequality indices, CF(γ) where parameter $0 < \gamma < 1$, plus limiting case for $\gamma = 0$
  - The smaller γ is, the more sensitive is CF(γ) to values of 'status' close to zero, i.e. those status values close to one

LSE

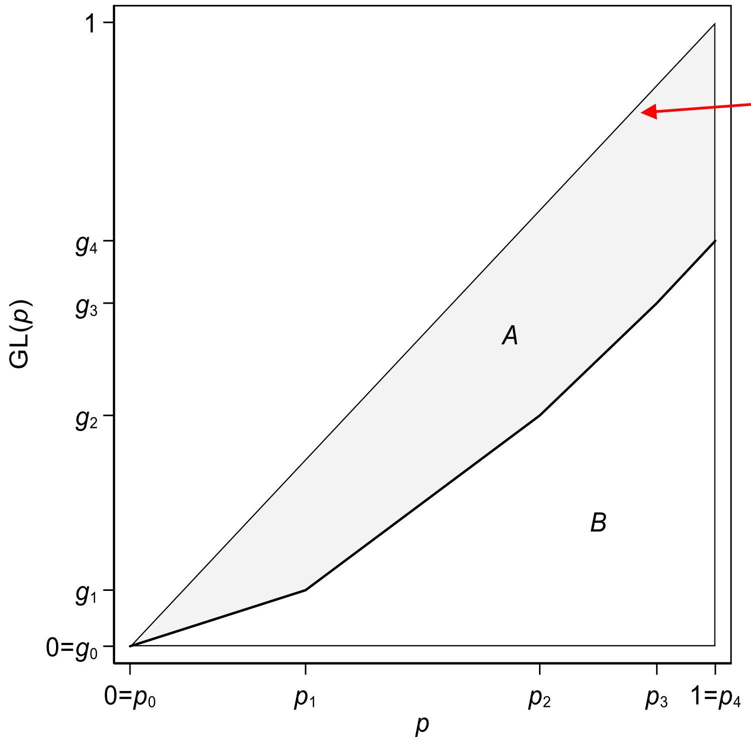
GL curve for distribution of 'status' (*s*)

[Jenkins, IZA DP 12811, 2019]

- $GL\left(\boldsymbol{s}, \frac{m}{n}\right) = \frac{1}{N}\sum_{i=1}^{m} s_i\,,\ m = 1, \ldots, N$
- $GL(\boldsymbol{s}, 0) = 0$; $GL(\boldsymbol{s}, 1)$ = arithmetic mean of $\boldsymbol{s}$
- Area-based inequality index $J = A/(A + B) = 1 - 2B$

Example: 4 category case

Curve has kinks because discrete distribution: vertices are at $\{p_0 = 0, 0\}$, and $\{p_k, \sum_{j=1}^{k} f_j F_j\}$ for each $k = 1, \ldots, K$ with $p_k = F_k$



45° ray is GL curve if complete equality

Inequality dominance results (non-polarization approaches), (1)

Jenkins (IZA DP 12811, 2019):

Equivalence of (a) and (b):

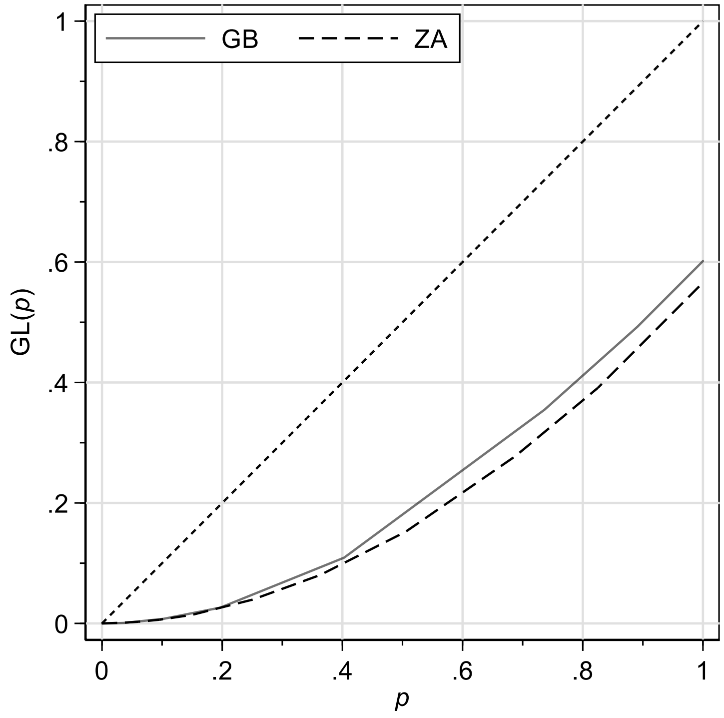
- (a) generalised Lorenz curve for 'status' distribution *A* everywhere on or above the generalised Lorenz curve for 'status' distribution *B* ⇔ (b) *A* is more equal than *B* according to all C-F inequality indices and the *J* index

- Result also goes through if C-F's peer-inclusive 'upward-looking status' definition is used, with appropriate re-definition of GL curve ordinates

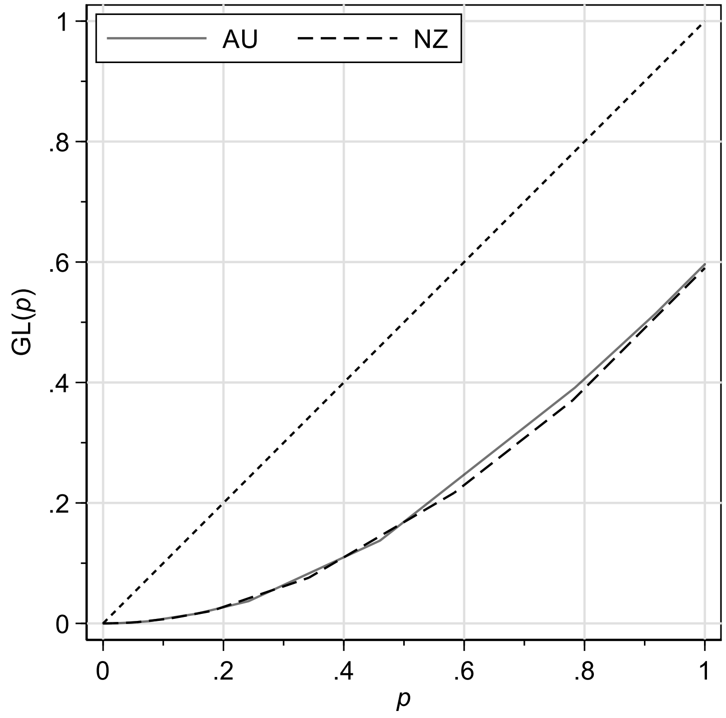
Comparison of GL curves for status

(Data from WVS, wave 5)

