Testing for Distributional Treatment Effects: A Set Identification Approach

Sung Jae Jun*, Yoonseok Lee† and Youngki Shin‡ §

July 2012

Abstract

We consider testing stochastic dominance between potential outcomes under partial identification, which uses dual representations of dominance hypotheses. Unlike the conventional approach of selection-on-observables, we allow for the presence of unobserved heterogeneity that can be arbitrarily correlated with a treatment variable. Using panel data, we establish the identified bounds of potential outcome distributions. Based on the bounds we formulate a dual expression for the hypothesis of interest, which we use to construct an easily–implementable test. As an empirical illustration, we investigate the effects of smoking during pregnancy on the infants' birthweight distribution.

JEL Classification: C12, C21, C23.

Keywords: treatment effects, partial identification, unobserved heterogeneity, stochastic dominance, panel data.

*Department of Economics and Center for the Study of Auctions, Procurements, and Competition Policy, Pennsylvania State University, 619 Kern Bldg, University Park, PA 16802, sjun@psu.edu.

†Department of Economics, University of Michigan, 611 Tappan St., Ann Arbor, MI 48109, yoolee@umich.edu

‡Department of Economics, University of Western Ontario, 1151 Richmond St., London, ON N6A 5C2, Canada, yshin29@uwo.ca

§We have benefited by many constructive comments and suggestions from Chris Bennett, Charles Brown, Matias Cattaneo, Yanqin Fan, Jin Hahn, Stefan Hoderlein, Han Hong, Kyoo il Kim, Simon Lee, Maria Ponomareva, Joris Pinkse, Christoph Rothe, Jeff Smith, Iván Fernández–Val. Also we would like to thank seminar participants at various universities and conferences. Jun thanks the Human Capital Foundation (www.hcfoundation.ru), especially Andrey P. Vavilov, for their support of CAPCP (http://capcp.psu.edu) at Penn State University; Lee acknowledges financial support from Livingston Research Scholar Award, University of Michigan; and Shin thanks the Social Science and Humanities Research Council of Canada (SSHRCC) for the research grant. All remaining errors are our own.

1. Introduction

Evaluating policy or treatment effects has been an important question in diverse disciplines. In cases where randomized experiments are not available as in many economic data sets, treatment variables are usually endogenous due to unobserved heterogeneity or self-selection. As well summarized by Imbens and Wooldridge (2009) and Heckman and Vytlacil (2007a,b), various econometric methods have been developed to address this issue, and a myriad of empirical studies were conducted based on them. Examples include approaches based on the assumption of unconfoundedness and matching (e.g., Dehejia and Wahba (2002)) and approaches based on instrumental variables (e.g., Angrist, Imbens, and Rubin (1996)).

Recently the importance of distributional treatment effects has been pointed out, and studies report accumulating evidence of them: see e.g., Heckman and Smith (1997), Djebbari and Smith (2008), Fan and Park (2010) and references therein. Particularly when there exist different treatment effects across outcome levels, comparing the distributions of potential outcomes is of interest. For example, Firpo (2007) used the unconfoundedness assumption to identify quantile treatment effects. Lee (2009) proposed nonparametric tests for the lack of distributional effects using the two-sample Mann–Whitney test. Abadie (2002) proposed bootstrap tests for stochastic dominance between potential outcomes using instrumental variables. Rothe (2010) discussed treatment effects when the counterfactual experiment is defined as changing the distribution of a treatment variable. Related to this, it has also received attention to test heterogeneous treatment effects conditional on covariates. For example, Crump, Hotz, Imbens, and Mitnik (2008) and Lee and Whang (2009) proposed testing methods for heterogeneous treatment effects under the unconfoundedness assumption.

In this paper we analyze heterogeneous treatment effects by testing stochastic dominance between potential outcomes. Unlike the aforementioned approaches, however, our method does not rely on unconfoundedness nor on instrumental variables. We instead consider a panel data model and deal with endogeneity by introducing unobserved heterogeneity that could be arbitrarily correlated with the treatment variable. Under the nonparametric/nonseparable setup, we obtain partial identification of the potential outcome distributions using time homogeneity assumptions. We then test for stochastic dominance between the potential outcome distributions using the partial identification result.

Examples of repeated (endogenous) treatments can be found in many places. First, the study on the wage premium of the union membership recognizes correlations between the union membership and individual unobserved ability. To deal with such an endogeneity problem, panel data such as the National Longitudinal Survey (NLS), the Current Population Survey (CPS), the Canadian Labor Market Activity Survey (LMAS), and the German Socio-economic Panel (GSOEP) are frequently used.[1] Second, Jung (2010) examined the effect of voluntary information disclosure of health insurance plans using panel data, where the decision to reveal additional information to the public is considered a treatment variable. Finally, many studies in health economics have analyzed the effect of smoking during pregnancy on birth outcomes, where the smoking behavior is believed to be correlated with other health-related unobserved factors such as nutrition status or drug addiction. For instance, Abrevaya (2006) estimated a fixed–effect model using a pseudo-panel data set constructed from the U.S. Natality Data Set.

It should be noted that these empirical panel studies apply the linear fixed effect model, although the maintained assumptions of it can be easily violated when heterogeneous treatment effects are allowed. For instance, low-ability workers might have different wage premia of the union membership from those of high-ability workers. In order to consider such general cases, we work in a nonparametric counterfactual framework, where unobserved heterogeneity is nonseparable: see also Chernozhukov, Fernandez-Val, and Newey (2009).

The main contribution of this paper is to provide a novel method of testing stochastic dominance between potential outcomes when their distributions are only partially identified. Although the identification result involves two potential outcome distributions, the hypothesis of stochastic dominance involves only the difference of the two distributions. This dimension reduction leads us into a dual expression of the hypothesis, to which many classical testing methods can be applied. Compared to the existing method, this new approach makes the implementation much simpler and gives more power. The idea of using a dual expression was first suggested by Hahn and Ridder (2009) in the context of inference for a single parameter when the whole vector of (finite-dimensional) parameters is partially identified. In our context, we treat the identified set of the potential outcome distributions as a model restriction,

[1] See, for example, Jones (1982), Blakemore, Hunt, and Kiker (1986), Robinson (1989), Lemieux (1998), Budd and Na (2000), and Beck and Fitzenberger (2004).

and test the compatibility of the model restriction with stochastic dominance. We could instead consider the existing inference procedures under partial identification such as Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Rosen (2008), Romano and Shaikh (2010), Andrews and Soares (2010), Kim (2009), Chernozhkukov, Lee, and Rosen (2009), and Ponomareva (2010). However, by the proposed method we avoid doing an inference on the entire vector of model parameters. As a result, we need not project the whole confidence region onto the subspace of interest, from which we naturally expect more power.

Finally, we review some related literature. Bhattacharya, Shaikh, and Vytlacil (2012) reanalyzed the effect of the Swan-Ganz catheterization on mortality outcomes under unobserved heterogeneity focusing on the mean treatment effect when an instrumental variable is available. Fan and Park (2010, 2011) derived sharp bounds on the distribution of the treatment effect defined as the difference of potential outcomes. They assumed randomized experiments but did not impose any structure on the joint distribution of potential outcomes. However, these studies constructed confidence regions using the standard method in Romano and Shaikh (2010) and in Chernozhukov, Hong, and Tamer (2007), respectively. In the recent literature on the stochastic dominance test, Linton, Song, and Whang (2010) and Donald and Hsu (2010) focused on the power improvement. Related to the endogeneity issue in the stochastic dominance test, Maier (2011) imposed the unconfoundedness assumption and Chernozhukov and Hansen (2006) used an instrumental variable. To the best of our knowledge, this is the first paper in the literature to test stochastic dominance under partial identification.

The remainder of the paper is organized as follows. In Section 2, we introduce the basic framework and derive the partial identification results. In Section 3, we compare two different approaches to test stochastic dominance under partial identification and obtain their dual representations. In Section 4, we construct the test statistics and derive its asymptotic properties. Section 5 presents an empirical illustration that investigates the effect of smoking on the infants' birthweight distribution. All mathematical proofs are collected in the appendix.

