PRELIMINARY AND INCOMPLETE
PLEASE DO NOT CIRCULATE OR CITE
BUT COMMENTS WELCOME!

Comparing latent inequality with ordinal health data

David M. Kaplan* Longhao Zhuo†

June 22, 2018

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. We contribute to three questions. First, what can certain latent relationships imply about the corresponding ordinal relationships, and vice-versa? Second, how can these ordinal relationships be tested statistically? Third, do Bayesian and frequentist statistical inferences differ significantly? Simulations and empirical examples illustrate our theoretical answers.

JEL classification: C11, C12, D30, I14

Keywords: Bayesian; nonparametric inference; partial identification; partial ordering; self-reported health status; shape restrictions

1 Introduction

One important, much-studied 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,

*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.

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 health inequality. The first type is inequality between two subpopulations (e.g., socioeconomic groups), where one has "better" health than the other. The second type is inequality within a population, where a widely dispersed distribution shows big health differences among individuals within the population. Additionally, dispersion increasing with age provides evidence of permanent (vs. transitory) health shocks, as studied by Deaton and Paxson (1998a,b).

Most generally, we wish to learn about inequality in a latent variable for which only ordinal data are available. This framework 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.

As detailed further below, the most important distinction from the literature is our framework of a continuous latent distribution that is not parametrically specified. Recently, in the context of happiness, Bond and Lang (2018) adopt this framework to show the impossibility of comparing latent means from ordinal data (their Theorem 1), pointing out that prior studies' conclusions about mean happiness 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. Within this more robust framework, we ask three research questions.

First, we ask: what do certain latent relationships imply about the corresponding ordinal relationships, and vice-versa? This is a question of identification: given the pair of latent distribution function bounds partially identified by the pair of ordinal distributions, what types of latent relationships apply to all (or none) of the latent distributions in the identified set? Compared to the parametric or discrete setups in the literature, our framework limits what may be learned, but certain latent relationships can still be refuted or verified by certain ordinal relationships.

Second, for the ordinal distribution relationships mentioned above, how can statistical inference be performed? With a finite sample, even the ordinal distributions are not perfectly known, so sampling uncertainty should be accounted for. We wish to statistically refute certain latent relationships: if latent relationship A implies ordinal relationship B, then rejection of B implies rejection of A (i.e., the contrapositive). Thus, we may learn about the latent relationship of interest even though statistically only the ordinal data and distributions are considered. We discuss how to apply recently developed frequentist moment inequality tests as well as a Bayesian Dirichlet–multinomial model with optional hypothesis-adjusted prior.

Third, do these frequentist and Bayesian inferences differ significantly? They do if the Bayesian prior on the probability vector parameter is uninformative, according to results from Kaplan and Zhuo (2018). For example, Kaplan and Zhuo (2018) explain why a Bayesian test of the null hypothesis of 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. The reason is partly due to the shape of the parameter subspace where stochastic dominance holds, but partly due to the prior: although it is uninformative for the parameter, it assigns much less than 1/2 probability to the null hypothesis of stochastic dominance. Here, we investigate whether the results from Kaplan and Zhuo (2018) still hold if the prior is adjusted to assign 1/2 probability to the null hypothesis. 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 The literature contains alternative approaches to assessing inequality with ordinal data, but our paper is different in two major ways. First, other papers take either a fully parametric latent model (like ordered probit) or else implicitly presume the latent distribution to be discrete, sharing the same categories as the ordinal distribution but with unknown cardinal values. 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 the semi/nonparametric ordered choice literature, our objects of interest are features of the latent distributions themselves. Second, 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 discussing uncertainty.

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. 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 they share the same median category and 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 certain cardinal measures of spread holding for all possible assignments 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. We also fill a gap in their work by providing statistical inference on the median-preserving spread and first-order stochastic dominance.

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, although only after a particular index is chosen (as well as weight parameters and functions). 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 any sort of statistical inference, but they simply report a jackknifed confidence interval without formal justification. In practice, it may be valuable both to compute inequality indexes and to more robustly compare distributions, as we do.

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) empirically examine whether the interquartile range of log wage has increased over time; although wage is continuously distributed and observable, missing data causes partial identification.

