Comparing Latent Inequality with Ordinal Data

David M. Kaplan* Longhao Zhuo

December 15, 2020

Abstract

We consider comparing two latent distributions when only ordinal data are available. Distinct from the literature, we allow a continuous latent distribution without imposing a parametric model. Primarily, we contribute identification results: if two known ordinal distributions' relationship satisfies certain properties, then the corresponding latent distributions' relationship must satisfy certain other properties related to inequality. We consider both between-group inequality (better/worse relationship) and within-group inequality (dispersion). These results also apply to conditional distributions. Secondarily, we discuss Bayesian and frequentist inference on the relevant ordinal relationships, which are various combinations of inequalities. Simulations and empirical examples illustrate our contributions.

JEL classification: C25,

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

1 Introduction

Our results help compare latent distributions when only ordinal data are available. As in most ordered choice models, we assume the ordinal variable's value depends on the latent variable's value relative to a set of thresholds. Ordinal examples include measures of health, bond ratings, political indices, subjective well-being, consumer confidence, and public school ratings.

We consider two types of inequality: within-group and between-group. Within-group inequality means dispersion, often quantified by interquantile ranges (differences between two quantiles). For example, to study whether "income inequality" in the U.S. has increased over time, the 90–10 interquantile range has been used to measure dispersion within the

*Corresponding author, kaplandm@missouri.edu. Many thanks to Alyssa Carlson, Denis Chetverikov, Yixiao Jiang, Jia Li, Matt Masten, Arnaud Maurel, Zack Miller, Peter Mueser, Shawn Ni, Adam Rosen, Dongchu Sun, and other participants from UConn, Duke, Yale, the 2018 Midwest Econometrics Group, and the 2019 Chinese Economists Society conference for helpful questions, comments, and references. Thanks also to the Cowles Foundation for their hospitality during part of this work.

income distribution in a given year. Between-group inequality means whether one group is better or worse than another. For example, "racial inequality" in health means one racial group tends to have better health than another.

We show how certain pairs of ordinal distributions provide evidence of latent within-group or between-group inequality. For within-group inequality (dispersion): if the ordinal CDFs cross, then certain latent interquantile ranges are larger for one distribution, even if one group's thresholds are shifted (by a constant) from the other group's thresholds. For between-group inequality (better/worse): assuming the groups have the same thresholds, certain ordinal CDF pairs imply latent "restricted stochastic dominance" in the sense of Atkinson (1987, Cond. I, p. 751), and more generally ordinal CDF differences imply certain quantiles are higher in one latent distribution. All our results are robust to arbitrary (increasing) transformations of the latent random variables.

These results can be interpreted in terms of partial identification. A pair of ordinal CDFs does not uniquely identify a pair of latent CDFs; there are an infinite number of latent CDF pairs (along with thresholds) in the identified set. Still, sometimes all such latent CDF pairs possess a particular property, such as restricted stochastic dominance or certain interquantile ranges being larger. Although we consider two CDFs that separately are not even partially identified (due to the unknown thresholds), our within-group inequality approach is similar in spirit to that of Stoye (2010), who derives bounds for dispersion parameters based on the "most compressed" and "most dispersed" CDFs in the identified set.

Distinct from the ordinal inequality literature (e.g., as surveyed by Jenkins, 2019, 2020), we allow a continuous (not discrete) latent distribution without imposing a parametric model. Continuous is more realistic than discrete because it allows latent differences within the same ordinal category: one person with "good" health can be healthier than another, one A-rated bond can have higher credit worthiness than another, one "pretty happy" person may be happier than another, etc. (Our results still apply to discrete or mixed latent distributions.) A discrete latent distribution is implicit in many of the proposed ordinal inequality indexes and explicit in the justification of the median-preserving spread of Allison and Foster (2004), which Madden (2014) calls "the breakthrough in analyzing inequality with [ordinal] data" (p. 206). Bond and Lang (2019, §II(B)) write, "Happiness researchers almost universally assume either that the ordered responses are measured on a discrete interval scale or that each group's latent happiness distribution is normal (i.e., ordered probit) or logistic (i.e., ordered logit)" (p. 1634). Such parametric models are unrealistic and yield fragile results. For example, Bond and Lang (2019) highlight several empirical happiness studies in which "parametric results are reversed using plausible transformations" (p. 1629).

More generally, Bond and Lang (2019) stress the near impossibility of comparing latent means in our framework (continuous, nonparametric). They provide conditions in Section II(A) that imply latent first-order stochastic dominance, but they note such conditions never hold in practice. Instead of latent means, we focus on latent quantiles, latent restricted stochastic dominance, and latent interquantile ranges, so we do not require unrealistic conditions.

With covariates, conditional distributions can be compared pairwise. This complements the latent median regression model of Chen, Oparina, Powdthavee, and Srisuma (2019).

We provide methods and code for both Bayesian and frequentist inference on the salient ordinal CDF conditions. We consider both because they can give very different conclusions with these general types of inequality conditions (Kaplan and Zhuo, forthcoming).

Our empirical examples show cases in which our results indicate evidence of latent inequality, even though latent means cannot be compared nonparametrically and latent medians are not informative. We interpret estimated differences as well as both frequentist and Bayesian inference.

Section 2 contains identification results. Section 3 describes frequentist and Bayesian inference. Section 4 discusses conditional distributions. Section 5 provides empirical illustrations. Appendix A has simulations comparing frequentist and Bayesian inference.

Acronyms used include those for cumulative distribution function (CDF), interquantile range (IQR), first-order stochastic dominance (SD1), restricted stochastic dominance (RSD), single crossing (SC), refined moment selection (RMS), and intersection–union test (IUT).

2 Identification of latent relationships

Assumption A1 describes the formal setting and notation. As in many ordered choice models, the ordinal random variables X and Y are derived from corresponding latent random variables X^* and Y^* using fixed (non-random) thresholds.

Assumption A1. The observable, ordinal random variables X and Y are derived from latent random variables X^* and Y^*. The J ordinal categories are denoted $1, 2, \ldots, J$. (These are category labels, not cardinal values.) 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 $F_X(j) = F_X^*(\gamma_j)$, where $F_X^*(\cdot)$ is the CDF of X^*. 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 $F_Y(j) = F_Y^*(\gamma_j + \Delta_\gamma)$, where $F_Y^*(\cdot)$ is the CDF of Y^*.

Figure 1 visualizes Assumption A1. It shows how the ordinal CDF values are particular points on the latent CDFs. Only $F_X(1)$, $F_X(2)$, $F_Y(1)$, and $F_Y(2)$ are observable; γ_1, γ_2, and

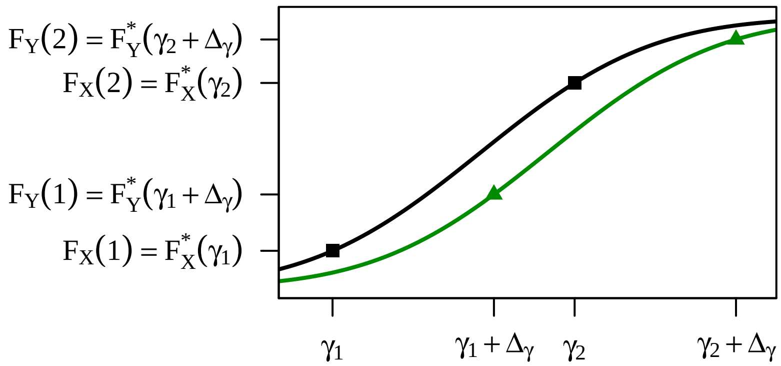

Figure 1: Example latent CDFs (lines) and ordinal CDFs (shapes).

Δ_γ are unknown.

The biggest restriction in Assumption A1 is that the thresholds for determining Y cannot differ arbitrarily from those determining X: they may differ, but only by a constant that does not depend on the ordinal category. This may be reasonable in some applications but not others. For example, for health, Lindeboom and van Doorslaer (2004) and Hernández-Quevedo, Jones, and Rice (2005) find evidence of a mix of "homogeneous reporting" ($\Delta_\gamma = 0$ in A1), "index shift" ($\Delta_\gamma \neq 0$ in A1), and "cut-point shift" (A1 violated), depending on the comparison groups; Lindeboom and van Doorslaer (2004) write, "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).

Many additional results can follow from assuming X^* and Y^* belong to the same (unknown) location–scale family,[1] but even though such an assumption is much weaker than a probit or logit model, it is not robust to transformations. That is, if $t(\cdot)$ is an increasing, nonlinear function, then $t(X^*)$ and $t(Y^*)$ are generally not in the same location–scale family even if X^* and Y^* are. Given the nature of the latent variables, it seems prudent to desire results robust to such transformations.

2.1 Within-group inequality

Figure 2 illustrates intuition for how an ordinal CDF crossing implies a relationship between certain latent interquantile ranges (IQRs). The black line shows the latent CDF of X^*, with the black squares showing $F_X^*(\gamma_1) = F_X(1)$ and $F_X^*(\gamma_2) = F_X(2)$. The green line shows the latent CDF of Y^*, with the green triangles showing $F_Y^*(\gamma_1) = F_Y(1)$ and $F_Y^*(\gamma_2) = F_Y(2)$, with $\Delta_\gamma = 0$ for simplicity. If instead $\Delta_\gamma \neq 0$, then the difference between thresholds still

[1]See the 2019 version of this paper.

remains $(\gamma_2 + \Delta_\gamma) - (\gamma_1 + \Delta_\gamma) = \gamma_2 - \gamma_1$, so the following intuition is unchanged.