(a) Dominance



(b) Non-dominance



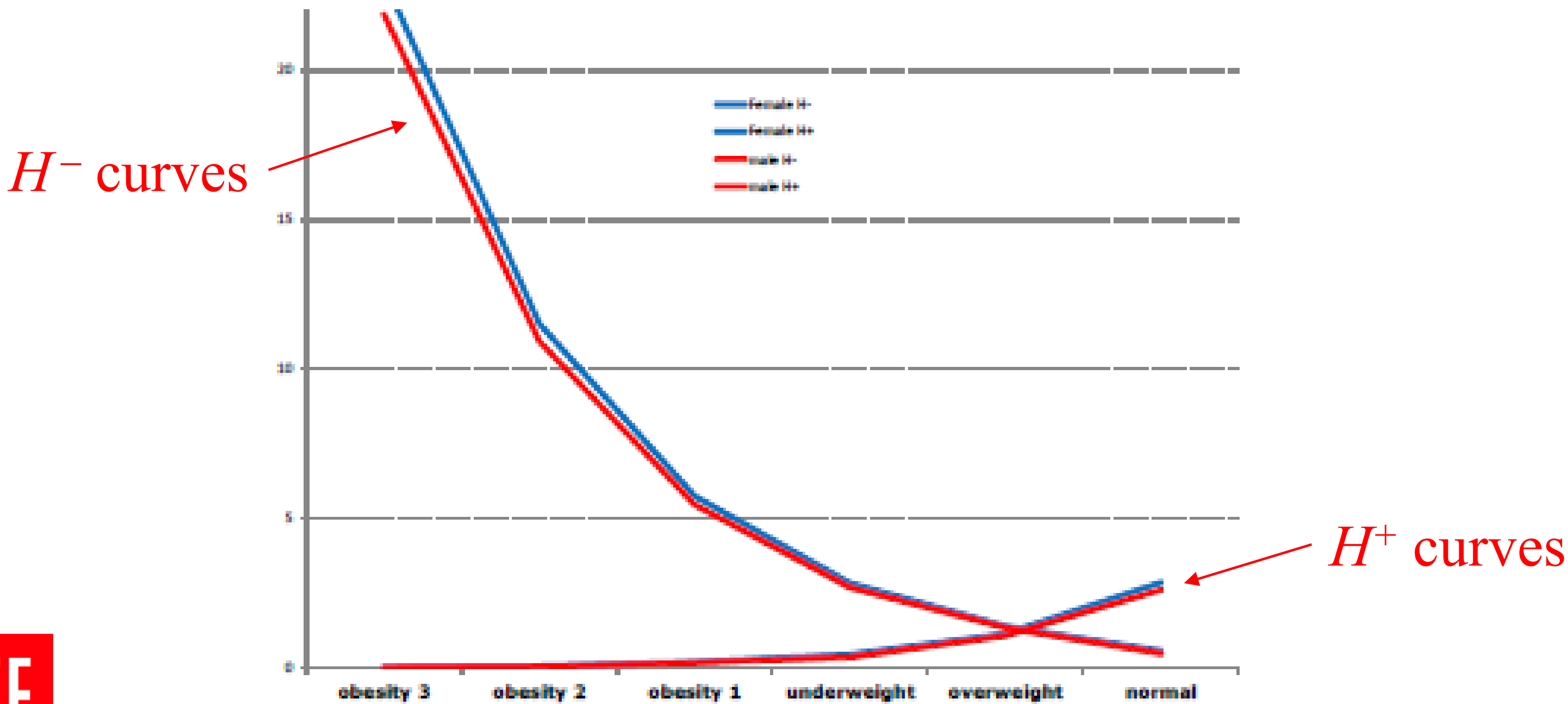
Inequality dominance results (non-polarization approaches), (2)

Gravel, Magdalou & Moyes (2015), revised:

- inequality increases if there is a shift in density mass away from a particular category (some shifted downward, some upward) – concept of a disequalising 'Hammond transfer'
  - Cf. regressive Pigou-Dalton transfer in the case of cardinal variables such as income
- "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." (p. 6)
- H^{+} and H^{-} criteria each involve specific cumulations across categories of CDF values (or survivor function values: 1–CDF)
- NB H^{+} criterion is a second-order 'welfare' criterion, so if there is F-dominance (first-order dominance), there is also H^{+} dominance
- Yet to be shown: whether the GM&M dominance criteria are consistent with unanimous rankings by C-F indices

H^+ and H^- dual-dominance criteria

- Distribution A 'more equal' than distribution B according to GM&M criteria (consistent with equalisation via Hammond transfers) ⇔ H^+ curve for A everywhere on or below H^+ curve for B **and** H^- curve for A everywhere on or below H^- curve for B
- Example: Discrete distributions for BMI, French men and women, 2008



4. Empirical illustrations using World Values Survey data

(For more information about the WVS, see here)

World Values Survey (WVS)

- "The WVS consists of nationally representative surveys conducted in almost 100 countries which contain almost 90 percent of the world's population, using a common questionnaire."
- One respondent adult per household (aged 18+)
- 6 waves to date:
  - 1: 1981–1984
  - 2: 1989–1993
  - 3: 1994–1998
  - 4: 1999–2004
  - 5: 2005–2009
  - 6: 2010–2014
- Only source with unit-record data for NZ that I can access
  - Data available for 1998 (wave 3), 2004 (wave 3), 2011 (wave 6)

LSE

WVS data: cross-national comparisons

Comparisons of distributions of life satisfactions for 5 countries

- NZ and 4 other countries: AU, GB, US, ZA
- WVS wave 5: mid-2000s
- Sample sizes are not large:

ISO 3166 alpha-2 code	Year survey 2004	2005	2006
AU	0	1,421	0
GB	0	1,041	0
NZ	954	0	0
US	0	0	1,249
ZA	0	0	2,988

LSE

Life Satisfaction across countries

NZ, AU, GB, US, ZA

WVS wave 5 (mid-2000s)