More generally, and most closely related to our approach, Stoye (2010) uses CDF bounds to derive bounds for two classes of dispersion (spread) parameters, including interquantile ranges as a special case. The biggest difference here is that each individual latent CDF is not even partially identified since the thresholds are unknown. Nonetheless, Stoye's (2010) consideration of the most compressed and most dispersed CDFs within the identified set is similar in spirit to our approach. The latent CDF bounds implied by ordinal data share the structure of his equation (3) (with zero mixture probability). However, neither bounding CDF has finite expectation, as he states to be required before (3), since ordinal data do not restrict the tails of the latent distributions. Thus, Theorem 2(i) does not apply, so we may not learn about "D_2-parameters" like variance from ordinal data without imposing additional restrictions. Despite the violation of this condition, it seems that the first part of Theorem 3(i) (with $p = 0$) should still apply as-is to interquantile ranges; in fact, it simplifies greatly. This simplification helps our results on dispersion (interquantile ranges) to be derived more directly, additionally 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 Bayesian and frequentist statistical inference. Section 4 highlights Bayesian and frequentist differences. 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. Then, we show what certain latent distribution relationships imply about the corresponding ordinal relationships, or vice-versa, under different sets of assumptions. That is, we derive testable implications of certain latent relationships, and we show when other latent relationships are identified. This section concerns only population distributions; data samples and statistical uncertainty are considered in Section 3.

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.

Sometimes, $\Delta_\gamma = 0$ (from A2) is further specified. 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 $\Delta_\gamma \neq 0$ is called an "index shift" by Lindeboom and van Doorslaer (2004), which they contrast with a "cut-point shift" allowing each γ_j to change arbitrarily between X and Y. 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, but the magnitude appears similar across socioeconomic groups.

The assumption of unbounded support in A1 actually matters for some results. For example, if the support is instead $[a, \infty)$ for $a \ge 0$, 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, but there may be cases where the opposite is true.

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 based 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 $F^*_X(r) \le F^*_Y(r)$ for all $r \in \mathbb{R}$; this is written $X^*\ \mathrm{SD}_1\ Y^*$ or $F^*_X\ \mathrm{SD}_1\ F^*_Y$.

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

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

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

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 SC F_B or A SC B; the second (switching F_A and F_B) is written F_B SC F_A or B 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 $\le$ with $<$) dominance; Davidson and Duclos (2013, p. 89) note that the distinction is statistically negligible in finite samples.

The single crossing property (Definition 5) has been useful in areas of economic theory like mechanism design, although this connection may not be insightful. 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" comes from Condition I of Atkinson (1987, p. 751) and is discussed in Davidson and Duclos (2000, 2013). Although it is not as strong as (unrestricted) SD1, it is still meaningful.

2.2 First-order stochastic dominance

First-order stochastic dominance (SD1) has long been valued for unambiguously ranking two cardinal 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. 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., $\Delta_\gamma = 0$. 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 $F_X^*(\gamma_j) \le F_Y^*(\gamma_j)$ for $j = 1, \ldots, J-1$ (note $\gamma_J = \infty$ provides no restriction); latent SD1 could easily be violated in the lower tail, for example, if $F_Y^*(\gamma_1 - \epsilon) = 0 < F_X^*(\gamma_1 - \epsilon)$ for some $\epsilon > 0$. 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.

Nonetheless, the result that latent SD1 implies ordinal SD1 provides a testable implication. There are certain violations of latent SD1 (e.g., in the tails) that cannot be detected with ordinal data, but some can be. Latent SD1 is refutable but not verifiable: ordinal distributions cannot provide conclusive evidence in favor of latent SD1, but they can provide conclusive evidence against latent SD1.

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. 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 $F_X^*(\gamma_j) \le F_Y^*(\gamma_j)$, but not enough structure to extrapolate SD1 into the tails.

Without $\Delta_\gamma = 0$, however, even latent SD1 implying ordinal SD1 fails to hold. Shifting thresholds cannot be distinguished from shifting of the latent distribution itself.

The above results are summarized in Proposition 2. Before that, we state Lemma 1, which helps prove parts of Propositions 2 and 3. Lemma 1 is essentially Proposition 1 from Bond and Lang (2018).