In Figure 2 (with a strictly increasing CDF), the difference between the $F_Y(2)$-quantile and $F_Y(1)$-quantile of Y^* is

$$Q^*_Y(F_Y(2)) - Q^*_Y(F_Y(1)) = (\gamma_2 + \Delta_\gamma) - (\gamma_1 + \Delta_\gamma) = \gamma_2 - \gamma_1.$$

Because $F_X(1) < F_Y(1)$ and $F_X(2) > F_Y(2)$, the corresponding IQR for X^* must be smaller. That is, the black $F^*_X(\cdot)$ reaches the value $F_Y(1)$ to the right of γ_1, but reaches $F_Y(2)$ to the left of γ_2, so

$$\overbrace{Q^*_X(F_Y(2))}^{<\gamma_2} - \overbrace{Q^*_X(F_Y(1))}^{>\gamma_1} < \gamma_2 - \gamma_1 = Q^*_Y(F_Y(2)) - Q^*_Y(F_Y(1)).$$

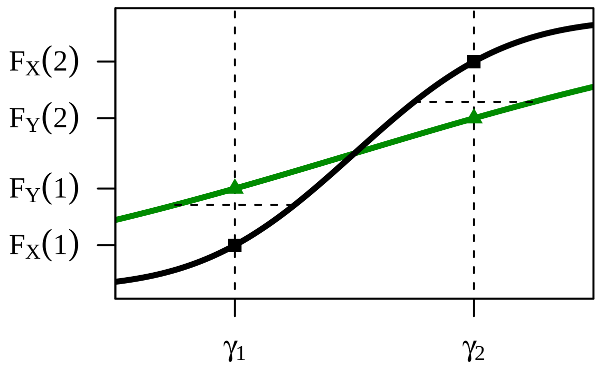

Figure 2: Illustration of Theorem 1 (with $\Delta_\gamma = 0$ for simplicity).

Further, consider any quantile indices τ_1 and τ_2 satisfying $F_X(1) < \tau_1 \le F_Y(1)$ and $F_Y(2) < \tau_2 \le F_X(2)$. As seen in Figure 2, the Y^* IQR is larger than $\gamma_2 - \gamma_1$ whereas the X^* IQR is smaller:

$$Q^*_Y(\tau_2) - Q^*_Y(\tau_1) > \gamma_2 - \gamma_1 > Q^*_X(\tau_2) - Q^*_X(\tau_1).$$

Theorem 1 formalizes and generalizes these arguments.

Theorem 1. *Let Assumption A1 hold. Assume a single crossing of the ordinal CDFs at category m ($1 \le m \le J-1$): $F_X(j) < F_Y(j)$ for $1 \le j \le m$ and $F_X(j) > F_Y(j)$ for $m < j \le J-1$. Let $Q^*_X(\cdot)$ and $Q^*_Y(\cdot)$ denote the quantile functions of X^* and Y^*, respectively. Then, $Q^*_X(\tau_2) - Q^*_X(\tau_1) < Q^*_Y(\tau_2) - Q^*_Y(\tau_1)$ for any combination of $\tau_2 \in \mathcal{T}_2$ and $\tau_1 \in \mathcal{T}_1$, where*

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

Proof. For any $\tau_1 \in \mathcal{T}_1$ and $\tau_2 \in \mathcal{T}_2$, there exist categories $j < k$ with $F_X(j) < \tau_1 \le F_Y(j)$

and $F_Y(k) < \tau_2 \leq F_X(k)$. From this and Assumption A1,

$$\begin{aligned}\tau_1 \leq F_Y(j) = F_Y^*(\gamma_j + \Delta_\gamma) &\implies Q_Y^*(\tau_1) \leq \gamma_j + \Delta_\gamma, \\ \tau_2 > F_Y(k) = F_Y^*(\gamma_k + \Delta_\gamma) &\implies Q_Y^*(\tau_2) > \gamma_k + \Delta_\gamma,\end{aligned}$$

so

$$Q_Y^*(\tau_2) - Q_Y^*(\tau_1) > (\gamma_k + \Delta_\gamma) - (\gamma_j + \Delta_\gamma) = \gamma_k - \gamma_j.$$

Similarly,

$$\begin{aligned}\tau_1 > F_X(j) = F_X^*(\gamma_j) &\implies Q_X^*(\tau_1) > \gamma_j, \\ \tau_2 \leq F_X(k) = F_X^*(\gamma_k) &\implies Q_X^*(\tau_2) \leq \gamma_k,\end{aligned}$$

so

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

Altogether,

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

□

Theorem 1 interprets the median-preserving spread of Allison and Foster (2004) when the latent distributions are continuous. Allison and Foster (2004) assume the latent distribution is discrete, with J possible values, and show the median-preserving spread implies second-order stochastic dominance. Because the median-preserving spread implies a single ordinal CDF crossing, our Theorem 1 applies. That is, with a continuous latent distribution, the median-preserving spread has a weaker interpretation but still provides evidence of larger latent dispersion.

2.2 Between-group inequality

For between-group inequality (better/worse), we consider quantiles and restricted stochastic dominance. Bond and Lang (2019) show full first-order stochastic dominance is essentially impossible to establish; quantiles and restricted stochastic dominance provide tractable alternatives that still provide evidence of between-group inequality.

Having a larger latent τ-quantile provides some evidence of being "better." Besides the evident intuition, this can also be interpreted in terms of quantile utility maximization (e.g., de Castro and Galvao, 2019; Manski, 1988; Rostek, 2010). For example, given strictly increasing utility function $u(\cdot)$, a τ-quantile utility maximizer strictly prefers X^* over Y^* if the τ-quantile of $u(X^*)$ is strictly greater than the τ-quantile of $u(Y^*)$, which is true if and only if $Q_X^*(\tau) > Q_Y^*(\tau)$. That is, learning $Q_X^*(\tau) > Q_Y^*(\tau)$ is equivalent to learning that X^* is preferred by all τ-quantile utility maximizers, regardless of utility function.

Restricted stochastic dominance (RSD) also provides evidence of being "better." The RSD concept in Definition 1 comes from Condition I of Atkinson (1987, p. 751) in the context of poverty. If X^* and Y^* were consumption or income, and v were a poverty line, "headcount poverty" is $F_X^*(v)$ and $F_Y^*(v)$. RSD on $\mathcal{V}$ means X^* is preferred to Y^* in terms of lower poverty headcount for all $v \in \mathcal{V}$. RSD on $\mathcal{V}$ can also be interpreted as X^* having a higher probability than Y^* of exceeding v for all $v \in \mathcal{V}$.

Definition 1 (latent restricted stochastic dominance)**.** There is "restricted stochastic dominance" (RSD) on $\mathcal{V}$ of X^* over Y^* iff $F_X^*(v) \leq F_Y^*(v)$ for all $v \in \mathcal{V}$.

Unlike the within-group inequality results, between-group inequality analysis requires $\Delta_\gamma = 0$. Whereas dispersion is invariant to pure location shifts, quantiles and RSD are not, so having the same thresholds γ_j for X and Y becomes critical.

The quantile intuition extends the known conclusion that latent medians can be compared if $\Delta_\gamma = 0$ and the median category differs. That is, if the median category of ordinal X is above that of Y, then the latent median of X^* is above that of Y^*. Specifically, if γ_j is the threshold between the two categories, then the median of Y^* is below γ_j whereas the median of X^* is above. The median is the 0.5-quantile, and this intuition extends to other quantiles.

RSD requires a stronger ordinal condition than quantile ranking. For intuition, consider $F_X^*(r) = F_Y^*(r)$ for all $|r - \gamma_j| > \epsilon$, for small $\epsilon > 0$. That is, they are identical except right around γ_j; moreover, $F_Y^*(\cdot)$ increases steeply from just before γ_j while $F_X^*(\cdot)$ increases steeply just after γ_j. Thus, X^* has a larger τ-quantile for the large range of τ satisfying $F_X^*(\gamma_j) < \tau \leq F_Y^*(\gamma_j)$, but there is only RSD on the very small interval $[\gamma_j - \epsilon, \gamma_j + \epsilon]$, which further is impossible to learn from ordinal data.

RSD can be inferred if the ordinal CDF of X is enough below that of Y, specifically $F_X(j+1) < F_Y(j)$ for at least one j. This is a weaker version of the condition from Bond and Lang (2019) that implies latent first-order stochastic dominance. However, their condition is so strong that they argue it practically never occurs because it requires at minimum $\mathrm{P}(Y = J) = \mathrm{P}(X = 1) = 0$. Our condition drops this requirement, at the expense of learning about the tails.

Theorems 2 and 3 formally state results for quantiles and RSD, respectively.

