Quantitative Economics 1 (2010), 323–357 1759-7331/20100323

Partial identification of spread parameters

Jörg Stoye
Department of Economics, Cornell University

This paper analyzes partial identification of parameters that measure a distribution's spread, for example, the variance, Gini coefficient, entropy, or interquartile range. The core results are tight, two-dimensional identification regions for the expectation and variance, the median and interquartile ratio, and many other combinations of parameters. They are developed for numerous identification settings, including but not limited to cases where one can bound either the relevant cumulative distribution function or the relevant probability measure. Applications include missing data, interval data, "short" versus "long" regressions, contaminated data, and certain forms of sensitivity analysis. The application to missing data is worked out in some detail, including closed-form worst-case bounds on some parameters as well as improved bounds that rely on nonparametric restrictions on selection effects. A brief empirical application to bounds on inequality measures is provided. The bounds are very easy to compute. The ideas underlying them are explained in detail and should be readily extended to even more settings than are explicitly discussed.

Keywords. Partial identification, nonparametric bounds, missing data, sensitivity analysis, variance, inequality.

JEL classification. C14, C24.

1. Introduction

This paper contributes to research on *partial identification*. A parameter is partially identified if the data generating process, together with assumptions a researcher is willing to make, reveals some nontrivial information about it but does not identify it in the conventional sense; that is, distinct parameter values may be observationally equivalent. Analysis of partial identification has become an active literature; see Manski (2003, 2007) for surveys. This paper extends its scope in two dimensions: First,

Jörg Stoye: stoye@cornell.edu
I am greatly indebted to Chuck Manski for his guidance and feedback on this project. For useful comments, I thank Elie Tamer and three extremely helpful referees as well as Adeline Delavande, Joel Horowitz, Peter Klibanoff, Francesca Molinari, seminar participants at Cal Tech, Chicago, Mannheim, NYU, Northwestern, Wisconsin–Madison, and conference audiences at Carlos III Madrid, Columbia, and Nürnberg-Erlangen universities. Steven Laufer provided able research assistance. Of course, any and all errors are mine. I gratefully acknowledge financial support through the Robert Eisner Memorial Fellowship, a Dissertation Year Fellowship, and a Graduate Research Travel Grant, all from Northwestern University, as well as through a University Research Challenge Grant from New York University.

DOI: 10.3982/QE24

it provides general identification results for most salient measures of a random variable's dispersion. To give a few examples, findings apply to joint bounds—that is, two-dimensional identification regions that are typically not rectangles—for the expectation and variance, expectation and Gini coefficient, or expectation and entropy, as well as joint bounds for the median and interquartile range or ratio. Between them, these and other parameters covered in the paper are of interest in applications ranging from analysis of income distributions to risk assessments. Second, findings apply to a number of rather general settings: The distribution of interest may be a mixture between an identified distribution and an unidentified one, where the mixture weight is either known or unknown, and where partial knowledge of the unidentified distribution can be expressed as bounds on either its cumulative distribution function (c.d.f.) or its measure. Between them, these scenarios encompass a wide range of problems, including missing data, interval data, "short" versus "long" regressions, contaminated data, and certain forms of sensitivity analysis.

Here is a basic intuition for how the bounds work. Assume for concreteness that one observes interval-valued income data; that is, for every household in a sample, one observes the income bracket to which the household belongs. Then one can generate upper [lower] worst-case bounds on expected income or a quantile of income by assuming that all true data points occupy the upper [lower] end of the respective bracket; intuitively, probability mass must be shifted to the extreme right [left] of its potential support. Let us now try to maximize the variance. Intuitively, this must be achieved by shifting probability mass to the edges of its potential support: indeed, the solution will be to assume that all data points below a threshold bracket occupy the lower end of their respective bracket and all data points above it occupy the upper end, with a possibly mixed allocation in the threshold bracket itself. The problem is that there are many configurations of incomes that fit this characterization, and it is not clear which of them maximizes variance. However, they all induce different means. For any given, hypothesized value of expected income, exactly one of these "dispersed" distributions is consistent with said value, and this distribution maximizes the variance among all distributions with that same expectation. Thus, one has bounds on the variance that depend on hypothetical values of the expectation. This leads to joint bounds on the expectation and variance, and, by implication, on any function of them, including unconstrained bounds on the variance. A similar statement holds true for "compressed" distributions that shift probability mass toward the center of the support and thereby minimize variance.

The gist of this paper's identification results is to demonstrate the level of generality to which this idea can be pushed. The answer is that it applies to joint bounds on the expectation and any parameter that increases in mean-preserving spreads, as well as joint bounds on any quantile and what will be called quantile contrasts. Furthermore, it applies in any of four general identification settings: The probability distribution of interest is a mixture between an identified and an unidentified distribution, where one can formulate bounds on either the unidentified c.d.f. or the corresponding measure, and where the mixture probability may or may not be known. Moreover, once the organizing principles that underlie these bounds are in place, it is expected that they can readily be extended to even more identification settings. Section 4 of this paper contains one example of this. In all cases, the bounds are easy to compute as long as it is easy to evaluate θ for a given c.d.f. or measure.

It is a sign of the wide applicability of these results that special cases emerged in different literatures. Vazques Alvarez, Melenburg, and van Soest (2003) identified bounds on the Gini coefficient and interquartile range, and Blundell, Gosling, Ichimura, and Meghir (2007) provided bounds on differences between quantiles, when some data are missing. Both are motivated by, and apply their results to, analysis of income or wage distributions. Neither considers joint identification of these statistics and expectations or quantiles nor extends the analysis to general bounds on c.d.f.'s (of which missing data are a special case) or on measures. The present treatment applies beyond these restrictions and also to many other parameters. Gastwirth (1972) and Cowell (1991) found worst-case bounds on the sample Gini coefficient under the assumption that one knows the income bracket but not the exact income of every household. Up to minor differences caused by discreteness, these results are special cases of identification analysis with bounds on distributions. Deriving bounds on parameters from bounds on measures, on the other hand, is formally similar to questions that arise in the literature on robust Bayesian inference, and some of my findings there are related to those in DeRobertis and Hartigan (1981) and Wasserman and Kadane (1992).

The remainder of this paper is structured as follows. In Section 2, I present the formal setting, define some relevant classes of parameters, and elucidate the identification scenarios. Section 3 contains the general identification analysis for the cases of bounds both on distributions and on measures. In Section 4, these results are employed toward a detailed analysis of certain missing-data problems, providing worst-case bounds but also showing how to refine them via partially identifying assumptions. The application is illustrated by a simple empirical example. While analysis in Sections 3 and 4 conditions on or ignores any existing covariate, Section 5 explicitly discusses the effect of covariates on bounds. Section 6 concludes. All proofs are relegated to the Appendix.[1]

2. Setting the stage

Abstracting from estimation issues, the identification problem can be described as follows. Let $Y \in \mathbb{R}$ and $X \in \mathcal{X}$ be random variables on probability space $(\mathbb{R} \times \mathcal{X}, \sigma_{\mathbb{R}} \times \sigma_{\mathcal{X}}, P)$, with Y the outcome of interest and X a covariate; here, $\sigma_{\mathbb{R}}$ is the Borel sigma algebra on $\mathbb{R}$ and $\sigma_{\mathcal{X}}$ is some sigma algebra on $\mathcal{X}$. The word "measurable" henceforth refers to these algebras.

Depending on what is convenient, I use either P or the corresponding c.d.f. F as the primitive object. The notation $\theta(P)$ (or $\theta(F)$) will be used for generic parameters of P

[1]This paper does not contain new contributions to estimation or inference for the identified sets it characterizes. Here are some (far from exhaustive) references. The only developed framework for estimation and inference that would apply at this paper's level of generality is the generalization of extremum estimators to set-valued extrema (Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2010)). Depending on the specifics of the case, other frameworks like moment inequalities (Andrews and Soares (2010), Chernozhukov, Hong, and Tamer (2007), Imbens and Manski (2004), Rosen (2008), Stoye (2009)) or set-valued random variables (Beresteanu and Molinari (2008)) may apply as well.

or F that take values on the (possibly extended) real line. Partial identification means that only some aspects of P are observable. Specifically, assume that the marginal distribution of X is known, but that the *identified set* $H(P(Y|X=x))$ for $P(Y|X=x)$ need not be a singleton for any x. As a result, a parameter θ of $P(Y|X=x)$ is partially identified and can merely be concluded to lie within its identified set

$$H^x(\theta)=\{\theta(P^*):P^*\in H(P(Y|X=x))\},$$

with point identification emerging as the special case where $H^x(\theta)$ is a singleton. In the subsequent text, $H^x(\theta)$ will be characterized via bounds on θ, that is, identified quantities $\underline{\theta}$ and $\overline{\theta}$ such that (s.t.) $H^x(\theta)\subseteq[\underline{\theta},\overline{\theta}]$. All bounds in this paper are best possible in the sense that they cannot be improved upon using only the assumptions made. Indeed, most bounds are *attainable*, that is, $\{\underline{\theta},\overline{\theta}\}\subseteq H^x(\theta)$; distributions achieving them will be explicitly characterized, so that computation will typically be easy; and the conditions where attainability can fail will be elaborated. An additional question is whether $H^x(\theta)$ can be concluded to be convex, so that $H^x(\theta)=[\underline{\theta},\overline{\theta}]$ if bounds are attainable. The answer depends on θ, but is in the affirmative for many parameters because all identified sets $H(P(Y|X=x))$ considered in this paper are closed under mixture. Throughout this paper, $H^x(\theta)$ is, therefore, convex for parameters θ s.t.

$$[\theta(P),\theta(P')]\subseteq\{\theta(\lambda P+(1-\lambda)P'):\lambda\in[0,1]\}$$

for all measures P, P'. This condition specifically holds if θ is such that $X_n\xrightarrow{d}X$ implies $\theta(P(X_n))\to\theta(P(X))$.

Characterizations of $\{H^x(\theta)\}_{x\in\mathcal{X}}$ immediately imply that the set $\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}}$ is bounded by the Cartesian product of covariate-wise identified sets:

$$\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}}\in\bigtimes_{x\in\mathcal{X}}\{H^x(\theta)\}.$$

In the presence of cross-covariate restrictions, this bound may be far from tight. I will initially analyze $H^x(\theta)$ for a wide range of scenarios and return to joint bounds on $\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}}$ in Section 5.

Some more notation is as follows: Write $E(Y)$ and $Q(\alpha)$ for the expectation and quantile function; thus, $Q(\alpha)=\inf\{y:F(y)\geq\alpha\}$ for any $\alpha\in(0,1]$. Define also $F^-(y)=\lim_{x\uparrow y}F(x)$. All of these will inherit markers applied to primitive objects; thus if a c.d.f. F_1 (say) was introduced, $E_1(Y)$ is the corresponding expectation. To avoid case distinctions, also define $Q(\alpha)=\inf\{\mathrm{supp}(Y)\}$ for any $\alpha\leq 0$ and $Q(\alpha)=\sup\{\mathrm{supp}(Y)\}$ for any $\alpha>1$, where these quantities may lie on the extended real line (but will not do so anywhere where this would cause difficulties).

The challenge is to characterize $H^x(\theta)$, which specifically requires extremizing θ over $H(P(Y|X=x))$. This paper's core insight is that this problem can be solved explicitly for many parameters and in many settings. I will first elaborate the "many parameters" and then the "many settings" part.

Definition 1 (Parameters). θ is a D_1-parameter if it increases with first-order stochastic dominance:

$$F(y) \leq G(y) \quad \forall y \quad \Longrightarrow \quad \theta(F) \geq \theta(G). \tag{1}$$

Examples of D_1-parameters (called D-parameters in Manski (2003)) include the expectation, any quantile, and any point of the c.d.f. Bounds on them will be derived, but mainly as a backdrop. Core interest is in *spread parameters*. By this, I mean two classes of parameters which jointly include most parameters that one would think of for measuring a variable's dispersion. The first of these is the following:

Definition 2 (Parameters). θ is a D_2-parameter if for distributions that have equal expectation, it increases with second-order stochastic dominance:

$$\begin{aligned} &\int y\,dF = \int y\,dG \quad \text{and} \quad \int_{-\infty}^{k} F(y)\,dy \leq \int_{-\infty}^{k} G(y)\,dy \quad \forall k \\ &\Longrightarrow \quad \theta(G) \geq \theta(F). \end{aligned} \tag{2}$$

Second-order dominance as just defined is one of several intuitive notions of increased riskiness. One could also think of G as more risky than F if it can be generated from F by the addition of a mean-preserving spread or if $\int u(y)\,dG \leq \int u(y)\,dF$ for every convex real-valued function u. These three notions are equivalent.[2] When applied to distributions with finite support, the partial ordering by second-order stochastic dominance furthermore coincides with the partial orderings induced by the Pigou–Dalton principle of transfers, Schur-convexity, or pointwise ordering of Lorenz curves.[3] As a result, the class of D_2-parameters is quite general: it includes the variance, any even (raw or centered) moment, the expectation of any convex (or, upon sign reversal, concave) utility function, and the vast majority of financial risk measures,[4] as well as a host of inequality measures, for example, the Gini coefficient, Herfindahl's index (better known as a measure of concentration of industries), any entropy-based index like Theil's, social welfare function-based indices like Atkinson's or Dalton's, the relative mean deviation, the equal shares coefficient, and the minimal majority index.[5] Finally, the distributions

[2]A good reference for this result is Shaked and Shantikumar (2007, Section 3), but its essence is much older. See Rothschild and Stiglitz (1970) for a well known reference from within economics, although their statement and proof apply only if Y is bounded.

[3]A function $f:\mathbb{R}^n \to \mathbb{R}$ is Schur-convex if for any bistochastic matrix Q and any $y \in \mathbb{R}^n$, $f(Qy) \leq f(y)$. A matrix with nonnegative elements is bistochastic if it is square, and all rows and columns sum to 1; intuitively, Qy therefore reflects a stochastic reallocation of incomes. The Pigou–Dalton principle (Dalton (1920)) states that if some income is redistributed from a richer to a poorer person without reversing their relative rank positions, then an inequality measure should decrease. Equivalence of these criteria to each other and the D_2 property was established by Atkinson (1970) and Dasgupta, Sen, and Starret (1973).