2. Model and Identification

In this section we discuss the basic framework and partial identification results. We consider the panel data $\{(Y_{it}, D_{it}, X'_{it})' : \ i = 1, 2, \cdots, n, t = 1, 2, \cdots, T\}$, where X_{it}

is a vector of covariates, D_{it} is a binary treatment variable, and Y_{it} is an outcome variable of interest. We focus on a short panel, and hence n is large relative to T.

The observed outcome Y_{it} depends on the treatment D_{it} in the following way:

$$Y_{it} = D_{it}Y_{it}^1 + (1 - D_{it})Y_{it}^0,$$

where Y_{it}^1 and Y_{it}^0 are potential outcomes when D_{it} is exogenously set to 1 and 0, respectively. For example, let D_{it} be an indicator of mother i's smoking status during pregnancy of the t-th baby. Then, the baby's counterfactual birthweight if the mother had and had not smoked would be denoted by Y_{it}^1 and Y_{it}^0, respectively. We observe only one of the potential outcomes, depending on the realized value of D_{it}.

Such a counterfactual setup is now standard, at least in the cross-section context, where the common objective is to compare (some features of) the distributions of the potential outcomes. This objective is usually achieved by assuming randomized treatment conditional on *observed* heterogeneity: see e.g., Rosenbaum and Rubin (1983), Hahn (1998), Hirano, Imbens, and Ridder (2003), and Firpo (2007). However, as we emphasized in the introduction, one of our main goals is to avoid this standard assumption, because it might be violated in some situations due to the presence of *unobserved* heterogeneity. Data on repeated treatments turn out to be useful for this purpose as we will further discuss below.

In addition to the observables $(Y_{it}, D_{it}, X_{it}')'$, we let α_i represent *unobserved* heterogeneity that is time-invariant and possibly correlated with $(D_{it}, X_{it}')'$.[2] We presume that α_i is the source of endogeneity so that full randomization is not achieved unless α_i is additionally controlled for. But the difficulty arises from the fact that α_i is not observed. We will deal with this difficulty by using panel data while not imposing α_i to be additively separable as in the linear fixed effect model.

We now make some assumptions required for our analysis. We will drop the subscript i unless there is a possibility of confusion.

Assumption 1 (Selection-On-Unobservables). *The treatment history $D = (D_1, \cdots, D_T)'$ is independent of the current potential outcomes (Y_t^0, Y_t^1) conditional on the current heterogeneity $(X_t', \alpha)'$.*

Assumption 2 (Time Homogeneity). *For $t \neq s$ and $j = 0, 1$, $(Y_t^j, X_t', \alpha)'$ has the same distribution as $(Y_s^j, X_s', \alpha)'$.*

[2]Unobserved heterogeneity can be allowed to be time-varying in a limited sense. We will discuss this issue in more detail at the end of this section.

Assumption 3 (Time Homogeneity of Heterogeneity)**.** *For $t \neq s$, the conditional distribution of $(X_t', \alpha)'$ given D is the same as that of $(X_s', \alpha)'$.*[3]

Assumption 1 is similar to but more flexible than the standard assumption of unconfoundedness (i.e., selection-on-observables), because an unobserved source of confounding is allowed. One practical restriction imposed by Assumption 1 is that there is no feedback from Y_{t-1} to D_t once α is controlled for. This requirement can be relaxed to the assumption of the independence of the current and past treatment given heterogeneity: i.e., (Y_t^0, Y_t^1) is independent of $(D_1, \cdots, D_t)'$ given $(X_t', \alpha)'$. To focus on the main idea of the paper, we would leave this extension as well as the analysis of other possible dynamics for future research.

Assumption 2 implies that the marginal distributions of the potential outcomes do not change over time, but they are allowed to be serially correlated. A similar time-homogeneity assumption can be found in Chernozhukov, Fernandez-Val, and Newey (2009) and Khan, Ponomareva, and Tamer (2011) in the context of a nonseparable panel regression model and a censored panel regression model, respectively.

Assumption 3 is a strong assumption, because it imposes restrictions on the time variability of X_t and the dependence between X_t and D_t. One way of avoiding this assumption is to reformulate Assumptions 1 and 2 using the history of the covariates $X = (X_1, \cdots, X_T)'$ as well as the treatment history D. This is an approach taken by Chernozhukov, Fernandez-Val, and Newey (2009), which we will discuss later in this section in more detail. For now we just note that Assumption 3 is (partly) testable and that it is trivially satisfied when observed heterogeneity is time-invariant. For variables that do not satisfy Assumption 3, we may restrict our analysis to an appropriate subpopulation that may satisfy it.

We now discuss the identification of the distribution functions of Y_t^j given X_t for $j = 0, 1$. We let $\mu(\cdot)$ and $\mu(\cdot|B = b)$ be the marginal and conditional distribution functions of α given a generic random variable $B = b$. Similarly, let $F^j(\cdot)$ and $F^j(\cdot|B = b)$ be the marginal and conditional distribution functions of Y_t^j. By Assumption 2, $F^j(\cdot|X_t = x)$ is invariant over t, and therefore we will simply write it as $F^j(\cdot|x)$.

[3] We will suppress the usual qualifier of "with probability one" whenever it is clear from the context.

Focusing on $j = 1$, we note that

$$F^1(y|x) = \mathbb{P}\left[D_t = 0|X_t = x\right]\mathbb{P}\left[Y_t^1 \leq y|D_t = 0, X_t = x\right] + \mathbb{P}\left[Y_t \leq y, D_t = 1|X_t = x\right], \tag{1}$$

by Assumption 1. The first term in (1) is not identified because Y_t^1 is not observed when $D_t = 0$. However, it is naturally bounded between 0 and $\mathbb{P}[D_t = 0|X_t = x]$, which results in interval identification of $F^1(y|x)$. Lemma 1 below shows that this interval can be further improved by using observations of multiple periods under Assumptions 1 through 3.

We introduce additional notations. We let $\mathcal{T}$ be the set of all permutations of $(1, 2, \cdots, T)$. For $j \in \{0, 1\}$ and $\ell = (t_1, t_2, \cdots, t_T) \in \mathcal{T}$, we define

$$\begin{aligned} p_{1\ell}^j(y|x) &= \mathbb{P}[Y_{t_1} \leq y, D_{t_1} = j|X_{t_1} = x], \\ p_{k\ell}^j(y|x) &= \mathbb{P}[Y_{t_k} \leq y, D_{t_1} = 1 - j, \cdots, D_{t_{k-1}} = 1 - j, D_{t_k} = j|X_{t_k} = x] \quad \text{for} \quad k \geq 2, \end{aligned}$$

and

$$\begin{aligned} L_\ell^j(x, y) &= \sum_{k=1}^{T} p_{k\ell}^j(y|x), \\ U_\ell^j(x, y) &= L_\ell^j(y|x) + \mathbb{P}[D_{t_1} = 1 - j, D_{t_2} = 1 - j, \cdots, D_{t_T} = 1 - j|X_{t_T} = x]. \end{aligned}$$

Note that all of these terms depend on the specific permutation ℓ without further assumptions. They turn out to be independent of ℓ under the exchangeability condition, which will be also useful to simplify the distribution theory for our inference procedure.

Assumption 4 (Exchangeability)**.** *For any $\ell = (t_1, t_2, \cdots, t_T) \in \mathcal{T}$ and $j \in \{0, 1\}$,*

$$\begin{aligned} &\mathbb{P}[D_1 = j|X_1 = x, \alpha = a] = \mathbb{P}[D_{t_1}|X_1 = x, \alpha = a] \quad \textit{and} \\ &\mathbb{P}[D_1 = 1 - j, D_2 = 1 - j, \cdots, D_k = j|X_k = x, \alpha = a] \\ &\qquad = \mathbb{P}[D_{t_1} = 1 - j, D_{t_2} = 1 - j, \cdots, D_{t_k} = j|X_k = x, \alpha = a] \quad \textit{for } k \geq 2 \end{aligned}$$

for all x and a in the support of X_k and α, respectively.

