Comparing latent inequality with ordinal data

David M. Kaplan* Longhao Zhuo†

February 14, 2019

Abstract

Using health as an example, we consider comparing two latent distributions when only ordinal data are available. Distinct from the literature, we assume a continuous latent distribution but not a parametric model. Primarily, we contribute (partial) identification results: given two known ordinal distributions, what can be learned about the relationship between the two corresponding latent distributions? Secondarily, we discuss Bayesian and frequentist inference on the relevant ordinal relationships, which are combinations of moment inequalities. Simulations and empirical examples illustrate our contributions.

JEL classification: C25, D30, I14

Keywords: health; nonparametric inference; partial identification; partial ordering; shape restrictions

1 Introduction

One common variable in health economics is self-reported health status (SRHS, a.k.a. self-assessed health). SRHS appears in important survey datasets like the Panel Study of Income Dynamics. Respondents rate their health from "excellent" to "poor" (or similar descriptions), often with five categories; i.e., SRHS is an ordinal variable. Despite its imprecision and subjectivity, SRHS has been valued for its synthesis of all dimensions of health, its strong correlation with more objective health measures, its broad availability, and its usefulness over a wide age range; e.g., see Deaton and Paxson (1998a, §I) or Allison and Foster (2004, p. 506).

SRHS aids the study of two types of inequality: between-group and within-group. The first type is inequality between two (sub)populations (e.g., socioeconomic groups), where one has "better" health than the other. The second type is inequality within a population,

*Corresponding author. Email: kaplandm@missouri.edu. Mail: Department of Economics, University of Missouri, 909 University Ave, 118 Professional Bldg, Columbia, MO 65211-6040, United States.

†University of Missouri. Email: longhao.zhuo@gmail.com.

where a widely dispersed distribution represents big health differences among individuals within the population.

More generally, our results help learning about inequality in any latent variable for which only ordinal data are available. Our results can apply to health, happiness, bond ratings, political indices, consumer confidence, public school ratings, and other topics. For health, SRHS is seen as an ordinal measure of a latent (unobserved) health variable, and latent inequality (not ordinal inequality) is of primary interest.

Distinct from most of the literature, our framework's latent distributions are continuous but not parametrically specified. We wish to avoid unrealistically strong parametric assumptions while still allowing latent variation within each ordinal category; any teenager (or parent) can confirm that the ordinal category "good" corresponds to a wide range of latent feelings. Distinct from most (but not all) of the semi/nonparametric ordered choice literature, our objects of interest are features of the latent distributions themselves (rather than marginal effects of regressors). Further, we treat statistical inference seriously from both frequentist and Bayesian perspectives. In contrast, many well-regarded papers like Allison and Foster (2004) only propose an ordinal relationship (or index) that can be computed in the sample, without quantifying uncertainty, although recently Gunawan, Griffiths, and Chotikapanich (2018) have discussed Bayesian inference for some of these.

Recently, in the context of happiness, Bond and Lang (2018) adopt a similar framework to show the impossibility of comparing latent means from ordinal data (their Theorem 1). They point out that prior empirical studies' conclusions about mean happiness differences were highly sensitive to their parametric specification (e.g., ordered probit). Complementing their negative result, we provide positive results about what can be learned about latent relationships.

First, we consider identification: given knowledge of two ordinal population distributions, what can be learned about the corresponding latent distributions? Without a parametric model, the latent distributions are not point identified. Further, the individual latent distributions are not even partially identified (barring arbitrary normalizations). However, under relatively weak conditions, the ordinal distributions rule out certain pairs of latent distributions. The resulting identified set of latent distributions can sometimes be characterized in terms of inequality. For example, sometimes the first latent distribution is more dispersed than the second, for every pair in the identified set. That is, we can learn from ordinal data that one latent distribution has more within-group inequality than another.

Second, we consider statistical inference on the informative ordinal relationships found in our identification results. These relationships hold in unions and/or intersections of parameter sets defined by moment inequalities. For frequentist hypothesis testing, we show how to apply the intersection–union test and/or recent moment inequality tests. For Bayesian inference, given a posterior over the ordinal category probabilities, it is straightforward to compute posterior probabilities of the ordinal relationships using our characterizations. With iid sampling, for which the Dirichlet–multinomial model can be used, we discuss two types of uninformative prior: one uninformative over the underlying ordinal probabilities, and a further adjustment that puts 1/2 prior probability on the relationship of interest.

Third, we compare frequentist and Bayesian inferences. Even asymptotically, these can differ significantly if the Bayesian prior on the parameters is uninformative, according to Kaplan and Zhuo (2018). For example, Kaplan and Zhuo (2018) explain why a Bayesian test of the null hypothesis of ordinal first-order stochastic dominance is more likely to reject than a frequentist test, but the Bayesian test is less likely to reject a null of non-dominance. In simulations, we find such results continue to hold even if the prior is adjusted to have equal probability on the null and alternative hypotheses. Such an adjustment is argued to be "objective" by authors like Berger and Sellke (1987, p. 113), although this is disputed by authors like Casella and Berger (1987, p. 344).

Literature Surveying related methodological papers, Madden (2014) writes, "The breakthrough in analyzing inequality with [ordinal] data came from Allison and Foster (2004)" (p. 206), referring to their median-preserving spread. The median-preserving spread has been applied to health data (as in Madden, 2010) as well as happiness data (Dutta and Foster, 2013) and could be applied to any ordinal data, as with our methods in this paper. Allison and Foster (2004) treat the latent distribution as discrete; only the cardinal value of each ordinal category is unknown. Further, they assume the cardinal values are universal, not differing among different groups. Inspired by the connection between mean-preserving spreads and second-order stochastic dominance, they define (p. 512) one distribution to be a median-preserving spread of another distribution if 1) they share the same median category and 2) the first distribution can be constructed from the second by moving probability mass away from the median. They show (Theorems 3 and 4) that the median-preserving spread is equivalent to the first distribution being more dispersed (under certain cardinal measures of spread) given any possible assignment of cardinal values to the ordinal categories. They also show (Theorem 1) that first-order stochastic dominance is equivalent to having a higher mean for every possible cardinal value assignment. Although insightful, these results all rely critically on the latent distribution being discrete, an assumption that we relax. For our dispersion results, we also relax the assumption that all groups map their latent values to ordinal values the same way. Further, we fill a gap in their work by providing frequentist statistical inference on the median-preserving spread, as well as variations on the Bayesian inference proposed in Gunawan et al. (2018).

Another strand of the literature contains a variety of inequality indexes. These summarize the ordinal probabilities into a single number measuring the magnitude of inequality. This provides a definitive comparison between any two ordinal distributions (i.e., a complete ordering), although only after choosing a particular index and weights. Recent work in SRHS-based inequality indexes includes Abul Naga and Yalcin (2008), Reardon (2009), Silber and Yalonetzky (2011), Lazar and Silber (2013), Lv, Wang, and Xu (2015), and Yalonetzky (2016). Of these, only Lazar and Silber (2013) mention statistical inference, although Gunawan et al. (2018) propose Bayesian inference on the index from Abul Naga and Yalcin (2008) as well as first-order stochastic dominance, the median-preserving spread of Allison and Foster (2004), and other ordinal health relationships.

The ordered choice literature is also related to our work, especially Section 4. Naturally, our partial identification results require weaker assumptions than point identification results for ordered choice models, even nonparametric ones. For example, to achieve nonparametric point identification in a binary threshold crossing model, Matzkin (1992) assumes the latent error is independent of the covariates (p. 241), which implies the covariates affect only the location (not scale or shape) of the latent distribution, and the covariates are continuously distributed (p. 241). Semiparametric models like in Klein and Sherman (2002) further restrict how the covariates can affect the latent distribution's location. Our results allow any relationship between covariates and the latent distribution, and they allow continuous, discrete, or even categorical covariates. The only case where our assumptions are not strictly weaker is with ordered choice models allowing stochastic thresholds (e.g., Gu, Jiang, and Yang, 2018).

In different settings, others have considered identified sets for measures of dispersion given partially identified cumulative distribution functions (CDFs). Blundell, Gosling, Ichimura, and Meghir (2007, §5.3) examine whether the interquartile range of log wage has increased over time; the partial identification is from missing data.

Most closely related to our approach, Stoye (2010) uses CDF bounds to derive bounds for dispersion (spread) parameters. Interquantile ranges are a special case. Besides our results on between-group inequality, the other biggest difference with Stoye (2010) is that we compare two CDFs, neither of which is even partially identified since the thresholds mapping latent values to ordinal categories are unknown. Stoye's (2010) consideration of the "most compressed" and "most dispersed" CDFs within the identified set is similar in spirit to our approach to within-group inequality. The latent CDF bounds implied by ordinal data share the structure of his equation (3) (with zero mixture probability). Although before his (3) it says bounding CDFs must have finite expectation (which fails with ordinal data), it seems that the first part of Theorem 3(i) (with $p = 0$) should still apply as-is to interquantile ranges. Our results on interquantile ranges are derived more directly, using features specific to the case of ordinal data, which is not mentioned by Stoye (2010).

Paper structure and notation Section 2 connects features of pairs of latent and ordinal distributions. Section 3 describes and compares Bayesian and frequentist statistical inference. Section 4 comments on extensions to conditional distributions ("regression"). Section 5 and Section 6 provide simulation and empirical illustrations, respectively.

Acronyms used include those for cumulative distribution function (CDF), Current Population Survey (CPS), interquantile range (IQR), median-decreasing spread (MDS), median-preserving spread (MPS), Panel Study of Income Dynamics (PSID), probability mass function (PMF), refined moment selection (RMS), self-reported health status (SRHS), and stochastic dominance (SD), as well as first-order SD (SD1) and second-order SD (SD2). Notationally, $\subseteq$ is subset and $\subset$ is proper subset. Random and non-random vectors are respectively typeset as, e.g., $\mathbf{X}$ and $\mathbf{x}$, while random and non-random scalars are typeset as X and x, and random and non-random matrices as $\underline{\mathbf{X}}$ and $\underline{\mathbf{x}}$; $\mathbb{1}\{\cdot\}$ is the indicator function. The Dirichlet distribution with parameters $a_1, \ldots, a_K$ is written $\text{Dir}(a_1, \ldots, a_K)$, the beta distribution $\text{Beta}(a, b)$, and the uniform distribution $\text{Unif}(a, b)$; in some cases these stand for random variables following such distributions.

2 Identification of latent relationships

We first present assumptions, notation, and definitions, followed by various identification results for latent between-group and within-group inequality.

2.1 Assumptions and definitions

Among the following, Assumptions A1 and A2 are always maintained; other assumptions may be added for specific results.

Assumption A1 (latent distributions)**.** The latent random variables X^* and Y^* have continuous CDFs $F_X^*(\cdot)$ and $F_Y^*(\cdot)$, respectively, each with unbounded support, i.e., $0 < F_X^*(r) < 1$ and $0 < F_Y^*(r) < 1$ for all $r \in \mathbb{R}$.

Assumption A2 (ordinal distributions)**.** The observable, ordinal random variables X and Y are derived from X^* and Y^* as follows. The J ordinal categories are dentoed $1, 2, \ldots, J$, with no cardinal meaning. The thresholds for X are $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_J = \infty$.

Using these, $X = j$ iff $\gamma_{j-1} < X^* \le \gamma_j$, also written $X = \sum_{j=1}^{J} j\,\mathbb{1}\{\gamma_{j-1} < X^* \le \gamma_j\}$, so the ordinal CDF is $F_X(j) = F_X^*(\gamma_j)$. The thresholds for Y are $\gamma_j + \Delta_\gamma$, and similarly $Y = \sum_{j=1}^{J} j\,\mathbb{1}\{\gamma_{j-1} + \Delta_\gamma < Y^* \le \gamma_j + \Delta_\gamma\}$ and $F_Y(j) = F_Y^*(\gamma_j + \Delta_\gamma)$.

Assumption A3 (location–scale model)**.** There exists a continuous CDF $F^*(\cdot)$ such that $F_X^*(r) = F^*((r - \mu_X)/\sigma_X)$ and $F_Y^*(r) = F^*((r - \mu_Y)/\sigma_Y)$; i.e., $F_X^*(\cdot)$ and $F_Y^*(\cdot)$ belong to the same location–scale family.

Assumption A4 (latent symmetry)**.** The latent distributions are symmetric: denoting their medians m_X and m_Y, respectively, $F_X^*(m_X+\delta)+F_X^*(m_X-\delta) = 1$ and $F_Y^*(m_Y+\delta)+F_Y^*(m_Y-\delta) = 1$ for all $\delta \in \mathbb{R}$.

Assumption A5 (latent unimodality)**.** The latent distributions are unimodal: the PDFs $f_X^*(\cdot)$ and $f_Y^*(\cdot)$ each have a single local (and global) maximum.

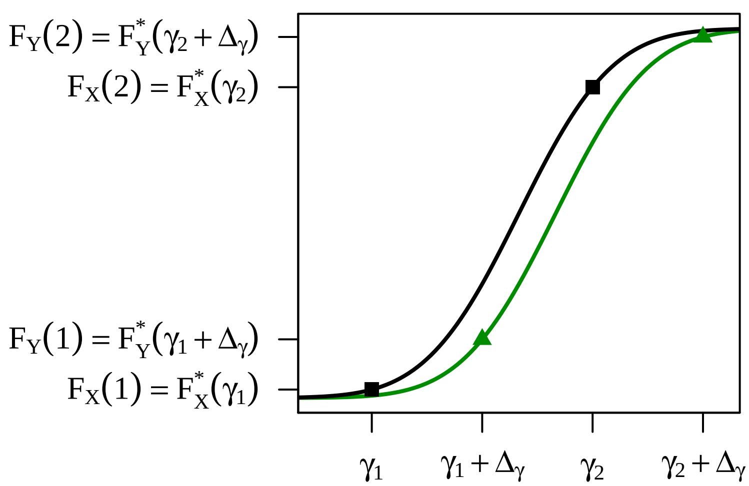

Figure 1: Example latent CDFs (lines) and ordinal CDFs (shapes).

Figures 1 and 2 illustrate the basic setup and implications of A1 and A2. Figure 1 shows two latent CDFs (lines) along with the observable ordinal CDF values (shapes). Figure 2 shows the latent CDF bounds implied by observed ordinal CDF values (points) given thresholds γ_j. If the γ_j were known somehow, then the latent CDF would be partially identified. The points $F_X^*(\gamma_j) = F_X(j)$ would be point identified, with the bounds coming from CDF monotonicity: for any $x \in (\gamma_j, \gamma_{j+1})$, $F_X^*(\gamma_j) \le F_X^*(x) \le F_X^*(\gamma_{j+1})$.