[4]Pedersen and Satchell (1998) proposed desirable properties of financial risk measures, two of which (their BP2 and BP4) jointly imply the D_2 property. Their list of risk measures that fulfil these (Theorem 9) includes most financial ones.

[5]This information is taken from Cowell (1995), who also provided definitions of all of these measures.

that attain the bounds generate extremal Lorenz curves for partially identified distributions.

I will distinguish between *constrained* and *unconstrained* bounds. A constrained bound applies to a partially identified distribution whose mean, $E(Y)$, is constrained to equal some preassigned value $\mu \in H(E(Y))$. This is convenient because subject to this constraint, all D_2-parameters are extremized by the same distributions. Also, it allows characterization of not just bounds on D_2-parameters, but joint bounds on them and the mean. These joint identification regions are of interest in their own right because they imply bounds on any joint function of the mean and a D_2-parameter; for example, one could perform partially identified mean-variance analysis. Unconstrained bounds can be computed by extremizing constrained bounds over feasible candidate values of the expectation and will, in general, be achieved by different distributions for different D_2-parameters.

The most salient measures of spread that are not D_2-parameters are what will be called quantile contrasts:

Definition 3 (Quantile Contrasts). θ is a quantile contrast if one can write $\theta(F) = f(Q(\alpha), Q(\beta))$, where $\alpha \leq \beta$, and the known, continuous function f is nonincreasing in the first argument and nondecreasing in the second argument.

Obvious examples of quantile contrasts are interquantile ranges and (for nonnegative variables) ratios.[6] Bounds on quantile contrasts will be derived too. The major difference is that they constrain some (user specified) quantile rather than the mean; hence one gets joint identification regions for the median and interquartile ratio, say. This paper does *not* provide joint bounds on D_2-parameters and quantiles, or on quantile contrasts and the expected value.

Next, the analysis covers all settings that fall under one of two scenarios. The label "bounds on c.d.f.'s" will refer to the case where one knows a collection of c.d.f.'s $\{F_1^x, F_L^x, F_U^x\}_{x\in\mathcal{X}}$ with finite expectations s.t.

$$\begin{aligned} &H(F(Y|X=x)) \\ &\quad = \{F^* : p^x F_1^x + (1-p^x)F_U^x \succeq_{\mathrm{FSD}} F^* \succeq_{\mathrm{FSD}} p^x F_1^x + (1-p^x)F_L^x\}, \end{aligned} \tag{3}$$

where $\succeq_{\mathrm{FSD}}$ denotes first-order dominance as defined in (1). Thus, the true conditional distributions are p^x-contaminations of known distributions by distributions whose c.d.f.'s can be bounded. The mixture probabilities $\{p^x\}_{x\in\mathcal{X}}$ may be known or unknown. By setting them equal to zero, this nests the case where one just knows bounds on conditional c.d.f.'s; it is strictly more general, however, because knowledge of $\{F_1^x\}_{x\in\mathcal{X}}$ implies some (relevant, as it turns out) information about densities.

[6]To see that these are not D_2, let $F(y) \equiv 1/2 + 1/2 \cdot 1\{y \geq 1/2\}$ and $G(y) \equiv 3/4 + 1/4 \cdot 1\{y = 1\}$ be different distributions for a random variable $Y \in [0, 1]$. Then G is a mean-preserving spread of F, yet F has an interquartile range of $1/2$ and G has an interquartile range of 0.

The label "bounds on measures" will refer to the case where one knows a collection of measures $\{P_1^x, P_L^x, P_U^x\}_{x\in\mathcal{X}}$ s.t.

$$H(P(Y|X=x)) = \{P^* : p^x P_1^x + (1-p^x)P_L^x \leq P^* \leq p^x P_1^x + (1-p^x)P_U^x\} \tag{4}$$

for all x, where $\leq$ indicates dominance of measures. To make the problem well defined and ensure existence of expectations, I impose that, for all $x \in \mathcal{X}$, (i) P_L^x and P_U^x are measurable, (ii) $P_U^x \geq P_L^x$, (iii) $P_L^x(\mathbb{R}) \leq 1 \leq P_U^x(\mathbb{R})$, (iv) $\int |y|\, dP_U^x(y) < \infty$, and (v) $\int |y|\, dP_1^x(y) < \infty$. Again, $\{p^x\}_{x\in\mathcal{X}}$ may or may not be known and could be zero. Note that P_L^x and P_U^x are not, in general, probability measures; the problem actually becomes trivial if either inequality in (iii) binds.

Between them, (3) and (4) cover a wide range of applications. Here are some examples:

- *Missing Data*. Say one observes not (Y, X), but (YZ, Z, X), where $Z \in \{0, 1\}$ and Y is observable only if $Z = 1$. Assuming that the distribution function of missing data can be bounded by c.d.f.'s (F_L^x, F_U^x), this maps onto (3) with $F_1^x = F(Y|Z=1, X=x)$ and $p^x = \Pr(Z=1|X=x)$. If the corresponding measure can be bounded, scenario (4) applies. Both cases will be encountered in Section 4.

An important special case of missing data occurs in treatment evaluation settings, namely when one is interested in the distribution of a potential outcome Y_t but only sees realizations of Y_t for subjects that receive treatment $T = t$. It is well known that worst-case bounds can be very wide here. They may be much tighter, however, if instruments are available, for example, because there is random assignment into treatment but nonrandom attrition. Recent work by Kitagawa (2009) extended results by Balke and Pearl (1997) to derive identified regions for $P(Y_t)$ for different instrumental variable assumptions. Many of these regions are defined by sandwich densities and, therefore, are special cases of (4). Separately, one may be willing to narrow bounds on $P(Y_t)$ by partially identifying assumptions. This is discussed in Section 4, where it leads to an application of (3).

- *Interval Data*. Say one observes random variables $(\underline{Y}, \overline{Y})$ s.t. $Y \in [\underline{Y}, \overline{Y}]$ a.s. (almost surely). This leads to bounds on distributions by setting $p = 1$, identifying F_U with the distribution of $\overline{Y}$, and identifying F_L with the distribution of $\underline{Y}$. The bounds will be tight whenever the observable intervals are ordered by strong set order; thus realizations $(\underline{y}_i, \overline{y}_i, \underline{y}_j, \overline{y}_j)$ are characterized by $(\underline{y}_i - \underline{y}_j)(\overline{y}_i - \overline{y}_j) \geq 0$ a.s., ensuring that (F_L, F_U) exhaust the partially identifying information. This holds if the observed intervals correspond to a preassigned partition of $\mathbb{R}$ (as with income data) or if one can bound measurement error by a number δ but is agnostic about its distribution; thus each measurement y^* implies that the corresponding realization of Y lies in $[y^* - \delta, y^* + \delta]$. This latter approach is common in the interval probabilities literature. The case of interval data that fail the ordering condition will be discussed later.

- *Estimating Games*. Haile and Tamer (2003) showed how moderate rationality restrictions on bidders' behavior yield quite narrow bounds (F_U, F_L) on the distribution functions of valuations. One can then conduct mean-variance (etc.) analysis subject to what is identified in their setting.

- *P-Boxes*. In computer science and related literatures, p-boxes, that is upper and lower bounds on c.d.f.'s, are a popular tool to model imprecise knowledge; see Ferson, Kreinovich, Ginzburg, Myers, and Sentz (2003) for a survey.

- *Contaminated and Corrupted Sampling*. Horowitz and Manski (1995) analyzed identification of D_1-parameters under contaminated and corrupted sampling. To extend their analysis, say one is interested in the measure P^* of $(Y|X=x)$, but only observes a contamination, that is, a mixture $P=\lambda P^*+(1-\lambda)\widetilde{P}$, where $\widetilde{P}$ is the contaminating measure. If one knows or imposes bounds on λ, one can learn about P^* because $P^*=(P-(1-\lambda)\widetilde{P})/\lambda \leq P/\lambda$; thus (4) applies with $p^x=0$, $P_L^x=0$, and $P_U^x=P/\lambda$. The analysis can be extended to the case of corrupted data, where data errors may arbitrarily depend on the true value of (Y,X). In this case, the object of interest is not P^*, but yet another measure $P^{**}=\lambda P^*+(1-\lambda)\overline{P}$. This leads to (looser) bounds of the form (4) on P^{**} if one can bound $\overline{P}$. Essentially the same bounds occur if one is interested in the measure of $(Y|X=x)$, the discrete regressor X is subject to misclassification, and one makes no assumptions other than bounding the probability of misclassification from above (Molinari (2008, Proposition 3)). More generally, however, misclassification of regressors will induce bounds on conditional distributions that do not fit this paper's framework.

- *Ecological Inference*. Suppose that X is discrete and that the marginal distributions of X and Y are identified, but that interest is in (elements of) $\{P(Y|X=x)\}_{x\in\mathcal{X}}$. Typical applications are "ecological inference" and "short" versus "long" regressions in the social sciences; see Manski (2007) for a discussion. The law of total probability,

$$P(Y)=\sum_{x\in\mathcal{X}} P(Y|X=x)\Pr(X=x), \tag{5}$$

implies (by reasoning very similar to the preceding bullet point) that $P(Y|X=x)\leq P(Y)/\Pr(X=x)$.[7] The implied bounds on $E(Y|X=x)$ were studied in detail by Cross and Manski (2002); Manski (2003) pointed out that they extend to D_1-parameters. A special feature of this setting is that the identified set for $H(\{P(Y|X=x)\}_{x\in\mathcal{X}})$ will be a small subset of $\bigtimes_{x\in\mathcal{X}}\{H(P(Y|X=x))\}$ because (5) generates rich cross-covariate restrictions. This will be elaborated in Section 5.

3. General identification results

In this and the next section, the covariate X will be dropped from notation. The resulting analysis applies both to unconditional bounds and to bounds that condition on X. Thus, in the case of bounds on c.d.f.'s, the identified set for F is

$$H(F)=\{F^*: pF_1+(1-p)F_U \succeq_{\text{FSD}} F^* \succeq_{\text{FSD}} pF_1+(1-p)F_L\}, \tag{6}$$

[7] Manski (2007) distinguished between a mixing covariate W and a conditioning covariate X. For simplicity and notational consistency, I relabel his W into X here and drop his X. (To reintroduce a covariate that plays the role of his X, simply condition every single expression in (5) on it.)

where (F_1, F_L, F_U) fulfil the regularity conditions listed before (3). In the case of bounds on measures, the identified set for P is

$$H(P) = \{P^* : pP_1 + (1-p)P_L \leq P^* \leq pP_1 + (1-p)P_U\} \tag{7}$$

with the regularity conditions listed after (4).

In either case, p may or may not be known.[8] The subsequent analysis presumes that it is. The correct adaptation to the case of unknown p depends on (F_1, F_L, F_U), respectively, (P_1, P_L, P_U). If $F_U \succeq_{\text{FSD}} F_1 \succeq_{\text{FSD}} F_L$, then $H(F)$ expands as p shrinks, and p should simply be set to its lowest possible value. If F_1 is not sandwiched in this way, then identified sets corresponding to different p need not be nested. One must then compute the union of the following bounds over all feasible p. Similarly, p should just be set to its lowest value if $P_U \geq P_1 \geq P_L$, and the analogous union must be formed otherwise.

Bounds on D_1-parameters in either setting are attained by shifting probability mass as far to the right or left as possible.

Lemma 1 (Bounds on D_1-Parameters). *Let θ be a D_1-parameter.*

(i) *Let $H(F)$ be as in (6) with p known. Then*

$$\theta(pF_1 + (1-p)F_L) \leq \theta(F) \leq \theta(pF_1 + (1-p)F_U).$$

(ii) *Let $H(P)$ be as in (7) with p known. Define the c.d.f.'s $(\underline{F}, \overline{F})$ by*

$$\underline{F}(y) = \min\bigl\{P_U((-\infty, y]), 1 - P_L((y, \infty))\bigr\},$$
$$\overline{F}(y) = \max\bigl\{P_L((-\infty, y]), 1 - P_U((y, \infty))\bigr\}.$$

Then

$$\theta(pF_1 + (1-p)\underline{F}) \leq \theta(F) \leq \theta(pF_1 + (1-p)\overline{F}).$$

All of these bounds are attainable.

The lemma is new in the form stated, but its intuition was anticipated in numerous special cases (DeRobertis and Hartigan (1981), Wasserman and Kadane (1992), Horowitz and Manski (1995), Manski (2003)). It mostly serves as a backdrop to the next and main finding, namely bounds on spread parameters. The intuition behind these is to push probability mass as far as possible to the edges of the support of Y. To formalize this, define the following.

Definition 4 (Compressed and Dispersed Distribution Functions). A c.d.f. F is compressed (relative to sandwich c.d.f.'s F_L, F_U) if there exists $a \in \mathbb{R} \cup \{-\infty, \infty\}$ s.t.

$$F(y) = \begin{cases} F_U(y), & y < a, \\ F_L(y), & y \geq a. \end{cases}$$

[8] If p is known, the definition of $H(P)$ may appear inefficient because one could simply drop p and P_1, and replace P_L with $pP_1 + (1-p)P_L$ and similarly for P_U. I use the above setup for notational unification, because p and P_1 may have substantive meaning (as in the next section), and to emphasize the possibility that p is unknown.