Assumption 4 is implied by the exchangeability of $(D_1, D_2, \cdots, D_k)$ given (X_k, α). Note that exchangeability does not require independence. More precisely, by the de Finetti's theorem, the exchangeability of $(D_1, \cdots, D_k)$ given (X_k, α) means that there is an additional latent variable such that $D_1, \cdots, D_k$ are *i.i.d.* given X_k, α and the latent variable. We now state our first lemma. We let $\mathcal{X}$ be the support of X_t.

Lemma 1. *Suppose that Assumptions 1, 2, and 3 are satisfied. For each $j \in \{0,1\}$, we have*

$$0 \leq \max_{\ell \in \mathcal{T}} L_\ell^j(x,y) \leq F^j(y|x) \leq \min_{\ell \in \mathcal{T}} U_\ell^j(x,y) \leq 1$$

for all $(x,y) \in \mathcal{X} \times \mathbb{R}$. Furthermore, if Assumption 4 is additionally made, then L_ℓ^j and U_ℓ^j are independent of ℓ.

Similar calculation used to derive Lemma 1 can be found in Manski (1990) and Chernozhukov, Fernandez-Val, and Newey (2009). Comparison with Chernozhukov, Fernandez-Val, and Newey (2009) is, however, more relevant here. The most apparent difference is that their analysis is based on conditioning on the history of both D_t and X_t, whereas our analysis uses contemporaneous observations X_t for the conditioning set. The object of interest of Chernozhukov, Fernandez-Val, and Newey (2009) is thus generally different from ours. The price that we pay for this is Assumption 3, which is a strong but (partly) testable assumption. Note also that our approach has a practical advantage over the alternative because of the smaller size of conditioning variables. For example, if X_{it} is a binary variable, the number of conditioning cells of our approach and that of the alternative are 2 and 2^T, respectively. The difference is larger when X_{it} has larger support.

For additional comparison, suppose that we have some structural function q such that

$$Y_t^j = q_j(X_t, \alpha, \epsilon_t^j) \quad \text{for} \quad j = 0, 1,$$

where ϵ_t^j is an unobserved idiosyncratic error. Lemma 1 targets the distribution of Y_t^j, because conditioning X_t can be averaged out. However, Chernozhukov, Fernandez-Val, and Newey (2009)'s counterfactual analysis targets a different object: using the time homogeneity of ϵ_t^j given the history of D_t and X_t, they showed the identification bounds of the distribution of $Y_t^j(x) = q_j(x, \alpha, \epsilon_t^j)$ for some x. Generally, we know that

$$\int \mathbb{P}[Y_t^j(x) \leq y] dF_X(x) \neq \int \mathbb{P}[Y_t^j(x) \leq y | X_t = x] dF_X(x) = \mathbb{P}[Y_t^j \leq y].$$

Therefore, Lemma 1 directly aims at the causal effect of D_t, whereas Chernozhukov, Fernandez-Val, and Newey (2009) studies the causal effect of (D_t, X_t). In the latter approach, integrating out X_t afterwards does not generally lead to the distribution of the potential outcome Y_t^j but it *defines* a different measure for the causality of D_t. Since, depending on the context, a treatment variable itself can be of more interest than other covariates, explicitly treating the treatment variable differently makes a good supplement for Chernozhukov, Fernandez-Val, and Newey (2009)'s approach. In fact, our discussions in the following sections can be easily modified after their approach, which will not require Assumption 3. The point is not in claiming that one approach dominates the other but in articulating the fact that distinguishing a treatment variable from other covariates could be subtle and add an extra challenge.

In order to make a further comparison of the two approaches, consider the simple case where there is no covariate (e.g. $X_t = 1$ for all t), in which case both the results of Chernozhukov, Fernandez-Val, and Newey (2009) and Lemma 1 provide the bounds for the distribution function of Y_t^j. In this case, Assumption 3 is automatically satisfied and our bounds under Assumption 4 coincide with those of Chernozhukov, Fernandez-Val, and Newey (2009). The reason why Assumption 4 is used as an extra assumption is that Assumption 2 requires only the time homogeneity of the marginal distribution of Y_t^j. In fact, our bounds can be shown to be independent of the permutation ℓ without Assumption 4 once the homogeneity of Y_t^j given the history of D_t is assumed, which is an assumption made in Chernozhukov, Fernandez-Val, and Newey (2009). Note, however, that this is as strong an assumption as Assumption 3 but that it is not testable.

We finish this section with a brief discussion of a time-varying unobserved heterogeneity. Until now, we assume that unobserved heterogeneity α is fixed over time. To allow a time-varying α_t in our setup, Assumptions 2 and 3 should be modified to include time homogeneity over α_t in addition to Y_t^j and X_t. However, the applications satisfying this restriction appears to be quite limited, and we leave it for future research to find alternative approaches.

3. Hypotheses

In this section, we formulate the hypotheses of interest, and propose two different approaches of testing them. Comparing these approaches, we argue that the proposed method based on the dual expression has some advantages over the projection based one.

We consider testing (conditional or uniform) stochastic dominance when the distributions of interest are partially identified. It is said that Y^0 first-order stochastically dominates Y^1 uniformly over X if and only if

$$F^0(y|x) \leq F^1(y|x) \quad \text{for all} \quad (x, y) \in \mathcal{X} \times \mathbb{R}. \tag{2}$$

For example, one might want to test whether or not the birthweight decreases over all range of birthweights and covariates when a mother smokes during pregnancy. In this case, if the birthweight increases for any quantiles given any sub-group characterized by $X = x$, then one should reject the null hypothesis.[4]

Note that testing the hypothesis in (2) is less conservative than testing the standard (unconditional) stochastic dominance or the existence of the average treatment effect, because both $F^0(y) \leq F^1(y)$ for all $y \in \mathbb{R}$ and $\mathbb{E}[Y_t^0] \geq \mathbb{E}[Y_t^1]$ are implied by the hypothesis in (2). The importance of testing conditional (or heterogeneous) treatment effects is also emphasized by Crump, Hotz, Imbens, and Mitnik (2008) and Lee and Whang (2009) under point-identified models.

Since $F^0(\cdot|\cdot)$ and $F^1(\cdot|\cdot)$ are partially identified in our setup, we cannot apply the existing methods of testing stochastic dominance. Instead, we need to consider some inference methods on the set of parameter values. It is also worthwhile to note that, different from the general set inference literature, our inference problem involves only a subset of the parameters. In this case, following Guggenberger, Hahn, and Kim (2008) and Hahn and Ridder (2009), we can construct a testing problem similar to the classical specification test. The original idea was developed for linear models with finite dimensional parameters, but we take a similar approach to extend it to the context of an infinite dimensional parameter space. To highlight the main idea, we first discuss how to test (2) by constructing a confidence region for the whole identified set. Next, we show how our alternative approach simplifies the inference procedure both conceptually and computationally.

To focus on the main idea, we first consider a finite dimensional parameter space. Suppose that we have a testing problem for a fixed (x, y) and suppress it from $L^j(x, y), U^j(x, y)$ and $F^j(y|x)$ for $j = 0, 1$. We will discuss the uniform case later. Now, the parameters F^0 and F^1 are identified by the following four moment inequalities:

$$L^0 \leq F^0 \leq U^0 \tag{3}$$

$$L^1 \leq F^1 \leq U^1 \tag{4}$$

[4]This interpretation is based on the so-called rank invariance assumption such that each individual maintain his/her rank in both treated and untreated distributions. It is controversial whether quantile treatment effects should be based on the joint distribution of the potential outcomes or the comparison of their marginal distributions. We do not discuss this issue here: see e.g. Heckman and Smith (1997), Firpo (2007), and Imbens and Wooldridge (2009) for more discussions.

where (L^j, U^j) for $j = 1, 2$ are (point-identified) expectations as we have obtained in Lemma 1. Since we want to test whether $F^1 - F^0 \geq 0$, it is useful to reparameterize $\gamma := F^1 - F^0$ and write the inequalities (3) and (4) as follows:

$$\begin{pmatrix} L^0 \leq F^1 - \gamma \leq U^0 \\ L^1 \leq F^1 \leq U^1 \end{pmatrix} \Longleftrightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ -1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F^1 \\ \gamma \end{pmatrix} \leq \begin{pmatrix} -L^0 \\ U^0 \\ -L^1 \\ U^1 \end{pmatrix}. \tag{5}$$

This is the standard linear inequality model. We let $\theta := (F^1, \gamma)'$ and Θ_I be the identified set such that all elements in Θ_I satisfy (5). Then, we can construct a confidence region for Θ_I following the recent development in the partial identification literature: see e.g. Andrews, Berry, and Jia (2004), Rosen (2008), Kim (2009), and Andrews and Soares (2010). See also Chernozhkukov, Lee, and Rosen (2009) for the case of conditional moment inequalities as in our original setup.