To learn anything, something must be assumed about the thresholds γ_j in A2. For certain results, we assume $\Delta_\gamma = 0$. This assumption is made (implicitly) in all the methodological papers cited in Section 1; e.g., Allison and Foster (2004) assume each category has the same cardinal value for both X and Y. Allowing $\Delta_\gamma \ne 0$ is called an "index shift" by Lindeboom

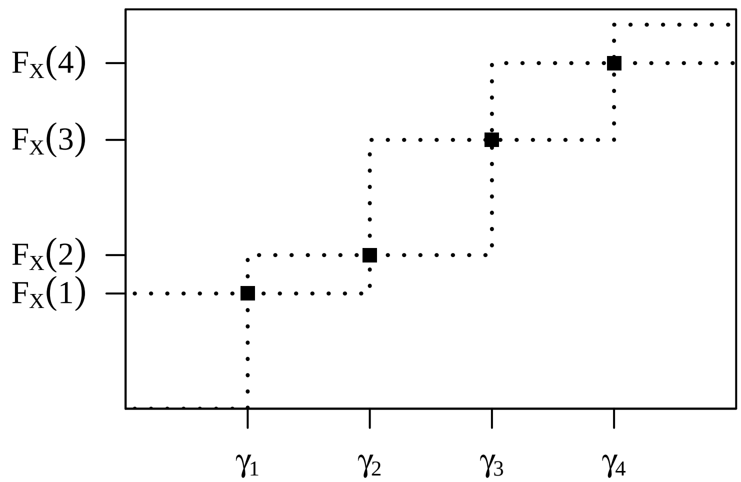

Figure 2: Latent CDF bounds given ordinal CDF and thresholds.

and van Doorslaer (2004), who contrast it with a "cut-point shift" that allows each γ_j to change arbitrarily (between X and Y).

Some empirical health papers have looked for evidence of threshold differences across groups. For example, in Canadian data, treating the McMaster Health Utility Index Mark 3 as latent health, Lindeboom and van Doorslaer (2004) find mixed evidence of different types of shifts across different types of groups (age, education, etc.); e.g., "For language, income and education, we find very few violations of the homogeneous reporting hypothesis, and in the few cases where it is violated, this appears almost invariably due to index rather than cut-point shift" (p. 1096). Other papers also report mixed evidence of no shift ($\Delta_\gamma = 0$), index shift ($\Delta_\gamma \neq 0$), and cut-point shift; e.g., Hernández-Quevedo, Jones, and Rice (2005) find evidence of an index shift (not cut-point shift) across waves of British survey data, and the magnitude appears similar across socioeconomic groups.

The assumption of unbounded support in A1 actually matters for some results based on the location–scale model definition in A3. For example, if the support is instead $[a, \infty)$ for $a \geq 0$, then increasing the scale parameter in the location–scale model (A3) may lead to latent SD1, which is not true if the latent distributions have unbounded support. For this reason (and others), we find unbounded support more reasonable.

Assumptions A3–A5 are additional semiparametric and nonparametric shape restrictions that can yield certain additional results below. We call A3 "semiparametric" since the location and scale parameters are sufficient to compare the latents distributions (in dispersion and first-order stochastic dominance), but the base distribution $F^*(\cdot)$ remains an unknown, infinite-dimensional nuisance parameter. Location–scale models have been widely used in economics and statistics, e.g., as motivation for mean–variance analysis of asset portfolios and by Bond and Lang (2018) for their ordinal happiness impossibility theorem. As in the latter (but not former) example, we do not require finite variance or well-defined mean. Assumptions A4 and A5 are primarily for results on dispersion.

We also use the following definitions.

Definition 1 (latent SD1)**.** Latent X^* first-order stochastically dominates Y^* iff $F_X^*(r) \leq F_Y^*(r)$ for all $r \in \mathbb{R}$; this is written $X^* \text{ SD}_1 \, Y^*$ or $F_X^* \text{ SD}_1 \, F_Y^*$.

Definition 2 (ordinal SD1)**.** Ordinal X first-order stochastically dominates Y iff $F_X(j) \leq F_Y(j)$ for $j = 1, \ldots, J$; this is written $X \text{ SD}_1 \, Y$ or $F_X \text{ SD}_1 \, F_Y$.

Definition 3 (restricted latent SD1)**.** There is restricted first-order stochastic dominance of X^* over Y^* iff $F_X^*(r) \leq F_Y^*(r)$ for all r in some interval $[r^-, r^+]$.

Definition 4 (latent SD2)**.** Latent X^* second-order stochastically dominates Y^* iff for all $u \in \mathbb{R}$, $\int_{-\infty}^{u} [F_X^*(r) - F_Y^*(r)] \, dr \leq 0$; this is written $X^* \text{ SD}_2 \, Y^*$ or $F_X^* \text{ SD}_2 \, F_Y^*$.

Definition 5 (single crossing)**.** CDFs $F_A(\cdot)$ and $F_B(\cdot)$ have a "single crossing" iff there exists a "crossing point" $m \in \mathbb{R}$ such that $F_A(r) < F_B(r)$ when $r < m$ and $F_A(r) > F_B(r)$ when $r > m$, or vice-versa (switching F_A and F_B). The first possibility is written $F_A \text{ SC } F_B$ or $A \text{ SC } B$; the second (switching F_A and F_B) is written $F_B \text{ SC } F_A$ or $B \text{ SC } A$.

Definition 6 (pure location shift)**.** The distribution of X^* is a pure location shift of the distribution of Y^* (and vice-versa) iff there exists $\delta \in \mathbb{R}$ such that $F_X^*(r) = F_Y^*(r - \delta)$ for all $r \in \mathbb{R}$.

Definition 7 ($\tau_2 - \tau_1$ IQR)**.** For a distribution with quantile function $Q(\cdot)$, the $\tau_2 - \tau_1$ interquantile range (IQR) is $Q(\tau_2) - Q(\tau_1)$.

The SD1 definitions are technically for "weak" instead of "strong" (replacing $\leq$ with $<$) dominance; Davidson and Duclos (2013, p. 89) note that the distinction is statistically indistinguishable in practice.

The single crossing property (Definition 5) has also been useful in areas of economic theory like mechanism design. Here, we presume m is a single point for simplicity, but it could be replaced by an interval. However, the strict inequalities cannot be replaced by weak inequalities (as it is sometimes defined) without changing some of our results.

The definition of "restricted SD1" is from Condition I of Atkinson (1987, p. 751) and is discussed in Davidson and Duclos (2000, 2013). Although not as strong as (unrestricted) SD1, it is still meaningful, especially if the interval $[r^-, r^+]$ is large. In our results, $r^- = \gamma_1$ and $r^+ = \gamma_{J-1}$ are the $F_X(1)$-quantile and $F_X(J-1)$-quantile of X^*, respectively.

2.2 Between-group inequality

First-order stochastic dominance (SD1) has long been valued for unambiguously ranking two distributions. For example, with income, an individual at some percentile of the dominant distribution has a higher income than the individual at the same percentile of the dominated distribution. Further, for any increasing utility function (of income), expected utility is higher for the first distribution. (That is, under a Rawlsian veil of ignorance, everybody would prefer ex ante to be born as a random person in the group/society with the dominant income distribution.) SD1 is similarly valuable for ranking latent health distributions, assuming it is latent (not ordinal) health that enters the utility function.

We discuss some SD1 intuition before stating formal results. It is clear that latent SD1 implies ordinal SD1 if X and Y have the same thresholds, i.e., $\Delta_\gamma = 0$. That is, ordinal SD1 is a testable implication of latent SD1. However, unlike in the discrete latent setup of Allison and Foster (2004), the converse is not true: ordinal SD1 does not imply latent SD1. With $\Delta_\gamma = 0$, ordinal SD1 only restricts $F_X^*(\gamma_j) \leq F_Y^*(\gamma_j)$ for $j = 1, \ldots, J-1$; latent SD1 could easily be violated in the lower tail, for example. To infer latent SD1 from ordinal SD1, we would need to impose enough structure that we could interpolate and extrapolate from only $J-1$ points on each CDF. Even Assumptions A3–A5 do not collectively provide enough structure to extrapolate, unless $\sigma_X = \sigma_Y$ was also assumed.

Ordinal SD1 can imply latent *restricted* SD1 (Definition 3) in two cases. The first case is when one ordinal CDF is far enough below the other that the CDF bounds like in Figure 2 do not overlap. The second case is with a location–scale model (A3). The location–scale model provides enough structure to interpolate the SD1 relationship between the γ_j where we know $F_X^*(\gamma_j) \leq F_Y^*(\gamma_j)$, but not enough structure to extrapolate SD1 into the tails.

Without $\Delta_\gamma = 0$, none of these results hold. A threshold shift $\Delta_\gamma \neq 0$ is observationally equivalent to a location shift of the latent distribution of Y^*.

The above results are summarized in Theorem 2. Before that, we state Lemma 1, which helps prove parts of Theorems 2 and 3. Lemma 1 is essentially Proposition 1 from Bond and Lang (2018). It also implies that A3 is violated if there are multiple ordinal CDF crossings.

Lemma 1. *Let Assumptions A1 and A3 hold. Assume $F^*(\cdot)$ is strictly increasing.*[1] *If $\sigma_X \neq \sigma_Y$, then the latent CDFs have a single crossing (Definition 5) with crossing point $m = (\sigma_Y \mu_X - \sigma_X \mu_Y)/(\sigma_Y - \sigma_X)$. If $\sigma_X = \sigma_Y$, then there is no crossing point (i.e., SD1 holds).*

Theorem 2. *Let Assumptions A1 and A2 hold.*

[1] This can be relaxed by replacing m in Definition 5 with an interval; the intuition is the same.

(i) *If* $\Delta_\gamma = 0$*, then* $X^* \text{ SD}_1 Y^* \implies X \text{ SD}_1 Y$*.*

(ii) *If* $\Delta_\gamma \neq 0$*, then* $X^* \text{ SD}_1 Y^* \not\Longrightarrow X \text{ SD}_1 Y$*.*

(iii) *If* $\Delta_\gamma = 0$ *and Assumption A3 holds with* $\sigma_X = \sigma_Y$*, then* $X^* \text{ SD}_1 Y^* \iff X \text{ SD}_1 Y$*.*

(iv) *Even if* $\Delta_\gamma = 0$ *and Assumptions A3–A5 hold,* $X \text{ SD}_1 Y \not\Longrightarrow X^* \text{ SD}_1 Y^*$*.*

(v) *If* $\Delta_\gamma = 0$*, then* $F_X(j+1) \leq F_Y(j)$ *for all* $j = 1, \ldots, J-2 \implies$ *restricted SD1 of* X^* *over* Y^* *on the interval* $[\gamma_1, \gamma_{J-1}]$*.*

(vi) *If* $\Delta_\gamma = 0$ *and A3 holds, then* $X \text{ SD}_1 Y \implies$ *restricted SD1 of* X^* *over* Y^* *on* $[\gamma_1, \gamma_{J-1}]$*.*

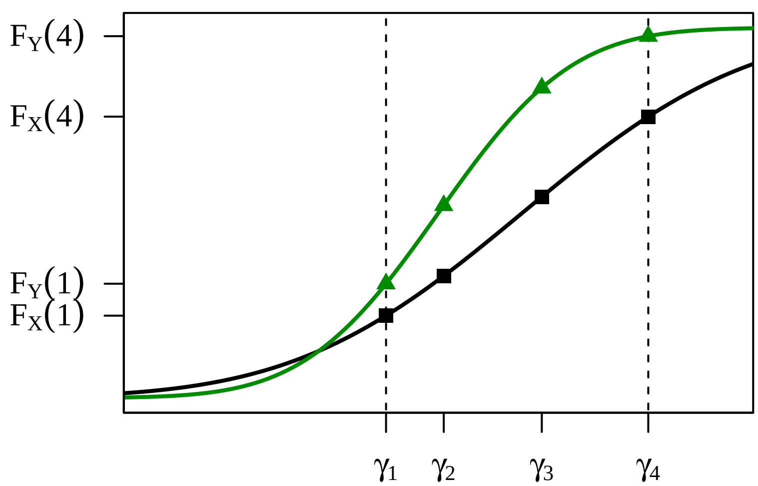

Figure 3: Illustration of Theorem 2(vi).

Figure 3 illustrates some results from Theorem 2. The latent CDFs satisfy the location–scale assumption, so at most one latent CDF crossing is possible. In Figure 3, the crossing is below γ_1, so there is restricted SD1 on $[\gamma_1, \gamma_4]$ (actually an even larger interval), and there is ordinal SD1. This helps illustrate the logic behind Theorem 2(vi). Theorem 2(iv) is also illustrated: even though there is ordinal SD1 and the latent distributions satisfy A3–A5, there is not latent SD1 since the latent CDFs cross in the lower tail. Some insight into Theorem 2(i) and Theorem 2(iii) may be gleaned by imagining Figure 3 modified so the latent CDFs did not cross.

2.3 Within-group inequality

There are multiple reasons to be interested in dispersion of the latent health distribution. First, evidence of dispersion increasing with age supports the existence of permanent "health shocks"; this is one focus of Deaton and Paxson (1998a,b), for example. Second, dispersion reflects inequality within a (sub)population. Third, dispersion is a component of welfare given a concave utility function, i.e., it is related to second-order stochastic dominance. Naturally, latent second-order stochastic dominance is more difficult to infer from ordinal distributions, requiring $\Delta_\gamma = 0$ (among other assumptions), whereas dispersion results allow $\Delta_\gamma \neq 0$.

In the location–scale model of A3, Lemma 1 says a latent CDF crossing implies unequal scales ($\sigma_X \neq \sigma_Y$). With $\Delta_\gamma = 0$, an ordinal CDF crossing implies a latent CDF crossing (though not vice-versa). In fact, an ordinal CDF crossing contains information about latent dispersion even without A3 or $\Delta_\gamma = 0$.

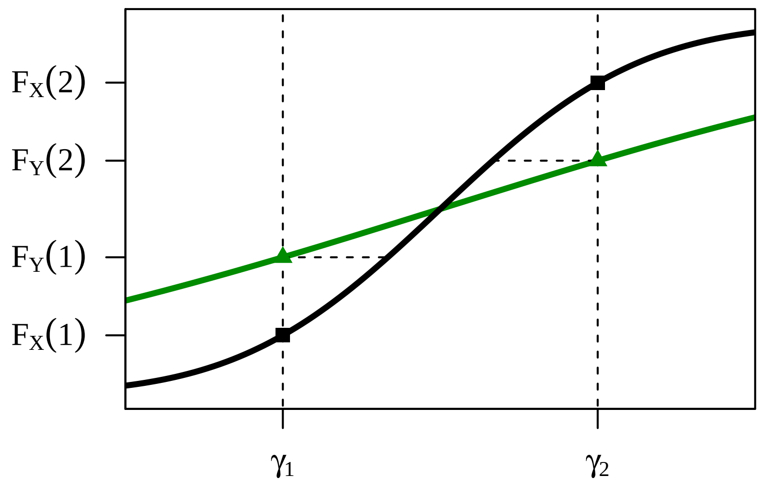

Figure 4: Illustration of Theorem 3(i).