Relative frequencies

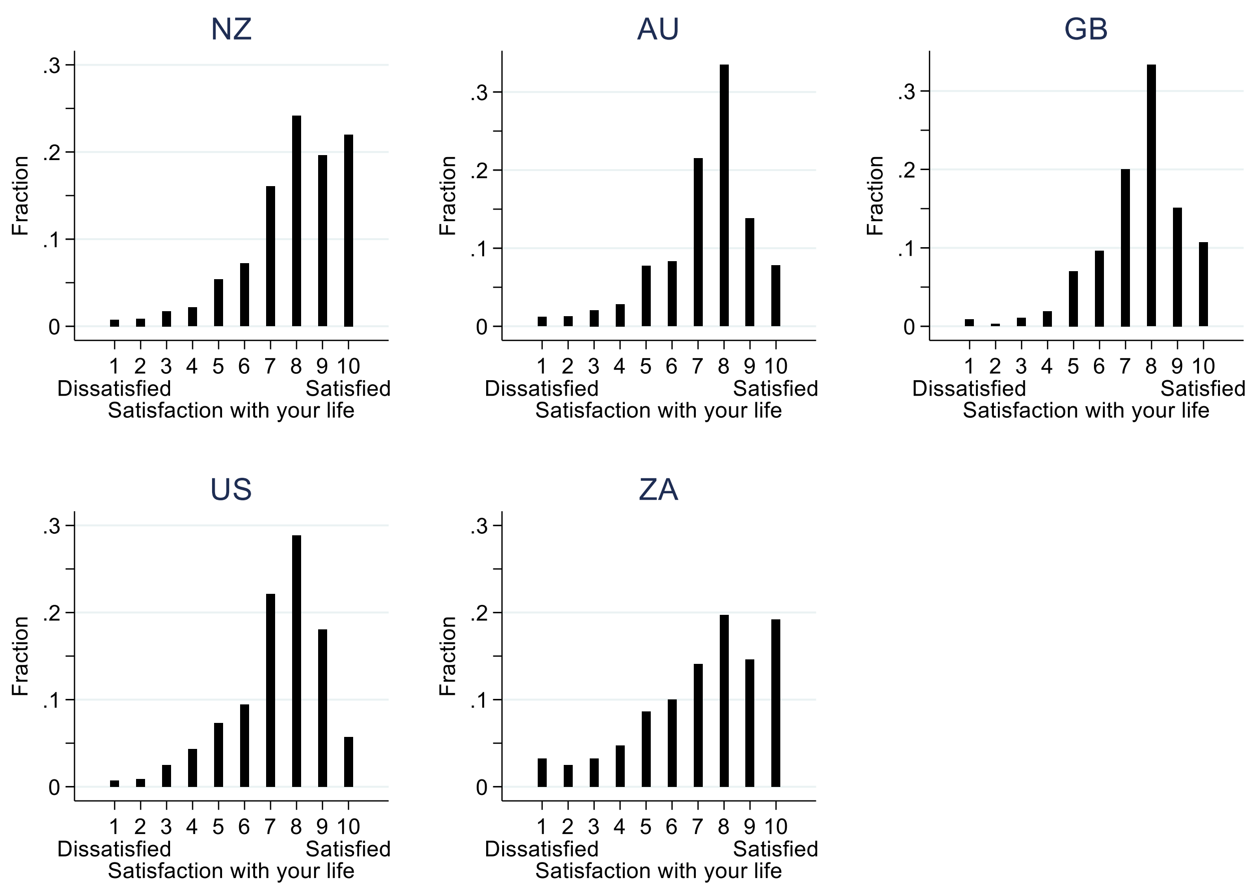
Dominance checks (F, S, GL, H^{+} and H^{-})

Inequality indices

Do you remember your answer to the LS Qn?

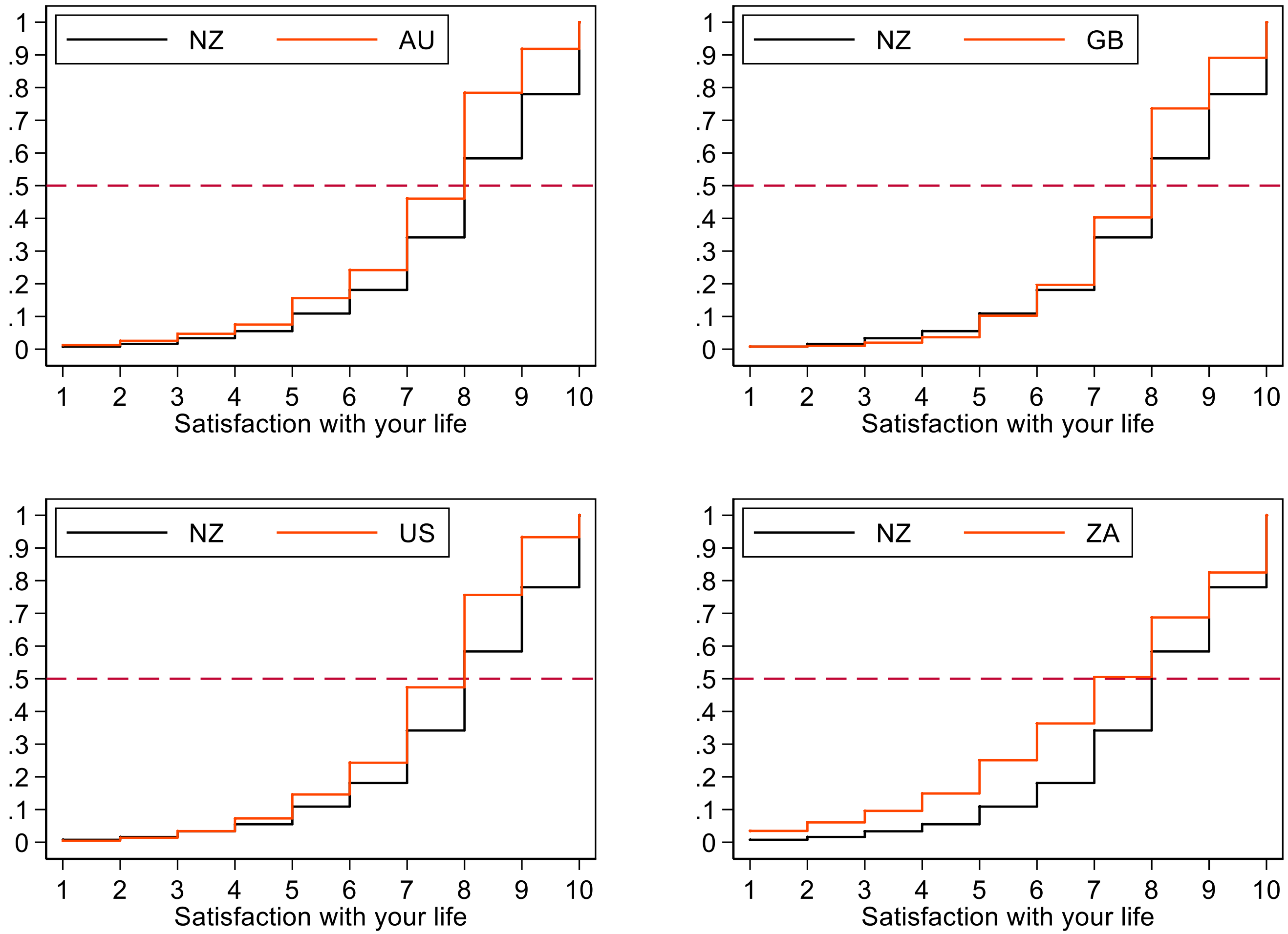
Relative frequency distributions for LS, across countries [nzae09]