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).*

Proof of Lemma 1. 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

$$(m - \mu_X)/\sigma_X = (m - \mu_Y)/\sigma_Y. \tag{1}$$

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

$$\frac{r - \mu_X}{\sigma_X} = \overbrace{\frac{m - \mu_X}{\sigma_X}}^{\text{apply (1)}} + \frac{r - m}{\sigma_X} = \frac{m - \mu_Y}{\sigma_Y} + \overbrace{\frac{r - m}{\sigma_X}}^{\text{apply } \sigma_X < \sigma_Y} > \frac{m - \mu_Y}{\sigma_Y} + \frac{r - m}{\sigma_Y} = \frac{r - \mu_Y}{\sigma_Y}.$$

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 $(r - m)/\sigma_X < (r - m)/\sigma_Y$. □

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

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

(ii) *If $\Delta_\gamma \neq 0$, then $X^*\ \mathrm{SD}_1\ Y^* \;\not\!\!\implies X\ \mathrm{SD}_1\ Y$.*

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

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

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

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

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

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

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

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

For Proposition 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 $F_X(1) \le F_Y(1)$).

For Proposition 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 $F_X^*(r) \le F_Y^*(r)$ iff $\mu_Y \le 0$ since $F^*(\cdot)$ is a non-decreasing function. That is, X^* SD$_1$ Y^* iff $\mu_Y \le 0$. Thus, $F_X(j) \le F_Y(j)$ is equivalent to $F_X^*(\gamma_j) \le F_Y^*(\gamma_j)$, implying $\mu_Y \le 0$ and thus X^* SD$_1$ Y^*.

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

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

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

For Proposition 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 $F_X(j) \le F_Y(j)$ for each $j = 1, \ldots, J-1$, then $F_X^*(\gamma_j) \le F_Y^*(\gamma_j)$ 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, $F_X^*(r) \le F_Y^*(r)$ for all $r \in [\gamma_1, \gamma_{J-1}]$, which is the definition of restricted latent SD1. □

2.3 Dispersion

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 (in addition to transitory) "health shocks"; this is a primary 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, second-order stochastic dominance is more difficult to infer, requiring $\Delta_\gamma = 0$, whereas dispersion results allow $\Delta_\gamma \neq 0$.

In the location–scale model of A3, equality of latent dispersions $\sigma_X = \sigma_Y$ implies latent SD1, so rejecting latent SD1 implies $\sigma_X \neq \sigma_Y$. From Proposition 2(i), under A1 and A2 with $\Delta_\gamma = 0$, latent SD1 implies ordinal SD1, so violation of ordinal SD1 implies violation of latent SD1, implying $\sigma_X \neq \sigma_Y$. Violation of ordinal SD1 requires at least one crossing of the ordinal CDFs. Below, we more precisely interpret an ordinal CDF crossing and explore what can be learned about dispersion without a crossing.

Most fundamentally, if the ordinal CDFs cross (once), then we know certain interquantile ranges (IQRs) are smaller for one of the latent distributions. Heuristically, imagine the crossing is at $j = 2$, so $F_X(1) < F_Y(1)$ and $F_X(3) > F_Y(3)$; then, $F^*_X(\gamma_1) < F^*_Y(\gamma_1 + \Delta_\gamma)$ and $F^*_X(\gamma_3) > F^*_Y(\gamma_3 + \Delta_\gamma)$. Note that $(\gamma_3 + \Delta_\gamma) - (\gamma_1 + \Delta_\gamma) = \gamma_3 - \gamma_1$, so it is irrelevant whether or not $\Delta_\gamma = 0$. Then, $\gamma_3 - \gamma_1$ is the $F_X(3) - F_X(1)$ IQR for X^* but the $F_Y(3) - F_Y(1)$ IQR for Y^*. Since $F_X(1) < F_Y(1)$ and $F_X(3) > F_Y(3)$, the $F_X(3) - F_X(1)$ IQR of Y^* is larger than the $F_Y(3) - F_Y(1)$ IQR of Y^*, so Y^* has a larger $F_X(3) - F_X(1)$ IQR than X^*. Similar logic determines other IQRs that must be larger for Y^*, as enumerated in our formal result. 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) is enough to make statements about 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, implying $\mu_Y < \mu_X$, and the CDF crossing implies $\sigma_Y > \sigma_X$, then SD2 may be inferred.

The above results are summarized in Propositions 3 and 4. Unlike for SD1, we do not need $\Delta_\gamma = 0$ to learn about dispersion since pure location shifts of the latent distributions do not affect their dispersion.