Most fundamentally, an ordinal CDF crossing implies certain interquantile ranges (IQRs) are smaller in one latent distribution. Figure 4 illustrates the following arguments. Heuristically, imagine a crossing between $j = 1$ and $j = 2$, so $F_X(1) < F_Y(1)$ and $F_X(2) > F_Y(2)$; then, $F_X^*(\gamma_1) < F_Y^*(\gamma_1 + \Delta_\gamma)$ and $F_X^*(\gamma_2) > F_Y^*(\gamma_2 + \Delta_\gamma)$. Note that $(\gamma_2 + \Delta_\gamma) - (\gamma_1 + \Delta_\gamma) = \gamma_2 - \gamma_1$, so it is irrelevant whether or not $\Delta_\gamma = 0$. Then, $\gamma_2 - \gamma_1$ is the $F_X(2) - F_X(1)$ IQR for X^* but the $F_Y(2) - F_Y(1)$ IQR for Y^*. Since $F_X(1) < F_Y(1)$ and $F_X(2) > F_Y(2)$, the $F_X(2) - F_X(1)$ IQR of Y^* is larger than the $F_Y(2) - F_Y(1)$ IQR of Y^*, so Y^* has a larger $F_X(2) - F_X(1)$ IQR than X^*. Similar logic determines other IQRs that must be larger for Y^*, as enumerated in Theorem 3. These results can also be seen in light of Theorem 3(i) from Stoye (2010), considering the latent CDF bounds implied by the ordinal data: even the most "compressed" possible $F_Y^*(\cdot)$ consistent with $F_Y(\cdot)$ still has a larger IQR than the most "dispersed" possible $F_X^*(\cdot)$.

Further imposing the location–scale model yields stronger results. In the location–scale model, the location parameter μ has no affect on IQRs, and the scale paramter σ affects all IQRs simultaneously. Imposing this model allows even a single larger IQR to imply a larger σ, which in turn implies all IQRs are larger; it provides enough structure to extrapolate from a set of IQRs to all IQRs. Note that while an ordinal CDF crossing implies a different σ, the lack of a crossing is ambiguous since the latent CDFs could cross in the tails, undetected by the ordinal distributions. Also, following from Lemma 1, multiple ordinal crossings imply that the location–scale model is misspecified.

Yet further imposing symmetry and $\Delta_\gamma = 0$ (fixed thresholds) identifies latent second-order stochastic dominance (SD2) in some cases. The location–scale model combined with symmetry simplifies SD2 to conditions on μ and σ. If the median category decreases to imply $\mu_Y < \mu_X$, and the CDFs cross to imply $\sigma_Y > \sigma_X$, then $X^* \text{ SD}_2 \, Y^*$.

The above results are formalized in Theorem 3.

Theorem 3. *Let Assumptions A1 and A2 hold. Assume $F_X^*(\cdot)$ and $F_Y^*(\cdot)$ are strictly increasing. Assume there is a single crossing of the ordinal CDFs at category m, i.e., $F_X(j) < F_Y(j)$ for $j \le m$ and $F_X(j) > F_Y(j)$ for $j > m$. Let $Q_X^*(\cdot)$ and $Q_Y^*(\cdot)$ denote the quantile functions of X^* and Y^*, respectively.*

(i) $Q_X^*(\tau_2) - Q_X^*(\tau_1) < Q_Y^*(\tau_2) - Q_Y^*(\tau_1)$ *for any combination of $\tau_2 \in \mathcal{T}_2$ and $\tau_1 \in \mathcal{T}_1$, where*

$$\mathcal{T}_1 \equiv \bigcup_{j=1}^{m} [F_X(j), F_Y(j)], \quad \mathcal{T}_2 \equiv \bigcup_{j=m+1}^{J-1} [F_Y(j), F_X(j)]. \tag{1}$$

(ii) *If Assumption A3 also holds, then $\sigma_X < \sigma_Y$ and $Q_X^*(\tau_2) - Q_X^*(\tau_1) < Q_Y^*(\tau_2) - Q_Y^*(\tau_1)$ for all $0 < \tau_1 < \tau_2 < 1$.*

(iii) *Let $\Delta_\gamma = 0$. Let Assumptions A3 and A4 hold, with $F^*(0) = 1/2$. If the median category of X is strictly above that of Y, then $X^* \text{ SD}_2 \, Y^*$. If there is the single crossing but instead X and Y have the same median, then it is ambiguous whether $X^* \text{ SD}_2 \, Y^*$ or not.*

Theorem 3 allows interpretation of the median-preserving spread (MPS) when the latent distributions are continuous, instead of discrete like in Allison and Foster (2004). Assuming the ordinal median is an interior category (not $j = 1$ or $j = J$), as seems implicit in Allison and Foster (2004), the MPS is a special case of a single crossing of ordinal CDFs. Theorem 3 says this is indeed evidence of different dispersion. However, Theorem 3(iii) says that MPS is not evidence of latent SD2, but that a median-decreasing spread can be (under certain assumptions). With a continuous latent distribution, there is too much ambiguity with equal ordinal medians; either latent median could be larger.

If the CDFs do not cross as in Theorem 3, then dispersion differences can be identified by a shape restriction: unimodal, symmetric latent distributions. Symmetry implies the mode is at the median, which we can approximately locate from the ordinal distribution. Unimodality implies the latent CDFs are concave above the median. Thus, if we observe a "fanning out" of CDFs above the median, then there is a certain interquantile range (IQR) that must be larger for the lower CDF.

Figure 5 illustrates the logic. For the upper distribution, the τ_1-quantile is γ_1. The figure shows the largest possible τ_2-quantile given the restriction of concavity on the CDF; the $\tau_2 - \tau_1$ IQR is maximized by the latent CDF that linearly connects $F_X^*(\gamma_1) = F_X(1) = \tau_1$ and $F_X^*(\gamma_2) = F_X(2)$. For the lower CDF, the τ_2-quantile is γ_2, and the largest possible τ_1-quantile (and thus smallest IQR) comes from the latent CDF that linearly connects the two points. In the figure, this smallest possible IQR of the lower distribution is still larger than the largest possible IQR of the upper distribution. This provides some evidence of (part of) the lower CDF being more dispersed.

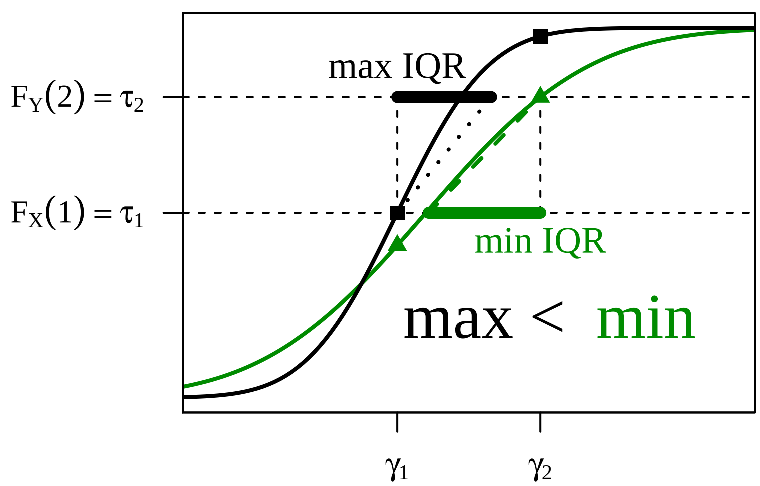

Figure 5: Illustration of Theorem 4(i).

Theorem 4. *Let Assumptions A1, A2, A4, and A5 hold.*

(i) *Let j denote any category for which $F_X(j) \geq F_Y(j) \geq 1/2$ and $F_X(j+1) \geq F_Y(j+1)$. If $F_Y(j+1) - F_Y(j) < F_X(j+1) - F_X(j)$, then the IQR between the $F_Y(j+1)$-quantile and $F_X(j)$-quantile of the Y^* distribution is larger than that of the X^* distribution. This holds whether $F_X(j) > F_Y(j+1)$ or $F_X(j) < F_Y(j+1)$.*

(ii) *Let $j+1$ denote any category for which $F_X(j+1) \leq F_Y(j+1) \leq 1/2$ and $F_X(j) \leq F_Y(j)$. If $F_Y(j+1) - F_Y(j) < F_X(j+1) - F_X(j)$, then the IQR between the $F_Y(j)$-quantile* *and $F_X(j+1)$-quantile of the Y^* distribution is larger than that of the X^* distribution. This holds whether $F_X(j+1) > F_Y(j)$ or $F_X(j+1) < F_Y(j)$.*

(iii) *If additionally Assumption A3 holds, with $F^*(0) = 1/2$, then the conclusions of Theorem 4(i) and Theorem 4(ii) strengthen to $\sigma_Y > \sigma_X$, implying the $\tau_2 - \tau_1$ IQR of Y^* is larger than that of X^*, for all $0 < \tau_1 < \tau_2 < 1$.*

3 Statistical inference on ordinal relationships

This section concerns statistical inference on the ordinal relationships in Section 2.[2] We describe how to statistically assess our (un)certainty of a certain relationship between two ordinal population distributions, given a sample of data from each. We characterize the relationships (hypotheses) of interest and then discuss both frequentist and Bayesian procedures, which are then compared.

3.1 Relationships of interest

Ordinal CDF and PMF values may be expressed as moments. For example, $F_X(j) = \mathrm{E}(\mathbb{1}\{X \leq j\})$ and $p_j^X = \mathrm{E}(\mathbb{1}\{X = j\})$. Consequently, all ordinal relationships from Section 2 can be written in terms of moment inequalities, with unions and/or intersections taken in some cases. These are now characterized. Since $F_X(J) = F_Y(J) = 1$ always, only $j = 1, \ldots, J-1$ are used.

The relationship $X \text{ SD}_1 Y$ is a set of moment inequalities:

$$X \text{ SD}_1 Y \iff \bigcap_{j=1}^{J-1} \{F_X(j) \leq F_Y(j)\}. \tag{2}$$

If the medians of X and Y are known, then Y MPS X (i.e., Y is a median-preserving spread of X) is also a set of moment inequalities. With m the shared median category,

$$Y \text{ MPS } X \iff \bigcap_{j=1}^{J-1} \Big\{[2\,\mathbb{1}\{j < m\} - 1][F_X(j) - F_Y(j)] \leq 0\Big\}. \tag{3}$$

Treating the median is known could be reasonable in large samples where, e.g., $F_X(3) = F_Y(3) = 0.3$, $F_X(4) = F_Y(4) = 0.7$, and $F_X(j)$ is close to $F_Y(j)$ for other j, so clearly $m = 4$ but MPS is not certain. If instead the medians are unknown, then MPS is a union of events

[2] Initial work on this section is found in Chapter 2 of the second author's dissertation (Zhuo, 2017).

over possible median values, checking (3) for each possible median:

$$Y \text{ MPS } X \iff \bigcup_{m=2}^{J-1} \Big\{ \{F_X(m-1) < 1/2 \le F_X(m)\} \cap \{F_Y(m-1) < 1/2 \le F_Y(m)\} \\ \cap \bigcap_{j=1}^{J-1} \Big\{ [2\,\mathbb{1}\{j < m\} - 1][F_X(j) - F_Y(j)] \le 0 \Big\} \Big\}. \tag{4}$$

Note that MPS with $m = 1$ or $m = J$ is equivalent to SD1, so only $m = 2, \ldots, J-1$ are included in (4).

A single crossing is similar to (4), except that the crossing point need not be the median, and the crossing point k must be between 2 and $J-1$. Using the definition and notation in Definition 5,

$$X \text{ SC } Y \iff \bigcup_{k=2}^{J-1} \bigcap_{j=1}^{J-1} \{[2\,\mathbb{1}\{j < k\} - 1][F_X(j) - F_Y(j)] < 0\}. \tag{5}$$

In practice, we cannot distinguish statistically between $<$ and $\le$.

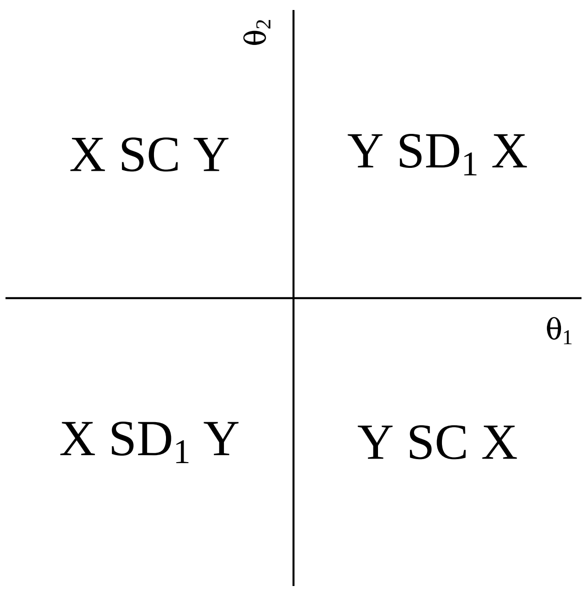

Figure 6: Mapping from (θ_1, θ_2) to ordinal relationships, $J = 3$.

Figure 6 shows which ordinal relationships are satisfied over different regions of (θ_1, θ_2) when $J = 3$. This is simpler than for SRHS, which has $J = 5$. In the upper-right quadrant, $\theta_1 > 0$ and $\theta_2 > 0$, meaning $F_X(1) > F_Y(1)$ and $F_X(2) > F_Y(2)$, so Y SD_1 X. Conversely, in the lower-left quadrant, $\theta_1 < 0$ and $\theta_2 < 0$, meaning $F_X(1) < F_Y(1)$ and $F_X(2) < F_Y(2)$, so X SD_1 Y. In the upper-left quadrant, $\theta_1 < 0$ and $\theta_2 > 0$, meaning $F_X(1) < F_Y(1)$ and $F_X(2) > F_Y(2)$, so X SC Y. Similarly, in the lower-right quadrant, Y SC X. With larger

$J > 3$, the SD1 regions look similar, but the SC regions are more complex since there are multiple possible crossing locations (unlike here), and there are also regions where neither SD1 nor SC holds.