It is dispersed if there exists $a \in [0, 1]$ s.t.

$$F(y) = \begin{cases} F_L(y), & y < Q_L(a), \\ a, & Q_L(a) \le y < Q_U(a), \\ F_U(y), & Q_U(a) \le y. \end{cases}$$

Definition 5 (Compressed and Dispersed Measures). A measure P is compressed (relative to sandwich measures P_L, P_U) if there exist $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$, $a \le b$,[9] s.t.

$$P(A) = \begin{cases} P_U(A), & A \subseteq (a, b), \\ P_L(A), & A \subseteq \mathbb{R} - [a, b]. \end{cases}$$

It is dispersed if there exist $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$, $a \le b$, s.t.

$$P(A) = \begin{cases} P_U(A), & A \subseteq \mathbb{R} - [a, b], \\ P_L(A), & A \subseteq (a, b). \end{cases}$$

Compressed and dispersed objects are illustrated in Figures 1 (for distribution functions) and 2 (for measures). Many of the figures' simplifying features (e.g., that $Y \in [0, 1]$ and that all sandwich objects are continuous with full support) are not essential for the definitions. Note also that each of the above four collections of compressed or dispersed objects is ordered by first-order dominance. That is, all measures which are compressed relative to (P_L, P_U), say, are ordered in this way. The terms "higher" and "lower," when applied to compressed or dispersed objects, henceforth refer to this ordering. The c.d.f.'s and measures that attain the bounds in Lemma 1 are simultaneously dispersed and compressed, and are the lowest and highest objects of their kind. This section's core

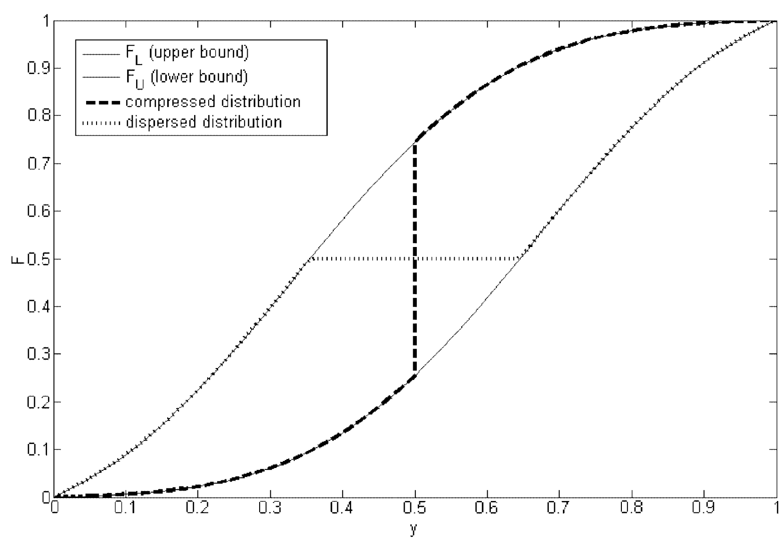

Figure 1. Compressed and dispersed c.d.f.'s.

[9]If $a = b$ (this can happen in examples involving large mass points), use the conventions $(a, b) = \emptyset$ and $[a, b] = \{a\}$.

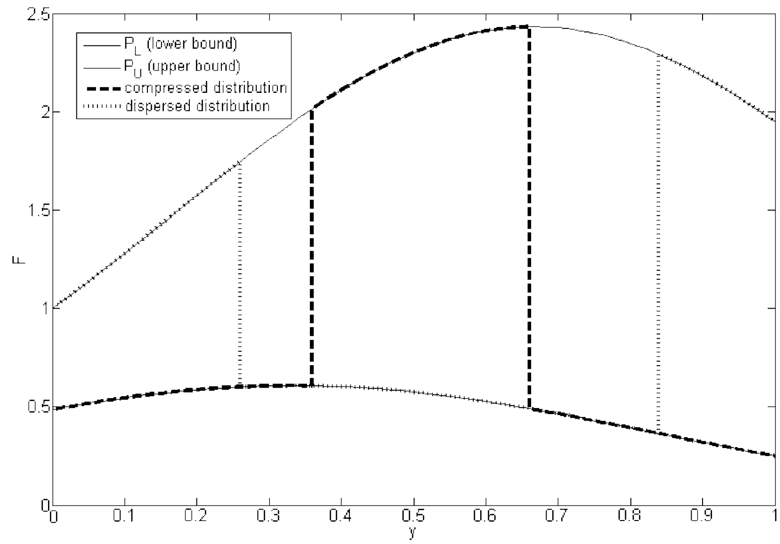

FIGURE 2. Compressed and dispersed measures.

insight is that intermediate compressed and dispersed objects trace out intermediate values of the expectation or any quantile of Y in a way that minimizes, respectively maximizes, spread. For bounds on D_2-parameters, a succinct statement obtains because each possible value of $E(Y)$ corresponds to exactly one compressed and dispersed object.

THEOREM 2 (Bounds on D_2-Parameters). *Let θ be a D_2-parameter and let $E(Y)=\mu$ for some preassigned $\mu \in H(E(Y))$.*

(i) *Let $H(F)$ be as in (6) with p known. Then*

$$\theta(pF_1+(1-p)\underline{F}_\mu) \leq \theta(F) \leq \theta(pF_1+(1-p)\overline{F}_\mu),$$

where the compressed c.d.f. $\underline{F}_\mu$ and the dispersed c.d.f. $\overline{F}_\mu$ are uniquely characterized by the condition that $\underline{E}_\mu(Y)=\overline{E}_\mu(Y)=(\mu-pE_1(Y))/(1-p)$.

(ii) *Let $H(P)$ be as in (7) with p known. Then*

$$\theta(pP_1+(1-p)\underline{P}_\mu) \leq \theta \leq \theta(pP_1+(1-p)\overline{P}_\mu),$$

where the compressed measure $\underline{P}_\mu$ and the dispersed measure $\overline{P}_\mu$ are uniquely characterized by the condition that $\underline{E}_\mu(Y)=\overline{E}_\mu(Y)=(\mu-pE_1(Y))/(1-p)$.

All of these bounds are attainable.

The corresponding result for quantile contrasts is more involved. I first give a precise statement and then explain the complications.

THEOREM 3 (Bounds on Quantile Contrasts). *Let $\theta=f(Q(\alpha),Q(\beta))$ be a quantile contrast, and let $Q(\gamma)=m$ for some preassigned $\gamma\in(\alpha,\beta)$ and $m\in H(Q(\gamma))$.*

(i) *Let $H(F)$ be as in (6) with p known. Then*

$$\begin{aligned}&\theta(pF_1+(1-p)\underline{F}_m)\leq\theta\\&\quad\leq\sup_{a\in[(\gamma-pF_1(m))/(1-p),(\gamma-pF_1^-(m))/(1-p)]\cap[0,1]}\theta(pF_1+(1-p)\overline{F}_a),\end{aligned}$$

where $\underline{F}_m$ is the compressed c.d.f. with threshold value $a=m$ and $\overline{F}_a$ is the dispersed c.d.f. with threshold value a. The lower bound is always attainable. If $p>0$ and F_1 has full support, the upper bound and all intermediate values are attainable as well. If F_1 is continuous, closed-form expressions for the bounds are

$$\begin{aligned}&f\left(\min\left\{Q_1\left(\frac{\alpha}{p}\right),m\right\},\max\left\{Q_1\left(1-\frac{1-\beta}{p}\right),m\right\}\right)\\&\quad\leq\theta\\&\quad\leq f\left(\max\left\{Q_1\left(F_1(m)-\frac{\gamma-\alpha}{p}\right),\underline{Q}(\alpha)\right\},\min\left\{Q_1\left(F_1(m)+\frac{\beta-\gamma}{p}\right),\overline{Q}(\beta)\right\}\right),\end{aligned}$$

where $\underline{Q}(\alpha)=\inf\{y: pF_1(y)+(1-p)F_L(y)\geq\alpha\}$ and $\overline{Q}(\beta)=\inf\{y: pF_1(y)+(1-p)\times F_U(y)\geq\beta\}$ are worst-case bounds on the respective quantiles.

(ii) *Let $H(P)$ be as in (7) with p known. Let $\underline{P}_u$ and $\overline{P}_u$ be the highest compressed and dispersed measures with*

$$\begin{aligned}&\underline{P}_u((-\infty,m])=\overline{P}_u((-\infty,m])\\&\qquad=\max\left\{\frac{\gamma-pP_1((-\infty,m])}{1-p},P_L((-\infty,m]),1-P_U((m,\infty))\right\},\end{aligned}$$

and let $\underline{P}_l$ and $\overline{P}_l$ be the lowest compressed and dispersed measures with

$$\begin{aligned}&\underline{P}_l((-\infty,m))=\overline{P}_l((-\infty,m))\\&\qquad=\min\left\{\frac{\gamma-pP_1((-\infty,m))}{1-p},P_U((-\infty,m)),1-P_L([m,\infty))\right\}.\end{aligned}$$

Then

$$\begin{aligned}&\inf_{\{P:P\text{ compressed},\ \underline{P}_u\succeq_{\mathrm{FSD}}P\succeq_{\mathrm{FSD}}\underline{P}_l\}}\theta(pP_1+(1-p)P)\\&\quad\leq\theta\leq\sup_{\{P:P\text{ dispersed},\ \overline{P}_u\succeq_{\mathrm{FSD}}P\succeq_{\mathrm{FSD}}\overline{P}_l\}}\theta(pP_1+(1-p)P).\end{aligned}$$

If the measure $(pP_1+(1-p)P_L)$ has full support, these bounds and all intermediate values are attainable. If $(pP_1+(1-p)P_U)$ is continuous, the compressed and dispersed measures generating the bounds are uniquely characterized by $P((-\infty,m])=(\gamma-pP_1((-\infty,m])/(1-p)$.

The added complications in this theorem arise as follows. Compressed and dispersed objects directly provide bounds on quantile contrasts subject to the constraint that $F(m) = \gamma$. But if distributions can have mass points and/or support gaps, then $F(m) = \gamma$ and $Q(\gamma) = m$ are nonnested conditions, and the above bounds are generated by evaluating all compressed or dispersed objects that ensure $F(m) \geq \gamma \geq F^{-}(m)$. In the presence of mass points, this may not pin down a unique compressed or dispersed object, necessitating the inf and sup operators. In the presence of support gaps, bounds may furthermore fail to be attainable for two reasons. First, θ may change discontinuously as one traces out compressed or dispersed objects; hence the sup and inf may not be attained, although the bound then remains best possible up to replacement of the weak with a strict inequality. Second, if m lies in a support gap, then $F(m) \geq \gamma = F^{-}(m)$ is consistent with $Q(\gamma) < m$; thus the object attaining the bound may induce a too low γ quantile. This second issue is entirely due to conventions about quantiles and disappears if one accepts a notion of set-valued quantiles. Furthermore, both problems disappear if Y is known to have full support; the sufficient conditions for attainability given in the theorem insure just that.

With the same exception of bounds on quantile contrasts in the presence of mass points and support gaps, joint identification regions are easily traced out by evaluating parameters on a one-dimensional grid of compressed and dispersed distributions. Constructing this grid by direct implementation of Definitions 4 and 5 is typically very fast, and repeated evaluation of parameters given c.d.f.'s or measures is trivial for many parameters.[10]

Of course, the tightness statements in Theorem 2 apply only if the information encoded in $H(P)$, respectively, $H(F)$, exhausts the available knowledge. Two examples where this fails are as follows:

- A randomized experiment identifies distributions of treatment outcomes Y_1 and control outcomes Y_0, but interest is in $\Delta = Y_1 - Y_0$. Fan and Park (forthcoming) imported results from the mathematical literature to provide bounds on the c.d.f. of Δ. Given that $E(\Delta)$ is identified, one can invoke Theorem 2 to bound D_2-parameters, but these bounds will not be tight because the relevant compressed and dispersed measures are not feasible for Δ. As Fan and Park (forthcoming, Lemma 2.2) pointed out, tight bounds are rather generated by assuming that Y_0 and Y_1 are perfectly positively (for minimal spread) or negatively (for maximal spread) dependent. Bounding quantile contrasts in this scenario appears harder. One can easily generate non-tight bounds by forming the set-valued difference of Fan and Park's (forthcoming) bounds for individual quantiles. Firpo and Ridder (2009) improved on this, but tight bounds are an open question.

- One observes intervals $[\underline{y}, \overline{y}]$ s.t. $y^* \in [\underline{y}, \overline{y}]$ a.s., but these need not be ordered, that is, some intervals may contain others. This problem has received attention in the interval probabilities literature. A formally equivalent problem that may be of interest to decision theorists is to bound parameters of distributions whose known features correspond to Dempster–Shafer structures. In this scenario, the problem of finding tight bounds becomes much harder. For the case of the variance, it was analyzed by Ferson, Ginzburg,

[10] MATLAB code implementing this paper's theorems is available from the author.

Kreinovich, Longpré, and Aviles (2005) and Kreinovich, Xiang, and Ferson (2006), who showed NP-completeness and provided algorithms.[11]

4. Application: Analysis of missing-data problems

This section illustrates the practical applicability of Theorems 2 and 3 (or, in one case, the adaptability of the underlying idea) by analyzing a missing-data problem. Thus, identify F_1 with the distribution of observable realizations of Y, identify F_0 with the distribution of unobservable ones, and identify $p = \Pr(Z = 1) \in (0, 1)$ with the fraction of observable data. The observability score p may or may not be identified. It is unidentified if Y is truncated, that is, missing data are not even recorded as missing. It is identified otherwise, a typical example being item nonresponse in surveys. This section's results apply directly if p is identified; remarks for the case of unknown p are as before.

The previous section's results do not apply if the unobserved distribution of Y is entirely unrestricted. I resolve this by assuming that $(Y|Z = 0)$ and $(Y|Z = 1)$ share a known, bounded support. While vacuous or extremely credible in many cases, this is a significant restriction in others. Its effect on (otherwise) worst-case bounds will be briefly discussed. Without further loss of generality, one can then project this support onto $[0, 1]$, and identify F_L and F_U with $F_L(y) = 1\{y \geq 0\}$ and $F_U(y) = 1\{y \geq 1\}$, where $1\{\cdot\}$ denotes the indicator function. This yields

$$H(F) = \{F^* : pF_1 + (1 - p)F_U \succeq_{\text{FSD}} F^* \succeq_{\text{FSD}} pF_1 + (1 - p)F_L\}$$

with (F_L, F_U) as just defined; thus Lemma 1(i), Theorem 2(i), and Theorem 3(i) yield bounds on D_1-parameters and spread parameters. The simple shape of $H(F)$ frequently allows for closed-form expressions, and for specific parameters, one may also be able to conclude that $H(\theta) = [\underline{\theta}, \overline{\theta}]$. Recalling the conventions that $Q(\alpha) = 0$ for $\alpha \leq 0$ and $Q(\alpha) = 1$ for $\alpha > 1$, one can state the following Corollary.