Proposition 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 with $F_X(j) < F_Y(j)$ for $j \leq m$ 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*

$$\mathcal{T}_1 \equiv \bigcup_{j=1}^{m} [F_X(j), F_Y(j)], \quad \mathcal{T}_2 \equiv \bigcup_{j=m+1}^{J-1} [F_Y(j), F_X(j)]. \tag{2}$$

(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^* SD_2 Y^*. If there is the single crossing but instead X and Y have the same median, then it is ambiguous whether X^* SD_2 Y^* or not.*

Proof of Proposition 3. For Proposition 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 $\mathbb{Q}_\tau(\cdot)$ as the τ-quantile operator (analogous to the expectation operator $\mathbb{E}(\cdot)$), so $\mathbb{Q}_\tau(W)$ is the τ-quantile of random variable W,

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

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 $j \leq m$, 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 $j \leq m$, 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 $j \leq m$, 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 Proposition 3(ii): from Proposition 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 $\tau \in (0, 1)$,

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

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 Proposition 3(iii): from Proposition 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 $ds = dr$),

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

$$= \int_{-\infty}^{0} [F^*(s/\sigma_X) - F^*(s/\sigma_Y)]\, ds - \int_{0}^{\infty} [F^*(-s/\sigma_X) - F^*(-s/\sigma_Y)]\, ds = 0.$$

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 $\mu_X \geq \mu_Y$.

If the median of X is strictly above the median of Y, then there exists k such that $F_Y(k) \geq 1/2$ but $F_X(k) < 1/2$. Given $\Delta_\gamma = 0$, this implies $F_Y^*(\gamma_k) \geq 1/2$ 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 $Y^*\ \mathrm{SD}_2\ X^*$. □

What does Proposition 3 say about the ordinal median-preserving spread (MPS) of 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. Our results suggest that this is indeed evidence of increased dispersion even with a latent continuous variable, with the strength of the evidence depending on the strength of the assumptions. Proposition 3(iii) suggests that a median-preserving spread is not strong evidence of latent SD2, but that a median-decreasing spread is, given the location–scale model. Specifically, with a continuous latent distribution, there is too much ambiguity when the medians fall into the same category; we cannot know which latent median is larger (without stronger assumptions, like a parametric model).

If the CDFs do not cross, we may still infer something about dispersion (IQRs) if we impose a further 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 after the median. If we observe a "fanning out" of CDFs after the median, there is a certain IQR that must be larger for the lower CDF if concavity holds. Again, further assuming a location–scale model would allow extrapolation of this IQR to σ and thus the entire distribution.

Proposition 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 Proposition 4(i) and Proposition 4(ii) strengthen to $\sigma_Y > \sigma_X$, implying all IQRs of Y^* are larger.*

Proof of Proposition 4. For Proposition 4(i): as in the proof of Proposition 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

$$Q_Y^*(\tau_2) - Q_Y^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_Y(j+1) - F_X(j)}{F_Y(j+1) - F_Y(j)}. \tag{3}$$

Equation (3) 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

$$Q_X^*(\tau_2) - Q_X^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_Y(j+1) - F_X(j)}{F_X(j+1) - F_X(j)}. \tag{4}$$

In (3) and (4), 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, (3) is larger than (4) iff $F_Y(j+1) - F_Y(j) < F_X(j+1) - F_X(j)$, which is the condition stated in Proposition 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

$$Q_Y^*(\tau_2) - Q_Y^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_X(j) - F_Y(j+1)}{F_Y(j+1) - F_Y(j)}. \tag{5}$$

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

$$Q_X^*(\tau_2) - Q_X^*(\tau_1) = (\gamma_{j+1} - \gamma_j)\frac{F_X(j) - F_Y(j+1)}{F_X(j+1) - F_X(j)}. \tag{6}$$

As before, only the denominator is different, so (5) is larger than (6) iff $F_Y(j+1) - F_Y(j) < F_X(j+1) - F_X(j)$, which is the condition stated in Proposition 4(i).

For Proposition 4(ii): the proof is symmetric to that of Proposition 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 Proposition 4(ii).

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

3 Statistical inference on ordinal relationships

In this section, we consider statistical inference on the ordinal distribution relationships in Section 2. That is, we want to statistically assess our (un)certainty that a certain relationship exists between two ordinal population distributions, given a sample of data from each. We characterize the hypotheses of interest and then discuss both frequentist and Bayesian procedures, which are further compared in Section 4.

3.1 Statistical setup and hypotheses of interest

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,

$$X_i \overset{iid}{\sim} F_X,\ i = 1, \ldots, n_X, \quad Y_i \overset{iid}{\sim} F_Y,\ i = 1, \ldots, n_Y. \tag{7}$$

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:

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

Both the CDFs and PMFs may be expressed as moments. For example, $F_X(j) = \mathbb{E}(\mathbb{1}\{X \le j\})$ and $p_j^X = \mathbb{E}(\mathbb{1}\{X = j\})$.