A median-decreasing spread (MDS), i.e., the combination of a single crossing and strictly smaller median, is

$$\begin{aligned} Y \text{ MDS } X \iff & \left\{ \bigcup_{k=2}^{J-1} \bigcap_{j=1}^{J-1} \{ [2\mathbb{1}\{j < k\} - 1][F_X(j) - F_Y(j)] < 0 \} \right\} \\ & \cap \left\{ \bigcup_{m=1}^{J-1} \{ F_X(m) < 1/2 \le F_Y(m) \} \right\}. \end{aligned} \tag{6}$$

The existence of multiple crossings is the lack of SD1 (i.e., zero crossings) or a single crossing:

$$\text{multiple crossings} \iff \text{not } \{X \text{ SD}_1\ Y \text{ or } Y \text{ SD}_1\ X \text{ or } X \text{ SC } Y \text{ or } Y \text{ SC } X\}. \tag{7}$$

Technically, the SC in (7) should be slightly weaker than in (5), allowing for weak inequalities (≤ 0), but the difference is again statistically indistinguishable in practice. The expression in (7) is the complement of a union of (a union of) an intersection of moment inequalities. There are alternative, equivalent characterizations, but they are no less complicated in form.

The presence of at least one "fanning out" of ordinal CDFs suggesting that Y has larger dispersion than X can be written as

$$Y \text{ FanOut } X \iff \bigcup_{j=2}^{J-1} \{A_j \cup B_j\}, \tag{8}$$

$$\begin{aligned} A_j \equiv & \{F_X(j-1) < F_Y(j-1)\} \cap \{F_X(j) \le 1/2\} \cap \{F_Y(j) \le 1/2\} \\ & \cap \{F_Y(j) - F_Y(j-1) < F_X(j) - F_X(j-1)\}, \\ B_j \equiv & \{F_X(j) > F_Y(j)\} \cap \{F_X(j-1) \ge 1/2\} \cap \{F_Y(j-1) \ge 1/2\} \\ & \cap \{F_Y(j) - F_Y(j-1) < F_X(j) - F_X(j-1)\}. \end{aligned}$$

The above characterization includes CDF crossing as a special case of fanning out.

3.2 Bayesian inference

With iid sampling, we show how to use the Dirichlet–multinomial model for Bayesian inference on all the ordinal relationships. We then discuss a possible adjustment to the prior. More generally, posterior probabilities of all relationships are easily computed from any posterior over the ordinal category probabilities, given the characterizations in Section 3.1. For example, one of our empirical examples has a complex sampling design. Code is provided on the first author's website.

Some of our results with the unadjusted prior are essentially the same as those in Gunawan et al. (2018). They also consider SD1 and MPS with an uninformative prior, albeit the improper rather than uniform prior; see their (7). They do not consider our adjusted prior, or our other relationships (SC, MDS, fanning out).

We consider iid sampling of two independent samples from two respective ordinal population distributions. No latent structure is imposed. Letting n_X and n_Y denote sample sizes, and continuing notation from earlier,

$$X_i \overset{iid}{\sim} F_X,\ i = 1, \ldots, n_X, \quad Y_i \overset{iid}{\sim} F_Y,\ i = 1, \ldots, n_Y. \tag{9}$$

Alternatively, sampling may be considered in terms of the category probabilities and observed counts. Let

$$\begin{aligned} \mathbf{p}^X &\equiv (p_1^X, \ldots, p_J^X), & \mathbf{p}^Y &\equiv (p_1^Y, \ldots, p_J^Y), \\ p_j^X &\equiv \mathrm{P}(X = j), & p_j^Y &\equiv \mathrm{P}(Y = j). \end{aligned}$$

That is, the vectors $\mathbf{p}^X$ and $\mathbf{p}^Y$ fully determine the probability mass functions (PMFs) of X and Y, respectively. The data samples can be summarized by counts of observations in each category, which follow a multinomial distribution:

$$\begin{aligned} \mathbf{N}^X &\equiv (N_1^X, \ldots, N_J^X), & \mathbf{N}^Y &\equiv (N_1^Y, \ldots, N_J^Y), \\ N_j^X &\equiv \sum_{i=1}^{n_X} \mathbb{1}\{X_i = j\}, & N_j^Y &\equiv \sum_{i=1}^{n_Y} \mathbb{1}\{Y_i = j\}, \\ \mathbf{N}^X &\sim \mathrm{Multinomial}(n_X, \mathbf{p}^X), & \mathbf{N}^Y &\sim \mathrm{Multinomial}(n_Y, \mathbf{p}^Y). \end{aligned}$$

We first consider the posteriors for the two ordinal distributions themselves. Here, the multinomial sampling distribution of category counts is convenient. The Dirichlet distribution is the conjugate prior for the multinomial, i.e., a Dirichlet prior results in a Dirichlet posterior. Informative and uninformative priors are both easy to use. We use the uniform (i.e., constant PDF) priors

$$\mathbf{p}^X \sim \mathrm{Dir}(\mathbf{1}), \quad \mathbf{p}^Y \sim \mathrm{Dir}(\mathbf{1}), \tag{10}$$

where $\mathbf{1} \equiv (1, 1, \ldots, 1)$ is a vector of J ones. If $J = 2$, then the priors simplify to $p_1^X \sim \mathrm{Unif}(0, 1)$ and $p_1^Y \sim \mathrm{Unif}(0, 1)$, with $p_2^X = 1 - p_1^X$ and $p_2^Y = 1 - p_1^Y$. More generally, $\mathbf{p}^X$ and $\mathbf{p}^Y$ are distributed uniformly over the unit simplex in $\mathbb{R}^J$, i.e., vectors whose J non-negative components sum to one.

By conjugacy, the posteriors are

$$\mathbf{p}^X \mid \mathbf{N}^X \sim \mathrm{Dir}(\mathbf{1}+\mathbf{N}^X), \quad \mathbf{p}^Y \mid \mathbf{N}^Y \sim \mathrm{Dir}(\mathbf{1}+\mathbf{N}^Y), \tag{11}$$

which are easy to sample from (without MCMC).

To compute the posterior probability of a particular ordinal relationship, one may take many draws from the Dirichlet posteriors in (11) and check whether or not the relationship holds in each draw. For example, given a draw of $\mathbf{p}^X$ and $\mathbf{p}^Y$,

$$X \ \mathrm{SD}_1 \ Y \iff \forall k = 1, \ldots, J-1, \ \sum_{j=1}^{k} p_j^X \leq \sum_{j=1}^{k} p_j^Y .$$

The proportion of total draws in which $X \ \mathrm{SD}_1 \ Y$ is the (approximate) posterior probability of $X \ \mathrm{SD}_1 \ Y$.

In some cases, a different type of "objective" prior may be desired. In (10), the prior is objective in the sense that it is uniform over all possible PMFs, which arguably reflects no prior information. However, if there is reason to suspect a certain relationship, then it may be desired to set the prior of the relationship to 1/2: $\mathrm{P}(H_0) = 1/2$, where H_0 is the relationship, like $H_0\colon X \ \mathrm{SD}_1 \ Y$.

To achieve this, the prior from (10) may be adjusted using an idea from Goutis, Casella, and Wells (1996). In their equation (7), they first compute $\gamma = \mathrm{P}^{\mathrm{orig}}(H_0)$, the original prior probability that H_0 is true. In their (8), they show the adjusted prior, computed by multiplying the original prior by the constant $\gamma/(1-\gamma)$ wherever H_0 is false, and then renormalizing (so that it integrates to one); this achieves 1/2 probability of H_0 under the adjusted prior, $\mathrm{P}^{\mathrm{adj}}(H_0) = 1/2$. In their (9), they show that the corresponding adjusted posterior may be computed from the unadjusted posterior as

$$\mathrm{P}^{\mathrm{adj}}(H_0 \mid \text{data}) = \frac{\mathrm{P}^{\mathrm{orig}}(H_0 \mid \text{data})}{\mathrm{P}^{\mathrm{orig}}(H_0 \mid \text{data}) + \frac{\gamma}{1-\gamma}[1 - \mathrm{P}^{\mathrm{orig}}(H_0 \mid \text{data})]}. \tag{12}$$

That is, we may compute the posterior probability using (11) and then adjust it using (12), once we know γ.

The value of γ can be pre-computed. Since γ depends only on the type of relationship (H_0) and the number of categories (J), we have simulated these for $J = 2, \ldots, 30$ categories and all relationships discussed earlier, using 10,000 draws each, and stored these values in our code. (If $J > 30$, additional values may be simulated in under a minute.) The computation of relationship probabilities is identical to before, except that draws are taken from the prior in (10) instead of the posterior in (11).

In sum, when an objective prior is desired, posterior probabilities of the ordinal distribution relationships in Section 2 may be computed as follows. These steps are implemented in our code.

1. Given the observed vectors of category counts, take R draws of $\mathbf{p}^X$ and $\mathbf{p}^Y$ from the Dirichlet distribution posteriors in (11).

2. In each of the R draws, check whether each relationship holds, using the characterizations in Section 3.1.

3. For each relationship, compute the proportion of the R draws in which that relationship held; this is the (approximate) posterior probability of that relationship.

4. If it is desired that the prior probability of a relationship is 1/2, then adjust the posterior using (12). The value of γ is computed by the above steps: it is the posterior probability of the relationship when $\mathbf{N}^X = \mathbf{N}^Y = \mathbf{0}$.

Usually, the posterior probabilities of different relationships are themselves the most informative summary of the data, but sometimes a more concrete decision must be made. In such a case, first a loss function must be determined, quantifying how "bad" a decision is given a true state of the world. Then, the Bayes decision rule chooses the decision that minimizes posterior expected loss.

One loss function for hypothesis testing takes value $1-\alpha$ for type I error, α for type II error, and zero otherwise. This is often called "generalized 0–1 loss" (e.g., Casella and Berger, 2002, eqn. (8.3.11)). This loss function is arguably implicit in frequentist hypothesis testing; e.g., Kaplan and Zhuo (2018, §2.1) describe connections like how an unbiased frequentist test with size α is the minimax risk decision rule under this loss function.

This generalized 0–1 loss function leads to a Bayesian hypothesis test that "rejects" a null hypothesis if and only if its posterior probability is below α. See Kaplan and Zhuo (2018, §2.1) for details. For example, if the posterior probability of $X \text{ SD}_1\ Y$ is 0.03 and $\alpha = 0.05$, then $H_0\colon X \text{ SD}_1\ Y$ is rejected. Computationally, this treats the posterior for a given H_0 like a p-value, but it does not imply that the posterior actually is a valid p-value; see Section 3.4.

3.3 Frequentist hypothesis testing

Frequentist hypothesis tests of first-order (and higher-order) stochastic dominance have been discussed more often for continuous random variables, as in Davidson and Duclos (2000) and Barrett and Donald (2003), for example. However, this is primarily because the discrete (or ordinal) case is much simpler, so there is less to be said.

In the frequentist framework, there is an important difference between testing a null hypothesis of dominance and testing a null of non-dominance. In particular, rejection of non-dominance is much stronger evidence of dominance than non-rejection of dominance. This distinction, which applies equally to continuous and discrete/ordinal variables, is discussed at length in Davidson and Duclos (2013). Earlier, Kaur, Prakasa Rao, and Singh (1994) also considered testing a null of non-SD2.

The following subsections concern testing different types of ordinal hypotheses, given the iid sampling setup from Section 3.2.

3.3.1 Null hypothesis: MPS (known medians) or SD1

Let

$$\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_{J-1}), \quad \theta_j \equiv F_X(j) - F_Y(j) = \mathrm{E}[\mathbb{1}\{X \le j\}] - \mathrm{E}[\mathbb{1}\{Y \le j\}]. \tag{13}$$

The null hypothesis $H_0\colon X \;\mathrm{SD}_1\; Y$ is equivalent to

$$H_0\colon \boldsymbol{\theta} \le \mathbf{0}, \tag{14}$$

i.e., $\theta_j \le 0$ for all $j = 1, \ldots, J-1$. If m is the known median of both X and Y, then the null hypothesis $H_0\colon Y$ MPS X is equivalent to

$$H_0\colon \underline{\mathbf{D}}\boldsymbol{\theta} \le \mathbf{0}, \tag{15}$$

where $\underline{\mathbf{D}}$ is a diagonal matrix ($D_{jk} = 0$ if $j \ne k$) with elements $D_{jj} = 1$ for $j = 1, \ldots, m-1$ and $D_{jj} = -1$ for $j = m, \ldots, J-1$.

The easiest way to test (14) or (15) is with a Bonferroni correction, but it will be conservative (i.e., size strictly below α). Using the Bonferroni approach, each individual hypothesis $H_{0j}\colon \theta_j \le 0$ or $H_{0j}\colon D_{jj}\theta_j \le 0$ is tested at an $\alpha/(J-1)$ significance level, where $J-1$ is the number of individual hypotheses. Then, the overall H_0 is rejected if any H_{0j} is rejected. This overall test's size is bounded above by α: the probability of a union of events (i.e., the H_{0j} rejection events) is bounded above by the sum of the event probabilities, which here is $\sum_{j=1}^{J-1} \alpha/(J-1) = \alpha$. The Bonferroni correction is related to the union–intersection test approach (e.g., Casella and Berger, 2002, §8.2.3,8.3.3). The tests for individual H_{0j} could be one-sided t-tests using the asymptotic distribution in (17) below, for example.

More sophisticated testing of null hypotheses like (14) and (15) has been considered in papers going back to Kodde and Palm (1986) and Perlman (1969). It is also the topic of many recent econometrics papers. For example, see Andrews and Barwick (2012), Romano, Shaikh, and Wolf (2014), McCloskey (2015), and references therein. Many of these papers try to improve power (while maintaining asymptotic size control) by determining which inequalities are "far" from binding and only testing the remainder, a procedure termed "moment selection" in the context of moment inequality testing. For example, if θ_1 is estimated to be very negative (e.g., 10 standard errors below zero), then we could test only $j = 2, \ldots, J-1$, which can be done with a smaller critical value and thus higher power; e.g., the Bonferroni correction would allow individual tests with level $\alpha/(J-2)$ instead of $\alpha/(J-1)$. Our case is simpler than moment inequality testing generally since our moment functions like $\mathbb{1}\{X \le j\} - \mathbb{1}\{Y \le j\}$ do not involve unknown parameters.

Implementation of any test above requires the sampling distribution of $\hat{\boldsymbol{\theta}}$, which is provided below in (17) for convenience. It is asymptotically multivariate normal, as seen in the following. Since $\sum_{i=1}^{n_X} \mathbb{1}\{X_i \le j\} \sim \text{Binomial}(n_x, F_X(j))$, the CDF estimators $\hat{F}_X(j) = n_X^{-1} \sum_{i=1}^{n_X} \mathbb{1}\{X_i \le j\}$ are scaled (by n_X^{-1}) binomial random variables, with mean $F_X(j)$ and variance $F_X(j)[1 - F_X(j)]/n_X$. In large samples, the central limit theorem provides the normal approximation

$$\begin{gathered} \sqrt{n_X}\big[(\hat{F}_X(1), \ldots, \hat{F}_X(J-1)) - (F_X(1), \ldots, F_X(J-1))\big] \xrightarrow{d} \text{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_X), \\ \Sigma_{X,jk} \equiv F_X(j)[1 - F_X(k)] \text{ for } j \le k, \quad \Sigma_{X,kj} = \Sigma_{X,jk}, \end{gathered} \tag{16}$$

and similarly for $\hat{F}_Y(\cdot)$. Without iid sampling, a different central limit theorem may still hold, albeit with different $\underline{\boldsymbol{\Sigma}}_X$ and $\underline{\boldsymbol{\Sigma}}_Y$.