Theorem 2. *Let Assumption A1 hold with* $\Delta_\gamma = 0$*. Let*

$$\mathcal{T}_X \equiv \bigcup_{j=1}^{J-1} (F_X(j), F_Y(j)]\, \mathbb{1}\{F_X(j) < F_Y(j)\},$$
$$\mathcal{T}_Y \equiv \bigcup_{j=1}^{J-1} (F_Y(j), F_X(j)]\, \mathbb{1}\{F_Y(j) < F_X(j)\}.$$

Then, $Q_X^*(\tau) > Q_Y^*(\tau)$ for all $\tau \in \mathcal{T}_X$, and $Q_Y^*(\tau) > Q_X^*(\tau)$ for all $\tau \in \mathcal{T}_Y$.

Proof. Consider any $\tau \in \mathcal{T}_X$, so $F_X(j) < \tau \leq F_Y(j)$ for some j. By A1 with $\Delta_\gamma = 0$, this is equivalent to $F_X^*(\gamma_j) < \tau \leq F_Y^*(\gamma_j)$. This implies $Q_Y^*(\tau) \leq \gamma_j < Q_X^*(\tau)$, i.e., $Q_X^*(\tau) > Q_Y^*(\tau)$. Switching the X and Y labels yields the result for $\mathcal{T}_Y$. $\square$

Theorem 3. *Let Assumption A1 hold with $\Delta_\gamma = 0$. Let $F_X(0) = F_Y(0) = 0$. Let*

$$\mathcal{V}_X \equiv \bigcup_{j=1}^{J} [\gamma_{j-1}, \gamma_j]\, \mathbb{1}\{F_X(j) \leq F_Y(j-1)\},$$
$$\mathcal{V}_Y \equiv \bigcup_{j=1}^{J} [\gamma_{j-1}, \gamma_j]\, \mathbb{1}\{F_Y(j) \leq F_X(j-1)\}.$$

Then, by Definition 1, there is restricted stochastic dominance on $\mathcal{V}_X$ of X^* over Y^*, and restricted stochastic dominance on $\mathcal{V}_Y$ of Y^* over X^*.

Proof. By Assumption A1 with $\Delta_\gamma = 0$, $F_X(j) = F_X^*(\gamma_j)$ and $F_Y(j-1) = F_Y^*(\gamma_{j-1})$. Because CDFs are non-decreasing, for any $v \in [\gamma_{j-1}, \gamma_j]$, $F_X^*(v) \leq F_X^*(\gamma_j)$ and $F_Y^*(\gamma_{j-1}) \leq F_Y^*(v)$. Thus, for any $v \in \mathcal{V}_X$, there exists j such that $\gamma_{j-1} \leq v \leq \gamma_j$ and

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

Switching the X and Y labels yields the result for $\mathcal{V}_Y$. $\square$

The following corollary states the quantile and RSD interpretations when there is particularly strong evidence of between-group inequality.

Corollary 4. *Let Assumption A1 hold with $\Delta_\gamma = 0$. If $F_X(j) < F_Y(j-1)$ for all $j = 2, \ldots, J-1$, then there is RSD on $[\gamma_1, \gamma_{J-1}]$ of X^* over Y^*, and $Q_X^*(\tau) > Q_Y^*(\tau)$ for all $F_X(1) < \tau \leq F_Y(J-1)$.*

Proof. First, apply Theorem 2. The intervals $(F_X(j), F_Y(j)]$ in $\mathcal{T}_X$ in Theorem 2 all overlap here because $F_X(j) < F_Y(j-1)$ for all $j = 2, \ldots, J-1$, and $\mathbb{1}\{F_X(j) < F_Y(j)\} = 1$ for all $j = 1, \ldots, J-1$ because $F_X(j) < F_Y(j-1) \leq F_Y(j)$. Thus, $\mathcal{T}_X = (F_X(1), F_Y(J-1)]$, whereas $\mathcal{T}_Y$ is empty.

Second, apply Theorem 3; $\mathcal{V}_Y$ is empty, whereas $\mathcal{V}_X = \cup_{j=2}^{J-1} [\gamma_{j-1}, \gamma_j] = [\gamma_1, \gamma_{J-1}]$ because $\mathbb{1}\{F_X(j) < F_Y(j-1)\} = 1$ for $j = 2, \ldots, J-1$. $\square$

3 Statistical inference on ordinal relationships

We describe frequentist hypothesis testing and Bayesian posterior probabilities for some of the salient ordinal distribution relationships from Section 2, given a sample of data from each of two ordinal populations.[2] We characterize the relationships as combinations of inequalities and then describe frequentist and Bayesian methods, which are then compared.

3.1 Relationships of interest

Notationally, let

$$\mathbf{F} \equiv (F_X(1), \ldots, F_X(J-1), F_Y(1), \ldots, F_Y(J-1))' \tag{2}$$

gather all relevant ordinal CDF values, and let

$$\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_{J-1}) \text{ with } \theta_j \equiv F_X(j) - F_Y(j) \tag{3}$$

gather the ordinal CDF differences, which can simplify notation.

Below, we characterize the subsets of the parameter space of $\mathbf{F}$ or $\boldsymbol{\theta}$ where various ordinal relationships hold, as well as the complements of these subsets where the relationships do not hold. These involve intersections and/or unions of more basic subsets. These characterizations help us in later sections to evaluate posterior probabilities, formulate frequentist hypothesis tests, and compare Bayesian and frequentist inference.

3.1.1 Between-group inequality

Ordinal first-order stochastic dominance $X \text{ SD}_1\ Y$ is equivalent to:

$$X \text{ SD}_1\ Y \iff \theta_j \le 0 \text{ for all } j = 1, \ldots, J-1 \iff \boldsymbol{\theta} \in \bigcap_{j=1}^{J-1} \{\boldsymbol{\theta} : \theta_j \le 0\}. \tag{4}$$

The opposite of (4) is

$$X\ non\text{SD}_1\ Y \iff \theta_j > 0 \text{ for some } j = 1, \ldots, J-1 \iff \boldsymbol{\theta} \in \bigcup_{j=1}^{J-1} \{\boldsymbol{\theta} : \theta_j > 0\}. \tag{5}$$

Ordinal SD1 has two interpretations; all assume $\Delta_\gamma = 0$. First, latent SD1 implies ordinal SD1, so rejecting ordinal SD1 implies rejecting latent SD1; i.e., ordinal SD1 is a testable implication of latent SD1. Second, by Theorem 2, $X \text{ SD}_1\ Y$ implies $\mathcal{T}_Y$ is empty

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

because $\mathbb{1}\{F_Y(j) < F_X(j)\} = 0$ for all j, whereas $\mathcal{X}$ may be non-empty, providing evidence of the latent X^* being better than Y^*.

Non-SD1 is considered because from a frequentist perspective, rejecting a null hypothesis of non-SD1 in favor of SD1 is stronger evidence of SD1 than non-rejection of a null of SD1 (e.g., Davidson and Duclos, 2013, p. 87).

A stronger form of ordinal SD1 is the premise of Corollary 4. Mathematically, it is the same as before after redefining $\theta_j \equiv F_X(j+1) - F_Y(j)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{J-2})'$. Then, as in (4) and (5), the relationship is satisfied in the intersection of the half-spaces $\{\boldsymbol{\theta} : \theta_j \leq 0\}$, the complement of which is the union of half-spaces $\{\boldsymbol{\theta} : \theta_j > 0\}$.

3.1.2 Within-group inequality

We consider an ordinal distribution relationship that suggests the latent Y^* is more dispersed (more within-group inequality) than X^*, as in Theorem 1. The premise of Theorem 1 is an ordinal CDF "single crossing" (SC) at category m ($1 \leq m \leq J-1$) with $F_X(j) < F_Y(j)$ for $j \leq m$ and $F_X(j) > F_Y(j)$ for $j > m$. That is, $\theta_j < 0$ for $j \leq m$ and $\theta_j > 0$ for $j > m$, which can be combined as $(2\,\mathbb{1}\{j \leq m\} - 1)\theta_j < 0$ for all $j = 1, \ldots, J-1$. SC holds if these inequalities hold jointly for any value of m between 1 and $J-2$, inclusive. With k representing possible values of m, SC is

$$X \text{ SC } Y \iff \boldsymbol{\theta} \in \bigcup_{k=1}^{J-2} \bigcap_{j=1}^{J-1} \{\boldsymbol{\theta} : (2\,\mathbb{1}\{j \leq k\} - 1)\theta_j < 0\}. \tag{6}$$

As with non-SD1, non-SC is simply the opposite, which then corresponds to an intersection of unions:

$$X \; non\text{SC } Y \iff \boldsymbol{\theta} \in \bigcap_{k=1}^{J-2} \bigcup_{j=1}^{J-1} \{\boldsymbol{\theta} : (2\,\mathbb{1}\{j \leq k\} - 1)\theta_j \geq 0\}. \tag{7}$$

3.2 Bayesian inference

Bayesian inference on the various ordinal relationships is straightforward. The posterior probability of any ordinal relationship can be calculated from the posterior distribution of the two ordinal CDFs. Further, this can be computed simultaneously for multiple relationships, and the result is "coherent" (following the basic laws of probability).

With independent, iid samples, the Dirichlet–multinomial model can be used. Gunawan, Griffiths, and Chotikapanich (2018) discuss Bayesian inference for ordinal SD1 with this model using the improper prior as in their (7), as well as inference for the median-preserving spread of Allison and Foster (2004).