Corollary 4 (Exact Identification Regions for Some Parameters). *Consider the missing-data scenario with $Y \in [0, 1]$ and p known.*

(i) *The identification regions for the expectation $E(Y)$ and α quantile $Q(\alpha)$ are*

$$H(E(Y)) = [pE_1(Y), pE_1(Y) + 1 - p],$$
$$H(Q(\alpha)) = \left[Q_1\left(1 - \frac{1 - \alpha}{p}\right), Q_1\left(\frac{\alpha}{p}\right)\right].$$

(ii) *If $E(Y) = \mu$ for some $\mu \in H(E(Y))$, the identification region for the variance $V(Y)$ is*

$$H(V(Y)) = [pE_1(Y^2) + (1 - p)\mu_0^2 - \mu^2, pV_1(Y) + \mu - \mu^2]$$

[11] It seems clear that their algorithm would also apply to the parameters considered here. Conversely, this paper's analysis shows that their problem much simplifies in some settings of interest to them.

and the identification region for the Gini coefficient $G(Y)$ *is*

$$\begin{aligned} H(G(Y)) = \Bigg[& p^2 G_1(Y) + \frac{1}{\mu} p(1-p) E_1(|Y-\mu_0|), \\ & p^2 G_1(Y) + \frac{1}{\mu}\{p(1-p)\big[(1-\mu_0)E_1(Y) + \mu_0(1-E_1(Y))\big] \\ & + (1-p)^2 \mu_0(1-\mu_0)\}\Bigg], \end{aligned}$$

where $\mu_0 = (\mu - pE_1(Y))/(1-p)$.

(iii) *Assume furthermore that* F_1 *is continuous with full support. If* $Q(\gamma) = m$ *for some preassigned* $\gamma \in (\alpha, \beta)$ *and* $m \in H(Q(\gamma))$, *then the identification region for a quantile contrast* $\theta = f(Q(\alpha), Q(\beta))$ *is*

$$\begin{aligned} H(\theta) = \Bigg[& f\left(\min\left\{Q_1\left(\frac{\alpha}{p}\right), m\right\}, \max\left\{Q_1\left(1-\frac{1-\beta}{p}\right), m\right\}\right), \\ & f\left(Q_1\left(F_1(m) - \frac{\gamma-\alpha}{p}\right), Q_1\left(F_1(m) + \frac{\beta-\gamma}{p}\right)\right)\Bigg]. \end{aligned}$$

The imputations of missing data that generate these bounds are very intuitive: Spread parameters are maximized by identifying F_0 with point masses at 0 and 1, and are minimized by identifying it with a single, appropriately placed point mass.

While Y is here assumed to be bounded, the same or other (larger, but still nontrivial) bounds may apply when this fails. For example, say that Y denotes income and one is willing to bound it from below (by zero) but not from above. Then upper bounds on the expectation and unconstrained upper bounds on variance and Gini coefficient are vacuous, but constrained lower bounds on spread parameters are unchanged. Furthermore, while there is no constrained upper bound on the variance, weaker but still informative constrained upper bounds on the Gini coefficient can be computed.[12] What is more, they increase in μ, which is helpful if one is willing to trade off some increase in inequality for an increase in the average; see the empirical illustration for more on this. Things look different yet for quantile contrasts, bounds on which are independent of any assumptions about $\operatorname{supp}(Y)$ if p is large enough.

These observations compare to the finding that inequality rankings which agree with second-order dominance are nonrobust in the classical sense, that is, their influence functions are not uniformly bounded (Cowell and Victoria-Feser (1996, 2002)), whereas interior quantiles, and hence quantile contrasts, are robust. Bounds analysis leads to a more differentiated picture: both variance and Gini coefficient are nonrobust, yet something can be said about upper bounds on the latter, and lower bounds are less affected anyway. Nonetheless, robustness and partial identification analyses are complementary in highlighting that popular measures of spread are highly sensitive to data quality problems.

[12]The closed-form expression is $G \le [p^2 E_1(|Y_i - Y_j|) + 2p(1-p)(\mu_1 + \mu_0) + 2(1-p)^2\mu_0]/2\mu$.

4.1 *Refining missing-data bounds with monotonicity assumptions*

Worst-case bounds are instructive but can be too wide to be of much use for a practitioner. At the same time, they are often generated by implausible imputations of missing data. One might, therefore, want to introduce partially identifying restrictions that reduce identified sets (although not necessarily to singletons) at the price of additional assumptions (but possibly quite weak ones).

I now analyze two such assumptions that capture positive [negative] selection into observation. The measurement of income distributions is an obvious application; consequently, assumptions in this spirit were investigated by authors interested in income inequality (Vazques Alvarez, Melenburg, and van Soest (1999), Blundell et al. (2007)). For example, the distribution of potential wages in the labor force might plausibly dominate its distribution in the population. Another salient example is analysis of potential outcomes when treatment (say, schooling) is selective along an unobserved covariate (say, ability) that influences Y (say, income). In such cases, one might be willing to impose the following assumptions.

Assumption 1 (Dominant Selection).

$$F_1 \succeq_{\text{FSD}} F_0.$$

Assumption 2 (Monotone Selection).

$$\Pr(Z = 1 | Y = y) \textit{ is nondecreasing in } y.$$

Both assumptions impose some kind of positive selection into observation; negative selection can be modelled by reversing them. Under dominant selection, the observed and unobserved distributions are ordered by first-order dominance. Monotone selection contrasts in several ways: First, it is easily seen to be stronger. In fact, it is equivalent to positive likelihood ratio dependence (in the language of copula theory), respectively, affiliation (in the language of auction theory), between Y and Z; in both contexts, that would be the most restrictive among numerous widely used notions of positive dependence (de Castro (2009), Nelsen (2006)). At the same time, monotone selection may be almost as plausible as dominant selection in many missing-data applications. It is directly expressed as an intuitive restriction on response probabilities, and it is easy to write out models of nonresponse that imply it. A symptom of this is that, as will be seen, imputations of missing data that are consistent with dominant but not monotone selection do not seem to capture positive selection.[13]

[13]Blundell et al. (2007) used dominant selection to refine unconditional bounds on the interquantile range, an analysis that is generalized here.

The implications of dominant selection follow from Lemma 1(i), Theorem 2(i), and Theorem 3(i) upon writing the new identified set for F as

$$H(F) = \{F^* : F_1 \succeq_{\text{FSD}} F^* \succeq_{\text{FSD}} pF_1 + (1-p)F_L\}.$$

This leads to closed-form results along the lines of Corollary 4, which are here omitted for brevity but are available from the author.

The upper and lower bounds on D_1-parameters are generated by identifying F_0 with F_L, respectively, F_1. Both of these imputations are consistent with monotone selection, and, hence, monotone selection does not lead to narrower upper and lower bounds on D_1-parameters.[14] Things change, however, when one looks at spread parameters. The implications of monotone selection for these do not follow from Theorems 2 and 3 because monotone selection introduces information that does not neatly fit either (6) or (7). The organizing principles advertised in this paper are still useful, however. Bounds on spread parameters will again be traced out by compressed and dispersed distribution functions; the challenge is to identify these. The solution is as follows.

Definition 6 (Compressed and Dispersed Distribution Functions With Monotone Selection). A c.d.f. F is compressed (relative to monotone selection) if

$$F(y) = \min\left\{\frac{1}{a}F_1(y), 1\right\}$$

for some $a \in [0, 1]$, where $a = 0$ is understood to represent $F = 1\{y \geq 0\}$.

It is dispersed (relative to monotone selection) if

$$F(y) = a + (1-a)F_1(y)$$

for some $a \in [0, 1]$.

Then one can state the following theorem:

Theorem 5 (Bounds on Spread Parameters With Monotone Selection). *Consider the missing-data scenario with $Y \in [0, 1]$ and p known. Assume monotone selection. Let compressed and dispersed c.d.f.'s be understood as just defined.*

(i) *Let $E(Y) = \mu$ for some preassigned $\mu \in H(E(Y))$ and let θ be a D_2-parameter. Then*

$$\theta(pF_1 + (1-p)\underline{F}_\mu) \leq \theta(F) \leq \theta(pF_1 + (1-p)\overline{F}_\mu),$$

where the compressed c.d.f. $\underline{F}_\mu$ and the dispersed c.d.f. $\overline{F}_\mu$ are uniquely characterized by the condition that $\underline{E}_\mu(Y) = \overline{E}_\mu(Y) = (\mu - pE_1(Y))/(1-p)$. These bounds are attainable.

[14] Monotone selection may lead to bounds that are smaller in the sense of excluding intermediate values. This is because it implies that P_0 is absolutely continuous with respect to P_1 except possibly on $\{0\}$. For example, only a point $y > 0$ that is *some* quantile of $(Y|Z = 1)$ can be *any* quantile of Y.

(ii) *Let $\theta = f(Q(\alpha), Q(\beta))$ be a quantile contrast, and let $Q(\gamma) = m$ for some preassigned $\gamma \in (\alpha, \beta)$ and $m \in H(Q(\gamma))$. Then*

$$\inf_{\substack{a \in [(1-p)F_1^-(m)/(\gamma - pF_1^-(m)),\\ ((1-p)F_1(m))/(\gamma - pF_1(m))] \cap [0,1]}} f\left(Q_1\left(\frac{a\alpha}{ap+1-p}\right), \max\left\{Q_1\left(\frac{a\beta}{ap+1-p}\right), Q_1\left(1-\frac{1-\beta}{p}\right)\right\}\right)$$
$$\leq \theta \leq \sup_{\substack{\pi \in [(\gamma - F_1(m))/(1-F_1(m)),\\ (\gamma - F_1^-(m))/(1-F_1^-(m))] \cap [0,1]}} f\left(Q_1\left(\frac{\alpha-\pi}{1-\pi}\right), Q_1\left(\frac{\beta-\pi}{1-\pi}\right)\right).$$

If F_1 is continuous, closed-form expressions for the bounds are

$$f\left(Q_1\left(\frac{\alpha F_1(m)}{\gamma}\right), \max\left\{Q_1\left(\frac{\beta F_1(m)}{\gamma}\right), Q_1\left(1-\frac{1-\beta}{p}\right)\right\}\right)$$
$$\leq \theta \leq f\left(Q_1\left(\frac{(1-\alpha)F_1(m)+\alpha-\gamma}{1-\gamma}\right), Q_1\left(\frac{(1-\beta)F_1(m)+\beta-\gamma}{1-\gamma}\right)\right)$$

and they are attainable. If F_1 has full support, the bounds as well as all intermediate values are attainable.

It is instructive to consider the selection models that correspond to compressed and dispersed c.d.f.'s. If F_0 is compressed, then $\Pr(Z=1|Y=y)$ equals some constant less than 1 for $y < Q_1(a)$ and equals 1 for $y > Q_1(a)$. If F_0 is dispersed, then $\Pr(Z=1|Y=y)$ is low (potentially zero) for $y=0$ and equal to a larger constant for all $y>0$. If one is committed to the general idea of positive selection but not to any parametric model, then these selection models might be hard to reject (give or take some smoothing). In contrast, compressed and dispersed c.d.f.'s under dominant selection correspond to "hump-shaped" selection models, where $\Pr(Z=1|Y=y)$ is low for high as well as low y and high for intermediate y or vice versa. This might not be what a researcher has in mind when specifying positive selection. Thus, the added power of monotone selection might come at a low cost in terms of plausibility.

4.2 *Refining missing-data bounds with sensitivity assumptions*

In this section, bounds will be refined using sensitivity assumptions, leading to an application of Theorems 2(ii) and 3(ii). Specifically, say a researcher is willing to bound a selection model's potential to distort the true probability measure P.

Assumption 3 (Limited Selectivity (LS(k))).

$$\frac{\Pr(Z=1|Y=y)}{\Pr(Z=1)} \geq \frac{1}{k}.$$

This assumption bounds from below the volatility of $\Pr(Z=1|Y=y)$ as a function of y; thus, it restricts the potential of selection effects to distort odds ratios between events.[15] For a slightly more "structural" way to assert it, one could formulate the assumption in terms of selection on unobservables. To do so, let there be a covariate W with values w and assume (i) limited selectivity on W, that is, $\Pr(Z=1|W=w)/\Pr(Z=1) \geq 1/k$, and (ii) that W does not affect Y except through Z, that is, $P(Y|W=w, Z=1) = P(Y|W=w, Z=0)$. It is easy to show that these conditions imply $\mathrm{LS}(k)$.

The restrictiveness of $\mathrm{LS}(k)$ is parameterized by k. For $k \to \infty$, the assumption becomes very weak and frequently vacuous.[16] For $k \to 1$, it imposes vanishing selectivity of observations. The exact interpretation of this limit depends on the treatment of X: If the analysis *conditions on* X, then $\mathrm{LS}(1)$ is equivalent to assuming that data are missing at random given X, that is, assumption MAR. If the analysis *ignores* X, then $\mathrm{LS}(1)$ is equivalent to assuming that data are missing completely at random, that is, assumption MCAR. In either case, small choices of k amount to local sensitivity analysis, whereas large choices turn it substantively into a partially identifying assumption. Note also that any statement of the form $\theta \in \Theta_0$ that is true of $\theta(P_1)$ will be true for every $\theta \in H(\theta)$ if $\mathrm{LS}(k)$ is imposed with k small enough (possibly $k=1$). For any such statement, one can therefore define a breakdown point k^* as the largest value of k s.t. $\mathrm{LS}(k)$ suffices to conclude $H(\theta) \subseteq \Theta_0$. Statements about θ can then be ranked by robustness to selectivity of observations.

Sensitivity assumptions that can be scaled from highly identifying to highly credible and notions of breakdown points have been entertained before. The best known precedent may be Rosenbaum's (2002) sensitivity analysis for propensity score methods; see Imbens (2003) for an application to economics. The bound λ on contamination probabilities in Horowitz and Manski (1995) can be interpreted in a similar fashion; see Kreider and Pepper (2007) for an application.

The implications of $\mathrm{LS}(k)$ for bounds on D_1 and spread parameters follow from Lemma 1(ii), Theorem 2(ii), and Theorem 3(ii) upon the following observation.