- Median = 8 in all countries except ZA (7). Mode = 8 in all 5 countries
- Means are NZ 7.9, 7.3 AU, 7.6 GB, 7.3 US, and 7.0 ZA
- NZ distribution appears to differ from AU, GB, US: relatively large fraction with 10
- ZA also with relatively high fraction with 10, but note also high fraction with very low LS



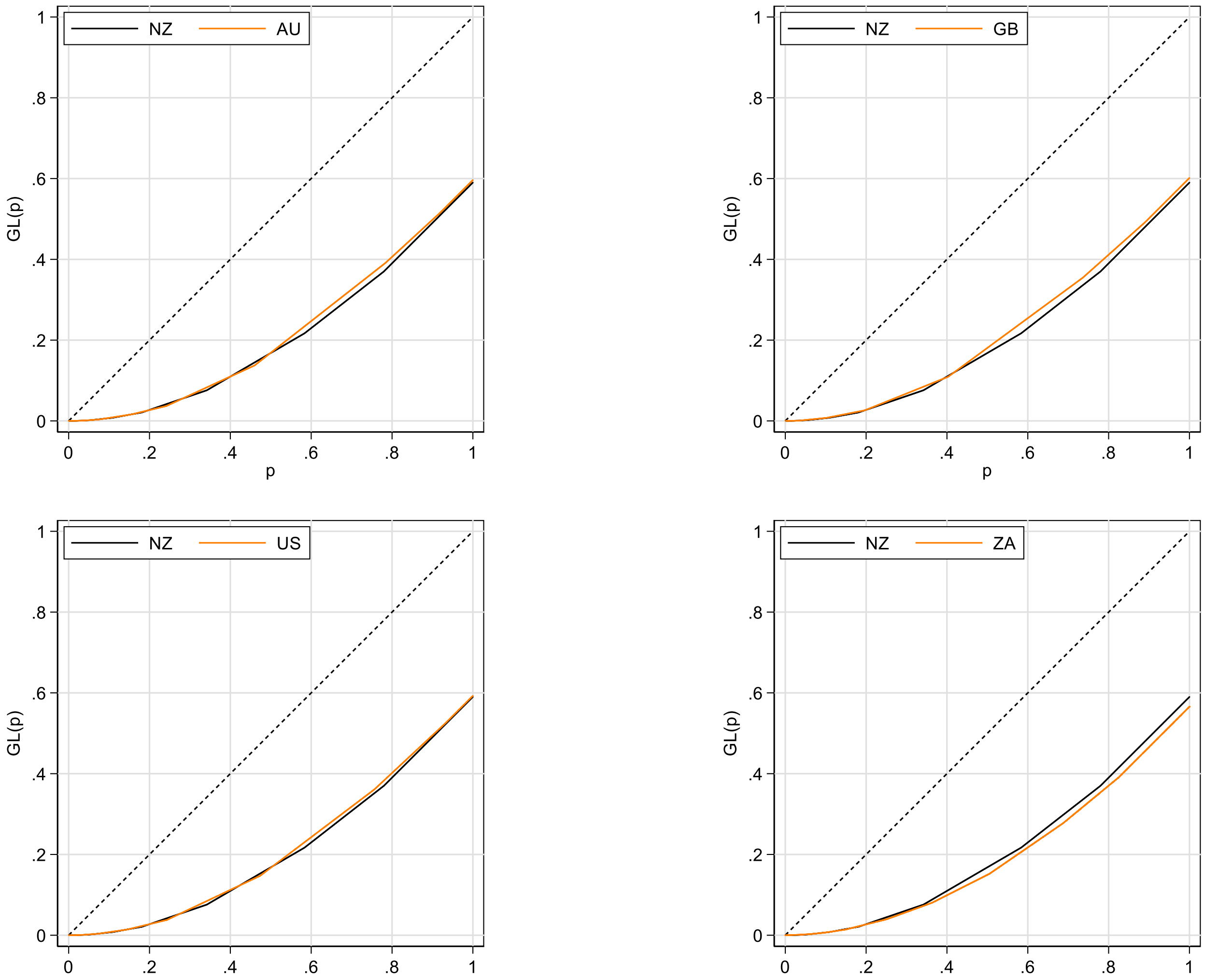
LS: CDF comparisons across countries [nzae02]

- Pairwise comparisons of NZ with each other country (paper has all comparisons)
- NZ *F*-dominates AU, US, and ZA , and hence NZ has higher mean, regardless of scale
- No *S*-dominance results at all (partly because: ZA has lower median; *F*-dominance)