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 $X\ \mathrm{SD}_1\ Y$ is a set of moment inequalities:

$$X\ \mathrm{SD}_1\ Y \iff \bigcap_{j=1}^{J-1} \{F_X(j) \le F_Y(j)\}. \tag{8}$$

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,

$$Y \text{ MPS } X \iff \bigcap_{j=1}^{J-1} \Big\{ [2\,\mathbb{1}\{j < m\} - 1][F_X(j) - F_Y(j)] \le 0 \Big\}. \tag{9}$$

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 over possible median values, checking (9) for each possible median:

$$\begin{aligned} Y \text{ MPS } X \iff \bigcup_{m=2}^{J-1} \Bigg\{ &\{F_X(m-1) < 1/2 \le F_X(m)\} \cap \{F_Y(m-1) < 1/2 \le F_Y(m)\} \\ &\cap \bigcap_{j=1}^{J-1} \Big\{ [2\,\mathbb{1}\{j < m\} - 1][F_X(j) - F_Y(j)] \le 0 \Big\} \Bigg\}. \end{aligned} \tag{10}$$

Note that MPS with $m = 1$ or $m = J$ is equivalent to SD1, so only $m = 2, \ldots, J-1$ are included in (10).

A single crossing is similar to (10), 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,

$$X \text{ SC } Y \iff \bigcup_{k=2}^{J-1} \bigcap_{j=1}^{J-1} \{[2\,\mathbb{1}\{j < k\} - 1][F_X(j) - F_Y(j)] < 0\}. \tag{11}$$

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

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

$$\begin{aligned} Y \text{ MDS } X \iff &\left\{ \bigcup_{k=2}^{J-1} \bigcap_{j=1}^{J-1} \{[2\,\mathbb{1}\{j < k\} - 1][F_X(j) - F_Y(j)] < 0\} \right\} \\ &\cap \left\{ \bigcup_{m=1}^{J-1} \{F_X(m) < 1/2 \le F_Y(m)\} \right\}. \end{aligned} \tag{12}$$

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

$$\text{multiple crossings} \iff \text{not } \{X \text{ SD}_1\ Y \text{ or } Y \text{ SD}_1\ X \text{ or } X \text{ SC } Y \text{ or } Y \text{ SC } X\}. \tag{13}$$

Technically, the SC in (13) should be slightly weaker than in (11), allowing for weak inequalities (≤ 0), but the difference is again statistically indistinguishable in practice. The expression in (13) 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

$$Y \text{ FanOut } X \iff \bigcup_{j=2}^{J-1} \{A_j \cup B_j\}, \tag{14}$$

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

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

3.2 Bayesian inference

Using the setup and notation of Section 3.1, we show how the Dirichlet–multinomial model can be used for Bayesian inference on all the ordinal relationships of interest. We then discuss a possible adjustment to the prior. Code is provided online.[2]

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

$$\mathbf{p}^X \sim \text{Dir}(\mathbf{1}), \quad \mathbf{p}^Y \sim \text{Dir}(\mathbf{1}), \tag{15}$$

where $\mathbf{1} \equiv (1, 1, \dots, 1)$ is a vector of J ones. If $J = 2$, then the priors simplify to $p_1^X \sim \text{Unif}(0, 1)$ and $p_1^Y \sim \text{Unif}(0, 1)$, 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.

Due to conjugacy, the posteriors are simply

$$\mathbf{p}^X \mid \mathbf{N}^X \sim \text{Dir}(\mathbf{1} + \mathbf{N}^X), \quad \mathbf{p}^Y \mid \mathbf{N}^Y \sim \text{Dir}(\mathbf{1} + \mathbf{N}^Y), \tag{16}$$

which are easy to sample from directly (i.e., without MCMC).

[2] https://faculty.missouri.edu/~kaplandm/code/ordinal_inequality_posteriors.R

To compute the posterior probability of a particular ordinal relationship, one may take many draws from the Dirichlet posteriors in (16) and check whether or not the relationship holds in each draw. For example, given a draw of $\mathbf{p}^X$ and $\mathbf{p}^Y$,

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

The proportion of total draws in which X SD$_1$ Y is the (approximate) posterior probability of X SD$_1$ Y. Increasing the number of draws makes the approximation error arbitrarily close to zero (with probability arbitrarily close to one).