Lemma 6. *Assumption* $\mathrm{LS}(k)$ *implies that scenario* (7) *applies with* $P_L = 0$ *and* $P_U = P_1 \cdot (k-p)/(1-p)$, *implying that* $H(P)$ *simplifies to*

$$H(P) = \{P^* : pP_1 \leq P^* \leq kP_1\}.$$

The simple form of $H(P)$ allows statement of bounds on numerous salient parameters in (tedious) closed form. These results are omitted here for brevity but are available from the author.

[15]Intuitively, one might be tempted to restrict $\Pr(Z=1|Y=y)$ independently from $\Pr(Z=1)$, that is, write $\Pr(Z=1|Y=y) \geq 1/k$. But this assumption's strength depends on p and, counterintuitively, eventually leads to stronger identification as *p decreases*. Indeed, as $p \downarrow 1/k$, the restriction that $\Pr(Z=1|Y=y) \geq 1/k$ becomes point identifying.

[16]More precisely, $\mathrm{LS}(k)$ with $k \to \infty$ approaches the claim that the $\mathrm{supp}(P_0) \subseteq \mathrm{supp}(P_1)$: For any measurable A, $P(A) > 0 \Leftrightarrow P_1(A) = (\Pr(Z=1|A)P(A))/\Pr(Z=1) \geq P(A)/k > 0$ for any finite k. Any other implication depends on limiting k. In many cases, $\mathrm{LS}(\infty)$ is, therefore, either vacuous (if the observed support of P_1 exhausts the logical range of Y) or was imposed for tractability anyway.

Table 1. Descriptive statistics for data used.

Data Set	CPS	ACS
Sample size	30,343	433,388
Percent allocated	1.28	4.65
Mean	55,874	61,453
Median	46,096	47,323
Gini coefficient	0.35	0.40
Interquartile ratio	2.48	3.04

4.3 *Numerical illustration with income data*

Here is a brief illustration of this section's analysis using real data. The substantial quest is to investigate what can be learned about income distributions from survey data if one is agnostic about missing data generated by item nonresponse. Imputations of income data typically rely on assumption MAR as the identifying restriction, yet one might easily imagine that even after conditioning on covariates, high income respondents are more likely to not respond to income items. Sensitivity to this will now be assessed through this section's bounds, demonstrating how the bounds work and revealing that some (not all) statistics are more tightly identified than one might have thought.

Bounds were computed for two different data sets, both of which collect the 2007 wage income of U.S. men aged 30–60 who were working for an employer (not self-employed, unemployed, or in the military) at the time. The first data set is from the 2007 March supplement of the Current Population Survey (CPS); the second set is from the 2007 American Community Survey (ACS). The first of these is the classic source for U.S. income data; the second one is relatively new and part of a new data generation program by the U.S. Census Bureau.[17] They are of very different size, with the CPS sample containing $n = 30{,}343$ observations and the ACS sample containing $n = 433{,}388$ observations. At the same time, the ACS has a larger fraction of allocated wage income data (4.65% allocated) than the CPS (1.28% allocated). Table 1 collects some simple descriptive statistics for both data sets.

Figures 3–6 display joint identification regions for the median and interquartile ratio, as well as the mean and Gini coefficient, if one takes allocated income data to be missing. All figures were extremely easy to compute. The outer bounds are worst-case bounds, which are then successively refined. The idea behind the refinements is twofold: For one thing, it is commonly believed that high income earners may be less likely to respond to income questions because these questions are more sensitive for them. This idea is captured by imposing first dominant and then monotone selection. Note that these are imposed in the sense of the unobserved distribution being higher than the observed one, the opposite of the definitions used in Section 4.1. Joint bounds on mean and Gini coefficient are refined using LS(10), LS(2), and LS(1.1). The first of these appears to be a very weak assumption that illustrates partial identification analysis; the last one may border on local sensitivity analysis with respect to LS(1), that is, assumption MCAR. The

[17] Both data sets were acquired via IPUMS, see King et al. (2009), respectively, Ruggles et al. (2009).

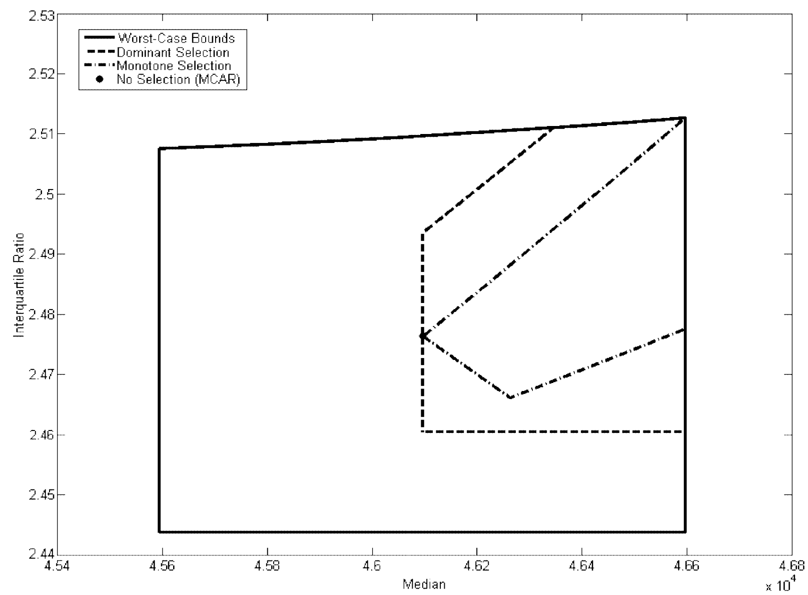

FIGURE 3. Joint identification of median and interquartile ratio (CPS data)

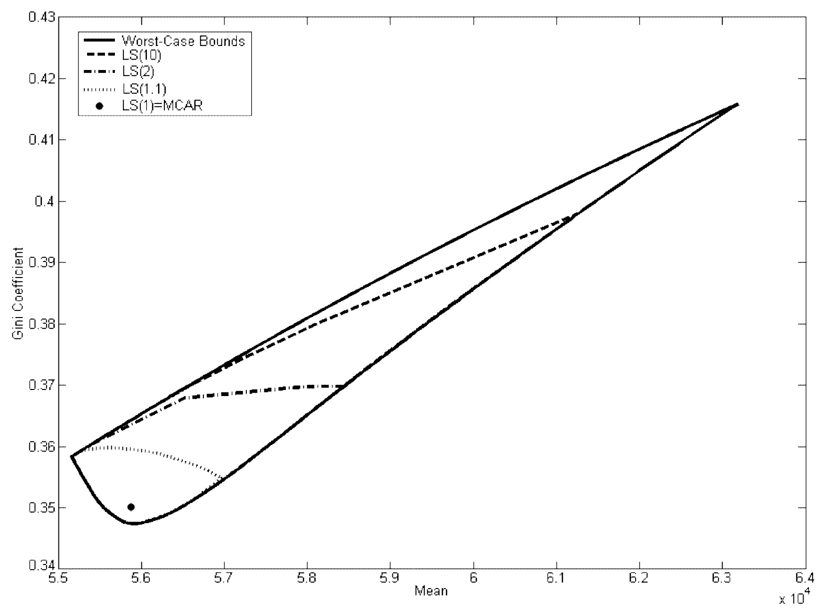

FIGURE 4. Joint identification of mean and Gini coefficient (CPS data).

implications of assumption MCAR are also displayed in each figure; they are simply the relevant parameters of F_1 (i.e., computed after discarding allocated data points).

Worst-case bounds on median and interquartile ratio (IQR) are reasonably informative, illustrating the comparative robustness of these statistics. Dominant and monotone selection obviously imply that the median of F must exceed the median of F_1, but

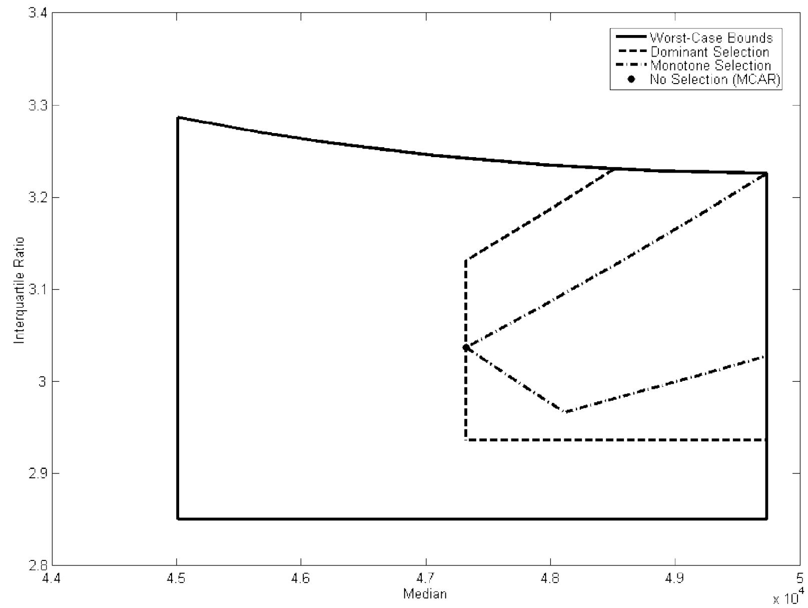

FIGURE 5. Joint identification of median and interquartile ratio (ACS data).

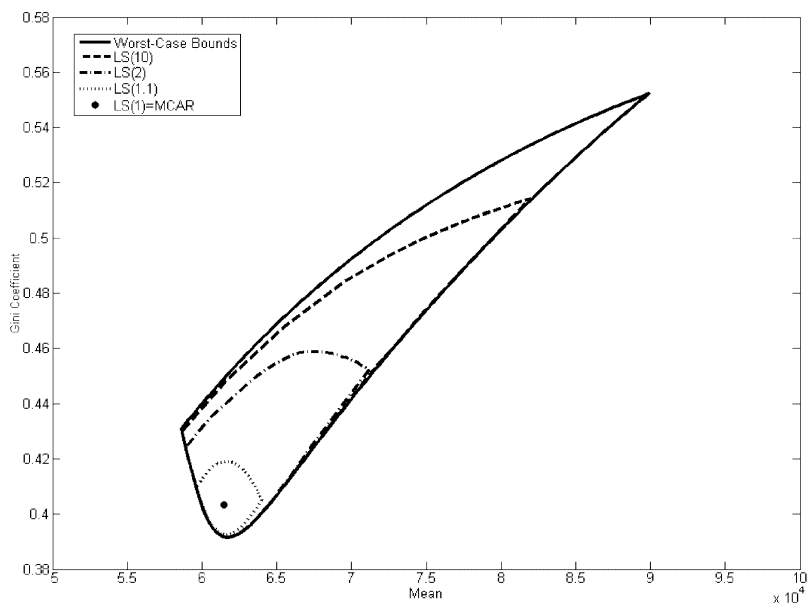

FIGURE 6. Joint identification of mean and Gini coefficient (ACS data).

they also substantially refine bounds on the interquartile ratio. Also, monotone selection, though perhaps only marginally less credible than dominant selection, has substantially more identifying power. Unconstrained bounds on the mean and Gini coefficient are discouraging (and include values that could be excluded through other data sources), reflecting the robustness issues discussed in Section 3. However, joint bounds on mean and Gini coefficient are a small subset of the Cartesian product of unconstrained bounds. This illustrates that joint bounds can add much information. In the particular case, high feasible values of the mean correspond to high values of the Gini coefficient, which is interesting because one would presumably be willing to trade off increases in one versus increases in the other at some rate (Klasen (1994), Sen (1976)). Furthermore, $\mathrm{LS}(k)$ has a large effect even with rather conservative choices of k. An intuition for this is that dispersed measures push probability mass into the tails, but if those tails are very thin in the observable distributions, then limited selectivity drastically limits the unobserved probability mass that can reside in them. This observation is expected to generalize, that is, limited selectivity can be a powerful refinement if bounds are driven by what is going on in a distribution's tails.

Two caveats to this exercise should be kept in mind. First, the displayed combinations of assumptions and parameters are selected for "what works." Dominant and monotone selection do not much affect the mean–Gini bounds because they fail to exclude the distributions that generate the top right region of those bounds. At the same time, limited selectivity does not much affect the median–IQR bounds because these are insensitive to the thickness of a distribution's tails.

Second, the figures display empirical analogs of identified sets. As this paper is about identification and because the example is meant to be merely illustrative, samples were taken to be the populations of interest. Insofar as one thinks of the displayed objects as plug-in estimators of identified sets, the impact of sampling uncertainty is not shown.[18] In particular, the highest observed income in either data set is taken to be the highest possible income. The shape of the top right "spikes" of bounds of mean and Gini coefficient is sensitive to this choice. It was made here to isolate the precise effect of limiting likelihood ratio distortions, but it means that the top right regions of the mean–Gini bounds are dubious as estimators of the corresponding regions of population bounds. (Estimating these regions with reasonable accuracy would require a high income oversample.) These considerations less affect the median–IQR bounds, which are insensitive to changes in the tails of the underlying distributions. More generally, ignoring estimation penalties favors the CPS, which delivers tighter bounds because of the smaller proportion of allocations, but would deliver larger standard errors due to its smaller sample size. Indeed, the figures quantitatively illustrate how, while the ACS has a much larger sample size, its larger fraction of imputations implies that conclusions drawn from it will be more dependent on identifying assumptions.