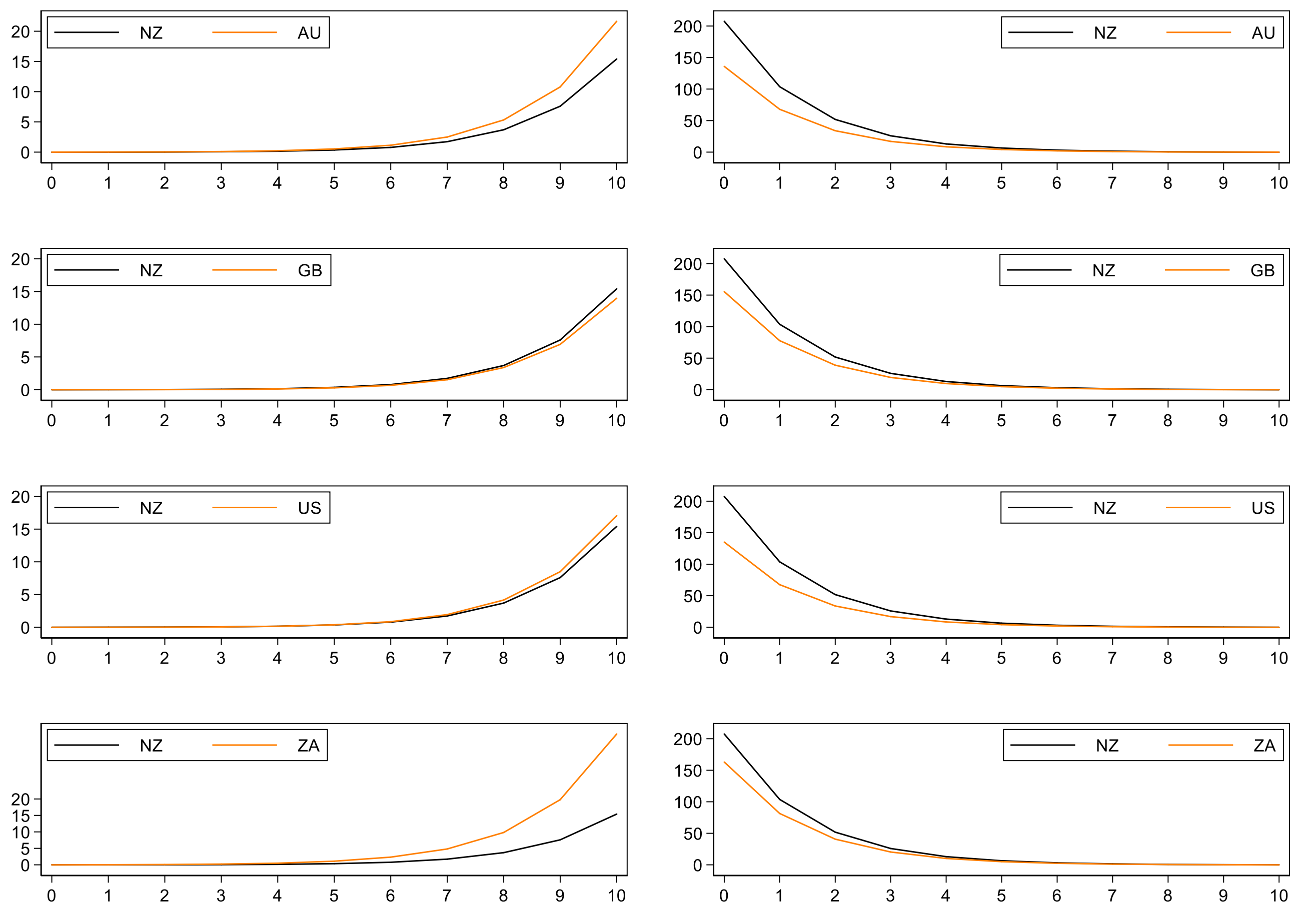
LS: GL dominance? [nzae08]

- More unequal according to all C-F and J indices: GL curve further from diagonal line
- NZ more equal than ZA and more unequal than GB, but differences small



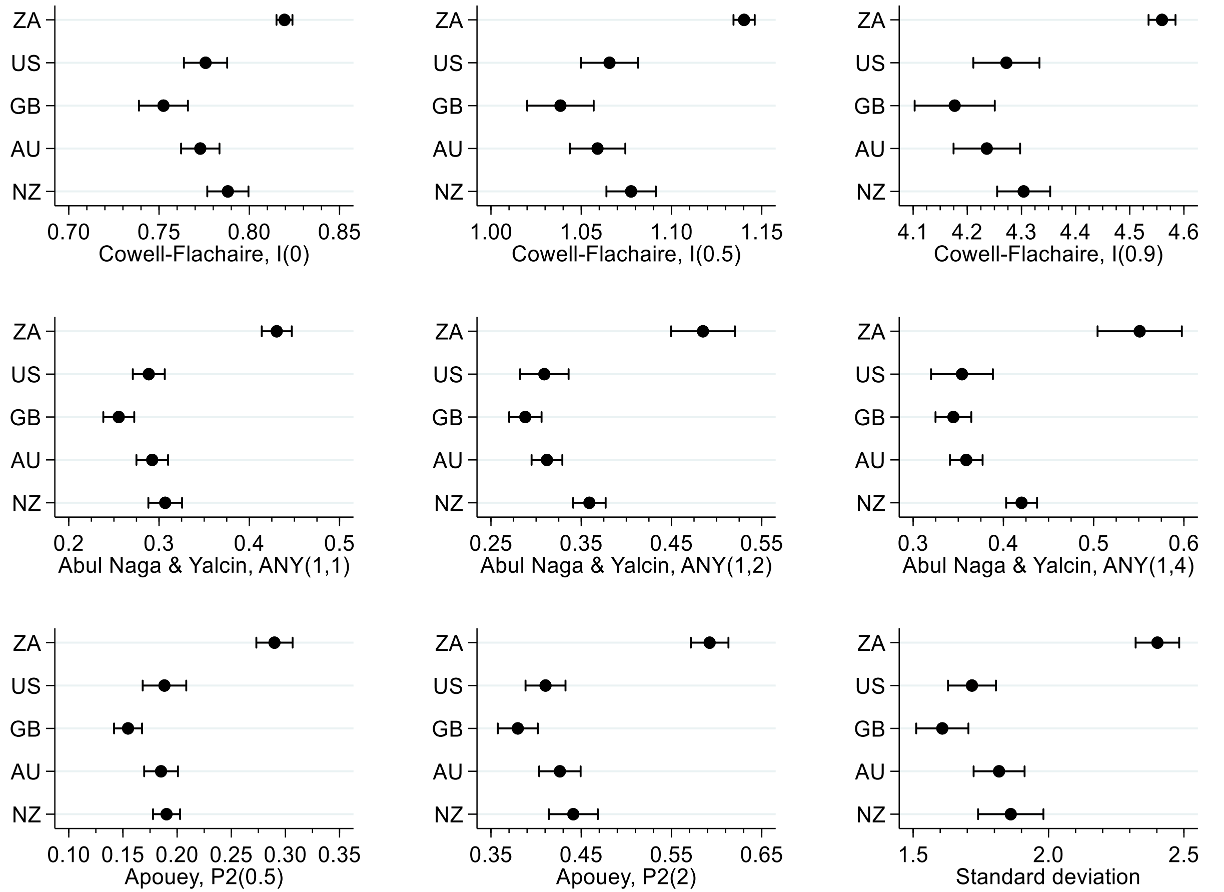
LS: H^{+} (lhs) and H^{-} (rhs) dominance? [nzae08]

- If curve for year *A* is everywhere below that for year *B* (*A* dominates *B*), *and* holds for both H^{+} and H^{-} comparisons, then *A* is more equal than *B*
- NZ appears more unequal than GB, but differences small; no other dual-dominance results



LS: Inequality indices (selection) [nzae12]

- Point estimates and 95% CIs (bootstrapped SEs, 500 replications)
- Inequality higher in ZA than in each of the other 4 countries (as per dominance)
- GB inequality point estimate is lowest (depending on index)
- NZ inequality point estimates close to those for US



5. Concluding remarks

(a) Analysis matters

- Don't apply methods for cardinal data to ordinal SWB data
- There's now a well-developed (and growing) 'toolbox' for undertaking distributional comparisons of SWB data, so they can be used for assessing social progress
  - For SWB levels and (less commonly-studied) SWB inequality
  - For 'showing the data' and dominance checks, and for summary indices of inequality
- Analytical challenges remain, e.g.
  - Which indices are consistent with GM&M's dual-dominance criterion?
    - E.g. are C-F indices or *J*?
  - Statistical inference
  - Combining information across SWB dimensions
    - Only life satisfaction and self-assessed health here; many more in the Living Standards Framework and OECD dashboard of indicators
- NB don't throw away income distribution comparisons!