Since the samples for X and Y are assumed independent, the sampling distributions of the corresponding CDF estimators are independent, so their covariance is zero. Thus, assuming $n_X/n_Y \to \delta \in (0, \infty)$,

$$\sqrt{n_X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \text{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\theta), \quad \underline{\boldsymbol{\Sigma}}_\theta \equiv \underline{\boldsymbol{\Sigma}}_X + \delta \underline{\boldsymbol{\Sigma}}_Y. \tag{17}$$

For MPS testing, by the continuous mapping theorem,

$$\sqrt{n_X}(\underline{\mathbf{D}}\hat{\boldsymbol{\theta}} - \underline{\mathbf{D}}\boldsymbol{\theta}) = \underline{\mathbf{D}}\sqrt{n_X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \underline{\mathbf{D}}\text{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}) = \text{N}(\mathbf{0}, \underline{\mathbf{D}}\underline{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}\underline{\mathbf{D}}'). \tag{18}$$

3.3.2 Null hypothesis: non-MPS (known medians) or non-SD1

The null hypothesis of non-SD1 (of X over Y) is

$$H_0\colon \theta_j > 0 \text{ for some } j = 1, \ldots, J-1, \tag{19}$$

with θ_j from (13). That is, (19) is true iff (14) is false. This can be written as a union:

$$H_0\colon \boldsymbol{\theta} \in \Theta_0, \quad \Theta_0 \equiv \bigcup_{j=1}^{J-1} \{\boldsymbol{\theta} : \theta_j > 0\}. \tag{20}$$

Thus, an intersection–union test may be used. Similarly, non-MPS (of Y versus X) can be tested by replacing θ_j with $D_{jj}\theta_j$.

The intersection–union test of (20) rejects when $H_{0j}\colon \theta_j > 0$ is rejected for all j. That is, the overall rejection region is the intersection of the rejection regions for individual tests of H_{0j}. If each H_{0j} test has size α, then the overall test has size α:

$$\begin{aligned}
\sup_{\boldsymbol{\theta}: H_0 \text{ true}} \mathrm{P}(\text{reject } H_0 \mid \boldsymbol{\theta}) &= \sup_{j\in\{1,\ldots,J-1\}} \sup_{\boldsymbol{\theta}: H_{0j} \text{ true}} \mathrm{P}(\text{reject } H_0 \mid \boldsymbol{\theta}) \\
&= \sup_{j\in\{1,\ldots,J-1\}} \sup_{\boldsymbol{\theta}: H_{0j} \text{ true}} \mathrm{P}(\text{reject all } H_{01},\ldots,H_{0J-1} \mid \boldsymbol{\theta}) \\
&\le \sup_{j\in\{1,\ldots,J-1\}} \overbrace{\sup_{\boldsymbol{\theta}: H_{0j} \text{ true}} \mathrm{P}(\text{reject } H_{0j} \mid \boldsymbol{\theta})}^{=\text{size of } H_{0j} \text{ test} \le \alpha} \\
&\le \sup_{j\in\{1,\ldots,J-1\}} \alpha \\
&= \alpha.
\end{aligned}$$

For more, see Theorem 8.3.23 and Sections 8.2.3 and 8.3.3 in Casella and Berger (2002), who also remark, "The IUT may be very conservative" (p. 306).

3.3.3 Null hypothesis: MPS (unknown medians) or single-crossing

To characterize the single-crossing null hypothesis $H_0\colon X \text{ SC } Y$, we follow the structure of (5). With $\theta_j \equiv F_X(j) - F_Y(j)$ as in (13),

$$H_0\colon \boldsymbol{\theta} \in \Theta_0, \quad \Theta_0 \equiv \bigcup_{c=2}^{J-1} \Theta_c, \quad \Theta_c \equiv \left\{\boldsymbol{\theta} : (2\,\mathbb{1}\{j<c\}-1)\theta_j \le 0,\ j=1,\ldots,J-1\right\}. \tag{21}$$

To test the H_0 in (21), we may combine the intersection–union approach of Section 3.3.2 with a method from Section 3.3.1. Each H_{0c} has the same structure as the null of MPS with known medians $m = c$, so any test from Section 3.3.1 may be used. That is, to test $H_0\colon X \text{ SC } Y$ at level α, there are two steps:

1. For $c = 2,\ldots,J-1$, using the definition in (21), test $H_{0c}\colon \boldsymbol{\theta} \in \Theta_c$ at level α using a method from Section 3.3.1.

2. Reject $H_0\colon X \text{ SC } Y$ if and only if all H_{0c} are rejected.

MPS with unknown medians is essentially a stricter version of single-crossing. MPS adds the condition that the crossing point is equal to the medians of both distributions. Consequently, tests of sets of inequalities cannot be used, but the simple (albeit conservative)

Bonferroni adjustment could still be used, adding the median equality hypotheses to the inequality hypotheses to be tested jointly. However, given the results in Section 2.3, it may make more sense to just test for single-crossing anyway.

3.3.4 Other null hypotheses

The other null hypotheses characterized in Section 3.1 are also combinations of unions and intersections of inequalities. In principle, these may be tested by combining the approaches of Sections 3.3.1 and 3.3.2.

It may be tempting to simplify the null hypotheses in Section 3.1 to make testing easier, but this may generate multiple testing problems. For example, with the "fanning out" hypothesis, since there are so many possible ways in which we could observe the phenomenon, it may be tempting to just find the largest empirical fanning out and only test that one. However, this is not a valid procedure. It is akin to running a regression with lots of regressors, taking (only) the one with the largest t-statistic, and concluding it is statistically significant at a 5% level if $|t| > 1.96$. This ignores the multiple testing problem: even if no regressor is significant, or in our case if no pairs of population CDF increments show a fanning out, there may be a high probability that at least one of them looks statistically significant in the sample. Testing only the biggest empirical fanning out would thus not control size at the desired level.

3.4 Bayesian and frequentist differences

Among the usual advantages and disadvantages of Bayesian and frequentist inference (e.g., coherence vs. calibration), some are particularly important for ordinal distribution relationships. Their relative importance may also depend on the empirical application.

The Bayesian approach seems particularly advantageous in this setting. First, it is easy to compute. Unlike in settings requiring MCMC, the posterior in (11) can be sampled from directly, and all relationships can be assessed simultaneously. In contrast, each null hypothesis requires a separate method for frequentist testing. Second, posterior probabilities are easy and intuitive to interpret. They reflect our beliefs about different possible relationships given the data, which is often how p-values are (incorrectly) interpreted. Third, the posterior probabilities of different possible ordinal relationships are coherent. That is, they obey the usual probability laws; e.g., the three posterior probabilities of $X \text{ SD}_1 Y$, $Y \text{ SD}_1 X$, and neither having SD1 sum to 100%. Coherence is more important than usual in this setting where a variety of relationships are considered simultaneously. Fourth, in cases when decisions must be made (not just reporting of posterior probabilities), it is easy to use more appropriate loss functions when computing the Bayes decision rule, and the choice of loss function is transparent and explicit. Fifth, assuming a finite loss function, any Bayes decision rule is admissible. That is, there is no other decision rule that has weakly better (frequentist) risk for every possible $F_X(\cdot)$ and $F_Y(\cdot)$ (and strictly better for at least some values). Admissibility is a frequentist property but is not attained by every frequentist hypothesis test.

The Bayesian approach is most often criticized for parametric likelihoods, subjective priors, and slow computation. However, here we do not have a latent parametric model, but rather a nonparametric model of the category probabilities. Subjective priors may be used if helpful, but we focus on using objective priors in Section 3.2. Computation, as noted above, is actually a Bayesian advantage in this setting.

The one remaining criticism is that the Bayesian hypothesis test from Section 3.2 may not control size at a chosen level α (if that is desired), even asymptotically. It is possible that using a different prior or a different loss function (when deciding whether to reject a hypothesis or not) may solve this problem. These remain interesting open questions. For now, we explain why the most obvious Bayesian hypothesis test suffers size distortion, i.e., why the posterior may not be interpreted as a p-value.

The frequentist size of Bayesian hypothesis tests of sets of inequality constraints (like in Section 3.1) has been studied by Kaplan and Zhuo (2018). They consider the Bayesian test that rejects some H_0 when the posterior probability of H_0 is below α, essentially treating the posterior like a p-value. This particular test has both practical and decision-theoretic motivation; see Section 3.2 and Kaplan and Zhuo (2018, §2.1). They consider a single-draw Gaussian experiment in which the sampling and posterior distributions are equivalent (motivated by the Bernstein–von Mises theorem). Their main results are: a) if the null hypothesis is that the (local) parameter belongs to a half-space of the unrestricted parameter space, then the Bayesian test's size equals the nominal α; b) if instead H_0 is a subset of a half-space, then the Bayesian test's size exceeds α; c) if instead H_0 cannot be contained in any half-space, then the Bayesian test's size may be above, below, or equal to α, possibly depending on the sampling/posterior distribution. For example, if $H_0\colon F_X(3) \leq 1/2$, then result (a) applies: asymptotically, the Bayesian test has size α. More relevant to inequality relationships, if $H_0\colon X \text{ SD}_1\ Y$ (or MPS with known median), then result (b) applies: size is above α. If H_0 is that X does not SD1 Y, then (c) applies: it may be possible that the Bayesian test's size is below α, but more information is needed. Result (a) comes partly from the shape of H_0 but also from the prior probability of H_0 being small.

However, the results of Kaplan and Zhuo (2018) do not apply to our Bayesian method with the prior probability of H_0 adjusted to be 1/2. Lacking general theoretical results, we provide simulation evidence in Section 5. In particular, the adjustment does not yield size control when the null is SD1 or SC.

In sum: if size control is supremely important, then frequentist tests may be preferred; otherwise, the Bayesian approach has advantages in both interpretation and computation.

4 Extension to conditional distributions

We briefly discuss extension of identification and inference to conditional distributions ("regression"). The previous sections' results essentially compare conditional distributions when the conditioning variable is binary. Instead of comparing Y^* with another population X^*, we can equivalently compare the conditional distributions $Y^* \mid Z = 0$ and $Y^* \mid Z = 1$. Assuming a sampling process (like iid) that allows consistent estimation of the conditional ordinal distributions $Y \mid Z = 0$ and $Y \mid Z = 1$, the identification results from Section 2 apply directly.

More generally, there can be a conditioning vector $\mathbf{Z}$ with categorical, discrete, and/or continuous elements. Given conditional ordinal distributions $Y \mid \mathbf{Z} = \mathbf{z}_1$ and $Y \mid \mathbf{Z} = \mathbf{z}_2$, the identification results of Section 2 apply to pairwise inequality comparisons of the corresponding latent conditional distributions $Y^* \mid \mathbf{Z} = \mathbf{z}_1$ and $Y^* \mid \mathbf{Z} = \mathbf{z}_2$, for any values $\mathbf{z}_1$ and $\mathbf{z}_2$. For statistical inference, like before, we can examine the conditional ordinal distributions directly, interpreting the results in light of the identification results.

Often the conditional ordinal distributions can be consistently estimated nonparametrically, even with continuous elements in conditioning vector $\mathbf{Z}$, even without iid sampling. For example, kernel regression of $\mathbb{1}\{Y \le j\}$ on $\mathbf{Z}$ (for all $j = 1, \ldots, J-1$) can do this, or other nonparametric "distribution regression" methods; e.g., see Frölich (2006), who also discusses semiparametric estimation. Regardless of the conditional ordinal CDF estimator, given asymptotic normality, the frequentist tests from Section 3 can be applied. Alternatively, given any posterior over any two conditional ordinal distributions, posterior probabilities of the different ordinal relationships of interest can be calculated.

With continuous $\mathbf{Z}$, there are infinite possible pairwise comparisons $(\mathbf{z}_1, \mathbf{z}_2)$, so thoughtful summaries are required. One possibility is to use a benchmark $\mathbf{z}_0$. Then, for a given relationship, the set of $\mathbf{z}$ where the relationship holds is identified and can be estimated. For example, we could estimate the set of $\mathbf{z}$ such that the ordinal distribution of $Y \mid \mathbf{Z} = \mathbf{z}$ has a single crossing of the benchmark distribution of $Y \mid \mathbf{Z} = \mathbf{z}_0$: $\{\mathbf{z} : \hat{F}_{Y\mid\mathbf{Z}}(\cdot \mid \mathbf{Z} = \mathbf{z}) \text{ SC } \hat{F}_{Y\mid\mathbf{Z}}(\cdot \mid \mathbf{Z} = \mathbf{z}_0)\}$. Other such summaries are left to future work.

5 Simulations

Our simulations investigate two questions about statistical inference on the ordinal relationships of interest. First, do the newer frequentist hypothesis tests (like those cited in Section 3.3.1) provide much improvement? Second, can the Bayesian posterior probabilities be treated as p-values? With the uninformative prior (over category probabilities), the answer should be "no" based on the theoretical results in Kaplan and Zhuo (2018), but they do not consider the adjusted prior that assigns 1/2 prior probability to the null hypothesis.

The null hypotheses are SD1 and SC, with DGPs as follows. In DGP 1, the ordinal distributions of X and Y are identical, with 1/5 probability on each of five categories. This is on the boundary of SD1 and SC, and all inequalities $F_X(j) \leq F_Y(j)$ for $j = 1, 2, 3, 4$ are binding for SD1. (Recall $F_X(5) = F_Y(5) = 1$ regardless.) DGP 2 is also on the boundary of both SD1 and SC, but only one inequality is binding for SD1: Y is the same, but X has only 1/10 probability on each of the first three categories, with 1/2 probability on the fourth category, so $F_X(j) < F_Y(j)$ for $j = 1, 2, 3$ but $F_X(4) = F_Y(4) = 4/5$.

The following methods are compared. First, for both SD1 and SC, posterior probabilities are computed from the Dirichlet–multinomial model, with both the unadjusted and adjusted priors from Section 3.2. The corresponding test rejects when the posterior is below α, as in Section 3.4. Second, also for both SD1 and SC, the recommended procedure from Section 2 of Andrews and Barwick (2012) is used, denoted RMS ("refined moment selection"). For SC, RMS is used as in Section 3.3.3. The basic idea of RMS (and related procedures) is to try to remove inequalities that are clearly slack in order to improve power, with some adjustment terms to ensure uniform size control. Third, for SD1, the Kolmogorov–Smirnov (KS) test is naively applied. The KS test is known to be conservative for discrete distributions, even under the least favorable configuration (DGP 1), but it is widely available and commonly used. It also lacks the power-improving "moment selection" of RMS. Other method and simulation parameters are given in Table 1.