Once we construct a confidence region $\mathcal{C}_n$ following any of the existing methods, it can be used for testing $\theta^* \in \Theta_I$ for any fixed point $\theta^* \in \Theta$, where Θ is the entire parameter space. However, recall that we are interested in testing whether one component of θ is positive, i.e. $\gamma \geq 0$. Thus, more specification is required for formulating hypotheses. In fact, there exist two sets of hypotheses that are potentially of interest:

$$\begin{cases} H^*_{0,a}: \ \forall \theta \in \Theta_I, \ \gamma \geq 0 \\ H^*_{1,a}: \ \exists \theta \in \Theta_I, \ \gamma < 0 \end{cases} \tag{6}$$

and

$$\begin{cases} H^*_{0,b}: \ \exists \theta \in \Theta_I, \ \gamma \geq 0 \\ H^*_{1,b}: \ \forall \theta \in \Theta_I, \ \gamma < 0. \end{cases} \tag{7}$$

Note that the null hypothesis in (6) is less conservative than that of (7) as $H^*_{0,a}$ implies $H^*_{0,b}$. In other words, the null hypothesis in (6) tests whether $\gamma \geq 0$ holds uniformly over all identified set Θ_I while that in (7) asks the existence of $\theta \in \Theta_I$ that satisfies $\gamma \geq 0$. Which formulation is more interesting depends on a specific research question at hand.

Suppose that one has decided the hypothesis to test between (6) and (7). Then, the next question is how to make a decision between the null and alternative. An immediate (and probably natural) approach is to project the confidence region $\mathcal{C}_n$ from (5) onto the space of γ and to check whether the projected confidence interval, say $\mathcal{C}_{n,\gamma}$, contains zero or not. We reject $H^*_{0,a}$ in favor of $H^*_{1,a}$ when $\mathcal{C}_{n,\gamma}$ contains any element that is less than zero, and we reject $H^*_{0,b}$ when all elements in $\mathcal{C}_{n,\gamma}$ are less than zero.

This projection approach is intuitive, but it can be unnecessarily conservative. As pointed out by Hahn and Ridder (2009), the confidence interval $\mathcal{C}_{n,\gamma}$ projected from an $(1-\alpha)$ confidence region $\mathcal{C}_n$ would be at least equal to or usually be larger than the true $(1-\alpha)$ confidence interval for γ. Furthermore, this problem would get worse as the dimension of θ increases. As proposed by Hahn and Ridder (2009), such problems could be solved using an alternative approach based on the model specification test. More precisely, we treat the identified set as a model restriction maintained both under the null and under the alternative; then we test for the compatibility of this restriction with the other restriction on the sub-parameters (e.g., stochastic dominance, $\gamma \geq 0$, in our case). We adopt this approach and provide the dual expressions for both (6) and (7) as summarized in the following lemma. Note that the dual expressions involve only point-identified parameters.

Lemma 2. *The hypotheses in (6) and (7) are equivalent to*

$$\begin{cases} H_{0,a} : U^0 \leq L^1 \\ H_{1,a} : U^0 > L^1 \end{cases} \tag{8}$$

and

$$\begin{cases} H_{0,b} : L^0 \leq U^1 \\ H_{1,b} : L^0 > U^1, \end{cases} \tag{9}$$

respectively.

We illustrate this duality in Figure 1. The outside box denotes the parameter space $\Theta = [-1, 1] \times [0, 1]$, and the inside area of dashed lines denotes two inequalities, $L^1 \leq F^1 \leq U^1$. For the other two inequalities involving $(F^1 - \gamma)$, we consider three different sets of (L^0, U^0), and represent them as dotted lines. Thus, Figure 1 shows three identified sets, $\Theta_{I,1}, \Theta_{I,2}$, and $\Theta_{I,3}$, where we can easily check if all or some of $\gamma \in \Theta_I$ is positive or not. Now it is clear that $H^*_{0,a}$ and $H_{0,a}$ are equivalent: all $\gamma \in \Theta_I$ are positive whenever U^0 is less than L^1 as $\Theta_{I,3}$ illustrates. Conversely, if U^0 is bigger than L^1, then there exist negative $\gamma \in \Theta_I$. Similarly, $H^*_{0,b}$ and $H_{0,b}$ are equivalent: there exist positive $\gamma \in \Theta_I$ whenever L^0 is less than U^1 as $\Theta_{I,2}$ illustrates, and all $\gamma \in \Theta_I$ are negative whenever L^0 is bigger than U^1 as $\Theta_{I,1}$ shows.

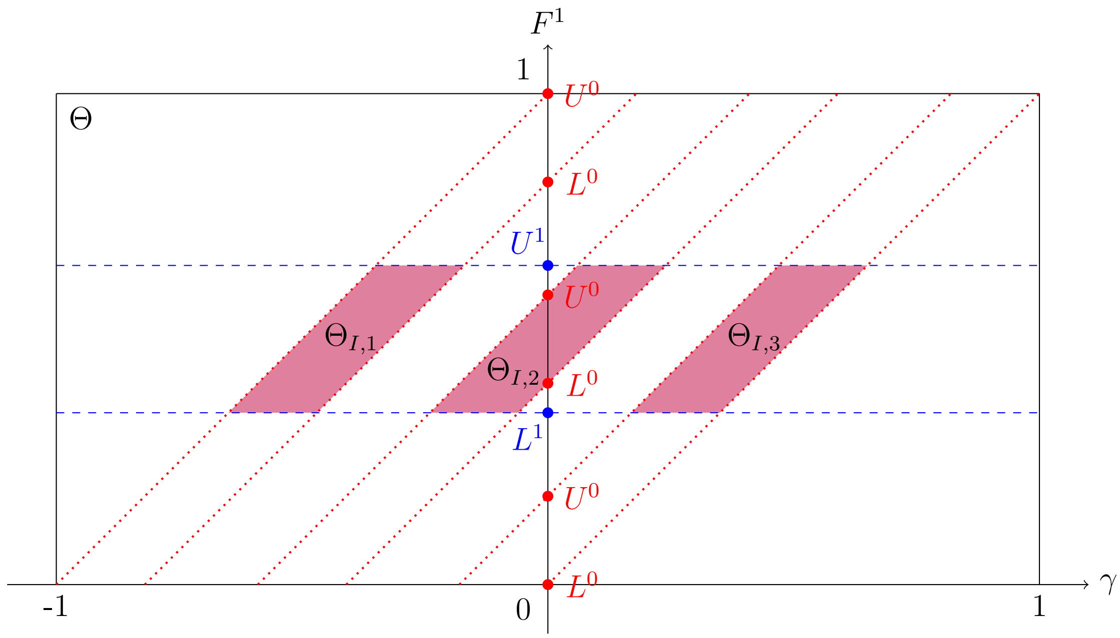


FIGURE 1. Graphical Illustration of the Duality

The dual property can be also understood in the following way. From the original moment inequalities (3) and (4), we can construct the bounds of $(F^1 - F^0)$:

$$L^1 - U^0 \leq F^1 - F^0 \leq U^1 - L^0. \tag{10}$$

Since all partially identified F^1 and F^0 satisfy this condition, $F^1 - F^0$ is positive for all such (F^1, F^0) whenever $L^1 - U^0 \geq 0$. This shows the equivalence of $H^*_{0,a}$ and $H_{0,a}$. Also, when $U^1 - L^0 \geq 0$, there exists (F^1, F^0) such that $0 \leq F^1 - F^0 \leq U^1 - L^0$ because $L^1 - U^0$ is always less than equal to $U^1 - L^0$ from the construction, which shows the equivalence of $H^*_{0,b}$ and $H_{0,b}$.