(b) Suggestive findings

According to WVS data:

- NZ has high LS levels relative to AU, US, ZA, but has more LS inequality than AU and US (but not ZA)
- The distribution of Life Satisfaction in NZ changed little between 1998 and 2011
- Patterns for health status differ from those for life satisfaction
- Comparisons of SWB across subgroups within a population are bedevilled by small sample sizes (large SEs)
  - Which underlines that looking at SEs is important
- Dominance checks may not be clear cut (curves intersect), but they are valuable for showing the data as well
  - Show and compare CDFs and relative frequency distributions

LSE

6. Estimation using **`ineqord`**

Stata users:

```
ssc install ineqord
```

Life satisfaction data used for the examples

- ‘APS’: Annual Population Survey Three-Year Pooled Dataset January 2015 - December 2017
- Data downloadable from the UK Data Service by researchers who register with them
- Nationally-representative survey of UK adults
- APSs are used by the UK’s Office for National Statistics to provide annual reports on personal well-being (see example on earlier slide)
- 530,300 (unweighted) observations of which 275,336 provide a non-missing response to the life satisfaction question: see over
  - Sample weights provided (variable **`PWTA17C`**), but not cluster or strata information

APS and life satisfaction (LS)

- Respondents 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'?"
  - Different LS scale from WVS (11 not 10 cats; scale starts at 0)
- Responses in the variable **`SATIS`**
- Missing values are recorded as –8 and –9
- All variables names are in upper case
- Before using **`ineqord`**, I convert the missing values to Stata missing values and, for convenience, put all variable names in lower case …

LSE

APS data management (1)

```
. use aps_3yr_jan15dec17_eul, clear
. rename _all, lower
. 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  [aw = pwta17c]

 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
```

LSE

APS data management (2)

- Scale runs from 0 to 10; it does not start at 1
- If **`ineqord`** were applied ignoring this, it would provide incorrect estimates
- 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 estimates on next slide for the UK adult population
- 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)

LSE

ineqord: example output

```
. 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           # distinct 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 | Allison-Foster    Average Jump           P2(2)           P2(1)           P2(.5)
----------+---------------------------------------------------------------------------------
          |        2.45139         0.24514         0.37056         0.24514         0.14555
--------------------------------------------------------------------------------------------
```

Continued over …

ineqord: example output (ctd.)

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) and J_u (upward-looking status)

All obs	Jd	Ju
	0.56744	0.55128

Comparisons of LS by marital status

- Create a new variable **mstat** collapsing the information held in the **masta** variable
  - Individuals in a cohabiting relationship are treated 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
```

Comparisons of LS by marital status (ctd)

```
. ta mstat [aw = pwta17c]

                          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
```

Ignore 'Other' henceforth

```
. ta ls mstat [aw = pwta17c], col nofreq
```

NB Now using derived variable **ls**

```
                      |              Marital status
          = satis + 1 | Single, n  Married,   Separated  Other (cu |     Total
----------------------+--------------------------------------------+----------
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
```

Dominance checks

- All the raw materials for the various dominance checks can be created by **`ineqord`** using the ‘cat’ and ‘gl’ options
- New variables are created that can then be **`list`**ed or displayed graphically

```
.          // single, never married
. ineqord ls [aw = pwta17c] if mstat == 1, alpha(.9) ///
>          catv(v_snm) catpr(f_snm) catcpr(F_snm) catspr(S_snm) ///
>          gldvar(gld_snm) gluvar(glu_snm)

<output not shown>

.          // married, living with spouse
. ineqord ls [aw = pwta17c] if mstat == 2, alpha(.9) ///
>          catv(v_m) catpr(f_m) catcpr(F_m) catspr(S_m) ///
>          gldvar(gld_m) gluvar(glu_m)

<output not shown>

.          // Separated, divorced, or widowed
. ineqord ls [aw = pwta17c] if mstat == 3, alpha(.9) ///
>          catv(v_sdw) catpr(f_sdw) catcpr(F_sdw) catspr(S_sdw) ///
>          gldvar(gld_sdw) gluvar(glu_sdw)

<output not shown>
```

Variables created (married group)

```
. sort v_m

. list v_m f_m F_m S_m gld_m glu_m if !missing(F_m)
```

CDF ordinates for *F*-dominance checks

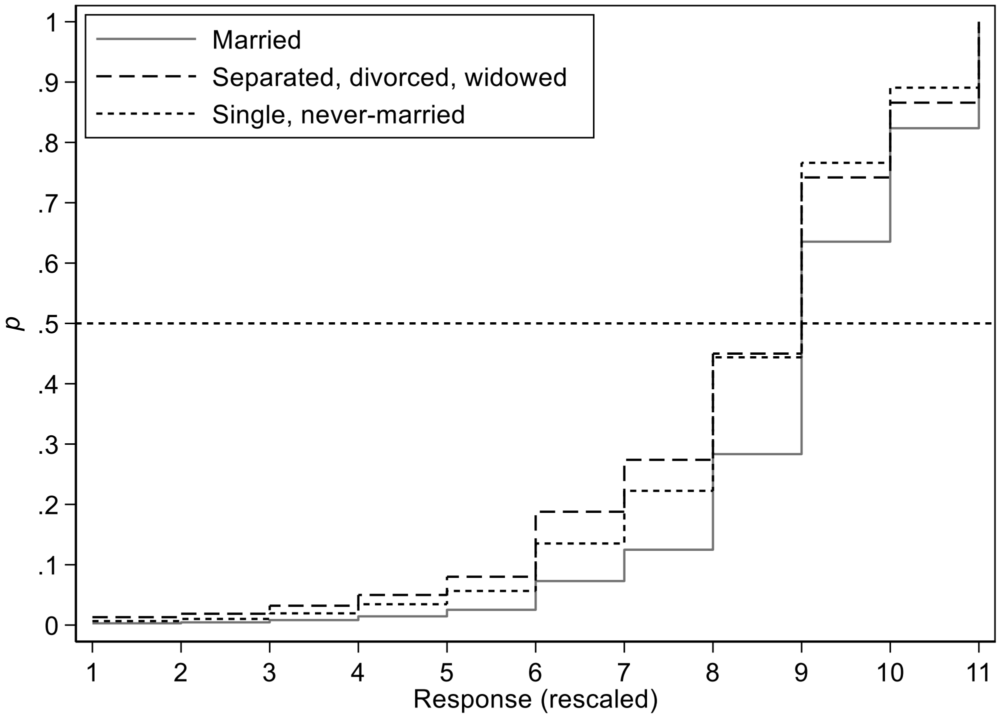
	v_m	f_m	F_m	S_m	gld_m	glu_m
1.	1	.0029368	.0029368	1	8.62e-06	.6103632
2.	2	.0015666	.0045034	.9970632	.0000157	.6074263
3.	3	.0038127	.0083161	.9954966	.0000474	.6058643
4.	4	.0061448	.014461	.9916838	.0001362	.6020688
5.	5	.0108831	.0253441	.985539	.0004121	.595975
6.	6	.0476428	.0729869	.9746559	.0038894	.5852493
7.	7	.0519418	.1249287	.9270132	.0103784	.5388139
8.	8	.1584475	.2833762	.8750713	.0552786	.4906632
9.	9	.352071	.6354473	.7166238	.2790012	.3520103
10.	10	.188099	.8235463	.3645527	.4339095	.0997078
11.	11	.1764537	1	.1764537	.6103632	.0311358
148810.	.	.	0	0	0	0

Zeros added by default; useful for graphs

GL curve ordinates: (downward-looking) status

F-dominance and *S*-dominance