With non-iid sampling, other Bayesian approaches can be used, like the nonparametric Bayesian approach to complex sampling design from Dong, Elliott, and Raghunathan (2014).

In the absence of prior information, there remains debate about the most "objective" prior to use. For example, one could use the improper prior like Gunawan, Griffiths, and Chotikapanich (2018), or the uniform prior, or a uniform prior adjusted to have 1/2 prior probability that a particular ordinal relationship holds and 1/2 prior probability that it does not, following equations (7)–(9) of Goutis, Casella, and Wells (1996). However, such debates are well beyond the scope of this paper.

3.3 Frequentist hypothesis testing

We describe frequentist tests of the possible null hypotheses from Section 3.1, assuming independent, iid ordinal samples of X_i $(i = 1, \ldots, n_X)$ and Y_i $(i = 1, \ldots, n_Y)$, which implies asymptotically normal estimators are available for $\mathbf{F}$ from (2) and $\boldsymbol{\theta}$ from (3).

3.3.1 Null hypothesis: ordinal SD1

Consider testing the null hypothesis $H_0\colon X \text{ SD}_1\, Y$, i.e., that (4) holds.

A Bonferroni approach is easy but conservative. For $j = 1, \ldots, J-1$, hypothesis $H_{0j}\colon \theta_j \le 0$ is tested at level $\alpha/(J-1)$, and the overall H_0 is rejected if any H_{0j} is rejected. The Bonferroni test controls size:

$$\mathrm{P}(\text{reject any } H_{0j} \mid H_0 \text{ true}) \le \sum_{j=1}^{J-1} \mathrm{P}(\text{reject } H_{0j} \mid H_0 \text{ true}) \le \sum_{j=1}^{J-1} \frac{\alpha}{J-1} = \alpha.$$

This relates to union–intersection tests (e.g., Casella and Berger, 2002, §8.2.3,8.3.3).

Recent methods like those of Andrews and Barwick (2012) and Romano, Shaikh, and Wolf (2014) improve power. Roughly speaking, instead of setting the critical value based on the least favorable configuration, they focus on the inequalities that seem "close" to binding in the sample. 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 approach would use individual level $\alpha/(J-2)$ instead of $\alpha/(J-1)$ for testing each H_{0j}.

Implementation of any test above requires the sampling distribution of estimator $\hat{\boldsymbol{\theta}}$. Letting $\hat{F}_X(j) \equiv n_X^{-1} \sum_{i=1}^{n_X} \mathbb{1}\{X_i \le j\}$,

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

and similarly for the vector of the $\hat{F}_Y(j)$. Because the X and Y samples are assumed independent, the corresponding CDF estimators have zero covariance. Thus, assuming $n_X/n_Y \to \delta \in (0,\infty)$,

$$\sqrt{n_X}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\theta), \quad \underline{\boldsymbol{\Sigma}}_\theta \equiv \underline{\boldsymbol{\Sigma}}_X + \delta \underline{\boldsymbol{\Sigma}}_Y. \tag{8}$$

3.3.2 Null hypothesis: non-SD1

Consider testing the null hypothesis $H_0\colon X$ *non*SD_1 Y, i.e., that (5) holds. Now, X SD_1 Y is the alternative hypothesis. Rejection of non-SD1 in favor of SD1 provides stronger evidence than non-rejection of SD1 because a false rejection of non-SD1 is a type I error whose rate is controlled at level α, whereas false non-rejection of SD1 is a type II error whose rate is not explicitly controlled.

The intersection–union test of $H_0\colon X$ *non*SD_1 Y 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 each H_{0j}. If each H_{0j} test has size α, then the overall test has size α:

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

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: single crossing

Consider testing the null hypothesis $H_0\colon X$ SC Y. Rewriting (6),

$$H_0\colon \boldsymbol{\theta} \in \Theta_0, \quad \Theta_0 \equiv \bigcup_{k=1}^{J-2} \Theta_k, \quad \Theta_k \equiv \big\{\boldsymbol{\theta} : (2\,\mathbb{1}\{j \le k\} - 1)\theta_j < 0,\ j = 1, \ldots, J-1\big\}. \tag{9}$$

This H_0 can be tested by combining the intersection–union approach of Section 3.3.2 with a method from Section 3.3.1. Each $H_{0k}\colon \boldsymbol{\theta} \in \Theta_k$ is equivalent to $H_{0k}\colon \underline{\mathbf{D}}\boldsymbol{\theta} < 0$ for diagonal matrix $\underline{\mathbf{D}}$ with elements $D_{jj} = 1$ for $j = 1, \ldots, k$ and $D_{jj} = -1$ for $j = k+1, \ldots, J-1$ (and $D_{jk} = 0$ if $j \ne k$). Applying the continuous mapping theorem to (8),

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

so $\underline{\mathbf{D}}\hat{\boldsymbol{\theta}}$ can be used to test $\underline{\mathbf{D}}\boldsymbol{\theta} < \mathbf{0}$ with the methods referenced in Section 3.3.1. That is, to test $H_0\colon X \text{ SC } Y$ at level α, there are two steps:[3]

1. For $k = 1, \dots, J-2$, using (9), test $H_{0k}\colon \boldsymbol{\theta} \in \Theta_k$ 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_{0k} are rejected.

3.4 Bayesian and frequentist differences

3.4.1 Advantages and disadvantages

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 with iid data using the Dirichlet–multinomial model. The Dirichlet posterior can be sampled from directly (without MCMC), 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 beliefs about different possible relationships given the data. Third, the posterior probabilities of different possible ordinal relationships are "coherent," meaning they obey the usual probability laws. For example, 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, if a binary decision is required (not just posterior probabilities), the Bayes decision rule makes the loss function transparent and explicit, and it is easy to use a variety of loss functions. Further, assuming finite loss, any Bayes rule is admissible.

The Bayesian approach is most often criticized for parametric likelihoods, subjective priors, and slow computation. However, this setting does not involve a latent parametric model, an objective prior can be used, and computation is fast (without MCMC).

3.4.2 Large-sample disagreement

Although frequentist confidence sets and Bayesian credible sets for $\boldsymbol{\theta}$ tend to agree in large samples, frequentist and Bayesian assessments of the evidence for or against SD1 or SC can differ. As a special case of more general results, SC and non-SC are discussed in detail

[3] This is essentially the same approach suggested in Remark 5.3 of Machado, Shaikh, and Vytlacil (2019).

by Kaplan and Zhuo (forthcoming, §5.1). Specifically, consider the Bayes rule using the "generalized 0–1 loss" (e.g., Casella and Berger, 2002, eqn. (8.3.11)) with value $1-\alpha$ for type I error, α for type II error, and zero otherwise. This loss function is arguably implicit in frequentist hypothesis testing (Kaplan and Zhuo, forthcoming, §2.1). Minimizing posterior expected loss leads to "rejection" of H_0 if and only if the posterior probability of H_0 is below α, i.e., it treats the posterior probability like a p-value.

Kaplan and Zhuo (forthcoming, Thm. 1 and §4.5) characterize the frequentist type I error rates and size of such a Bayesian test applied to $H_0\colon \boldsymbol{\theta} \in \Theta_0$ for arbitrary set Θ_0. If Θ_0 is a half-space of the parameter space, then (asymptotically) the Bayes rule is also a valid frequentist test. If Θ_0 is convex, then the Bayes rule's asymptotic size is above α. Otherwise, it depends on the sample size and the local shape of Θ_0 near the true value; non-convexity can result in type I error rates well below α, as with DGP 1 and H_0: *non*SC in Table 1 of Kaplan and Zhuo (forthcoming).

These results can be applied more specifically to our ordinal hypotheses. If only a single inequality is near binding, then the test essentially reduces to a one-dimensional one-sided test, in which case the Bayes rule's frequentist asymptotic size is α. If multiple inequalities are involved with $H_0\colon X \text{ SD}_1\ Y$, then Θ_0 is convex, so the Bayes rule's size is above α; it is not conservative enough by conventional frequentist standards (or, frequentist testing is too conservative by Bayesian standards). This also matches the results of Kline (2011). However, if instead $H_0\colon X\ non\text{SD}_1\ Y$, then there is non-convexity that can make the Bayes rule too conservative by frequentist standards. For $H_0\colon X \text{ SC } Y$, Θ_0 is a subset of a half-space, but it has some locally non-convex regions. This makes the Bayes rule tend to have size above α in smaller samples, but it can have size well above or below α in large samples depending on whether Θ_0 is locally convex or non-convex; see Section 5.1.2 of Kaplan and Zhuo (forthcoming).

The simulation results in Appendix A illustrate some of these points.

4 Extension to conditional distributions

We briefly discuss extension 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 $\mathbf{z}_2$. Often the conditional ordinal distributions can be consistently estimated nonparametrically. For example, kernel regression of $\mathbb{1}\{Y \leq j\}$ 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. Bayesian inference follows from the posterior over the two conditional ordinal distributions, and frequentist inference follows from asymptotic normality of the estimators.