This dual approach has some advantages over the projection-based approach. First, the hypotheses involve only point-identified parameters, so that we can easily develop test statistics and their distribution theory. Second, this approach will give no worse (and likely better) power as is discussed above and in Hahn and Ridder (2009). Finally, the proposed method is computationally less demanding and does not require any tuning parameter for constructing the confidence region $\mathcal{C}_n$.

We are now ready to provide the general expression of the stochastic dominance test in our setup. Now uniformity over $(x, y) \in \mathcal{X} \times \mathbb{R}$ is explicit.

Test of Stochastic Dominance

$$\begin{cases} H^*_{0,a}: \ \forall F^1(\cdot|\cdot), F^0(\cdot|\cdot) \text{ satisfying the inequalities in Lemma 1, we have} \\ \qquad F^0(y|x) \le F^1(y|x) \text{ for all } (x,y) \in \mathcal{X} \times \mathbb{R}, \\ H^*_{1,a}: \ \exists F^1(\cdot|\cdot), F^0(\cdot|\cdot) \text{ satisfying the inequalities in Lemma 1 and} \\ \qquad F^0(y|x) > F^1(y|x) \text{ for some } (x,y) \in \mathcal{X} \times \mathbb{R}. \end{cases} \tag{11}$$

$$\begin{cases} H^*_{0,b}: \ \exists F^1(\cdot|\cdot), F^0(\cdot|\cdot) \text{ satisfying the inequalities in Lemma 1 and} \\ \qquad F^0(y|x) \le F^1(y|x) \text{ for all } (x,y) \in \mathcal{X} \times \mathbb{R}, \\ H^*_{1,b}: \ \forall F^1(\cdot|\cdot), F^0(\cdot|\cdot) \text{ satisfying the inequalities in Lemma 1, we have} \\ \qquad F^0(y|x) > F^1(y|x) \text{ for some } (x,y) \in \mathcal{X} \times \mathbb{R}. \end{cases} \tag{12}$$

We let $\Delta_a(x,y) := U^0(x,y) - L^1(x,y)$ and $\Delta_b(x,y) := L^0(x,y) - U^1(x,y)$. By the same logic as above, the hypotheses in (11) and (12) then have the dual characterizations as follows.

Lemma 3. *The hypotheses in (11) and (12) are equivalent to*

$$\begin{cases} H_{0,a}: \ \Delta_a(x,y) \le 0 \textit{ for all } (x,y) \in \mathcal{X} \times \mathbb{R} \\ H_{1,a}: \ \Delta_a(x,y) > 0 \textit{ for some } (x,y) \in \mathcal{X} \times \mathbb{R} \end{cases} \tag{13}$$

and

$$\begin{cases} H_{0,b}: \ \Delta_b(x,y) \le 0 \textit{ for all } (x,y) \in \mathcal{X} \times \mathbb{R} \\ H_{1,b}: \ \Delta_b(x,y) > 0 \textit{ for some } (x,y) \in \mathcal{X} \times \mathbb{R}, \end{cases} \tag{14}$$

respectively.

Again, note that both $\Delta_a(\cdot,\cdot)$ and $\Delta_b(\cdot,\cdot)$ are point-identified and our testing problem becomes the standard inequality test. As a result, the test statistic and the asymptotic properties can be derived in a standard way as we will see in the next section.

4. Test Statistics and Asymptotic Theory

The test statistics we consider use the sample analog of $\Delta_h(\cdot,\cdot)$ for $h = a, b$. We will focus on the case where the covariate X is discrete to highlight the main idea.[5] We

[5]An analysis of the continuous X case is highly nonstandard because of the lack of tightness in a typical nonparametric estimator like the one using a kernel. Lee and Whang (2009) showed that

assume that a sample $\{(Y_{it}, D_{it}, X_{it}')' : \ i = 1, 2, \cdots, n, \ t = 1, 2, \cdots, T\}$ is available, which is $i.i.d.$ over individuals. T is assumed to be fixed.

Let $\sum_{(t_1,\cdots,t_T)}$ be the summation over all the $T!$ permutations $(t_1, \cdots, t_T)$ of $(1, 2, \cdots, T)$. Using the binary indicator $\mathbb{I}\{\cdot\}$, we let $\widehat{\mathbb{P}}[X_{it_k} = x] := n^{-1} \sum_{i=1}^{n} \mathbb{I}\{X_{it_k} = x\}$ be the usual sample analog of $\mathbb{P}[X_{it_k} = x]$. Further, for each individual i and $j = 1, 0$, we define

$$1_{iT}\{x, y; j\} = \frac{1}{T!} \sum_{(t_1,\cdots,t_T)} \sum_{k=1}^{T} 1_i\{x, y; j, k\},$$

$$1_{iT}\{x; j\} = \frac{1}{T!} \sum_{(t_1,\cdots,t_T)} \mathbb{I}\{D_{it_1} = 1 - j, \cdots, D_{it_T} = 1 - j, X_{it_T} = x\}/\widehat{\mathbb{P}}[X_{it_T} = x],$$

where

$$1_i\{x, y; j, k\} = \begin{cases} \mathbb{I}\{Y_{it_1} \le y, D_{it_1} = j, X_{it_1} = x\}/\widehat{\mathbb{P}}[X_{it_1} = x], & \text{and} \\ \mathbb{I}\{Y_{it_k} \le y, D_{it_1} = 1 - j, \cdots, D_{it_{k-1}} = 1 - j, D_{it_k} = j, X_{it_k} = x\}/\widehat{\mathbb{P}}[X_{it_k} = x] & \text{for } k \ge 2. \end{cases}$$

Then the sample analogs of $L^j(x, y)$ and $U^j(x, y)$ are

$$\widehat{L}^j(x, y) = \frac{1}{n} \sum_{i=1}^{n} 1_{iT}\{x, y; j\} \quad \text{and} \quad \widehat{U}^j(x, y) = \frac{1}{n} \sum_{i=1}^{n} \big(1_{iT}\{x, y; j\} + 1_{iT}\{x; j\}\big).$$

Let $\widehat{\Delta}_a(x, y) = \widehat{U}^0(x, y) - \widehat{L}^1(x, y)$ and $\widehat{\Delta}_b(x, y) = \widehat{L}^0(x, y) - \widehat{U}^1(x, y)$. Then, for $h = a, b$, our statistics are defined as

$$\boldsymbol{T}_h^{KS} = \sup_{(x,y) \in \mathcal{X} \times \mathbb{R}} \sqrt{n} \widehat{\Delta}_h(x, y),$$

$$\boldsymbol{T}_h^{CV} = \sum_{x \in \mathcal{X}} \int_{\mathbb{R}} \max\{\sqrt{n} \widehat{\Delta}_h(x, y), \ 0\} w(x, y) dy,$$

where w is an integrable (nonnegative) weight function on $\mathcal{X} \times \mathbb{R}$.

These statistics are of the Kolmogorov–Smirnov type and the Cramér–von Mises type, respectively. They are simple summations of indicator functions; computing the whole confidence region $\mathcal{C}_n$ is not required. Therefore, even with simulating p-values as

a Poissonization technique is useful to analyze a Cramér-von Mises type statistic similar to the statistic considered in this section. In case of the Kolmogorov-Smirnov test, one may apply the strong approximation technique in Chernozhkukov, Lee, and Rosen (2009). A complete extension of these analyses to our context is beyond the scope of this paper.

explained below, the test procedures based on these statistics require less computation time than other projection-based alternatives.