```
. 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(s1color) graphregion(color(white)) ///
>         saving(aps01_Fdom_m-sdw-snm.gph, replace)
(file aps01_Fdom_m-sdw-snm.gph saved)
```



M *F-dom* SDW, and
M *F-dom* SNM

CDF(SNM) and
CDF(SDW) cross

median(LS) = 9

SDW *S-dom* SNM:
below the median,
CDF(SDW)
further from the median than
CDF(SNM) and the reverse
is the case above the median

Unanimous rankings by C-F and *J* indices?

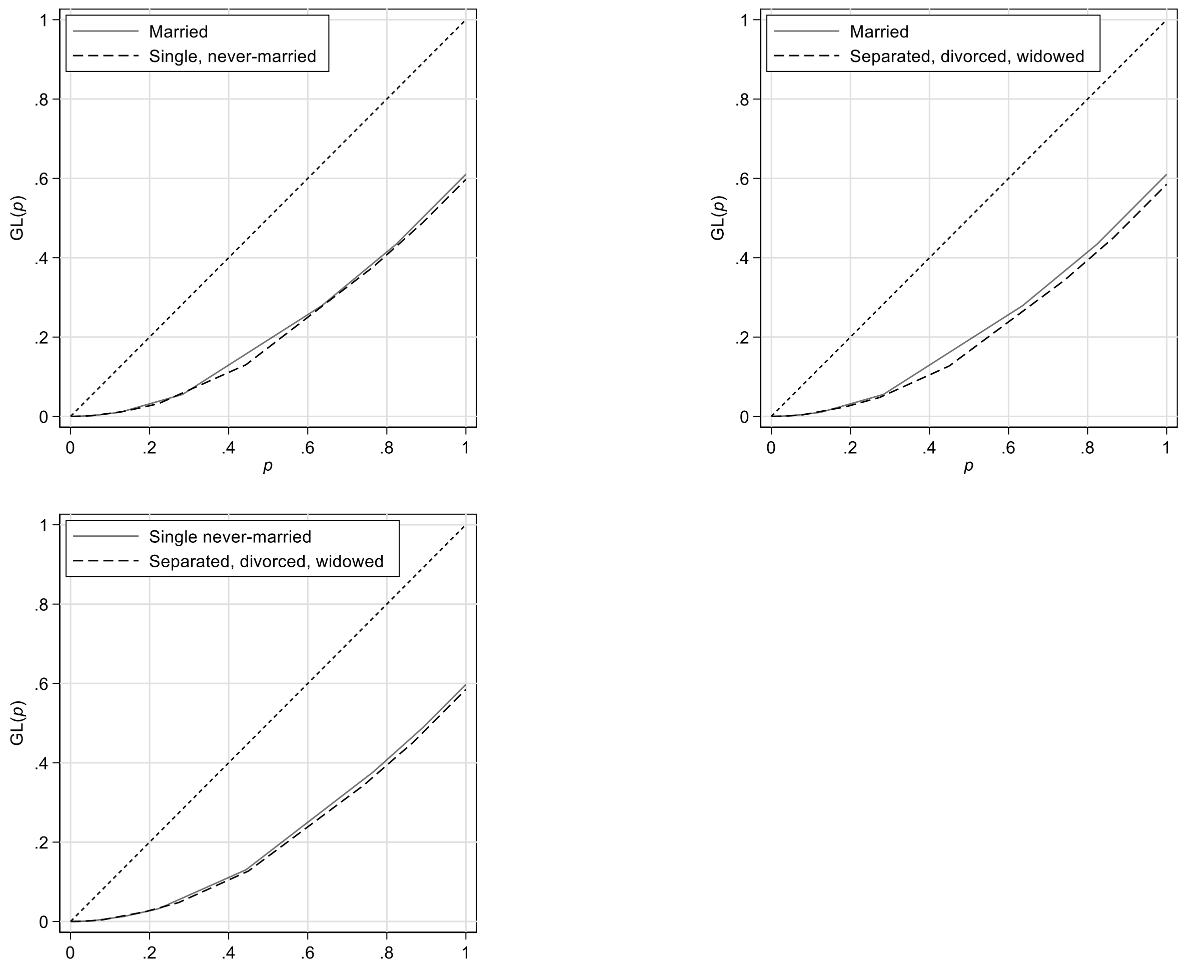
- Code used for the Married and SDW groups' comparison is below