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 $\mathbf{z}_0$, and consider the set of $\mathbf{z}$ for which a particular relationship holds. This set is identified and can be estimated. For example, we could estimate the set of $\mathbf{z}$ such that the ordinal CDF of $Y \mid \mathbf{Z} = \mathbf{z}$ has a single crossing of the benchmark CDF of $Y \mid \mathbf{Z} = \mathbf{z}_0$: $\{\mathbf{z} : \hat{F}_{Y|\mathbf{Z}}(\cdot \mid \mathbf{Z} = \mathbf{z}) \text{ SC } \hat{F}_{Y|\mathbf{Z}}(\cdot \mid \mathbf{Z} = \mathbf{z}_0)\}$.

5 Empirical illustrations

The following empirical examples illustrate our theoretical results. Data is publicly available and/or provided alongside the replication code on Kaplan's website.[4] Code is in R (R Core Team, 2020), with help from packages `readstata13` (Garbuszus and Jeworutzki, 2018) and `quadprog` (Weingessel, 2019). For between-group inequality, we assume $\Delta_\gamma = 0$ throughout.

For estimation, we use sampling weights (when provided) to compute appropriately weighted empirical ordinal CDFs, and then we interpret the differences in terms of latent quantiles and latent restricted stochastic dominance (RSD) using Theorems 2 and 3.

For inference on ordinal first-order stochastic dominance (SD1), we use both frequentist and Bayesian methods. The Bayesian results use the posterior probability of ordinal SD1 from a Dirichlet–multinomial model with uniform prior. For the null of SD1, we use the refined moment selection (RMS) test of Andrews and Barwick (2012) as described in Section 3.3.1. For the null of non-SD1, we use the intersection–union test (IUT) as described in Section 3.3.2. Given some level like $\alpha = 0.05$, we say there is "statistically significant" evidence in favor of $X \text{ SD}_1\ Y$ if the posterior probability is above $1 - \alpha$ (Bayesian) or the IUT rejects $H_0\colon X\ non\text{SD}_1\ Y$ at level α (frequentist), and evidence against $X \text{ SD}_1\ Y$ if the posterior probability is below α (Bayesian) or RMS rejects $H_0\colon X \text{ SD}_1\ Y$ at level α (frequentist). For inference, we use sampling weights when provided but (though not ideal)

[4] `https://faculty.missouri.edu/kaplandm`

otherwise treat sampling as iid. We use $\alpha = 0.05$ unless otherwise noted.

5.1 Happiness

We revisit examples from the online empirical appendix of Bond and Lang (2019), using the data they provide.

5.1.1 Declining relative female happiness in U.S.

We use the 1972–2006 General Social Survey (GSS) data analyzed in Supplemental Appendix A.3.7 of Bond and Lang (2019). They first note that no latent means can be compared nonparametrically. To illustrate this further, their Figure A-10 shows that although an ordered probit model finds that females' latent mean happiness has declined over 1972–2006 relative to males', the opposite result can be found by specifying a parametric latent distribution with enough left-skewness.

Although the latent mean results are not robust, we find some alternative evidence of declining relative female happiness in the U.S.

Table 1 shows the empirical results. Within each year (row), the female and male ordinal happiness CDFs are shown, as well as whether there is first-order stochastic dominance of female over male ($F\ \mathrm{SD}_1\ M$) or male over female ($M\ \mathrm{SD}_1\ F$). (There are only 3 categories, so the CDFs must have $F(3) = 1$, which is not shown.) Bayesian and frequentist inference results are also shown. If there is SD1 in the sample ("yes"), then + indicates posterior probability of SD1 above $1 - \alpha$, and * indicates the IUT rejecting non-SD1 at level α. If there is not SD1 in the sample ("no"), then + indicates posterior probability of SD1 below α, and * indicates RMS rejecting SD1 at level α. In accord with the theoretical results of Kaplan and Zhuo (forthcoming), we sometimes see no+ and no+* but never no*.

Table 1 shows some evidence in favor of female happiness declining relative to male happiness over 1972–2006 in the U.S. In the first half of the table, the female ordinal sample distribution is dominant (SD1) seven times, and female dominance is never rejected by either Bayesian or frequentist measures; this remains true even with $\alpha = 0.1$ (not shown). In contrast, the male sample distribution is dominant only once, and its dominance can be rejected multiple times. Because latent SD1 implies ordinal SD1, rejecting male ordinal SD1 implies rejecting male latent SD1, too. Overall, in the earlier half, there is some evidence of female happiness being higher than male happiness, even though there is no evidence of latent RSD and the latent quantile ranges that can be compared (by Theorem 2) are small.

In contrast, the latter half of the table shows essentially the opposite pattern. In the sample, the male ordinal distribution dominates six times, whereas the female distribution

Table 1: Female vs. male happiness in U.S., 1972–2006.

	Female		Male			
Year	$F_F(1)$	$F_F(2)$	$F_M(1)$	$F_M(2)$	F SD$_1$ M?	M SD$_1$ F?
1972	0.149	0.682	0.181	0.713	yes	no+
1973	0.136	0.622	0.125	0.663	no	no+
1974	0.122	0.585	0.142	0.662	yes	no+*
1975	0.137	0.661	0.123	0.685	no	no
1976	0.124	0.651	0.127	0.670	yes	no
1977	0.124	0.634	0.113	0.673	no	no+
1978	0.091	0.659	0.102	0.653	no	no
1980	0.131	0.638	0.136	0.690	yes	no+
1982	0.144	0.682	0.146	0.712	yes	no
1983	0.127	0.689	0.129	0.688	no	no
1984	0.119	0.632	0.145	0.683	yes	no+*
1985	0.114	0.717	0.113	0.709	no	yes
1986	0.114	0.663	0.114	0.697	yes	no
1987	0.129	0.703	0.142	0.717	yes	no
1988	0.103	0.674	0.079	0.642	no+	yes
1989	0.101	0.670	0.092	0.678	no	no
1990	0.092	0.671	0.087	0.659	no	yes
1991	0.129	0.700	0.084	0.675	no+*	yes
1993	0.119	0.669	0.101	0.705	no	no
1994	0.129	0.720	0.113	0.701	no+	yes
1996	0.127	0.709	0.114	0.680	no+	yes
1998	0.130	0.674	0.109	0.694	no+	no
2000	0.104	0.679	0.107	0.688	yes	no
2002	0.151	0.715	0.094	0.676	no+*	yes
2004	0.120	0.671	0.151	0.704	yes	no+
2006	0.134	0.689	0.126	0.696	no	no

[+] Bayesian statistical significance: yes+ means posterior probability of SD1 above 0.95, no+ means posterior probability of SD1 below 0.05.

[*] frequentist statistical significance at level $\alpha = 0.05$: no* means RMS rejects SD1, yes* means IUT rejects non-SD1.

dominates only three times (one of which is the very first year). Further, male ordinal SD1 is never rejected by RMS and only has below α posterior probability once, whereas female ordinal SD1 (and thus latent SD1) is rejected multiple times by both frequentist and Bayesian measures. That said, again the latent quantile ranges that can be compared by Theorem 2 are small.

Overall, although the magnitudes are modest, the statistical evidence suggests female happiness has indeed declined relative to male happiness over 1972–2006 in the U.S.

5.1.2 Happiness over the life cycle

We use the 12-country Eurobarometer extract analyzed by Bond and Lang (2019) in Appendix A.3.2. The literature has suggested a happiness life-cycle U-shape is common: happiness decreases from the beginning of adulthood through mid-life and then increases again through retirement and beyond. Bond and Lang (2019) point out that latent means cannot be compared nonparametrically for any pair of age groups within any country, and they show (Figure A-4) that simply adding skewness to a Gaussian parameterization often changes the apparent qualitative life-cycle pattern.

We apply our results to examine the U-shape in the same samples of men from 12 European countries as in Bond and Lang (2019). We consider three age groups: 20–39 (young), 40–59 (mid), and 60–79 (old). Within each country, we check whether the young ordinal distribution dominates (SD1) the mid distribution, and similarly whether the old dominates the mid. Rejecting ordinal SD1 implies rejecting latent SD1. Of course, one latent distribution may still have a higher mean even if it is not stochastically dominant; we do not examine a latent mean U-shape (which is impossible to do nonparametrically), but rather we consider a latent "stochastic U-shape" pattern.

Table 2 shows the results. First, considering the sample ordinal SD1 relationships, only 4 of the 12 countries satisfy the U-shape pattern (Belgium, Spain, France, and Netherlands), although Denmark is also extremely close (the deviation cannot even be seen in the table due to rounding). Second, nonetheless, some of the U-shape relationships are statistically significant with $\alpha = 0.1$: in both Belgium and France, the IUT rejects non-SD1 for both young/mid and old/mid. Third, more often the U-shape hypothesis is rejected: ordinal SD1 is rejected, which implies latent SD1 is also rejected. Ordinal SD1 of young over mid is rejected by RMS for West Germany (FRG), Great Britain, and Luxembourg, and the posterior probability is below α for all these countries as well as Ireland. Further, ordinal SD1 of old over mid is rejected by RMS for Portugal, and the posterior probability is below α additionally for Greece and Italy.

Overall, although some countries are consistent with the U-shape, there is also strong

Table 2: Life cycle happiness for men.