5. Using cross-covariate restrictions

The preceding results restrict parameters conditionally on (or absent) X, using only observable features of the corresponding conditional distributions. They immediately im-

[18]The mean–Gini bounds literally use the sample distributions. For the median–IQR bounds, these distributions were kernel smoothed to remove uninstructive wiggliness in the identified sets' boundaries and to render constrained bounds convex. Confidence regions could, in principle, be constructed using the methodology of Chernozhukov, Hong, and Tamer (2007), but carrying this out is beyond the scope of this paper.

ply that the identified set for $\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}}$ is constrained as

$$H\big(\big\{\theta(P(Y|X=x))\big\}_{x\in\mathcal{X}}\big)\subseteq \bigtimes_{x\in\mathcal{X}}\{H^x(\theta)\}.$$

If there are no cross-covariate restrictions, this set inclusion will be weak. In many other cases, it will be strict. In particular, strict inclusion would mean that set-valued differences between this paper's bounds on $\theta(P(Y|X=x))$ and $\theta(P(Y|X=x'))$ need not be tight for the contrast $\theta(P(Y|X=x))-\theta(P(Y|X=x'))$ or other functions of the two. A good example is partial identification of treatment effects with random assignment but selective noncompliance, where the set-valued difference between worst case bounds for $E(Y_0)$ and $E(Y_1)$ is not, in general, tight for $E(Y_1-Y_0)$ (Balke and Pearl (1997)). Another important example is given by short versus long regressions, where the law of total probability yields intricate adding-up constraints on $\{P(Y|X=x)\}_{x\in\mathcal{X}}$. In this case, numerical characterization of $H(\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}})$ poses no conceptual problems. Letting $\mathcal{X}=\{x_1,\dots,x_J\}$, one first identifies $H(\theta(P(Y|X=x_1)))$, then $H(\theta(P(Y|X=x_2)))$ as a function of $\theta(P(Y|X=x_1)$, and so on. However, the computational cost of this procedure may be high. Also, one may wonder if there is a useful characterization of $H(\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}})$ that goes beyond stating this procedure. The answer is affirmative for some well behaved statistics, for example, expectations of convex functions, that are subject to the findings in Molinari and Peski (2005), but I am not aware of an interesting characterization of $H(\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}})$ that would apply at this paper's level of generality.

In addition to cross-covariate restrictions that are logically implied by the identification problem, one may want to exploit the presence of covariates to refine bounds. Most interestingly, cross-covariate restrictions may refine conditional bounds, that is, they may shrink $H(\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}})$ to the point that implied bounds on $\theta(P(Y|X=x))$ are smaller than $H^x(\theta)$. Two examples are as follows. First, assume that $P(Y|X=x)$ does not depend on x; thus X is an instrumental variable in the sense of Manski (1990). This easily implies that $H(\{\theta(P(Y|X=x))\}_{x\in\mathcal{X}})=\{\{\theta\}_{x\in\mathcal{X}}:\theta\in\bigcap_{x\in\mathcal{X}}H^x(\theta)\}$; thus conditional identified sets can be tightened to

$$H^x(\theta)=\bigcap_{x\in\mathcal{X}}H^x(\theta).$$

The assumption is testable in the sense of being potentially inconsistent with observable aspects of P because $\bigcap_{x\in\mathcal{X}}H^x(\theta)$ could be the empty set. It can also be weakened, for example, by assuming that $\theta(P(Y|X=x))$ weakly increases in x. (This requires the set $\mathcal{X}$ to be ordered.) Then one has

$$\begin{aligned}&H\big(\big\{\theta(P(Y|X=x))\big\}_{x\in\mathcal{X}}\big)\\&\quad=\Bigg\{\big\{\theta(P(Y|X=x))\big\}_{x\in\mathcal{X}}\in\bigtimes_{x\in\mathcal{X}}H^x(\theta):\\&\qquad\theta(P(Y|X=x))\text{ is nondecreasing in }x\Bigg\};\end{aligned}$$

thus nonmonotonicities in the conditional bounds can be "ironed out." For bounds on expectations, this was pointed out by Manski and Pepper (2000). The assumption is again testable; furthermore, one can easily think of extensions like monotone concave instrumental variables and so on.

This paper's findings can also be combined with cross-covariate restrictions in more intricate ways. Consider a treatment evaluation scenario; thus the covariate is a (binary, for simplicity) treatment $T \in \{0, 1\}$. Identification of potential outcome distributions is subject to a missing-data problem because all counterfactual realizations are missing. But one could refine worst-case bounds by adapting assumptions like monotone treatment response (Manski (1997)) and monotone treatment selection (Manski and Pepper (2000)). An easy adaptation would be along the lines of the preceding paragraph, defining monotone treatment response as $\theta(P(Y_1|T = t)) \geq \theta(P(Y_0|T = t))$ and monotone treatment selection as $\theta(P(Y_t|T = 1)) \geq \theta(P(Y_t|T = 0))$, where these expressions are understood to hold for all t. Refinements of bounds would follow easily. These assumptions might not be well motivated, however; why would one impose that treatment increases the spread of a variable, say? For a more plausible adaptation that also relates more interestingly to this paper's results, define monotone treatment response as $F(Y_1|T = t) \succeq_{\text{FSD}} F(Y_0|T = t)$ and monotone treatment selection as $F(Y_t|T = 1) \succeq_{\text{FSD}} F(Y_t|T = 0)$. Alternatively, one might want to assume that subjects select into the treatment that is better for them; thus $F(Y_t|T = t) \succeq_{\text{FSD}} F(Y_t|T = 1 - t)$. The implied refinements of bounds on spread parameters of $P(Y_0)$ and $P(Y_1)$ would follow from Theorem 2 because these assumptions' content is exhausted by the implied upper and/or lower bounds on $F(Y_1|T = 0)$ and $F(Y_0|T = 1)$. Indeed, the "Roy model" assumption of selection into the individually better treatment would recover dominant selection. In the absence of instrumental variables or additional restrictions, the implied bounds on parameter contrasts would be tight as well.

6. Conclusion

This paper investigated problems of partial identification, which occur whenever a parameter of interest is not fully determined by a data generating process and the assumptions a researcher is willing to make. The main contribution was to show how a few general ideas can be used to much extend the scope of this literature by providing "off-the-shelf" and closed-form or easily computed bounds for many parameters. The "recipe" for these bounds can be summarized in three steps: First, recognize that mean-preserving spreads and quantile contrasts are powerful organizing principles. Second, combine these principles with appropriate constraints. For mean-preserving spreads, this entails fixing a distribution's mean; for quantile contrasts, it entails fixing some quantile. Third, it may now be possible to discover compressed and dispersed distributions that extremize the parameters of interest. The general results in Section 3 characterized these objects for two rather general settings. The result on monotone selection suggests that it may be feasible to do so in other, trickier scenarios as well.

Special emphasis was placed on missing-data problems as a potential application. I provided worst-case bounds for such problems if a random variable is bounded and refined the bounds by restricting the direction (dominant selection, monotone selection) or the extent (limited selectivity) of selection into observation. The findings were illustrated in a brief empirical example. It is hoped that they further enhance the appeal of partial identification analysis.

Appendix: Proofs

Proof of Lemma 1. (i) Straightforward.

(ii) To see validity of the bounds, note that

$$P_L((-\infty, y]) \leq P_0((-\infty, y]) \leq P_U((-\infty, y]),$$
$$1 - P_U((y, \infty]) \leq 1 - P_0((y, \infty]) \leq 1 - P_L((y, \infty]).$$

Thus $\overline{F} \succeq_{\text{FSD}} F_0 \succeq_{\text{FSD}} \underline{F}$. The bounds are attainable because the implied measures $\underline{P}$ and $\overline{P}$ are sandwiched between P_L and P_U. □

Proof of Theorem 2. A word on notation: For any measures P and P' and event E, the phrase "$P \geq P'$ on E" means that $P(A) \geq P'(A)$ for any event $A \subseteq E$. The following fact will be used in this and subsequent proofs.

Lemma 7. *Suppose that X and Y are random variables with $E(X) = E(Y)$, c.d.f.'s F_X and F_Y, and measures P_X and P_Y s.t. one of the following statements holds*:

(i) *There exists $a \in \mathbb{R}$ s.t.*

$$F_X(x) \leq F_Y(x) \quad \forall x < a,$$
$$F_X(x) \geq F_Y(x) \quad \forall x > a.$$

(ii) *There exist $a, b \in \mathbb{R}$ s.t.*

$$P_Y \leq P_X \quad \text{on } (a, b),$$
$$P_Y \geq P_X \quad \text{on } \mathbb{R} - [a, b].$$

Then $F_X \succeq_{\text{SSD}} F_Y$, where $\succeq_{\text{SSD}}$ denotes second-order dominance as defined in (2).

For the proof see Shaked and Shantikumar (2007, Theorem 3.A.44).

(i) F_L and F_U are simultaneously compressed and dispersed, and uniquely attain the lower and upper bounds on $E(Y)$; thus the claim is immediate if μ equals one of these bounds. Now let μ assume an intermediate value. To see that $\underline{F}_\mu$ is well defined, let F_a be the compressed c.d.f. with threshold value $a \in \mathbb{R}$. Recall that if Y is nonnegative, then $E(Y) = \int_0^\infty (1 - F(y))\, dy$ (e.g., Billingsley (1995)). Defining $Y^+ = \max\{Y, 0\}$ and $Y^- = \min\{Y, 0\}$, one, therefore, has $E(Y^+) = \int_0^\infty (1 - F(y))\, dy$ and $E(Y^-) = -\int_{-\infty}^0 F(y)\, dy$. Thus,

$$E_a(Y) = -\int_{-\infty}^{\min\{a,0\}} F_U(y)\, dy - \int_{\min\{a,0\}}^0 F_L(y)\, dy$$
$$+ \int_0^{\max\{a,0\}} (1 - F_U(y))\, dy + \int_{\max\{a,0\}}^\infty (1 - F_L(y))\, dy,$$

which is finite by assumptions on F_L and F_U, and is easily seen to be continuous and nondecreasing in a as well as to converge to $E_L(Y)$ $[E_U(Y)]$ as $a \to -\infty$ $[a \to \infty]$. Existence of $\underline{F}_\mu$ follows immediately; uniqueness obtains because any two different compressed c.d.f.'s are ordered by first-order stochastic dominance and, therefore, have different expectations. (The threshold a characterizing $\underline{F}_\mu$ may not be unique, namely if $F_L = F_U$ over some interval.)

$\underline{F}_\mu$ attains the lower bound on θ; it remains to show that the bound is valid. Thus, compare $pF_1 + (1-p)\underline{F}_\mu$ to some other distribution $pF_1 + (1-p)F_0 \in H(F)$. Then $pF_1(y) + (1-p)\underline{F}_\mu(y) \leq pF_1(y) + (1-p)F_0(y)$ for all $y < a$ and $pF_1(y) + (1-p)\underline{F}_\mu(y) \geq pF_1(y) + (1-p)F_0(y)$ for all $y > a$, where a is the threshold characterizing $\underline{F}_\mu$. If furthermore $E_0(Y) = (\mu - pE_1(Y))/(1-p)$, then $pF_1(y) + (1-p)\underline{F}_\mu(y) \succeq_{\text{SSD}} pF_1(y) + (1-p)F_0(y)$ by Lemma 7(i). The argument for the upper bound is similar.

(ii) $\underline{P}$ and $\overline{P}$ as defined in Lemma 1(ii) are simultaneously compressed and dispersed, and uniquely attain the bounds on $E(Y)$; thus the claim is immediate if μ equals one of these bounds. Now let μ assume an intermediate value. Any compressed measure P induces a c.d.f. F, where

$$\begin{aligned} F(y) = {} & P_L\big((-\infty, \min\{a, y\})\big) + P(\{a\}) \cdot 1\{y \geq a\} + P_L\big((a, \min\{b, y\})\big) \cdot 1\{y > a\} \\ & + P(\{b\}) \cdot 1\{y \geq b\} + P_L((b, y]) \cdot 1\{y > b\} \end{aligned}$$

with slight abuse of notation (some intervals in the display may be ill defined, but only if the corresponding indicator is 0). An argument very similar to the one in part (i) now reveals that there exists exactly one compressed measure $\underline{P}_\mu$ with expectation $\underline{E}_\mu = (\mu - pE_1(Y))/(1-p)$. This establishes attainability of the lower bound; it remains to show validity. Let (a, b) be the thresholds that partially characterize $\underline{P}_\mu$ and let P_0 be an arbitrary probability measure s.t. $P_U \geq P_0 \geq P_L$. Then $\underline{P}_\mu \geq P_0$ on (a, b) and $P_0 \geq \underline{P}_\mu$ on $\mathbb{R} - [a, b]$. If also $E_0(Y) = (\mu - pE_1(Y))/(1-p)$, then $pP_1 + (1-p)P_0 \succeq_{\text{SSD}} pP_1 + (1-p)\underline{P}_\mu$ follows by Lemma 7(ii). The argument for the upper bound is similar. $\square$

Proof of Theorem 3. (i) Regarding the lower bound, observe first that $Q(\alpha) \leq Q(\gamma) = m$; hence $Q(\alpha)$ is bounded from above by the minimum between m and the worst-case upper bound on $Q(\alpha)$. Similarly, $Q(\beta)$ is bounded from below by the maximum between its worst-case lower bound and m.

$\underline{F}_m$ achieves these bounds and furthermore induces $Q(\gamma) = m$. To see this, observe first that $pF_1(y) + (1-p)\underline{F}_m(y) = pF_1(y) + (1-p)F_L(y)$ for every $y \geq m$; hence every α quantile of $pF_1 + (1-p)\underline{F}_m$ for $\alpha \geq pF_1(m) + (1-p)F_L(m)$ achieves its lower bound. On the other hand, $pF_1(y) + (1-p)\underline{F}_m(y) = pF_1(y) + (1-p)F_U(y)$ for every $y < m$; hence every α quantile of $pF_1 + (1-p)\underline{F}_m$ for $\alpha < pF_1^-(m) + (1-p)F_U^-(m)$ achieves its upper bound. Every other quantile of $pF_1 + (1-p)\underline{F}_m$ equals m. This establishes that the bounds are attained. Furthermore, $m \in H(Q(\gamma))$ implies that $pF_1(y) + (1-p)F_U(y) < \gamma$ for every $y < m$ and that $pF_1(m) + (1-p)F_L(m) \geq \gamma$; thus the γ quantile of $pF_1 + (1-p)\underline{F}_m$ equals m.

Regarding the upper bound, fix any feasible F_0 s.t. $pF_1 + (1-p)F_0$ has γ quantile m; note in particular that this implies $F_0(m) \geq \frac{\gamma - pF_1(m)}{1-p}$ and $F_0^-(m) \leq \frac{\gamma - pF_1^-(m)}{1-p}$. Let $\overline{F}$ be the dispersed c.d.f. with parameter $a = \min\{F_0(m), \frac{\gamma - pF_1^-(m)}{1-p}\}$. Then it follows that $\overline{F}$ is in the set that the upper bound maximizes over. Note also that (i) $F_0(y) \leq a$ for every $y < m$ because otherwise, one would have $F_0^-(m) > \frac{\gamma - pF_1^-(m)}{1-p}$, and (ii) $F_0(y) \geq F_0(m) \geq a$ for every $y \geq m$. Thus, $\overline{F}(y) = \min\{F_L(y), a\} \geq F_0(y)$ for all $y < m$ and $\overline{F}(y) = \max\{F_U(y), a\} \leq F_0(y)$ for all $y > m$. It follows that $pF_1 + (1-p)\overline{F}$ has a lower α quantile and higher β quantile than $pF_1 + (1-p)F_0$.