Table 1 shows the answer to our first question is yes, the recent RMS testing procedure improves considerably over KS. For DGP 1 and especially DGP 2, KS has type I error rate well below α. Although this is not itself problematic, it implies (by continuity) that power is low for deviations from these DGPs; recall that both are on the boundary of H_0, so even infinitessimal deviations can violate H_0. In contrast, RMS has type I error rate near α for DGP 1, for all sample sizes. Though somewhat lower, its type I error rate is also much closer to α for DGP 2. The combined intersection–union RMS test for SC also appears reasonable, although as expected it is more conservative than the SD1 test. Overall, RMS appears to control size without being overly conservative like KS.

Table 1: Simulated type I error rate, nominal $\alpha = 0.10$.

		H_0: X SD_1 Y				H_0: X SC Y		
DGP	n	KS	RMS	Bayes	Bayes (adj)	RMS	Bayes	Bayes (adj)
1	50	0.038	0.089	0.436	0.204	0.032	0.439	0.175
1	100	0.022	0.084	0.430	0.205	0.029	0.359	0.142
1	500	0.027	0.092	0.447	0.199	0.034	0.428	0.171
1	1000	0.032	0.079	0.454	0.228	0.032	0.408	0.155
2	50	0.004	0.057	0.127	0.032	0.031	0.125	0.034
2	100	0.002	0.068	0.105	0.031	0.085	0.133	0.041
2	500	0.006	0.087	0.098	0.029	0.095	0.114	0.032
2	1000	0.003	0.074	0.084	0.025	0.060	0.084	0.018

Notes: sample sizes $n_X = n_Y = n$, 1000 simulation replications, 1000 posterior draws, 200 RMS bootstrap draws.

Table 1 shows the answer to our second question is no, the posteriors cannot be treated as p-values, even with the prior adjusted to have 1/2 probability on H_0. For DGP 1, for SD1, the type I error rate for Bayes (adj) stays around 0.20 as n increases, and it stays around 0.15 for SC. These are both clearly above the nominal $\alpha = 0.10$. For DGP 2, the adjusted prior instead leads to rejection rates well *below* α because the unadjusted Bayes test already has type I error rates very close to the nominal level. This is due to DGP 2 having only a single binding inequality, which (in large enough samples) reduces the problem to essentially a single inequality test, for which the unadjusted posterior behaves like a p-value; see Kaplan and Zhuo (2018) for further discussion.

In principle, a different prior adjustment could lead to a posterior that behaves like a p-value. For SD1, the prior could be adjusted so that the type I error rate is α for the least favorable configuration, to ensure size control. However, a different adjustment may be required for different α, in which case no single adjustment would lead to the posterior behaving like a p-value. Additionally, the interpretation of the adjustment may seem arbitrary, as opposed to the uninformative (unadjusted) prior or the prior with 1/2 probability on H_0 as advocated in (parts of) the literature. Still, it may be practically valuable and theoretically insightful for future work to determine the prior adjustment leading to size control for different hypotheses.

6 Empirical illustrations

Section 6.1 examines selection into a wellness incentive program, comparing sample distributions and hypothesis tests. Section 6.2 uses PSID data to compute posteriors for different relationships between U.S. states' health distributions.

6.1 Selection into wellness incentive programs

Jones, Molitor, and Reif (2018) implement and study a wellness incentive program for employees of the University of Illinois. They produce many results on both treatment effects and selection effects, for a variety of variables and incentive structures.

We focus on selection into participation based on initial health. Within the "treatment group" of employees allowed to participate in the program, actual participation was voluntary. There were three opportunities to participate: the initial screening, fall activity, and spring activity. Comparing participant and non-participant baseline characteristics shows which types the incentive program is likely to attract.

Jones et al. (2018) find that participants are generally "better" or "less dispersed" than non-participants. These patterns generally hold across different participation opportunities and variables. Examples of "better" are lower average medical spending and more healthy behaviors like gym use; see page 3 and Tables A.1b and A.1d. Examples of "less dispersed" are baseline medical spending and salary; see page 3 and Figures 4 and 5.

Our dataset comes from the paper's authors. Of 3300 total individuals eligible to participate, 1848 did the screening, 903 did the fall activity, and 740 did the spring activity. Baseline self-reported health status (SRHS) has the usual five-category scale from poor to excellent.

There is no evidence of selection into the initial screening based on health. The null hypothesis that the two SRHS distributions (i.e., participant and non-participant) are identical cannot be rejected; the p-value for the Pearson's chi-squared test is 0.48. Further, the sample SRHS CDFs cross multiple times, suggesting neither distribution is better or less dispersed.

There is mildly suggestive evidence of selection into the fall activity. This is seen in Figure 7. The null of equality again cannot be rejected ($p = 0.19$). However, participants' sample SRHS distribution first-order stochastically dominates (SD1) that of non-participants.

The strongest evidence of selection on health is into the spring activity. This is again seen in Figure 7. Equality of the two SRHS distributions is rejected at a 5% level ($p = 0.014$). Even accounting for multiple testing of three pairs of distributions (screening, fall, spring), this test still rejects at a 5% familywise error rate level since 0.014 is below the Bonferroni-adjusted $0.05/3 = 0.017$. As with the fall activity, the sample SRHS distribution

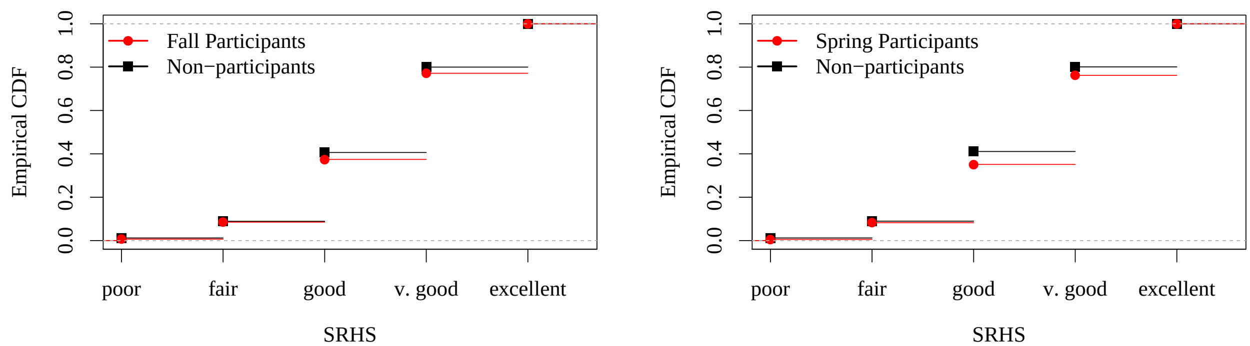

Figure 7: Empirical CDFs of baseline SRHS for participants (circle/red) and non-participants (square/black), for fall (left) and spring (right) activity.

of participants dominates (SD1) that of non-participants. Sample SD1 implies the null hypothesis of SD1 is not rejected at any common level.

The spring activity has further evidence for ordinal SD1. Let $F_P(\cdot)$ denote the SRHS (ordinal) CDF for the participants, and let $F_N(\cdot)$ denote the SRHS CDF for non-participants. Let $j = 1$ denote the "poor" category, up to $j = 5$ for "excellent." Since $F_P(5) = F_N(5) = 1$, only $j = 1, 2, 3, 4$ are relevant. Full SD1 of participant over non-participant SRHS is characterized by $F_P(j) \leq F_N(j)$ for $j = 1, 2, 3, 4$. Rejecting $F_P(j) \geq F_N(j)$ in favor of $F_P(j) < F_N(j)$ thus provides partial evidence of SD1. Indeed, this happens for some of the categories: the one-sided t-test p-values are 0.058, 0.28, 0.0018, and 0.011, for $j = 1, 2, 3, 4$, respectively. Further, for $j = 1, 2, 3$, define null and alternative hypotheses

$$\begin{aligned} H_0^j &: \{F_P(j) \geq F_N(j) \text{ or } F_P(j+1) \geq F_N(j+1)\}, \\ H_a^j &: \{F_P(j) < F_N(j) \text{ and } F_P(j+1) < F_N(j+1)\}. \end{aligned} \tag{22}$$

Full ordinal SD1 is equivalent to all H_a^j being true. Rejecting any H_0^j thus provides partial evidence toward SD1, without being fully conclusive, i.e., it is evidence of restricted SD1. Each individual H_0^j can be tested with an intersection–union test (IUT), as discussed earlier for testing (19). For example, to test H_0^3, p-values for one-sided t-tests of $F_P(3) \geq F_N(3)$ and of $F_P(4) \geq F_N(4)$ are separately computed; the IUT rejects at level α if both p-values are below α. Here, the p-values are 0.0018 and 0.011, so the IUT rejects H_0^3 at a 0.011 (or higher) level. However, neither H_0^2 nor H_0^1 is rejected since the p-value for the $j = 2$ t-test is 0.28. Using a Bonferroni correction (which is conservative) to account for testing three H_0^j simultaneously, controlling familywise error rate at level $\alpha = 0.05$ requires each H_0^j to be tested at level $\alpha/3 = 0.017$. Even then, H_0^3 is still rejected since $0.011 < 0.017$. Altogether, the sample SD1, non-rejection of SD1, and rejection of restricted non-SD1 form strongly suggestive evidence, though not definitive.

If there is ordinal SRHS SD1, then Theorem 2 offers different latent interpretations. For example, with a latent location–scale model, Theorem 2(vi) says ordinal SD1 implies latent restricted SD1. Although weaker than full SD1, this still provides evidence of selection, suggesting generally healthier individuals tend to select into participation. This comports with evidence from Jones et al. (2018).

6.2 Comparisons between U.S. states

Our new methodology is applied to the 2011 Panel Study of Income Dynamics (PSID).[3] Specifically, we consider Bayesian posterior probabilities of different relationships between health distributions (of heads of household) in different U.S. states. Due to the complex sampling design of the PSID (stratification, clustering, weights), we use the nonparametric methodology from Dong, Elliott, and Raghunathan (2014), as implemented in IVEware.[4] After taking 400 draws from the posterior, we compute the proportions of draws in which SD1, SC, and fanning out hold in each direction; see Definitions 2 and 5 and Theorem 4. The possible interpretations of these ordinal relationships in terms of the latent distributions are given in Section 2. Roughly speaking, $X \text{ SD}_1 Y$ provides some evidence that X corresponds to a healthier latent distribution than Y, while $X \text{ SC } Y$ provides some evidence that X corresponds to a less-dispersed latent distribution, as does "Y fans out." Other implementation details may be seen in our provided code.

Table 2: Empirical results from PSID 2011.

		Posterior probability (%)					
X	Y	$X \text{ SD}_1 Y$	$Y \text{ SD}_1 X$	$X \text{ SC } Y$	$Y \text{ SC } X$	X fansout	Y fansout
AZ	MO	0	90	4	2	34	11
NY	UT	0	3	0	94	70	1
IL	NY	20	0	66	1	3	92
MN	NY	24	0	57	0	3	96
IA	MO	0	10	2	16	98	42
LA	NY	10	0	10	1	23	97

Notes: probabilities rounded to the nearest percent. Observations per state: AZ (127), IL (287), IA (159), LA (141), MN (135), MO (253), NY (319), UT (87).

[3]The collection of this PSID data was partly supported by the National Institutes of Health under grant numbers R01 HD069609 and R01 AG040213, and the National Science Foundation under award numbers SES 1157698 and 1623684.

[4]Version 0.3; software developed by the Researchers at the Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan, available at `https://www.src.isr.umich.edu/software/`

Table 2 shows results for a few state pairs and three types of relationship. For SD1, usually one direction has probability near zero, while the other direction may have either small or large probability. For example, while there is 0% posterior probability for both AZ SD1 MO and NY SD1 UT, there is only 3% probability that UT SD1 NY, but 90% posterior probability that MO SD1 AZ.

For dispersion, it is more common to see large posterior probabilities of "fanning out" than SC. This is partly due to SD1 and SC being mutually exclusive, whereas fanning out may occur concurrently with SD1 and/or SC or even fanning out in the opposite direction. For example, with IL and NY, there is 92% posterior probability of IL being less dispersed in terms of fanning out, but only 66% probability that IL SC NY, partly because there is 20% probability of IL SD1 NY. Also, more assumptions are required to link fanning out to latent dispersion, so SC provides stronger evidence of a dispersion difference. The 94% probability of UT SC NY provides very strong evidence that UT is less dispersed. The 96% probability of some fanning out of NY relative to MN alone is less strong evidence that MN is less dispersed than NY, but this is bolstered by the small 3% probability of fanning out in the opposite direction, as well as the 57% probability of MN SC NY.

Table 3: Posterior probabilities (%) of SD1 and SC from PSID 2011.

	X SD_1 Y; Y is:					X SC Y; Y is:				
X	MO	KS	NE	IA	IL	MO	KS	NE	IA	IL
MO	—	0	10	10	0	—	7	67*	16	30
KS	34*	—	20*	10	3	6	—	24	6	16
NE	3	0	—	6	0	2	0	—	3	0
IA	0	0	6	—	0	2	0	66*	—	6
IL	40*	4	18*	43*	—	4	14	48	15	—

Notes: probabilities rounded to the nearest percent. Observations per state: IL (287), IA (159), KS (55), MO (253), NE (70). Asterisk (*) indicates that relationship holds in the sample (i.e., for the survey-weighted empirical distributions).

Table 3 shows posterior probabilities and sample relationships among Missouri (MO) and some of its midwestern neighbors, Kansas (KS), Nebraska (NE), Iowa (IA), and Illinois (IL). There are some cases where SD1 holds in the sample but the posterior probability shows great uncertainty. For example, IL SD1 NE in the sample, but the posterior is only 18%. There are other cases where SD1 or SC does not hold in the sample, but the possibility cannot be ruled out statistically. For example, IL SC NE does not hold in the sample but has a substantial 48% posterior probability.

7 Conclusion

We have provided new ways to compare health inequality in continuous latent distributions when only ordinal data are available, without making parametric assumptions. We have characterized implications of various latent and ordinal relationships under different sets of shape restrictions. We have discussed and compared both frequentist and Bayesian statistical inference on the relevant ordinal relationships.

Future research could consider: additional identification results, under alternative assumptions; multidimensional ordinal variables as in Yalonetzky (2013); whether any prior lets the Bayesian posterior probabilities be interpreted as p-values; stochastic thresholds; and further extension to conditional distributions and structural models like ordered choice.