In some cases, a different type of "objective" prior may be desired. In (15), 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 $H_0\colon X \text{ SD}_1\, Y$. In this case, the prior 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

$$\mathrm{P}^{\mathrm{adj}}(H_0 \mid \text{data}) = \frac{\mathrm{P}^{\mathrm{orig}}(H_0 \mid \text{data})}{\mathrm{P}^{\mathrm{orig}}(H_0 \mid \text{data}) + \frac{\gamma}{1-\gamma}[1 - \mathrm{P}^{\mathrm{orig}}(H_0 \mid \text{data})]} . \tag{17}$$

That is, we may compute the posterior probability using (16) and then adjust it using (17), once we know γ. 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 is identical to before, except that draws are taken from the prior in (15) instead of the posterior in (16).

The Bayesian approach is summarized in Method 1.

Method 1 (Bayesian inference)**.** Posterior probabilities of the ordinal distribution relationships in Section 2 may be computed as follows.

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 (16).

2. In each of the R draws, check whether each relationship holds, using the characteriza-

tions 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 (17). 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, a "generalized 0–1 loss" (e.g., Casella and Berger, 2002, eqn. (8.3.11)). This loss function is arguably implicit in frequentist hypothesis testing; 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 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 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. The iid sampling assumption continues to be maintained.

3.3.1 Null hypothesis: MPS (known medians) or SD1

For testing $H_0\colon X \text{ SD}_1\ Y$, let

$$\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_{J-1}), \quad \theta_j \equiv F_X(j) - F_Y(j). \tag{18}$$

For testing $H_0\colon Y$ MPS X with known median m for both X and Y, instead

$$\theta_j \equiv \begin{cases} F_X(j) - F_Y(j) & \text{if } j < m \\ F_Y(j) - F_X(j) & \text{if } j \geq m. \end{cases} \tag{19}$$

Note $F_X(j) - F_Y(j) = \mathbb{E}[\mathbb{1}\{X \leq j\}] - \mathbb{E}[\mathbb{1}\{Y \leq j\}] = \mathbb{E}[\mathbb{1}\{X \leq j\} - \mathbb{1}\{Y \leq j\}]$, a moment of the joint distribution of X and Y. For both SD1 and MPS, the null hypothesis is equivalent to

$$H_0\colon \boldsymbol{\theta} \leq \mathbf{0}, \tag{20}$$

i.e., $\theta_j \leq 0$ for all $j = 1, \ldots, J-1$.

The easiest way to test (20) is with a Bonferroni correction, but it will be conservative (i.e., size strictly below α). Using the Bonferroni approach, each individual hypothesis $H_{0j}\colon \theta_j \leq 0$ 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 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 (22) below, for example.

More sophisticated testing of null hypotheses like (20) 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 McCloskey (2015) and references therein, including Andrews and Barwick (2012) and Romano, Shaikh, and Wolf (2014). 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 the moment functions in (20) like $\mathbb{1}\{X \leq j\} - \mathbb{1}\{Y \leq j\}$ do not involve unknown parameters.

Implementation of any test above requires the sampling distribution of $\hat{\boldsymbol{\theta}}$, which is provided below in (22) 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

$$\begin{gathered}\sqrt{n_X}\left[(\hat{F}_X(1), \ldots, \hat{F}_X(J-1)) - (F_X(1), \ldots, F_X(J-1))\right] \xrightarrow{d} \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_X), \\ \Sigma_{X,jk} \equiv F_X(j)[1 - F_X(k)] \text{ for } j \le k, \quad \Sigma_{X,kj} = \Sigma_{X,jk}.\end{gathered} \tag{21}$$

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 $n_X/n_Y \to \delta \in (0, \infty)$,

$$\sqrt{n_X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\theta), \quad \underline{\boldsymbol{\Sigma}}_\theta \equiv \underline{\boldsymbol{\Sigma}}_X + \delta \underline{\boldsymbol{\Sigma}}_Y. \tag{22}$$

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

The null hypothesis of non-SD1 (of X over Y) or non-MPS (of Y vs. X) is

$$H_0\colon \theta_j > 0 \text{ for some } j = 1, \ldots, J-1, \tag{23}$$

where the θ_j remain as in (18) or (19). That is, (23) is true iff (20) is false. This can be written as a union:

$$H_0\colon \boldsymbol{\theta} \in \Theta_0, \quad \Theta_0 \equiv \bigcup_{j=1}^{J-1} \{\boldsymbol{\theta} : \theta_j > 0\}. \tag{24}$$

Thus, an intersection–union test may be used.