Country	Age group	$F(1)$	$F(2)$	$F(3)$	Y SD_1 M?	O SD_1 M?
BEL	Young	0.029	0.132	0.715		
BEL	Mid	0.040	0.158	0.733	yes+*	yes+*
BEL	Old	0.030	0.140	0.699		
DNK	Young	0.005	0.032	0.418		
DNK	Mid	0.008	0.041	0.427	yes	no
DNK	Old	0.008	0.039	0.392		
ESP	Young	0.039	0.225	0.807		
ESP	Mid	0.043	0.254	0.817	yes	yes
ESP	Old	0.038	0.202	0.764		
FRA	Young	0.071	0.259	0.882		
FRA	Mid	0.080	0.294	0.889	yes *	yes+*
FRA	Old	0.046	0.187	0.811		
FRG	Young	0.018	0.146	0.819		
FRG	Mid	0.023	0.136	0.794	no+*	yes+*
FRG	Old	0.017	0.113	0.739		
GBR	Young	0.038	0.140	0.741		
GBR	Mid	0.044	0.156	0.721	no+*	yes+*
GBR	Old	0.033	0.126	0.624		
GRC	Young	0.110	0.368	0.859		
GRC	Mid	0.142	0.411	0.867	yes+*	no+
GRC	Old	0.149	0.424	0.845		
IRL	Young	0.060	0.176	0.707		
IRL	Mid	0.054	0.168	0.699	no+	yes+*
IRL	Old	0.031	0.114	0.602		
ITA	Young	0.065	0.278	0.873		
ITA	Mid	0.069	0.276	0.879	no	no+
ITA	Old	0.073	0.286	0.857		
LUX	Young	0.014	0.075	0.650		
LUX	Mid	0.016	0.070	0.618	no+*	yes
LUX	Old	0.009	0.062	0.532		
NLD	Young	0.009	0.060	0.593		
NLD	Mid	0.016	0.088	0.649	yes+*	yes
NLD	Old	0.014	0.069	0.517		
PRT	Young	0.056	0.258	0.930		
PRT	Mid	0.079	0.338	0.950	yes+*	no+*
PRT	Old	0.091	0.356	0.952		

[+] Bayesian statistical significance: yes+ means posterior probability of SD1 above 0.90, no+ means posterior probability of SD1 below 0.10.

[*] frequentist statistical significance at level $\alpha = 0.10$: no* means RMS rejects SD1, yes* means IUT rejects non-SD1.

evidence against a latent SD1 U-shape in 4 countries and moderate evidence against it in 3 others.

5.2 Adult health

We study ordinal measures of mental health and overall health from the popular NHIS data, available through IPUMS (Blewett, Rivera Drew, King, and Williams, 2019). We use the 2006 and 2008 waves because the latter is in the Great Recession while the former is not. Besides the health variables, we also use the appropriate sampling weights and the provided measures of poverty, education, and race. Details can be seen in the provided replication code.

5.2.1 Mental health

For mental health, we compare non-recession and recession distributions separately for men in poverty and men not in poverty. The goal is to assess whether the recession is associated with worse mental health, and whether the association is stronger for those in poverty. Although we cannot determine the 2006 poverty status of individuals observed in 2008 (these are repeated cross-sections, not panel data), the proportion in poverty is very similar in both years (12.0%, 11.7%). More potentially problematic is the significant proportions of the samples (23%, 11%) for which the poverty measure is unavailable; we do not attempt to assess possible sample selection bias.

The mental health variable is based on the Kessler-6 scale for nonspecific psychological distress introduced by Kessler, Andrews, Colpe, Hiripi, Mroczek, Normand, Walters, and Zaslavsky (2002). We code the worst mental health as 1 and the best as 25 (i.e., we subtract the raw K6 score from 25). As a rough guide, values of 1–12 help predict serious mental illness (Kessler, Barker, Colpe, Epstein, Gfroerer, Hiripi, Howes, Normand, Manderscheid, Walters, and Zaslavsky, 2003, p. 188).

Our interpretations in terms of latent quantiles and latent RSD (from Theorems 2 and 3) are more helpful than considering the latent mean or median. The latent means cannot be compared nonparametrically, as noted by Bond and Lang (2019), and the median is always a very high value indicating good mental health. It is arguably more important to understand differences in the lower part of the mental health distribution, which our results help with.

Figure 3 shows the ordinal (weighted) empirical CDFs for mental health. For men in poverty (left graph), the CDF for 2006 generally lies below that for 2008, indicating worse mental health in 2008 during the recession. Visually, it is clear that Theorem 3 applies over a large range of values, indicating latent RSD. Specifically, writing X for 2006 and Y for

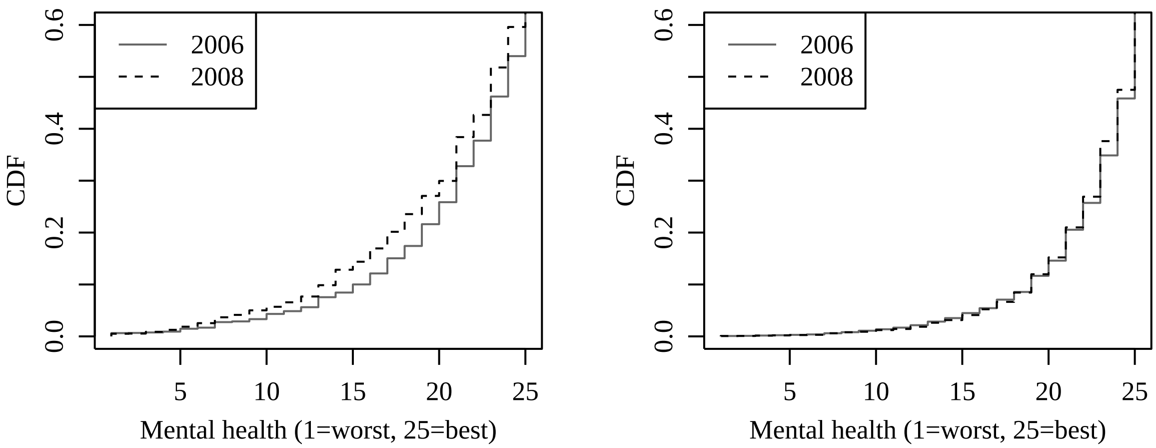

Figure 3: Men in poverty (left) and not (right).

2008, $F_X(j) \leq F_Y(j-1)$ for $8 \leq j \leq 20$ as well as $j \in \{6, 22\}$, so there is latent RSD over $[\gamma_7, \gamma_{20}]$ as well as $[\gamma_5, \gamma_6] \cup [\gamma_{21}, \gamma_{22}]$. This includes part of the range (1–12) that predicts serious mental illness. Interpreting the results by Theorem 2 instead, the latent mental health distribution during the recession is worse over a broad set of quantile indices $\mathcal{T}_X$ that includes $[0.03, 0.30]$ as well as $[0.33, 0.42]$ and other values above 0.42 and below 0.03. That is, the 2008 latent mental health distribution is worse than 2006 over most of the lower part of the distribution.

Figure 3 also shows the 2006 and 2008 mental health ordinal CDFs for men not in poverty (right graph). These appear nearly identical for most of the distribution, especially the lower half. The only categories at which the CDFs differ by at least 0.01 are $22 \leq j \leq 24$, and the only category where $F_X(j)/F_Y(j) < 0.95$ is $j = 23$. That is, the changes are mostly within the part of the distribution corresponding to good mental health.

In all, although we only have ordinal data, we can still see that the mental health costs of the recession are concentrated on those in poverty, and Theorems 2 and 3 let us interpret the decline in mental health in terms of a broad range of the latent distribution.

5.2.2 Overall health

For overall health, we compare 2006 to 2008 for men in poverty, as well as comparing across racial and education groups within 2006. The overall health variable has values poor, fair, good, very good, and excellent, which we code as 1, 2, 3, 4, and 5, respectively.

Table 3 shows the 2006/2008 comparison in the first two rows, using the sample of men in poverty as in the mental health analysis. The first row shows the 2006 (weighted empirical) CDF evaluated at poor (1), fair (2), good (3), and very good (4); the CDF at excellent always equals 1 by definition. The second row shows the 2008 CDF, which is higher at $1 \leq j \leq 2$

Table 3: General health comparisons.

Sample	Group	$F(1)$	$F(2)$	$F(3)$	$F(4)$	SD1?
men in poverty	2006	0.044	0.156	0.456	0.706	no+
men in poverty	2008	0.049	0.166	0.449	0.671	no+
2006	low edu	0.029	0.121	0.391	0.673	no+*
2006	high edu	0.016	0.073	0.296	0.643	yes+*
2006	white	0.021	0.092	0.330	0.646	yes+*
2006	Black	0.028	0.126	0.415	0.686	no+*

+ Bayesian statistical significance; yes+ means posterior probability of SD1 (of this group over the other group) above 0.95, no+ means posterior probability of SD1 below 0.05.

* frequentist statistical significance at level 0.05; no* means RMS rejects SD1 (of this group over the other group), yes* means IUT rejects non-SD1.

but lower at $3 \leq j \leq 4$. Thus, in the sample, there is not SD1 in either direction. Further, the posterior probability of SD1 of 2006 over 2008 is below α. The pattern seen for mental health is not apparent for overall health.