If F_1 is continuous, then a as used in the preceding paragraph is uniquely given by $a = \frac{\gamma - pF_1(m)}{1-p}$ and the theorem's closed-form expression can be computed by substituting closed-form expressions for $(\underline{F}_m, \overline{F}_a)$ into the definition of quantile contrasts. Finally, $pF_1(m) + (1-p)\underline{F}_m(m) \geq \gamma \geq pF_1^-(m) + (1-p)\underline{F}_m^-(m)$ as well as $pF_1(m) + (1-p)\overline{F}_a(m) \geq \gamma \geq pF_1^-(m) + (1-p)\overline{F}_a^-(m)$ for any a that can emerge from the procedure in the preceding paragraph. This property is inherited by any mixture of $\underline{F}_m$ and $\overline{F}_a$. If F_1 has full support and $p > 0$, it is sufficient for $Q(\gamma) = m$. Furthermore, as $\theta(pF_1 + (1-p)(\lambda \underline{F}_m + (1-\lambda)\overline{F}_a))$ is continuous in both λ and a in this case, the supremum in the upper bound and all intermediate values are attainable by mixing compressed and dispersed distribution functions.

(ii) Fix any feasible P_0 s.t. $pP_1 + (1-p)P_0$ has γ quantile m. Let $\underline{P}$ be the compressed measure s.t. $\underline{P}((-\infty, m]) = P_0((-\infty, m])$. This measure may not be unique if $P_0((-\infty, m]) = P_L((-\infty, m])$ or $P_0((m, \infty)) = P_L((m, \infty))$. In the former case, take the lowest candidate measure; in the latter case, take the highest one. Let (a, b) be the thresholds that partially characterize $\underline{P}$. Then the construction implies that $m \in [a, b]$, and also that $\underline{P} = P_L$ on $(-\infty, a)$ and $\underline{P} = P_U$ on $(a, m]$ if the latter interval exists. For future use, note the implication that $\underline{P}((-\infty, m)) \leq P_0((-\infty, m))$. Also write

$$\underline{P}((-\infty, y]) = P_L((-\infty, y]) \leq P_0((-\infty, y])$$

if $y < a$ and

$$\begin{aligned}\underline{P}((-\infty, y]) &= \underline{P}((-\infty, m]) - \underline{P}((y, m]) \\ &= P_0((-\infty, m]) - P_U((y, m]) \leq P_0((-\infty, y])\end{aligned}$$

if $a \leq y < m$; thus $\underline{P}((-\infty, y]) \leq P_0((-\infty, y])$ for any $y \leq m$. It follows that $pP_1 + (1-p)\underline{P}$ has a weakly higher α quantile than $pP_1 + (1-p)P_0$. By a similar argument, $pP_1 + (1-p)\underline{P}$ has a weakly lower β quantile than $pP_1 + (1-p)P_0$; thus $\theta(pP_1 + (1-p)\underline{P}) \leq \theta(pP_1 + (1-p)P_0)$.

It remains to show that $\underline{P}$ is in the set that the lower bound extremizes over. If $pP_1 + (1-p)P_0$ has γ quantile m, then $P_0((-\infty, m]) \geq \frac{\gamma - pP_1((-\infty, m])}{1-p}$ and $P_0((-\infty, m)) \leq \frac{\gamma - pP_1((-\infty, m))}{1-p}$. Also recalling that $\underline{P}((-\infty, m]) = P_0((-\infty, m])$ by construction and $\underline{P}((-\infty, m)) \leq P_0((-\infty, m))$ as shown above, one finds $\underline{P}((-\infty, m]) \geq \frac{\gamma - pP_1((-\infty, m])}{1-p}$ and $\underline{P}((-\infty, m)) \leq \frac{\gamma - pP_1((-\infty, m))}{1-p}$. Finally, $\underline{P}$ is sandwiched between P_L and P_U; thus $\underline{P}((-\infty, m]) \geq \max\{P_L((-\infty, m]), 1 - P_U((m, \infty))\}$ and $\underline{P}((-\infty, m)) \leq \min\{P_U((-\infty,$

$m)), 1 - P_L([m, \infty))\}$ in analogy to Lemma 1(ii). All in all,

$$\underline{P}((-\infty, m]) \geq \max\left\{\frac{\gamma - pP_1((-\infty, m])}{1-p}, P_L((-\infty, m]), 1 - P_U((m, \infty))\right\},$$
$$\underline{P}((-\infty, m)) \leq \min\left\{\frac{\gamma - pP_1((-\infty, m))}{1-p}, P_U((-\infty, m)), 1 - P_L([m, \infty))\right\},$$

as required.

For the upper bound, fix any feasible P_0 s.t. $pP_1 + (1-p)P_0$ has γ quantile m and let $\overline{P}$ be the dispersed measure s.t. $\overline{P}((-\infty, m]) = P_0((-\infty, m])$ or $\overline{P}((-\infty, m)) = \frac{\gamma - pP_1((-\infty,m))}{1-p}$, whichever is the higher measure in terms of first-order dominance. This measure may not be unique if $P_0((-\infty, m]) = P_U((-\infty, m])$ or $P_0((m, \infty)) = P_U((m, \infty))$. In the former case, take the highest candidate measure; in the latter case, take the lowest one. Then the proof that $\theta(pP_1 + (1-p)\overline{P}) \geq \theta(pP_1 + (1-p)P_0)$ is much as before. To see that $\overline{P}$ is in the set that the upper bound extremizes over, note (in addition to observations from the preceding paragraph) that if $\overline{P}((-\infty, m]) = P_0((-\infty, m])$, then one can write $\overline{P}((-\infty, m]) = P_0((-\infty, m]) \geq \frac{\gamma - pP_1((-\infty,m])}{1-p}$, whereas $\overline{P}((-\infty, m)) \leq \frac{\gamma - pP_1((-\infty,m))}{1-p}$ will then hold by construction of $\overline{P}$. If $\overline{P}((-\infty, m)) = \frac{\gamma - pP_1((-\infty,m))}{1-p}$, then $\overline{P}((-\infty, m)) \leq \frac{\gamma - pP_1((-\infty,m))}{1-p}$ holds trivially and one can also write $\overline{P}((-\infty, m]) \geq \overline{P}((-\infty, m)) = \frac{\gamma - pP_1((-\infty,m))}{1-p} \geq \frac{\gamma - pP_1((-\infty,m])}{1-p}$. The statement about attainability follows much as before. □

Proof of Theorem 5. (i) To see that $(\underline{F}_\mu, \overline{F}_\mu)$ are well defined, let $\mu_0 = (\mu - pE_1(Y))/(1-p)$ as before. Then $\underline{F}_\mu$ must induce $\underline{E}_\mu = \int_0^1 \max\{1 - F_1(y)/a, 0\}\, dy = \mu_0$. The integral's value continuously increases from 0 to $E_1(Y)$ as a increases from 0 to 1; thus it equals μ_0 for some a. Any two different compressed distributions have different expectations; thus $\underline{F}_\mu$ is unique (although, as before, there may be a range of parameter values a characterizing the same $\underline{F}_\mu$). $\overline{F}_\mu$ is characterized by parameter value $a = (E_1(Y) - \mu)/((1-p)E_1(Y))$.

The measure P_0 must induce $E_0(Y) = \mu_0$ as well. Furthermore, simple probability calculus yields

$$P_0(A) = \frac{p}{1-p} \frac{1 - \Pr(Z = 1|Y \in A)}{\Pr(Z = 1|Y \in A)} P_1(A) \tag{8}$$

for any event $A \subseteq [0, 1]$ with $\Pr(Z = 1|Y \in A) > 0$. The proof will show that together with $\Pr(Z = 1|Y = y)$ being nondecreasing, this implies $\underline{P}_\mu \succeq_{\mathrm{SSD}} P_0 \succeq_{\mathrm{SSD}} \overline{P}_\mu$. As second-order dominance is preserved under mixture with P_1, this establishes validity of the bounds, which are furthermore attainable because $\underline{P}_\mu$ and $\overline{P}_\mu$ are consistent with monotone selection. The proof will freely use the following, simple fact: If two probability measures P and P' are s.t. there exists y with $P \leq P'$ on $(-\infty, y)$ and $P \geq P'$ on (y, ∞), then $P \succeq_{\mathrm{FSD}} P'$.

To show $P_0 \succeq_{\mathrm{SSD}} \overline{P}_\mu$, note that

$$\overline{P}_\mu = \frac{p}{1-p} \frac{1-q}{q} P_1 \tag{9}$$

on $(0,1]$, where q is some constant.[19] Suppose by contradiction that $\Pr(Z=1|Y=y)\leq q$ for all $y\in(0,1]$. By comparison of (8) and (9), it would follow that $P_0\geq\overline{P}_\mu$ on $(0,1]$. As both P_0 and $\overline{P}_\mu$ integrate to 1, this would imply $P_0(\{0\})\leq\overline{P}_\mu(\{0\})$; thus $P_0\succeq_{\text{FSD}}\overline{P}_\mu$, which is consistent with $E_0(Y)=\overline{E}_\mu(Y)$ only if $P_0=\overline{P}_\mu$. A similar argument excludes the possibility that $\Pr(Z=1|Y=y)\geq q$ for all $y\in(0,1]$. Recalling that $\Pr(Z=1|Y=y)$ is nondecreasing, it now follows that there exists $y^*\in(0,1]$ s.t. $\Pr(Z=1|Y=y)\leq q$ for all $y\in(0,y^*)$ and $\Pr(Z=1|Y=y)\geq q$ for all $y\in(y^*,1]$ (with ill defined intervals understood to be empty). Again comparing (8) and (9), one finds $\overline{P}_\mu\leq P_0$ on $(0,y^*)$ and $\overline{P}_\mu\geq P_0$ on $(y^*,1]$. If also $\overline{P}_\mu(\{0\})<P_0(\{0\})$, it would follow that $\overline{P}_\mu\succeq_{\text{FSD}}P_0$ and the measures are not equal, contradicting $E_0(Y)=\overline{E}_\mu(Y)$. Hence, $\overline{P}_\mu(\{0\})\geq P_0(\{0\})$. $P_0\succeq_{\text{SSD}}\overline{P}_\mu$ now follows from Lemma 7(ii), applied with threshold values $(0,y^*)$.

To show $\underline{P}_\mu\succeq_{\text{SSD}}P_0$, let a be the parameter value characterizing $\underline{P}_\mu$, and observe that

$$\underline{P}_\mu=\frac{p}{1-p}\frac{1-r}{r}P_1 \tag{10}$$

on $[0,Q_1(a))$, where $r=p/(p+(1-p)a)$, and also that $\underline{P}_\mu((Q_1(a),1])=0$. Suppose that $\Pr(Z=1|Y=0)\geq r$. As $\Pr(Z=1|Y=y)$ is nondecreasing, it follows that $\Pr(Z=1|Y=y)\geq r$ everywhere and, therefore, by comparison of (8) and (10), that $P_0\leq\underline{P}_\mu$ on $[0,Q_1(a))$. As also trivially $P_0\geq\underline{P}_\mu$ on $(Q_1(a),1]$, one finds $P_0\succeq_{\text{FSD}}\underline{P}_\mu$, which is consistent with $E_0(Y)=\underline{E}_\mu(Y)$ only if $P_0=\underline{P}_\mu$. Now suppose that $\Pr(Z=1|Y=0)<r$; hence there exists $y^*\leq Q_1(a)$ s.t. $\Pr(Z=1|Y=y)\leq r$ for $y\in[0,y^*)$ and $\Pr(Z=1|Y=y)\geq r$ for $y\in(y^*,Q_1(a))$ (with ill defined intervals understood to be empty). Again comparing (8) and (10), it follows that $P_0\geq\underline{P}_\mu$ on $[0,y^*)$ and that $P_0\leq\underline{P}_\mu$ on $(y^*,Q_1(a))$. Recalling again that $P_0\geq\underline{P}_\mu$ on $(Q_1(a),1]$, $\underline{P}_\mu\succeq_{\text{SSD}}P_0$ follows from Lemma 7(ii), applied with threshold values $(y^*,Q_1(a))$.

(ii) Consider any distribution F_0, consistent with monotone selection, s.t. $pF_1+(1-p)F_0$ has γ quantile m. Let $g=F_0(m)$. Equation (8) implies

$$\frac{F_1(y)}{F_0(y)}=\frac{1-p}{p}\frac{\Pr(Z=1|Y\leq y)}{1-\Pr(Z=1|Y\leq y)}$$

for any y. Nondecreasingness in y of $\Pr(Z=1|Y=y)$ easily implies nondecreasingness in y of $\Pr(Z=1|Y\leq y)$. This first yields $\Pr(Z=1|Y\leq y)\leq p$ for all y; thus the denominator in the above display cannot vanish. More importantly, it means that $F_1(y)/F_0(y)$ is nondecreasing, in particular,

$$F_0(y)\geq g\cdot\frac{F_1(y)}{F_1(m)} \tag{11}$$

for all $y\leq m$ and

$$F_0(y)\leq\min\left\{g\cdot\frac{F_1(y)}{F_1(m)},1\right\}$$

[19] Specifically, $q=p(1-p)\mu/(\mu-p^2E_1(Y))$, but this value is not important.

for all $y \geq m$. These lower [upper] bounds on F_0 for $y \leq m$ [$y \geq m$] imply upper [lower] bounds on $Q(\alpha)$ for any $\alpha \leq \gamma$ [$\alpha \geq \gamma$]. The bounds are attained by the compressed distribution with parameter $a = F_1(m)/g$, and some algebra yields the closed-form expressions for implied quantiles (in terms of a) provided in the theorem.