The intersection–union test of (24) 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 (11). Letting $\theta_j \equiv F_X(j) - F_Y(j)$,

$$H_0\colon \boldsymbol{\theta} \in \Theta_0, \quad \Theta_0 \equiv \bigcup_{c=2}^{J-1} \Theta_c, \quad \Theta_c \equiv \{\boldsymbol{\theta} : (2\,\mathbb{1}\{j < c\} - 1)\theta_j \leq 0, j = 1, \ldots, J-1\}. \tag{25}$$

To test the H_0 in (25), we may combine the intersection–union approach of Section 3.3.2 with a method from Section 3.3.1, as stated in Method 2. 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.

Method 2 (SC test)**.** To test $H_0\colon X \text{ SC } Y$ at level α:

1. For $c = 2, \ldots, J-1$, using the definition in (25), 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 that the regressors are jointly significant if the t-statistic exceeds (in absolute value) the standard 1.96 critical value. 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.

4 Bayesian and frequentist differences

There are advantages and disadvantages of Bayesian and frequentist inference generally, although some are more important for ordinal distribution relationships related to inequality. 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 (16) 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%, and the fact that MPS is a special case of single-crossing means the former always has a smaller posterior probability than the latter. 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. 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. This particular test has both practical and decision-theoretic motivation; see Section 3.2 and Kaplan and Zhuo (2018, §2.1). They consider a limit experiment in which the sampling and posterior distributions are equivalent, i.e., a Bernstein–von Mises theorem holds. 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. Although we do not provide theoretical results on this case here, we provide some simulation evidence in Section 5.

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.

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 posterior probabilities from the Bayesian Method 1 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 of Method 1 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 using Method 1, with both the unadjusted and adjusted priors. The corresponding test rejects if and only if the posterior is below α; that is, we treat the posterior as a p-value. Further decision-theoretic motivation for studying this test is provided by Kaplan and Zhuo (2018). Second, also for both SD1 and SC, the recommended procedure from Section 2 of Andrews and Barwick (2012) is used, denoted RMS (for "refined moment selection"). For SC, RMS is used within Method 2. 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: 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 first question is yes, the newer 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. RMS appears to control size without being overly conservative like KS.

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 application

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 (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 the following

[3]The collection of data used in this study was partly supported by the National Institutes of Health under grant number 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/`

mutually exclusive relationships hold: first-order stochastic dominance (in each direction), single crossing (in each direction), and multiple crossings. Other implementation details may be seen in our provided code.

Table 2: Empirical results from PSID 2011.

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

Note: 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).

Table 2 shows results for a few state pairs and three types of relationship. The state pairs are those involving either Missouri or New York (MO or NY, current and former residences of the first author) where there was strong evidence of some relationship (posterior above 90%); other states are Arizona (AZ), Illinois (IL), Iowa (IA), Louisiana (LA), Minnesota (MN), and Utah (UT). For SD1, usually one direction has probability very near zero, while the other direction may have large or small 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 single-crossing (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. As another example, the 98% posterior probability of there being some fanning out of IA compared to MO is not as strong evidence since there is 42% probability of some fanning out in the opposite direction (implying greater dispersion in MO); in contrast, a single crossing by definition may only go in one direction. 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, although this is bolstered by the small 3% probability of any fanning out in the opposite direction, as well as the 57% probability of MN SC NY and 24% probability of MN SD1 NY;

in all, there is strong evidence of MN having a less-dispersed latent health distribution than 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 and its midwestern neighbors, Kansas (KS), Nebraska (NE), Iowa (IA), and Illinois (IL). The numbers show the posterior probability (%), and an asterisk indicates that the relationship holds in the sample. As in any application with sample sizes that are not enormous, it is important to consider uncertainty, not just whether a relationship holds in the sample or not. The posterior probabilities add such detail. There are cases where SD1 holds in the sample, as in IL SD1 NE, but the posterior probability is only 18%, so we should not feel too confident in the SD1 conclusion. Conversely, there are cases where SD1 or SC does not hold in the sample, but we should not necessarily reject it, as with IL SC NE, which does not hold in the sample but has 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; and extensions to settings like regression.

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.

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

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

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

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

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

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 Mathematical 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`