We next turn our attention to the distributions of our statistics. For $x_1, x_2 \in \mathcal{X}$ and $y_1, y_2 \in \mathbb{R}$, we let $H_a(x_1, y_1, x_2, y_2)$ and $H_b(x_1, y_1, x_2, y_2)$ be defined as

$$\mathrm{Cov}\big(1^*_{iT}\{x_1, y_1; 0\} + 1^*_{iT}\{x_1; 0\} - 1^*_{iT}\{x_1, y_1; 1\}, 1^*_{iT}\{x_2, y_2; 0\} + 1^*_{iT}\{x_2; 0\} - 1^*_{iT}\{x_2, y_2; 1\}\big),$$
$$\mathrm{Cov}\big(1^*_{iT}\{x_1, y_1; 0\} - 1^*_{iT}\{x_1, y_1; 1\} - 1^*_{iT}\{x_1; 1\}, 1^*_{iT}\{x_2, y_2; 0\} - 1^*_{iT}\{x_2, y_2; 1\} - 1^*_{iT}\{x_2; 1\}\big),$$

respectively, where 1^*_{iT}'s are the population versions of 1_{iT}'s, i.e. using $\mathbb{P}[X_{it} = x]$ in lieu of $\widehat{\mathbb{P}}[X_{it} = x]$. Let $\mathbb{G}_a$ and $\mathbb{G}_b$ be Gaussian processes in $\ell^\infty(\mathcal{X} \times \mathbb{R})$ with the covariance kernels H_a and H_b, respectively.[6] We need to state one more assumption.

Assumption 5. $\inf_{x\in\mathcal{X}} \min_{1\le t\le T} \mathbb{P}[X_{it} = x] > 0$.

Assumption 5 is the usual regularity condition to exclude the possibility that the denominators of the statistics are too small. If this assumption is not satisfied, trimming will be needed.

Theorem 1 (Null Distributions)**.** *Suppose that Assumptions 1-5 hold. Let $h \in \{a, b\}$. Under $H_{0,h}$, there exist sequences of random variables $\{Z^{KS}_{n,h}\}$ and $\{Z^{CV}_{n,h}\}$ such that*

$$\boldsymbol{T}^{KS}_h \le Z^{KS}_{n,h} \xrightarrow{d} \sup_{(x,y)\in\mathcal{X}\times\mathbb{R}} \mathbb{G}_h(x, y),$$
$$\boldsymbol{T}^{CV}_h \le Z^{CV}_{n,h} \xrightarrow{d} \sum_{x\in\mathcal{X}} \int_{\mathbb{R}} \max\{\mathbb{G}_h(x, y),\ 0\} w(x, y) dy$$

as $n \to \infty$, where the inequalities hold with probability one and they hold as equalities when $\Delta_h(x, y) = 0$ for all $(x, y) \in \mathcal{X} \times \mathbb{R}$.

Since our test is a one-sided test, the inequalities in Theorem 1 are natural. Specifically, the inequalities become equalities when $\Delta_h(x, y) = 0$ for all $(x, y) \in \mathcal{X} \times \mathbb{R}$. Therefore, we can use quantiles of $Z^{KS}_{n,h}$ or $Z^{CV}_{n,h}$ to control for the maximum probability of the type I error, where those quantiles can be obtained by either simulations or the bootstrap as Barrett and Donald (2003).[7] In particular, simulating p-values can be done by adopting the p-value transformation method of Hansen (1996). The key idea is that the Gaussian process $\mathbb{G}_h$ can be simulated by $\hat{\mathbb{G}}_{s,h}$ defined in (15) below,

[6] $\ell^\infty(\mathcal{X} \times \mathbb{R})$ is the collection of all bounded functions from $\mathcal{X} \times \mathbb{R}$ to $\mathbb{R}$.

[7] When w is a probability density, the integral can be computed by a Monte Carlo method, because the integral can be approximated by discretizing the support of w due to the stochastic equicontinuity of $\mathbb{G}_h$.

where $\hat{\mathbb{G}}_{s,h}$ weakly converges in probability to $\mathbb{G}_h$ in $\ell^\infty(\mathcal{X} \times \mathbb{R})$ in the sense of Giné and Zinn (1990), for example. We illustrate the simulation procedure in more detail below.

Let $s \in \{1, 2, \cdots, S\}$ denote each simulation, where S is the number of replications we consider. For each s, let $\{u_{si} : \ i = 1, 2, \cdots, n\}$ be *i.i.d.* draws from the standard normal distribution that is independent of the data. For $(x, y) \in \mathcal{X} \times \mathbb{R}$, we define

$$\hat{\mathbb{G}}_{s,h}(x, y) = n^{-1/2} \sum_{i=1}^{n} \Big(1_{iT,h}\{x, y\} - \widehat{\Delta}_h(x, y) \Big) u_{si}, \tag{15}$$

where $1_{iT,a}\{x, y\} = 1_{iT}\{x, y; 0\} + 1_{iT}\{x; 0\} - 1_{iT}\{x, y; 1\}$ and $1_{iT,b}\{x, y\} = 1_{iT}\{x, y; 0\} - 1_{iT}\{x, y; 1\} - 1_{iT}\{x; 1\}$. Further, we define

$$\widehat{\boldsymbol{Z}}^{KS}_{s,h} = \sup_{(x,y) \in \mathcal{X} \times \mathbb{R}} \hat{\mathbb{G}}_{s,h}(x, y) \quad \text{and} \quad \widehat{\boldsymbol{Z}}^{CV}_{s,h} = \sum_{x \in \mathcal{X}} \int_{\mathbb{R}} \max\{\hat{\mathbb{G}}_{s,h}(x, y), \ 0\} w(x, y) dy.$$

Then, the simulated p-values are given by

$$\hat{p}^{KS}_h = \frac{1}{S} \sum_{i=1}^{S} 1\big\{ \widehat{\boldsymbol{Z}}^{KS}_{s,h} \geq \boldsymbol{T}^{KS}_h \big\} \quad \text{and} \quad \hat{p}^{CV}_h = \frac{1}{S} \sum_{i=1}^{S} 1\big\{ \widehat{\boldsymbol{Z}}^{CV}_{s,h} \geq \boldsymbol{T}^{CV}_h \big\}.$$

We now discuss the power properties of the tests. Basically, all the standard properties of the Kolmogorov–Smirnov test and the Cramér–von Mises test apply in our context.

Theorem 2 (Power)**.** *Suppose that Assumptions 1-5 hold. Let C be any constant and let $h \in \{a, b\}$. Then, under $H_{1,h}$,*

$$\mathbb{P}[\boldsymbol{T}^{KS}_h > C] \to 1 \quad \textit{and} \quad \mathbb{P}[\boldsymbol{T}^{CV}_h > C] \to 1$$

as $n \to \infty$, provided that

$$\sup_{(x,y) \in \mathcal{X} \times \mathbb{R}} \Delta_h(x, y) > 0 \quad \textit{and} \quad \sum_{x \in \mathcal{X}} \int_{\mathbb{R}} \max\{\Delta_h(x, y), \ 0\} w(x, y) dy > 0,$$

respectively.

Theorem 2 shows that our tests are consistent against any fixed alternatives. Our tests also have non-trivial local power against $\sqrt{n}$–local alternatives. For this discussion, we consider local deviations from the least favorable null:

$$H_{1,h,n} : \ \Delta_h(x, y) = \delta_h(x, y)/\sqrt{n} \tag{16}$$

for some nonnegative function $\delta_h(\cdot,\cdot)$. The theorem below shows the local power properties of our tests.

Theorem 3 (Local Power). *Suppose that Assumptions 1-5 hold. Let $c_h^{KS}(\beta)$ and $c_h^{CV}(\beta)$ be critical values for size $\beta \in (0,1)$ based on the null distributions given in Theorem 1. Let $h \in \{a,b\}$. Under $H_{1,h,n}$ in (16),*

$$\lim_{n\to\infty} \mathbb{P}\left[\boldsymbol{T}_h^{KS} > c_h^{KS}(\beta)\right] \geq \beta \quad \textit{and} \quad \lim_{n\to\infty} \mathbb{P}\left[\boldsymbol{T}_h^{CV} > c_h^{CV}(\beta)\right] \geq \beta,$$

where the inequalities are strict if

$$\sup_{(x,y)\in\mathbb{R}} \delta_h(x,y) > 0 \quad \textit{and} \quad \sum_{x\in\mathcal{X}} \int_{\mathbb{R}} \delta_h(x,y) w(x,y) dy > 0,$$

respectively.