```
. 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(s1color) ///
>          saving(aps01_gld_sdw-m.gph, replace)
(file aps01_gld_sdw-m.gph saved)
```

- Analogous code for the other 2 pair-wise comparisons followed by a **`graph combine`** produced the figure on next slide

Unanimous rankings by C-F and *J* indices?

- All 3 pairwise comparisons reveal dominance: 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



Dual-dominance (GM&M criteria)

- I have not undertaken dual-dominance checks using these data
- Calculations of ordinates of H^{+} and H^{-} curves are not available as an option in **`ineqord`** at present (perhaps in future version?)
- But they can be derived post-estimation using estimates of the CDF

Indices of polarization and inequality

- Selection: inequality indices CF(α) for $\alpha = 0, 0.25, 0.5, 0.75, 0.9$; and *J*; and polarization indices, *ANY*(1, 1), the top-sensitive *ANY*(4, 1), and the bottom-sensitive *ANY*(1, 4)
- How how one can derive standard errors for the indices using Saigo et al.'s (2001) repeated half-sample bootstrap and Van Kerm's (2013) **`rhsbsample`** (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 WVS examples earlier)
  - Hence this code may be usefully applied in other contexts

Indices: derivations and summary

For each group:

1. drop observations with missing values
2. use **`rhsbsample`** to create the bootstrap sample weights
3. **`svyset`** the data
   - If survey design variables other than weights had been available, this is where they would have been cited
4. Call **`ineqord`** using the **`svy bootstrap`** prefix command
   - **`alpha(0.9)`** option used to derive estimates of C-F for values of α spanning its range
   - Estimates for more polarization indices than I cited earlier, just in case
   - 'd' suffix on the estimates' names reminds us that I am using C-F's peer-inclusive downward-looking status definition
5. Save the estimates of indices, standard errors and confidence intervals to a dataset using Newson's **parmest** (SSC)
6. [Repeat steps 1–5 for the other 2 groups; combine the data sets using **`append`**; draw the graphs summarizing the results]

LSE

Derivations and summary: code (married)

```
. use aps_3yr_jan15dec17_eul if MARSTA == 2, clear
< code creating "ls" for Married group omitted >
. drop if missing(ls)   // wise to run bootstrap with no missing obs in file
(164,851 observations deleted)

. rename pwta17c wgt  // for convenience
 . drop if missing(wgt)
(0 observations deleted)

. loc R = 500
. forvalues i = 1/`R' {
  2.          qui gen rhsbrw`i' = .
  3.          qui rhsbsample, 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>
```

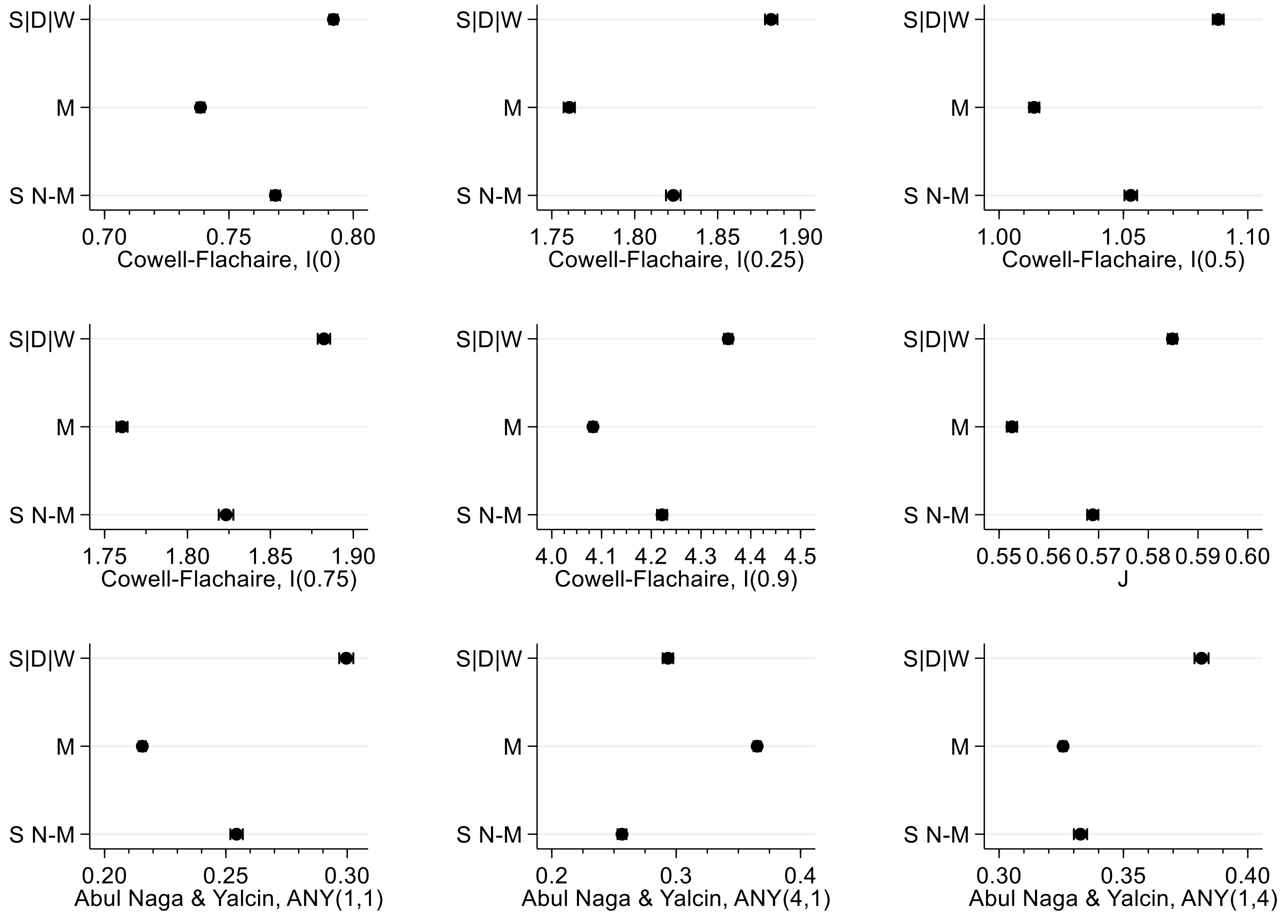
Derivations and summary: code (married)

```
. 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
```

Indices, by marital status group

[Bars show 95% CIs]

- Recall that no *S*-dominance for comparisons between the Married and each of the other 2 groups
- Figure shows that inequality ranking depends on the index: for *ANY*(1, 1) and *ANY*(1, 4), the ranking is the same as for the inequality indices; but for top-sensitive index, *ANY*(4, 1), Married shows the greatest polarization rather than the lowest (NB Married group has relatively large fractions of responses at the 2 top two LS scale points
- Magnitudes of inequality and polarisation differences across groups depends on the index chosen



LSE

Envoi

Empirical analysis = principles + data + estimation

1. Review of the *principles* of how to compare distributions of ordinal data (concepts + results)

plus

2. Examples of *empirical implementation* to illustrate how to apply the principles (data + estimation)
   - Detailed study of data and data sources is essential in all applied work – in addition to 'pure' statistical issues
   - Writing programmes (and Stata ado-files in particular) are a very good way to learn about practical issues of real-world applications of the principles
     - General: 'Messy' but crucial issues – e.g. proper error handling, weighting and other sample design features, proper handling of missing values, etc. – are all too often neglected
     - Specific: e.g. **`ineqord`** deals with complications arising if no responses for a level; and non-standard scale labels
   - Graphs are often much better than tables for summaries

LSE