Table 3 next compares low and high education groups with the 2006 data, with "high" meaning any post-secondary education. Again, the two CDFs are shown; the high education CDF is below the low education CDF at all $1 \leq j \leq 4$, i.e., there is ordinal SD1 in the sample. This generally indicates better health. More specifically, Theorem 2 says the latent high-education health distribution is better at the τ-quantile for (rounding to nearest 0.01)

$$\tau \in [0.02, 0.03] \cup [0.07, 0.12] \cup [0.30, 0.39] \cup [0.64, 0.67].$$

This is weaker than latent SD1, but it has a clear interpretation. Letting X represent the low education ordinal health distribution and similarly Y for high education, Bayesian and frequentist methods both reject $X \text{ SD}_1 Y$. Further, there is positive evidence of $Y \text{ SD}_1 X$: the posterior probability is above $1 - \alpha$, and $Y \; non\text{SD}_1 \; X$ is rejected by the IUT at level α in favor of $Y \text{ SD}_1 X$.

Table 3 shows similar results for the Black/white comparison as for low/high education. There is SD1 in the sample: the white ordinal CDF is lower at each $1 \leq j \leq 4$. Again, more strongly, SD1 of Black over white ordinal health is rejected by both frequentist (RMS) and Bayesian analysis, whereas SD1 of white over Black ordinal health is supported by a posterior probability above $1 - \alpha$ and the IUT's rejection of non-SD1 in favor of SD1.

5.3 Neonatal health

A common measure of neonatal health is the Apgar score, an ordinal measure coded as integers from 0 through 10; we add one to match our paper's notation with categories $j = 1, \ldots, J = 11$. It is named after Dr. Virginia Apgar but also a backronym for the five components (appearance, pulse, grimace, activity, respiration) that comprise the total score. The dataset comes from Wooldridge (2020) via the R package `wooldridge` (Shea, 2018).

We compare infants of mothers based on smoking, race, and education. Often birthweight is studied as a proxy for health, because it is continuously distributed and thus easier to compare. Our analysis complements such studies: although our statistical conclusions are weaker, we find qualitatively similar empirical evidence by directly examining a measure of health (Apgar) and interpreting the results in terms of the continuous latent health distribution. Details can be seen in the provided replication code.

In all cases, the latent means cannot be compared nonparametrically, and the medians are all identical (equal to the second-highest category). For neonatal health, the lower quantiles are arguably much more important than the mean or median anyway.

Table 4: Apgar score (scaled 1–11) ordinal CDFs.

Group	$F(3)$	$F(4)$	$F(5)$	$F(6)$	$F(7)$	$F(8)$	$F(9)$	$F(10)$	RSD?
smoked	0.014	0.048	0.061	0.061	0.068	0.129	0.490	1.000	
did not	0.008	0.012	0.016	0.024	0.045	0.088	0.401	0.998	$[\gamma_3, \gamma_7]$
non-white	0.010	0.029	0.034	0.053	0.062	0.111	0.375	1.000	
white	0.008	0.014	0.019	0.026	0.046	0.088	0.403	0.998	$[\gamma_4, \gamma_7]$

Table 4 shows the ordinal CDFs for mothers who reported smoking or not smoking, respectively, in the first two rows. The bottom two categories are omitted because they have such low counts (and to save space); the top category is omitted because $F(11) = 1$ by definition. The non-smoking ordinal CDF is below the smoking CDF for all $j = 3, \ldots, 10$, which can be interpreted in terms of latent quantiles with Theorem 2. For example, from $j = 9$, the non-smoking latent distribution is better between the 40th and 49th percentiles, and similarly for quantiles $\tau \in [0.008, 0.068]$, among others. Further, Theorem 3 applies, finding a range in the lower tail $[\gamma_3, \gamma_7]$ where there is latent RSD.

Table 4 then shows the ordinal CDFs for non-white and white mothers. The white CDF is lower at most j but not $j = 9$. Again, Theorems 2 and 3 provide interpretations in terms of latent health. For example, the white latent distribution is better at quantile indices $\tau \in [0.014, 0.062]$, among others, and there is latent RSD over $[\gamma_4, \gamma_7]$ in the lower tail.

Table 5 restricts attention to only white mothers. The first two rows show the CDFs

Table 5: Apgar score (scaled 1–11) ordinal CDFs, white mothers.

Group	$F(3)$	$F(4)$	$F(5)$	$F(6)$	$F(7)$	$F(8)$	$F(9)$	$F(10)$	RSD?
smoked	0.015	0.037	0.051	0.051	0.059	0.118	0.478	1.000	
did not	0.007	0.012	0.015	0.022	0.045	0.087	0.407	0.998	$[\gamma_2, \gamma_7]$
edu≤ 12	0.012	0.022	0.028	0.037	0.053	0.104	0.422	0.999	
edu> 12	0.005	0.009	0.013	0.019	0.042	0.078	0.389	0.998	$[\gamma_3, \gamma_6]$

for mothers who reported smoking and those who did not. Just as in the overall sample, there are differences in the lower part of the distribution. Here, the latent RSD extends over $[\gamma_2, \gamma_7]$, and the latent non-smoking τ-quantile is better for $\tau \in [0.003, 0.058]$ among other quantile ranges. The next rows compare mothers without and with any post-secondary education. There are differences, but they are smaller. The latent RSD is only over $[\gamma_3, \gamma_6]$, and the quantile index set includes $\tau \in [0.009, 0.037]$ and smaller ranges elsewhere in the distribution. There is a clear difference by education, but it appears smaller than the neonatal health differences by race or smoking.

6 Conclusion

We compare continuous latent distributions nonparametrically when only ordinal data are available. Our identification results interpret certain ordinal patterns as evidence of between-group inequality, in terms of quantiles or restricted stochastic dominance, while other ordinal patterns indicate differences in latent within-group inequality, in terms of interquantile ranges. We discuss and compare frequentist and Bayesian inference for the relevant ordinal relationships. Empirical examples with different ordinal measures of health and happiness show how our results provide insight, even when latent means cannot be compared. Our approach can be applied similarly with ordinal measures from education, politics, finance, and other areas. In future work, we plan to apply the inner and outer confidence sets from Kaplan (2020) to this ordinal setting.

A Simulations

The following simulations illustrate properties of some of the possible inference methods. Generally, we examine the benefit of more recent frequentist approaches over more simple frequentist approaches, as well as the difference between frequentist and Bayesian assessments. Although the simulations report frequentist rejection probabilities, we do not mean to implicitly favor frequentist over Bayesian inference; it is simply a convenient way to compare which approach is "more conservative" in different cases.

Four null hypotheses are tested with $\alpha = 0.1$ using various methods. The four null hypotheses are $H_0\colon X\ \mathrm{SD}_1\ Y$, $H_0\colon X\ \mathrm{SC}\ Y$, $H_0\colon X\ non\mathrm{SD}_1\ Y$, and $H_0\colon X\ non\mathrm{SC}\ Y$, as characterized in Section 3.1. Each H_0 is tested using the Bayesian test in Section 3.4.2 that rejects H_0 when its posterior probability is below α, with label "Bayes" referring to a uniform prior and "Bayes (adj)" the adjusted prior with $\mathrm{P}(H_0) = 1/2$. SD1 is also tested with a one-sided Kolmogorov–Smirnov test ("KS"), as well as the refined moment selection ("RMS") recommended in Section 2 of Andrews and Barwick (2012). SC is also tested with RMS embedded into the IUT approach described in Section 3.3.3. Non-SD1 is also tested using the IUT described in Section 3.3.2.

The two DGPs (and sample sizes $n_X = n_Y = n \in \{20, 100, 500\}$) are the same as in Section 5.1.3 of Kaplan and Zhuo (forthcoming), who only consider Bayesian tests of SC and non-SC. Recall the notation $\theta_j \equiv F_X(j) - F_Y(j)$ and $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3, \theta_4)'$ from (3). As in Section 3.1, the null hypotheses can also be written as $H_0\colon \boldsymbol{\theta} \in \Theta_0$.

DGP 1 has the true $\boldsymbol{\theta}$ at a locally convex corner of Θ_0, for SD1 and SC alike: $F_X(1) = F_Y(1) = 0.2$, $F_X(2) = 0.39 < 0.4 = F_Y(2)$, $F_X(3) = 0.5 < 0.6 = F_Y(3)$, and $F_X(4) = F_Y(4) = 0.8$. That is, given $\theta_2 < 0$ and $\theta_3 < 0$, in terms of the (θ_1, θ_4) parameter subspace, SD1 is satisfied only in the third quadrant, SC is satisfied only in the second quadrant, and DGP 1 is at the origin $(\theta_1, \theta_4) = (0, 0)$, right on the corner of Θ_0.

DGP 2 has SD1 violated (not on the boundary) and is at a locally non-convex boundary point of SC: $F_X(1) = 0.18 < 0.22 = F_Y(1)$, $F_X(2) = F_Y(2) = 0.4$, $F_X(3) = F_Y(3) = 0.6$, $F_X(4) = 0.82 > 0.78 = F_Y(4)$. Frequentist tests should reject non-SD1 with probability below α, but they should reject SD1 with increasing probability as the sample size increases.

Table 6: Simulated rejection probability, nominal $\alpha = 0.1$.