A Proofs

A.1 Proof of Lemma 1

Proof. As a constructive proof, we solve for the crossing point m. Since $F_X^*(r) = F^*((r - \mu_X)/\sigma_X)$ and $F_Y^*(r) = F^*((r - \mu_Y)/\sigma_Y)$, the crossing point m that solves $F_X^*(m) = F_Y^*(m)$ satisfies

$$(m - \mu_X)/\sigma_X = (m - \mu_Y)/\sigma_Y. \tag{23}$$

Solving for m yields $m = (\sigma_Y \mu_X - \sigma_X \mu_Y)/(\sigma_Y - \sigma_X)$. If $\sigma_X = \sigma_Y$, then the denominator is zero and m does not exist. Otherwise (when $\sigma_X \neq \sigma_Y$), the formula provides a unique value. To verify that m is a crossing point, let $\sigma_X < \sigma_Y$ (wlog). For any $r > m$,

$$\frac{r - \mu}{\sigma} = \frac{m - \mu}{\sigma} + \frac{r - m}{\sigma},$$

so

$$\frac{r - \mu_X}{\sigma_X} = \overbrace{\frac{m - \mu_X}{\sigma_X}}^{\text{apply (23)}} + \frac{r - m}{\sigma_X} = \frac{m - \mu_Y}{\sigma_Y} + \overbrace{\frac{r - m}{\sigma_X}}^{\text{apply } \sigma_X < \sigma_Y} > \frac{m - \mu_Y}{\sigma_Y} + \frac{r - m}{\sigma_Y} = \frac{r - \mu_Y}{\sigma_Y}.$$

Since $F^*(\cdot)$ is strictly increasing, $(r - \mu_X)/\sigma_X > (r - \mu_Y)/\sigma_Y$ implies $F_X^*(r) > F_Y^*(r)$. Thus, $F_X^*(r) > F_Y^*(r)$ for all $r > m$. Similarly, $F_X^*(r) < F_Y^*(r)$ for all $r < m$ since then $(r - m)/\sigma_X < (r - m)/\sigma_Y$. □

A.2 Proof of Theorem 2

Proof. For Theorem 2(i): for each $j = 1, \ldots, J$,

$$\begin{aligned} F_X(j) &= F_X^*(\gamma_j) && \text{by A2} \\ &\leq F_Y^*(\gamma_j) && \text{since } X^* \text{ SD}_1\ Y^* \\ &= F_Y(j) && \text{since } \Delta_\gamma = 0, \end{aligned}$$

i.e., $F_X(j) \leq F_Y(j)$, meaning X SD$_1$ Y.

For Theorem 2(ii): a counterexample suffices. Let $X^* \sim \text{N}(0,1)$ and $Y^* \sim \text{N}(-1,1)$, so X^* SD$_1$ Y^*. Let $\gamma_1 = 0$ and $\Delta_\gamma = -2$. Let $\Phi(\cdot)$ be the standard normal CDF. Then, using A2, $F_X(1) = F_X^*(\gamma_1) = \Phi(0)$, and $F_Y(1) = F_Y^*(\gamma_1 + \Delta_\gamma) = \Phi(-1)$, so $F_Y(1) < F_X(1)$ and thus X does not SD1 Y (which would require $F_X(1) \leq F_Y(1)$).

For Theorem 2(iii): with A3 and $\sigma_X = \sigma_Y$, then let $\sigma_X = \sigma_Y = 1$ wlog (by rescaling $F^*(\cdot)$ if necessary). Similarly, let $\mu_X = 0$ wlog (by shifting $F^*(\cdot)$), so we have a pure location shift (Definition 6). For any $r \in \mathbb{R}$, $F_X^*(r) = F^*(r)$ and $F_Y^*(r) = F^*(r - \mu_Y)$, so $F_X^*(r) \leq F_Y^*(r)$ iff $\mu_Y \leq 0$ since $F^*(\cdot)$ is a non-decreasing function. That is, X^* SD$_1$ Y^* iff $\mu_Y \leq 0$. Thus, $F_X(j) \leq F_Y(j)$ is equivalent to $F_X^*(\gamma_j) \leq F_Y^*(\gamma_j)$, implying $\mu_Y \leq 0$ and thus X^* SD$_1$ Y^*.

For Theorem 2(iv): a counterexample suffices. Let $X^* \sim \text{N}(0,1)$ and $Y^* \sim \text{N}(\mu_Y, \sigma_Y^2)$, satisfying A3–A5. If $\sigma_Y \neq 1$, then $F_X^*(\cdot)$ and $F_Y^*(\cdot)$ have a single crossing (Definition 5) at crossing point $m \in \mathbb{R}$. If $m < \gamma_1$ and $\sigma_Y < 1$, then $F_X^*(\gamma_j) < F_Y^*(\gamma_j)$ for $j = 1, \ldots, J-1$, so X SD$_1$ Y, even though X^* does not SD1 Y^* since $F_X^*(m - \epsilon) > F_Y^*(m - \epsilon)$ for all $\epsilon > 0$.

For Theorem 2(v): consider $r \in [\gamma_j, \gamma_{j+1}]$. Since CDFs are non-decreasing, $F_Y^*(r) \geq F_Y^*(\gamma_j) = F_Y(j)$, and $F_X^*(r) \leq F_X^*(\gamma_{j+1}) = F_X(j+1)$. Thus, if $F_X(j+1) \leq F_Y(j)$, then

$$F_X^*(r) \leq F_X^*(\gamma_{j+1}) = F_X(j+1) \leq F_Y(j) = F_Y^*(\gamma_j) \leq F_Y^*(r).$$

This holds for $j = 1, \ldots, J-2$, so $F_X^*(r) \leq F_Y^*(r)$ for all $r \in [\gamma_1, \gamma_{J-1}]$, i.e., restricted SD1.

For Theorem 2(vi): given A3, from Lemma 1, there is at most one crossing point m of the latent CDFs. (Let $m = \infty$ if no crossing point.) If $F_X(j) \leq F_Y(j)$ for each $j = 1, \ldots, J-1$, then $F_X^*(\gamma_j) \leq F_Y^*(\gamma_j)$ for each j, so the corresponding γ_j must all be on the same side of the crossing point m. Thus, regardless of whether they are all above or below m, $F_X^*(r) \leq F_Y^*(r)$ for all $r \in [\gamma_1, \gamma_{J-1}]$, which is the definition of restricted latent SD1. □

A.3 Proof of Theorem 3

Proof. For Theorem 3(i): first, we show $\Delta_\gamma \neq 0$ does not affect IQRs. Using thresholds $\gamma_j + \Delta_\gamma$ for Y^* is equivalent to using thresholds γ_j for $Y^* - \Delta_\gamma$. Since Δ_γ is a constant, the quantiles of $Y^* - \Delta_\gamma$ are equal to the quantiles of Y^* minus Δ_γ (by equivariance of the quantile function), so the Δ_γ cancels out when taking differences of quantiles: writing $\mathrm{Q}_\tau(\cdot)$ as the τ-quantile operator (analogous to the expectation operator $\mathrm{E}(\cdot)$), so $\mathrm{Q}_\tau(W)$ is the τ-quantile of random variable W,

$$\mathrm{Q}_{\tau_2}(Y^* - \Delta_\gamma) - \mathrm{Q}_{\tau_1}(Y^* - \Delta_\gamma) = [\mathrm{Q}_{\tau_2}(Y^*) - \Delta_\gamma] - [\mathrm{Q}_{\tau_1}(Y^*) - \Delta_\gamma] = \mathrm{Q}_{\tau_2}(Y^*) - \mathrm{Q}_{\tau_1}(Y^*).$$

Thus, we proceed as if $\Delta_\gamma = 0$, which can always be accomplished by a location shift of Y^* that does not affect IQRs.

Second, for any $\tau_1 = F_X(j) = F_X^*(\gamma_j)$ for $j \le m$, we know $F_X(j) < F_Y(j)$, so $\tau_1 = F_X^*(\gamma_j) < F_Y^*(\gamma_j)$. By continuity and the assumption of strictly increasing latent CDFs, this implies $Q_Y^*(\tau_1) < \gamma_j$, so $Q_Y^*(\tau_1) < Q_X^*(\tau_1)$. Similarly, if $\tau_1 = F_Y(j)$ for $j \le m$, then $\tau_1 = F_Y^*(\gamma_j) > F_X^*(\gamma_j)$ and $Q_X^*(\tau_1) > \gamma_j = Q_Y^*(\tau_1)$. Similarly again, if τ_1 is strictly between $F_X(j)$ and $F_Y(j)$ for $j \le m$, then $Q_Y^*(\tau_1) < \gamma_j$ and $Q_X^*(\tau_1) > \gamma_j$, so again $Q_Y^*(\tau_1) < Q_X^*(\tau_1)$. Altogether, for all $\tau_1 \in \mathcal{T}_1$, $Q_Y^*(\tau_1) < Q_X^*(\tau_1)$. Similarly, for all $\tau_2 \in \mathcal{T}_2$, $Q_Y^*(\tau_2) > Q_X^*(\tau_2)$. For example, if $\tau_2 \in (F_Y(j), F_X(j))$ for $j > m$, then $Q_X^*(\tau_2) < \gamma_j$ and $Q_Y^*(\tau_2) > \gamma_j$, so $Q_Y^*(\tau_2) > Q_X^*(\tau_2)$.

Third, combining the above results, for any $\tau_1 \in \mathcal{T}_1$ and $\tau_2 \in \mathcal{T}_2$,

$$Q_X^*(\tau_2) - Q_X^*(\tau_1) < Q_Y^*(\tau_2) - Q_X^*(\tau_1) < Q_Y^*(\tau_2) - Q_Y^*(\tau_1).$$

For Theorem 3(ii): from Theorem 3(i), at least one IQR is strictly smaller for X^* than Y^*. This may be extrapolated to the whole distribution because given A3, the latent distributions' IQRs are either all identical (when $\sigma_X = \sigma_Y$), all smaller for X^* (when $\sigma_X < \sigma_Y$), or all larger for X^* (when $\sigma_X > \sigma_Y$). This is seen by the following. Since IQRs are unaffected by the locations μ_X and μ_Y of the latent distributions, let $\mu_X = \mu_Y = 0$ for simplicity. Since the latent CDFs are assumed strictly increasing, the latent quantile functions are simply inverse CDFs, so for any $\tau \in (0, 1)$,

$$\begin{aligned} \tau = F_Y^*(Q_Y^*(\tau_2)) &= F^*(Q_Y^*(\tau)/\sigma_Y), \\ (F^*)^{-1}(\tau) &= Q_Y^*(\tau)/\sigma_Y, \\ \sigma_Y Q^*(\tau) &= Q_Y^*(\tau), \end{aligned}$$

and similarly $Q_X^*(\tau) = \sigma_X Q^*(\tau)$. Thus, for any $0 < \tau_1 < \tau_2 < 1$,

$$\begin{aligned} Q_X^*(\tau_2) - Q_X^*(\tau_1) &= \sigma_X Q^*(\tau_2) - \sigma_X Q^*(\tau_1) = \sigma_X[Q^*(\tau_2) - Q^*(\tau_1)], \\ Q_Y^*(\tau_2) - Q_Y^*(\tau_1) &= \sigma_Y Q^*(\tau_2) - \sigma_Y Q^*(\tau_1) = \sigma_Y[Q^*(\tau_2) - Q^*(\tau_1)], \end{aligned}$$

so all IQRs are strictly smaller for X^* iff $\sigma_X < \sigma_Y$, strictly larger for X^* iff $\sigma_X > \sigma_Y$, and identical iff $\sigma_X = \sigma_Y$. Thus, altogether, the existence of even a single strictly smaller IQR of X^* implies that $\sigma_X < \sigma_Y$ and thus all other IQRs are smaller, too.

For Theorem 3(iii): from Theorem 3(ii), the ordinal CDF single crossing implies $\sigma_X < \sigma_Y$. From Definition 4, latent SD2 requires $\int_{-\infty}^{u}[F_X^*(r) - F_Y^*(r)]\,dr \le 0$ for all $u \in \mathbb{R}$. Since $F_X^*(r) - F_Y^*(r) \ge 0$ for all $r \ge m$, where m is the latent CDF crossing point, the integral is maximized at $r = \infty$, so it suffices to check $\int_{-\infty}^{\infty}[F_X^*(r) - F_Y^*(r)]\,dr \le 0$. Since $F^*(\cdot)$ is assumed symmetric with median zero, then $F^*(\delta) + F^*(-\delta) = 1$ and $F^*(\delta) = 1 - F^*(-\delta)$. If $\mu_X = \mu_Y \equiv \mu$, then with a change of variables to $s = r - \mu$ (so $ds = dr$),

$$
\begin{aligned}
&\int_{-\infty}^{\infty}[F_X^*(r) - F_Y^*(r)]\,dr \\
&= \int_{-\infty}^{\infty}[F^*((r-\mu)/\sigma_X) - F^*((r-\mu)/\sigma_Y)]\,dr \\
&= \int_{-\infty}^{\infty}[F^*(s/\sigma_X) - F^*(s/\sigma_Y)]\,ds \\
&= \int_{-\infty}^{0}[F^*(s/\sigma_X) - F^*(s/\sigma_Y)]\,ds + \int_{0}^{\infty}[F^*(s/\sigma_X) - F^*(s/\sigma_Y)]\,ds \\
&= \int_{-\infty}^{0}[F^*(s/\sigma_X) - F^*(s/\sigma_Y)]\,ds + \int_{0}^{\infty}\{1 - F^*(-s/\sigma_X) - [1 - F^*(-s/\sigma_Y)]\}\,ds \\
&= \int_{-\infty}^{0}[F^*(s/\sigma_X) - F^*(s/\sigma_Y)]\,ds - \int_{0}^{\infty}[F^*(-s/\sigma_X) - F^*(-s/\sigma_Y)]\,ds = 0.
\end{aligned}
$$

If we increase μ_X to some $\mu_X > \mu_Y$, then $F_X^*(r)$ decreases for all r since $(r - \mu_X)/\sigma_X$ decreases, so the integral becomes negative; latent SD2 holds. Similarly, if we decrease μ_X to some $\mu_X < \mu_Y$, then $F_X^*(r)$ increases everywhere and the integral becomes positive; latent SD2 does not hold. That is: in the symmetric location–scale model, if $\sigma_X < \sigma_Y$, then SD2 holds iff $\mu_X \ge \mu_Y$.

If the median of X is strictly above the median of Y, then there exists k such that $F_Y(k) \ge 1/2$ but $F_X(k) < 1/2$. Given $\Delta_\gamma = 0$, this implies $F_Y^*(\gamma_k) \ge 1/2$ and $F_X^*(\gamma_k) < 1/2$, so the median of Y^* is weakly below γ_k while the median of X^* is strictly above γ_k, i.e., $\mu_X > \mu_Y$. Altogether, the ordinal CDF single crossing is sufficient for $\sigma_X < \sigma_Y$, and the strictly larger ordinal median is sufficient for $\mu_X > \mu_Y$, which together are sufficient for $X^*\ \mathrm{SD}_2\ Y^*$.