Several extensions of the results above are possible. First, allowing for the case of continuous X_{it} is important. As we quickly discussed in Footnote 5, the poissonization technique in Lee and Whang (2009) or the strong approximation in Chernozhkukov, Lee, and Rosen (2009) can be considered. Second, the idea of a contact set as in Linton, Song, and Whang (2010) can improve the power of the tests. The idea is that it is not the entire $\mathcal{X} \times \mathbb{R}$ but only a subset of it that plays a relevant role in distinguishing the null from the alternative. To be more specific, modifying $\boldsymbol{T}_h^{KS}$ and $\boldsymbol{T}_h^{CV}$ by taking supremum or integrating over the contact set $\mathcal{C}_h = \{(x,y) \in \mathcal{X} \times \mathbb{B} : \ \Delta_h(x,y) = 0\}$ (and adjusting critical values accordingly) can improve the power. However, this improvement requires that some tuning parameter b_n be introduced, because the contact set needs to estimated as $\hat{\mathcal{C}}_h = \{(x,y) \in \mathcal{X} \times \mathbb{R} : \ |\widehat{\Delta}_h(x,h)| \leq b_n\}$. Third, we have focused on first-order stochastic dominance, but the idea can be also extended to higher order dominance.

5. Empirical Illustration: The Effect of Smoking on Birthweight

As an empirical illustration, we analyze the effect of smoking during pregnancy on infant's birthweight (e.g., Permutt and Hebel (1989), Evans and Ringel (1999), Abrevaya (2006), and Abrevaya and Dahl (2008)). We test for the presence of first-order stochastic dominance between potential birthweight outcomes. Our analysis is based on the data set constructed by Abrevaya (2006) from the U.S. Natality Data Set in 1990–1998. We select the "matched panel #3" as it is constructed in the most conservative way. The same data set (but only with the switchers) is also used by Arellano and Bonhomme (2011) in the random coefficients panel model.

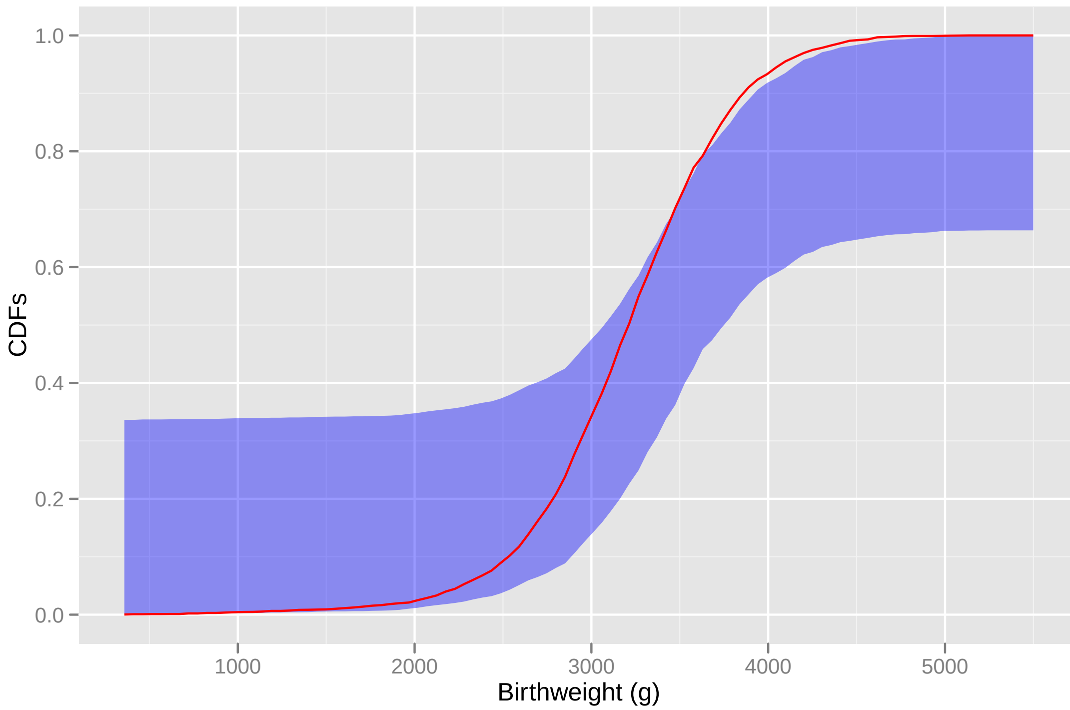
We focus on the $n = 2,113$ sample of those who had *three births*, had *ever smoked* during pregnancy (ever-smokers), and gave at least one birth when they were *between 18 and 35 years old*. Now that an observation is partly missing when a mother gave a birth outside the age range, we have unbalanced panel data with $(n_1, n_2, n_3) = (1979, 2038, 1942)$, where n_t is the number of observations at the t-th birth. The reason for omitting those who never smoked (never-smokers) is that the sample size of 'never-smokers' is too large (more than 82% of the entire sample) to obtain any meaningful bounds. Moreover, from the policy perspective, ever-smokers make a more relevant population under the presumption that ever-smokers may quit smoking in the future but never-smokers are unlikely to start smoking during pregnancy. As a result, $F^1(\cdot|\cdot)$, the (conditional) distribution of the potential birthweight when a mother smokes, is point-identified while the other distribution $F^0(\cdot|\cdot)$ is only partially identified. Also, note that all birthweight observations are mean-adjusted as there is a tendency that, among siblings, the first born are lighter than those born in the later order on average.

Figure 2 shows the estimation results without conditioning. The solid line stands for the estimate of $F^1(\cdot)$ and the shaded area for the bound estimates of $F^0(\cdot)$. We can find that $\widehat{F}^1(\cdot)$ is located above the lower bound estimate of $F^0(\cdot)$ over all birthweights, which suggests that there might exist F^0 that stochastically dominates F^1 in the first order. However, $\widehat{F}^1(\cdot)$ is clearly above the upper bound estimate of $F^0(\cdot)$ around the birthweight near $4,000$ grams, suggesting that there might not exist $F^0(\cdot)$ that is stochastically dominated by $F^1(\cdot)$.

This graphical intuition is formally confirmed by the test results in Table 1. The table reports p-values for several hypotheses. The p-values are calculated by the simulated p-value method discussed in Section 4 with the number of simulations S equal to $1,000$. For the weight $w(x, y)$ in $\boldsymbol{T}_b^{CV}$, we use the product of empirical densities given by

$$\begin{aligned} w(x, y_m) &= \widehat{\mathbb{P}}\{X_{it} = x\}\widehat{\mathbb{P}}\{y_{m-1} < Y_{it} \le y_m\} \\ &= \frac{1}{n}\sum_{i=1}^{n} 1\{X_{it} = x\} \times \frac{1}{n}\sum_{i=1}^{n} 1\{y_{m-1} < Y_{it} \le y_m\}, \end{aligned}$$

FIGURE 2. Distributions of Potential Birthweights



Note: The solid line stands for the estimate of $F^1(\cdot)$, and the shaded areas for the bound estimates of $F^0(\cdot)$.

where x is a point in the support of the discrete X_{it} and $y_{\min} = y_0 < y_1 < \cdots < y_m < \cdots < y_{M-1} < y_M = y_{\max}$ is some grid points over the support of Y_{it}: x is used only for *Conditional Test*. We use equi-spaced grid points with $M = 100$.