Thus, a lower bound on θ is given by its minimum value over the set of all compressed distributions $\underline{F}$ that can emerge from the preceding paragraph's procedure. To see that this is just the set indicated in the theorem, note that if $pF_1 + (1-p)F_0$ has γ quantile m, then

$$\begin{aligned} &pF_1(m) + (1-p)F_0(m) \geq \gamma \quad \Rightarrow \quad g \geq \frac{\gamma - pF_1(m)}{1-p} \\ &\quad \Rightarrow \quad a \leq \frac{(1-p)F_1(m)}{\gamma - pF_1(m)}, \end{aligned} \tag{12}$$

assuming that $F_1(m) \leq \gamma$ and, therefore, $\gamma - pF_1(m) > 0$; else, the definitional constraint $a \leq 1$ binds. Similarly,

$$pF_1^-(m) + (1-p)F_0^-(m) \leq \gamma$$

and simultaneously

$$F_0^-(m) \geq g \cdot \frac{F_1^-(m)}{F_1(m)}$$

in analogy to (11); thus substituting in yields

$$g \leq \frac{\gamma - pF_1^-(m)}{1-p} \cdot \frac{F_1(m)}{F_1^-(m)} \quad \Rightarrow \quad a \geq \frac{(1-p)F_1^-(m)}{\gamma - pF_1^-(m)}. \tag{13}$$

Regarding the upper bound, (8) implies

$$\frac{1 - F_0^-(y)}{1 - F_1^-(y)} = \frac{p}{1-p} \frac{1 - \Pr(Z = 1|Y \geq y)}{\Pr(Z = 1|Y \geq y)}.$$

Nondecreasingness in y of $\Pr(Z = 1|Y = y)$ easily implies nondecreasingness in y of $\Pr(Z = 1|Y \geq y)$. As also $\Pr(Z = 1|Y \geq 0) = p$, the above denominator cannot vanish. More importantly, $\frac{1-F_0^-(y)}{1-F_1^-(y)}$ and, by an analogous argument, $\frac{1-F_0(y)}{1-F_1(y)}$, are nonincreasing in y. Let $\overline{F}$ be the dispersed distribution s.t. $\overline{F}(m) = F_0(m)$ or $\overline{F}^-(m) = (\gamma - pF_1^-(m))/(1-p)$, whichever is the higher (in terms of first-order dominance) measure. For any $y \geq m$, one can then write

$$\frac{1 - \overline{F}(y)}{1 - F_1(y)} = \frac{1 - \overline{F}(m)}{1 - F_1(m)} \geq \frac{1 - F_0(m)}{1 - F_1(m)} \geq \frac{1 - F_0(y)}{1 - F_1(y)},$$

where the equality follows from (9), the first inequality uses that $\overline{F}(m) \leq F_0(m)$ by construction, and the last step uses that $\frac{1-F_0(y)}{1-F_1(y)}$ is nonincreasing. Hence, $F_0(y) \geq \overline{F}(y)$ for any $y \geq m$, which implies that $pF_1 + (1-p)\overline{F}$ has a higher β quantile than $pF_1 + (1-p)F_0$. Now consider any $y < m$. If $\overline{F}^-(m) = (\gamma - pF_1^-(m))/(1-p)$, then use the implication that $\overline{F}^-(m) \geq F_0^-(m)$ to write

$$\frac{1-\overline{F}^-(y)}{1-F_1^-(y)} = \frac{1-\overline{F}^-(m)}{1-F_1^-(m)} \leq \frac{1-F_0^-(m)}{1-F_1^-(m)} \leq \frac{1-F_0^-(y)}{1-F_1^-(y)}$$

in exact analogy to the previous display. If $\overline{F}(m) = F_0(m)$, rather write

$$\frac{1-\overline{F}(y)}{1-F_1(y)} = \frac{1-\overline{F}(m)}{1-F_1(m)} = \frac{1-F_0(m)}{1-F_1(m)} \leq \frac{1-F_0(y)}{1-F_1(y)}.$$

In either case, it follows that $pF_1 + (1-p)\overline{F}$ has a lower α quantile than $pF_1 + (1-p)F_0$. Finally, $Q(\gamma) = m$ implies $pF_1(m) + (1-p)F_0(m) \geq \gamma \geq pF_1^-(m) + (1-p)F_0^-(m)$, and the construction of $\overline{F}$ then implies $pF_1(m) + (1-p)\overline{F}(m) \geq \gamma \geq pF_1^-(m) + (1-p)\overline{F}^-(m)$.

To see the theorem's closed-form expression, it is easiest to circumvent the parameterization of dispersed distributions by a and rather observe that the implied distribution of Y is described by $F(y) = \pi + (1-\pi)F_1(y)$, where $\pi = (1-p)a$. The constraints on π given in the theorem obtain through $\pi + (1-\pi)F_1(y) \geq \gamma \geq \pi + (1-\pi)F_1^-(y)$, and the closed-form expressions in the objective function obtain through

$$\begin{aligned} Q(\alpha) &= \inf\{y : \pi + (1-\pi)F_1(y) \geq \alpha\} \\ &= \inf\left\{y : F_1(y) \geq \frac{\alpha-\pi}{1-\pi}\right\} = Q_1\left(\frac{\alpha-\pi}{1-\pi}\right) \end{aligned}$$

and similarly for $Q(\beta)$.

Monotone selection and $0 \in \text{supp}(Y|Z=1)$ imply that P is absolutely continuous with respect to P_1 except possibly at 0. Thus, $m \in H(Q(\gamma))$ implies that m is not in a support gap of P_1. If F_1 is continuous, one can then write $Q(\gamma) = m \Leftrightarrow F(m) = F^-(m) = \gamma$, which uniquely characterizes the relevant compressed and dispersed c.d.f.'s and also ensures that they induce the targeted γ quantile. Thus, the bounds are attainable and can be computed in closed form. Remarks for the case of full support are as before. □

Proof of Lemma 6. $P_0 \geq 0$ is obvious. To see that $P_0 \leq P_1 \cdot (k-p)/(1-p)$, write

$$\begin{aligned} \frac{P_0}{P_1} &= \frac{P(Y|Z=0)}{P(Y|Z=1)} = \frac{P(Z=0|Y)P(Y)/P(Z=0)}{P(Z=1|Y)P(Y)/P(Z=1)} \\ &= \frac{1-P(Z=1|Y)}{P(Z=1|Y)} \frac{P(Z=1)}{1-P(Z=1)} \\ &\leq \frac{1-p/k}{p/k} \frac{p}{1-p} = \frac{k-p}{1-p}, \end{aligned}$$

where the inequality arises from using twice that $P(Z=1|Y) \leq p/k$. □

References

Andrews, D. W. K. and G. Soares (2010), "Inference for parameters defined by moment inequalities using generalized moment selection." *Econometrica*, 78, 119–157. [325]

Atkinson, A. B. (1970), "On the measurement of inequality." *Journal of Economic Theory*, 2, 244–263. [327]

Balke, A. and J. Pearl (1997), "Bounds on treatment effects from studies with imperfect compliance." *Journal of the American Statistical Association*, 92, 1171–1176. [329, 346]

Beresteanu, A. and F. Molinari (2008), "Asymptotic properties for a class of partially identified models." *Econometrica*, 76, 763–814. [325]

Billingsley, P. (1995), *Probability and Measure*, third edition. Wiley, New York. [348]

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007), "Changes in the distribution of male and female wages accounting for employment composition using bounds." *Econometrica*, 75, 323–363. [325, 338]

Chernozhukov, V., H. Hong, and E. Tamer (2007), "Parameter set inference in a class of econometric models." *Econometrica*, 75, 1243–1284. [325, 345]

Cowell, F. A. (1991), "Grouping bounds for inequality measures under alternative informational assumptions." *Journal of Econometrics*, 48, 1–14. [325]

Cowell, F. A. (1995), *Measuring Inequality*, second edition. Harvester Wheatsheaf, London, U.K. [327]

Cowell, F. A. and M.-P. Victoria-Feser (1996), "Robustness properties of inequality measures." *Econometrica*, 64, 77–101. [337]

Cowell, F. A. and M.-P. Victoria-Feser (2002), "Welfare rankings in the presence of contaminated data." *Econometrica*, 70, 1221–1233. [337]

Cross, P. J. and C. F. Manski (2002), "Regressions, short and long." *Econometrica*, 70, 357–368. [330]

Dalton, H. (1920), "The measurement of inequality of incomes." *Economic Journal*, 30, 348–361. [327]

Dasgupta, P., A. K. Sen, and D. Starret (1973), "Notes on the measurement of inequality." *Journal of Economic Theory*, 6, 180–187. [327]

de Castro, L. I. (2009), "Affiliation and positive dependence in economic models." Unpublished manuscript, University of Illinois (Urbana–Champaign). [338]

DeRobertis, L. and J. A. Hartigan (1981), "Bayesian inference using intervals of measures." *The Annals of Statistics*, 9, 235–244. [325, 331]

Fan, Y. and S. Park (forthcoming), "Sharp bounds on the distribution of treatment effects and their statistical inference." *Econometric Theory*. [335]

Ferson, S., L. Ginzburg, V. Kreinovich, L. Longpré, and M. Aviles (2005), "Exact bounds on finite populations of interval data." *Reliable Computing*, 11, 207–233. [336]

Ferson, S., V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz (2003), "Constructing probability boxes and Dempster–Shafer structures." SAND Report 2002-4015, Sandia National Laboratories. [330]

Firpo, S. and G. Ridder (2009), "Bounds on functionals of the distribution of treatment effects." Unpublished manuscript, São Paulo School of Economics. [335]

Gastwirth, J. L. (1972), "The estimation of the Lorenz curve and Gini index." *Review of Economics and Statistics*, 54, 306–316. [325]

Haile, P. A. and E. Tamer (2003), "Inference with an incomplete model of English auctions." *Journal of Political Economy*, 111, 1–51. [329]

Horowitz, J. and C. F. Manski (1995), "Identification and robustness with contaminated and corrupted data." *Econometrica*, 63, 281–302. [330, 331, 341]

Imbens, G. (2003), "Sensitivity to exogeneity assumptions in program evaluation." *American Economic Review (Papers and Proceedings)*, 93, 126–132. [341]

Imbens, G. and C. F. Manski (2004), "Confidence intervals for partially identified parameters." *Econometrica*, 72, 1845–1857. [325]

King, M., S. Ruggles, T. Alexander, D. Leicach, and M. Sobek (2009), *Integrated Public Use Microdata Series, Current Population Survey: Version 2.0*. Minnesota Population Center, Minneapolis, MN. Available at http://cps.ipums.org. [342]

Kitagawa, T. (2009), "Identification region of the potential outcome distribution under instrument independence." Working Paper 30/09, CeMMAP. [329]

Klasen, S. (1994), "Growth and well-being: Introducing distribution-weighted growth rates to reevaluate U.S. post-war economic performance." *Review of Income and Wealth*, 40, 251–272. [345]

Kreider, B. and J. V. Pepper (2007), "Disability and employment: Reevaluating the evidence in light of reporting errors." *Journal of the American Statistical Association*, 102, 432–441. [341]

Kreinovich, V., G. Xiang, and S. Ferson (2006), "Computing mean and variance under Dempster–Shafer uncertainty: Toward faster algorithms." *International Journal of Approximate Reasoning*, 42, 212–227. [336]

Manski, C. F. (1990), "Nonparametric bounds on treatment effects." *American Economic Review (Papers and Proceedings)*, 80, 319–323. [346]

Manski, C. F. (1997), "Monotone treatment response." *Econometrica*, 65, 1311–1334. [347]

Manski, C. F. (2003), *Partial Identification of Probability Distributions*. Springer-Verlag, New York. [323, 327, 330, 331]

Manski, C. F. (2007), *Identification for Prediction and Decisions*. Harvard University Press. [323, 330]

Manski, C. F. and J. V. Pepper (2000), "Monotone instrumental variables: With an application to the returns to schooling." *Econometrica*, 68, 997–1010. [347]

Molinari, F. (2008), "Partial identification of probability distributions with misclassified data." *Journal of Econometrics*, 144, 81–117. [330]

Molinari, F. and M. Peski (2005), "Generalization of a result on 'regressions, short and long'." *Econometric Theory*, 22, 159–163. [346]

Nelsen, R. B. (2006), *An Introduction to Copulas*. Springer-Verlag, New York. [338]

Pedersen, C. S. and S. E. Satchell (1998), "An extended family of financial-risk measures." *Geneva Papers on Risk and Insurance Theory*, 23, 89–117. [327]

Romano, J. P. and A. M. Shaikh (2010), "Inference for the identified set in partially identified econometric models." *Econometrica*, 78, 169–211. [325]

Rosen, A. (2008), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities." *Journal of Econometrics*, 146, 107–117. [325]

Rosenbaum, P. R. (2002), *Observational Studies*, second edition. Springer-Verlag, New York. [341]

Rothschild, M. and J. E. Stiglitz (1970), "Increasing risk: I. A definition." *Journal of Economic Theory*, 2, 225–243. ("Addendum." *Journal of Economic Theory*, 5, 306.) [327]

Ruggles, S., M. Sobek, T. Alexander, C. A. Fitch, R. Goeken, P. K. Hall, M. King, and C. Ronnander (2009), *Integrated Public Use Microdata Series: Version 4.0*. Minnesota Population Center, Minneapolis, MN. Available at http://usa.ipums.org/usa/. [342]

Sen, A. K. (1976), "Real national income." *Review of Economic Studies*, 43, 19–36. [345]

Shaked, M. and J. G. Shantikumar (2007), *Stochastic Orders*. Springer-Verlag, New York. [327, 348]

Stoye, J. (2009), "More on confidence regions for partially identified parameters." *Econometrica*, 77, 1299–1315. [325]

Vazques Alvarez, R., B. Melenburg, and A. van Soest (1999), "Bounds on quantiles in the presence of full and partial item nonresponse." Unpublished manuscript, CentER Tilburg. [338]

Vazques Alvarez, R., B. Melenburg, and A. van Soest (2003), "Selection bias and measures of inequality." Unpublished manuscript, University College Dublin. [325]

Wasserman, L. and J. B. Kadane (1992), "Computing bounds on expectations." *Journal of the American Statistical Association*, 87, 516–522. [325, 331]

Submitted October, 2009. Final version accepted October, 2010.