		$H_0\colon X\ \mathrm{SD}_1\ Y$				$H_0\colon X\ \mathrm{SC}\ Y$		
DGP	n	KS	RMS	Bayes	Bayes (adj)	RMS	Bayes	Bayes (adj)
1	20	0.023	0.095	0.376	0.150	0.020	0.401	0.165
1	100	0.013	0.075	0.350	0.144	0.047	0.501	0.208
1	500	0.015	0.083	0.347	0.143	0.066	0.525	0.250
2	20	0.034	0.106	0.430	0.193	0.005	0.249	0.067
2	100	0.034	0.152	0.518	0.262	0.003	0.117	0.029
2	500	0.156	0.415	0.732	0.481	0.000	0.014	0.003

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

Table 6 shows rejection probabilities for all DGPs, n, and methods for the first two null hypotheses, revealing a few patterns. First, RMS always has a higher rejection probability than KS but still has type I error rates below α. When the SD1 null is true (DGP 1), the KS rejection probability is close to zero; the RMS rejection probability is significantly higher, but still below α. When the SD1 null is false (DGP 2), RMS has significantly better power than KS. Despite the improvement, RMS still has type I error rates near zero when the null is SC, even though both DGPs are on the boundary of SC. Second, the Bayesian test has type I error rate above α when the DGP is on a convex "corner" of Θ_0, i.e., for DGP 1. This illustrates the theoretical results of Kaplan and Zhuo (forthcoming). Of course, from the Bayesian perspective, frequentist tests are too conservative in this case. Naturally, the Bayesian test has better power than the frequentist tests against SD1 in DGP 2. Third, for the null of SC with DGP 2, when the DGP is at a locally non-convex point on the boundary of Θ_0 (which globally is smaller than a half-space), the Bayesian test's rejection probability is above α with smaller n (when the global shape of Θ_0 dominates) but falls well below α with larger n (when the local shape matters more). See Section 5.1.3 of Kaplan and Zhuo (forthcoming) for details. Fourth, although the adjusted prior (with $\mathrm{P}(H_0) = 1/2$) tends to make the Bayesian test closer to a frequentist test, it does not fully do so, and sometimes it makes the Bayesian test worse by frequentist standards.

Table 7: Simulated type I error rate, nominal $\alpha = 0.1$.

		$H_0\colon X$ *non*SD_1 Y			$H_0\colon X$ *non*SC Y	
DGP	n	IUT	Bayes	Bayes (adj)	Bayes	Bayes (adj)
1	20	0.015	0.006	0.044	0.000	0.016
1	100	0.034	0.010	0.047	0.000	0.005
1	500	0.046	0.007	0.045	0.000	0.007
2	20	0.005	0.001	0.017	0.003	0.049
2	100	0.019	0.003	0.018	0.014	0.144
2	500	0.033	0.000	0.002	0.147	0.506

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

Table 7 shows type I error rates for the other null hypotheses, non-SD1 and non-SC, illustrating additional points. First, non-SD1 is true for both DGPs, and the IUT's type I error rates are always below α, but it is indeed conservative: DGP 1 is on the boundary of SD1, yet type I error rate is only around $\alpha/2$. Second, because non-SD1 corresponds to a non-convex Θ_0, the Bayesian test's type I error rate is now well below α, in some cases zero (up to simulation error). Third, the pattern for DGP 2 and the null of non-SC is the opposite for the Bayesian test as for the null of SC: now the type I error rate is well below α for small n but increases above α for larger n. Fourth, most clearly in DGP 2 with the null of non-SC and largest n, the adjusted prior can make the type I error rate worse.

References

Allison, R. Andrew and James E. Foster. 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, Donald W. K. and Panle Jia Barwick. 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://www.jstor.org/stable/1911028.

Blewett, Lynn A., Julia A. Rivera Drew, Miriam L. King, and Kari C. W. Williams. 2019. “IPUMS Health Surveys: National Health Interview Survey, Version 6.4.” https://doi.org/10.18128/D070.V6.4.

Bond, Timothy N. and Kevin Lang. 2019. “The Sad Truth About Happiness Scales.” *Journal of Political Economy* 127 (4):1629–1640. URL https://doi.org/10.1086/701679.

Casella, George and Roger L. Berger. 2002. *Statistical Inference*. Pacific Grove, CA: Duxbury, 2nd ed.

Chen, Le-Yu, Ekaterina Oparina, Nattavudh Powdthavee, and Sorawoot Srisuma. 2019. “Have Econometric Analyses of Happiness Data Been Futile? A Simple Truth About Happiness Scales.” Working paper available at https://arxiv.org/pdf/1902.07696.pdf.

Davidson, Russell and Jean-Yves Duclos. 2013. “Testing for Restricted Stochastic Dominance.” *Econometric Reviews* 32 (1):84–125. URL https://doi.org/10.1080/07474938.2012.690332.

de Castro, Luciano and Antonio F. Galvao. 2019. “Dynamic Quantile Models of Rational Behavior.” *Econometrica* 87 (6):1893–1939. URL https://doi.org/10.3982/ECTA15146.

Dong, Qi, Michael R. Elliott, and Trivellore E. Raghunathan. 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.

Frölich, Markus. 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.

Garbuszus, Jan Marvin and Sebastian Jeworutzki. 2018. *readstata13: Import ‘Stata’ Data Files*. URL https://CRAN.R-project.org/package=readstata13. R package version 0.9.2.

Goutis, Constantinos, George Casella, and Martin T. Wells. 1996. “Assessing Evidence in Multiple Hypotheses.” *Journal of the American Statistical Association* 91 (435):1268–1277. URL https://www.jstor.org/stable/2291745.

Gunawan, David, William E. Griffiths, and Duangkamon Chotikapanich. 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, Cristina, Andrew M. Jones, and Nigel Rice. 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`.

Jenkins, Stephen P. 2019. "Better off? Distributional comparisons for ordinal data about personal well-being." *New Zealand Economic Papers* 54 (3):1–28. URL `https://doi.org/10.1080/00779954.2019.1697729`.

———. 2020. "Comparing distributions of ordinal data." *Stata Journal* 20 (3):505–531. URL `https://doi.org/10.1177/1536867X20953565`.

Kaplan, David M. 2020. "Inference on Consensus Ranking of Distributions." Working paper available at `https://faculty.missouri.edu/kaplandm`.

Kaplan, David M. and Longhao Zhuo. forthcoming. "Frequentist properties of Bayesian inequality tests." *Journal of Econometrics* XXX (XXX):XXX. URL `https://doi.org/10.1016/j.jeconom.2020.05.015`.

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. T. Normand, E. E. Walters, and A. M. Zaslavsky. 2002. "Short screening scales to monitor population prevalences and trends in non-specific psychological distress." *Psychological Medicine* 32 (6):959–976. URL `https://doi.org/10.1017/S0033291702006074`.

Kessler, Ronald C., Peggy R. Barker, Lisa J. Colpe, Joan F. Epstein, Joseph C. Gfroerer, Eva Hiripi, Mary J. Howes, Sharon-Lise T. Normand, Ronald W. Manderscheid, Ellen E. Walters, and Alan M. Zaslavsky. 2003. "Screening for Serious Mental Illness in the General Population." *Archives of General Psychiatry* 60 (2):184–189. URL `https://doi.org/10.1001/archpsyc.60.2.184`.

Kline, Brendan. 2011. "The Bayesian and frequentist approaches to testing a one-sided hypothesis about a multivariate mean." *Journal of Statistical Planning and Inference* 141 (9):3131–3141. URL `https://doi.org/10.1016/j.jspi.2011.03.034`.

Lindeboom, Maarten and Eddy van Doorslaer. 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`.

Machado, Cecilia, Azeem M. Shaikh, and Edward J. Vytlacil. 2019. "Instrumental variables and the sign of the average treatment effect." *Journal of Econometrics* 212 (2):522–555. URL `https://doi.org/10.1016/j.jeconom.2018.04.007`.

Madden, David. 2014. "Dominance and the Measurement of Inequality." In *Encyclopedia of Health Economics*, vol. 1, edited by Anthony J. Culyer. Elsevier, 204–208. URL `https://doi.org/10.1016/B978-0-12-375678-7.00725-2`.

Manski, Charles F. 1988. "Ordinal utility models of decision making under uncertainty." *Theory and Decision* 25 (1):79–104. URL `https://doi.org/10.1007/BF00129169`.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `https://www.R-project.org`.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2014. "A Practical Two-Step Method for Testing Moment Inequalities." *Econometrica* 82 (5):1979–2002. URL `https://www.jstor.org/stable/24029299`.

Rostek, Marzena. 2010. "Quantile Maximization in Decision Theory." *Review of Economic Studies* 77 (1):339–371. URL `https://doi.org/10.1111/j.1467-937X.2009.00564.x`.

Shea, Justin M. 2018. *wooldridge: 111 Data Sets from "Introductory Econometrics: A*

Modern Approach, 6e" by Jeffrey M. Wooldridge. URL `https://CRAN.R-project.org/package=wooldridge`. R package version 1.3.1.

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

Weingessel, Andreas. 2019. *quadprog: Functions to Solve Quadratic Programming Problems*. URL `https://CRAN.R-project.org/package=quadprog`. R package version 1.5-8; S original by Berwin A. Turlach; Fortran contributions from Cleve Moler (dpodi/LINPACK).

Wooldridge, Jeffrey M. 2020. *Introductory Econometrics: A Modern Approach*. Cengage, 7th ed.

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