If X and Y have the same median category k, then the latent medians are only bounded by γ_{k-1} and γ_k, so it is possible to have $\mu_Y > \mu_X$. Then, neither $X^*\ \mathrm{SD}_2\ Y^*$ nor $Y^*\ \mathrm{SD}_2\ X^*$. □

A.4 Proof of Theorem 4

Proof. For Theorem 4(i): as in the proof of Theorem 3(i), IQRs are unaffected by pure location shifts, so we let $\Delta_\gamma = 0$ for simplicity of notation. First, consider when $F_Y(j+1) > F_X(j)$; let $\tau_2 = F_Y(j+1)$ and $\tau_1 = F_X(j)$. We show that the smallest possible $\tau_2 - \tau_1$ IQR of Y^* (consistent with the ordinal distribution and assumptions) is still larger than the largest possible $\tau_2 - \tau_1$ IQR of X^*. Since we know $Q_Y^*(\tau_2) = \gamma_{j+1}$ (from A2), minimizing the $\tau_2 - \tau_1$ IQR for Y^* is equivalent to maximizing $Q_Y^*(\tau_1)$. From the combination of A4 and A5, $F_Y^*(\cdot)$ is concave after the median $Q_Y^*(1/2)$; since $F_Y(j) \geq 1/2$, $F_Y^*(\cdot)$ is concave on (at least) the interval $[\gamma_j, \gamma_{j+1}]$. Given this concavity constraint, $Q_Y^*(\tau_1)$ is maximized if $F_Y^*(\cdot)$ is a straight line over the interval $[\gamma_j, \gamma_{j+1}]$, i.e., having as little concavity as possible. Given this linearity, the (smallest possible) $\tau_2 - \tau_1$ IQR of Y^* is

$$Q_Y^*(\tau_2) - Q_Y^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_Y(j+1) - F_X(j)}{F_Y(j+1) - F_Y(j)}. \tag{24}$$

Equation (24) can be viewed as linear interpolation, taking the overall interval length $\gamma_{j+1} - \gamma_j$ and multiplying by the proportion determined by the probability ratio, or it can be viewed as taking the (constant) quantile function slope $(\gamma_{j+1} - \gamma_j)/[F_Y(j+1) - F_Y(j)]$ and multiplying by the quantile index difference $F_Y(j+1) - F_X(j)$. For X^*, we know $Q_X^*(\tau_1) = \gamma_j$, so the IQR is maximized by maximizing $Q_X^*(\tau_2)$. Given the concavity constraint, this is achieved when $F_X^*(\cdot)$ is a straight line over the interval, in which case the (largest possible) $\tau_2 - \tau_1$ IQR of X^* is

$$Q_X^*(\tau_2) - Q_X^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_Y(j+1) - F_X(j)}{F_X(j+1) - F_X(j)}. \tag{25}$$

In (24) and (25), both IQRs are expressed as a proportion of the length of the interval, $\gamma_{j+1} - \gamma_j$. The difference is (only) in the denominator of the proportion. Thus, (24) is larger than (25) iff $F_Y(j+1) - F_Y(j) < F_X(j+1) - F_X(j)$, which is the condition stated in Theorem 4(i).

Second, consider when $F_Y(j+1) < F_X(j)$; now let $\tau_1 = F_Y(j+1)$ and $\tau_2 = F_X(j)$, so interest is again in the $\tau_2 - \tau_1$ IQR. The approach is similar above: given concavity, straight-line CDFs minimize the Y^* IQR and maximize the X^* IQR, and then the linear interpolation (now linear extrapolation) formula provides the IQRs. For Y^*, now $Q_Y^*(\tau_1) = \gamma_{j+1}$, so the IQR is minimized by minimizing $Q_Y^*(\tau_2)$, which occurs when $F_Y^*(\cdot)$ is linear. Thus, the (smallest possible) $\tau_2 - \tau_1$ IQR of Y^* is

$$Q_Y^*(\tau_2) - Q_Y^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_X(j) - F_Y(j+1)}{F_Y(j+1) - F_Y(j)}. \tag{26}$$

For X^*, now $Q_X^*(\tau_2) = \gamma_j$, so the IQR is maximized by minimizing $Q_X^*(\tau_1)$, which occurs when $F_X^*(\cdot)$ is linear. (Even though we are extrapolating to the left of γ_j, we must still be above the median of X^* since $\tau_1 \geq 1/2$, so the concavity constraint still holds.) Thus, the (largest possible) $\tau_2 - \tau_1$ IQR of X^* is

$$Q_X^*(\tau_2) - Q_X^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_X(j) - F_Y(j+1)}{F_X(j+1) - F_X(j)}. \tag{27}$$

As before, only the denominator is different, so (26) is larger than (27) iff $F_Y(j+1) - F_Y(j) < F_X(j+1) - F_X(j)$, which is the condition stated in Theorem 4(i).

For Theorem 4(ii): the proof is symmetric to that of Theorem 4(i). That is, the structure is identical, but now we have an upper endpoint where before we had a lower endpoint, and we have a convexity instead of concavity constraint on the latent CDFs, but the IQR difference is still always minimized when both latent CDFs are straight lines (on the relevant intervals). For example, when $\tau_2 = F_X(j+1) > F_Y(j) = \tau_1$, the smallest possible Y^* IQR and largest possible X^* IQR are

$$Q_Y^*(\tau_2) - Q_Y^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_X(j+1) - F_Y(j)}{F_Y(j+1) - F_Y(j)},$$
$$Q_X^*(\tau_2) - Q_X^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_X(j+1) - F_Y(j)}{F_X(j+1) - F_X(j)}.$$

As before, only the denominator differs, so the criterion simplifies to which denominator is smaller, as stated in Theorem 4(ii).

For Theorem 4(iii): the proof is the same as that of Theorem 3(ii), which shows how a single larger IQR implies a larger σ. □

References

Abul Naga, R. H., Yalcin, T., 2008. Inequality measurement for ordered response health data. Journal of Health Economics 27 (6), 1614–1625.
URL `https://doi.org/10.1016/j.jhealeco.2008.07.015`

Allison, R. A., Foster, J. E., 2004. Measuring health inequality using qualitative data. Journal of Health Economics 23 (3), 505–524.
URL `https://doi.org/10.1016/j.jhealeco.2003.10.006`

Andrews, D. W. K., Barwick, P. J., 2012. Inference for parameters defined by moment inequalities: A recommended moment selection procedure. Econometrica 80 (6), 2805–2826.
URL `https://www.jstor.org/stable/23357242`

Atkinson, A. B., 1987. On the measurement of poverty. Econometrica 55 (4), 749–764.
URL `https://doi.org/10.2307/1911028`

Barrett, G. F., Donald, S. G., 2003. Consistent tests for stochastic dominance. Econometrica 71 (1), 71–104.
URL https://www.jstor.org/stable/3082041

Berger, J. O., Sellke, T., 1987. Testing a point null hypothesis: The irreconcilability of p values and evidence. Journal of the American Statistical Association 82 (397), 112–122.
URL https://doi.org/10.1080/01621459.1987.10478397

Blundell, R., Gosling, A., Ichimura, H., Meghir, C., 2007. Changes in the distribution of male and female wages accounting for employment composition using bounds. Econometrica 75 (2), 323–363.
URL https://www.jstor.org/stable/4501993

Bond, T. N., Lang, K., 2018. The sad truth about happiness scales. Journal of Political Economy, forthcoming.
URL https://doi.org/10.1086/701679

Casella, G., Berger, R. L., 1987. Testing precise hypotheses: Comment. Statistical Science 2 (3), 344–347.
URL https://www.jstor.org/stable/2245777

Casella, G., Berger, R. L., 2002. Statistical Inference, 2nd Edition. Duxbury, Pacific Grove, CA.

Davidson, R., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68 (6), 1435–1464.
URL https://www.jstor.org/stable/3003995

Davidson, R., Duclos, J.-Y., 2013. Testing for restricted stochastic dominance. Econometric Reviews 32 (1), 84–125.
URL https://doi.org/10.1080/07474938.2012.690332

Deaton, A., Paxson, C., 1998a. Aging and inequality in income and health. American Economic Review (Papers and Proceedings) 88 (2), 248–253.
URL https://www.jstor.org/stable/116928

Deaton, A., Paxson, C., 1998b. Health, income, and inequality over the life cycle. In: Frontiers in the Economics of Aging. University of Chicago Press, pp. 431–462.
URL https://www.nber.org/chapters/c7309

Dong, Q., Elliott, M. R., Raghunathan, T. E., 2014. A nonparametric method to generate synthetic populations to adjust for complex sampling design features. Survey Methodology 40 (1), 29.
URL https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X201400114003

Dutta, I., Foster, J., 2013. Inequality of happiness in the U.S.: 1972–2010. Review of Income and Wealth 59 (3), 393–415.
URL https://doi.org/10.1111/j.1475-4991.2012.00527.x

Frölich, M., 2006. Non-parametric regression for binary dependent variables. Econometrics Journal 9 (3), 511–540.
URL https://doi.org/10.1111/j.1368-423X.2006.00196.x

Goutis, C., Casella, G., Wells, M. T., 1996. Assessing evidence in multiple hypotheses. Journal of the American Statistical Association 91 (435), 1268–1277.
URL https://www.jstor.org/stable/2291745

Gu, Z., Jiang, Y., Yang, S., 2018. Ordered-response models with correlated thresholds, working paper, available at https://yixiaojiang-ethan.weebly.com/research.html.

Gunawan, D., Griffiths, W. E., Chotikapanich, D., 2018. Bayesian inference for health inequality and welfare using qualitative data. Economics Letters 162, 76–80.
URL https://doi.org/10.1016/j.econlet.2017.11.005

Hernández-Quevedo, C., Jones, A. M., Rice, N., 2005. Reporting bias and heterogeneity in self-assessed health. evidence from the British Household Panel Survey. HEDG Working Paper 05/04, Health, Econometrics and Data Group, The University of York.
URL https://ideas.repec.org/p/yor/hectdg/05-04.html

Jones, D., Molitor, D., Reif, J., 2018. What do workplace wellness programs do? evidence from the Illinois Workplace Wellness Study. NBER Working Paper 24229, National Bureau of Economic Research.
URL http://www.nber.org/papers/w24229

Kaplan, D. M., Zhuo, L., 2018. Frequentist size of Bayesian inequality tests, working paper, available at https://faculty.missouri.edu/~kaplandm.

Kaur, A., Prakasa Rao, B. L. S., Singh, H., 1994. Testing for second-order stochastic dominance of two distributions. Econometric Theory 10 (5), 849–866.
URL https://doi.org/10.1017/S0266466600008884

Klein, R. W., Sherman, R. P., 2002. Shift restrictions and semiparametric estimation in ordered response models. Econometrica 70 (2), 663–691.
URL https://doi.org/10.1111/1468-0262.00299

Kodde, D. A., Palm, F. C., 1986. Wald criteria for jointly testing equality and inequality restrictions. Econometrica 54 (5), 1243–1248.
URL https://www.jstor.org/stable/1912331

Lazar, A., Silber, J., 2013. On the cardinal measurement of health inequality when only ordinal information is available on individual health status. Health Economics 22 (1), 106–113.
URL https://doi.org/10.1002/hec.1821

Lindeboom, M., van Doorslaer, E., 2004. Cut-point shift and index shift in self-reported health. Journal of Health Economics 23 (6), 1083–1099.
URL https://doi.org/10.1016/j.jhealeco.2004.01.002

Lv, G., Wang, Y., Xu, Y., 2015. On a new class of measures for health inequality based on ordinal data. Journal of Economic Inequality 13 (3), 465–477.
URL https://doi.org/10.1007/s10888-014-9289-4

Madden, D., 2010. Ordinal and cardinal measures of health inequality: An empirical comparison. Health Economics 19 (2), 243–250.
URL https://doi.org/10.1002/hec.1472

Madden, D., 2014. Dominance and the measurement of inequality. In: Culyer, A. J. (Ed.), Encyclopedia of Health Economics. Vol. 1. Elsevier, pp. 204–208.
URL https://doi.org/10.1016/B978-0-12-375678-7.00725-2

Matzkin, R. L., 1992. Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. Econometrica 60 (2), 239–270.
URL https://www.jstor.org/stable/2951596

McCloskey, A., 2015. On the computation of size-correct power-directed tests with null hypotheses characterized by inequalities, working paper, available at http://www.brown.edu/Departments/Economics/Faculty/Adam_McCloskey/Research.html.

Perlman, M. D., 1969. One-sided testing problems in multivariate analysis. Annals of Math-

ematical Statistics 40 (2), 549–567.
URL https://projecteuclid.org/euclid.aoms/1177697723

PSID, 2018. Panel Study of Income Dynamics, public use dataset. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI; available at https://simba.isr.umich.edu/data/data.aspx.

Reardon, S. F., 2009. Measures of ordinal segregation. In: Flückiger, Y., Reardon, S. F., Silber, J. (Eds.), Occupational and Residential Segregation. Vol. 17 of Research on Economic Inequality. Emerald Group Publishing Limited, pp. 129–155.
URL https://doi.org/10.1108/S1049-2585(2009)0000017011

Romano, J. P., Shaikh, A. M., Wolf, M., 2014. A practical two-step method for testing moment inequalities. Econometrica 82 (5), 1979–2002.
URL https://www.jstor.org/stable/24029299

Silber, J., Yalonetzky, G., 2011. Measuring inequality in life chances with ordinal variables. In: Rodríguez, J. G. (Ed.), Inequality of Opportunity: Theory and Measurement. Vol. 19 of Research on Economic Inequality. Emerald Group Publishing Limited, Ch. 4, pp. 77–98.
URL https://doi.org/10.1108/S1049-2585(2011)0000019007

Stoye, J., 2010. Partial identification of spread parameters. Quantitative Economics 1 (2), 323–357.
URL https://doi.org/10.3982/QE24

Yalonetzky, G., 2013. Stochastic dominance with ordinal variables: Conditions and a test. Econometric Reviews 32 (1), 126–163.
URL https://doi.org/10.1080/07474938.2012.690653

Yalonetzky, G., 2016. Robust ordinal inequality comparisons with Kolm-independent measures. Working Paper 401, ECINEQ, Society for the Study of Economic Inequality.
URL https://ideas.repec.org/p/inq/inqwps/ecineq2016-401.html

Zhuo, L., 2017. Essays on decision making under uncertainty: Stochastic dominance. PhD dissertation, University of Missouri.
URL https://hdl.handle.net/10355/63757