The first panel *Unconditional Test* in the table reports p-values for the null hypothesis that there exist $F^0(\cdot)$ and $F^1(\cdot)$ in the identified region such that $F^0(\cdot)$ first-order stochastically dominates (FSD, hereafter) $F^1(\cdot)$, i.e. type-b hypothesis in Section 3.[8] We next switch the direction and test if $F^1(\cdot)$ FSD $F^0(\cdot)$. As we expected from Figure 2, neither $\boldsymbol{T}_b^{KS}$ nor $\boldsymbol{T}_b^{CV}$ can reject the null of 'F^0 FSD F^1' with yielding very large p-values. However, the p-values for the flipped null are clearly smaller and it is at the margin of a 5% confidence level in case of $\boldsymbol{T}_b^{KS}$.

This result is further reinforced when we apply the conditional (or uniform) stochastic dominance test. We consider three conditioning variables: *Age*, *High-school*, and

[8]Note that we do not report the test results for the type-a hypotheses since they are obviously rejected for both directions.

TABLE 1. p-values of the Type-b Tests (ever-smoker, 3-birth, age 18-35)

Null Hypothesis	F^0 FSD F^1		F^1 FSD F^0	
	$\boldsymbol{T}_b^{KS}$	$\boldsymbol{T}_b^{CV}$	$\boldsymbol{T}_b^{KS}$	$\boldsymbol{T}_b^{CV}$
Unconditional Test	0.995	0.484	0.054	0.217
Conditional Test				
Age	1.000	0.485	0.001	0.027
High-school	1.000	0.500	0.012	0.073
Marriage	1.000	0.488	0.047	0.114

Note: All results are based on the ever-smoker samples with 3 births, who gave births at the age between 18 and 35, where the birthweight observations are mean-adjusted by the amounts of mean shifts for each t. FSD stands for 'first-order stochastically dominates.' $\boldsymbol{T}_b^{KS}$ is the Kolmogorov-Smirnov type test and $\boldsymbol{T}_b^{CV}$ is the Cramer-von Mises type test for the type-b hypotheses.

Marriage. *Age* divides the sample into two groups depending on whether a baby was born when a mother was between 18–26 years old or between 27–35 years old. *High-school* and *Marriage* denote if she has a high-school degree and if she is married, respectively. From the original construction of the data set, *Age* is a time-varying covariate but the others are time-invariant. The second panel of Table 1 under *Conditional Test* shows the testing results. Again, we cannot reject the null hypotheses of 'F^0 FSD F^1' in any cases, but the p-values from 'F^1 FSD F^0' are much lower than those from *Unconditional Test*. This result coincides with our intuition that the conditional test provides more information by detecting heterogenous treatment effects that could be just averaged out by the unconditional test.

In sum, these results illustrate that the proposed method helps derive robust empirical results under flexible model restrictions. Also, this empirical application shows how easily the testing procedure can be implemented into the real data.

Appendix A. Proofs

A.1. **Proof of Lemma 1.** We only consider $T = 2$ because it can be shown similarly using induction arguments when $T > 2$. Consider a particular permutation $(t_1, t_2) = (1, 2)$ and $j = 1$ for notational simplicity. Then, we want to show the following:

$$\mathbb{P}[Y_1^1 \leq y | X_1 = x] = \mathbb{P}[Y_1 \leq y, D_1 = 1 | X_1 = x] + \mathbb{P}[Y_2 \leq y, D_1 = 0, D_2 = 1 | X_2 = x] + \mathbb{P}[Y_2^1 \leq y | D_1 = 0, D_2 = 0, X_2 = x] \cdot \mathbb{P}[D_1 = 0, D_2 = 0 | X_2 = x], \tag{17}$$

where the first two terms in (17) are identified but the last term is bounded by an interval of

$$\big[0, \mathbb{P}[D_1 = 0, D_2 = 0 | X_2 = x]\big]. \tag{18}$$

First, note that

$$\mathbb{P}[Y_1^1 \leq y | X_1 = x] = \mathbb{P}[Y_1 \leq y, D_1 = 1 | X_1 = x] + \mathbb{P}[Y_1^1 \leq y, D_1 = 0 | X_1 = x], \tag{19}$$

where the first term corresponds to the first term on the right hand side of (17). For the second term in (19), we have

$$\mathbb{P}[Y_1^1 \leq y, D_1 = 0 | X_1 = x] = \mathbb{P}[Y_1^1 \leq y, D_1 = 0, D_2 = 1 | X_1 = x] + \mathbb{P}[Y_1^1 \leq y, D_1 = 0, D_2 = 0 | X_1 = x]. \tag{20}$$

Now we focus on the first term of (20):

$$\mathbb{P}[Y_1^1 \leq y, D_1 = 0, D_2 = 1 | X_1 = x] \tag{21}$$

$$= \int \mathbb{P}[Y_1^1 \leq y, D_1 = 0, D_2 = 1 | X_1 = x, \alpha = a] d\mu(a | X_1 = x) \tag{22}$$

$$= \int \mathbb{P}[Y_1^1 \leq y | X_1 = x, \alpha = a] \mathbb{P}[D_1 = 0, D_2 = 1 | X_1 = x, \alpha = a] d\mu(a | X_1 = x) \tag{23}$$

$$= \int \mathbb{P}[Y_2^1 \leq y | X_2 = x, \alpha = a] \mathbb{P}[D_1 = 0, D_2 = 1 | X_2 = x, \alpha = a] d\mu(a | X_2 = x) \tag{24}$$

$$= \int \mathbb{P}[Y_2^1 \leq y, D_1 = 0, D_2 = 1 | X_2 = x, \alpha = a] d\mu(a | X_2 = x) \tag{25}$$

$$= \int \mathbb{P}[Y_2 \leq y, D_1 = 0, D_2 = 1 | X_2 = x, \alpha = a] d\mu(a | X_2 = x) \tag{26}$$

$$= \mathbb{P}[Y_2 \leq y, D_1 = 0, D_2 = 1 | X_2 = x], \tag{27}$$

which is equal to the second term on the right hand side of (17). Note that equalities in (23) and (25) hold from Assumption 1 and equality in (24) from Assumptions 2–3 and the Bayes theorem. Similarly, it can be shown that

$$\mathbb{P}[Y_1^1 \leq y, D_1 = 0 | X_1 = x] = \mathbb{P}[Y_2^1 \leq y | D_1 = 0, D_2 = 0, X_2 = x] \cdot \mathbb{P}[D_1 = 0, D_2 = 0 | X_2 = x],$$

which establishes equation (17).

Now, we show that the bounds do not depend on a particular permutation using Assumption 4. For a fixed $\ell = (t_1, t_2, \cdots, t_T) \in \mathcal{T}$, we focus on $p^1_{k\ell}(y|x)$ for $k \geq 2$, where the case of $k = 1$ is similar but simpler. Let $\bar{D}_k = (D_{t_1}, \cdots, D_{t_k})$ and $d_k = (0, \cdots, 0, 1)$ corresponding to $\bar{D}_k$. Then

$$\begin{aligned} p^1_{k\ell}(y|x) &= \int \mathbb{P}[Y_{t_k} \leq y, \bar{D}_k = d_k | X_{t_k} = x, \alpha = a] d\mu(a | X_{t_k} = x) \\ &= \int \mathbb{P}[Y^1_{t_k} \leq y | X_{t_k} = x, \alpha = a] \mathbb{P}[\bar{D}_k = d_k | X_{t_k} = x, \alpha = a] d\mu(a | X_{t_k} = x) \\ &= \int \mathbb{P}[Y^1_k \leq y | X_k = x, \alpha = a] \mathbb{P}[\bar{D}_k = d_k | X_{t_k} = x, \alpha = a] d\mu(a | X_k = x), \end{aligned} \tag{28}$$

where these equalities hold by Assumptions 1 and 2. Note here that

$$\mathbb{P}[\bar{D}_k = d_k | X_{t_k} = x, \alpha = a] = \mathbb{P}[\bar{D}_k = d_k | X_k = x, \alpha = a] \tag{29}$$

by Assumptions 2, 3, and the Bayes theorem. Applying Assumption 4 to the right hand side of (29) and combining it with (28) completes the proof. □

A.2. **Proof of Lemma 2.** The proof is omitted since the general version is proved in Lemma 3 below. □

A.3. **Proof of Lemma 3.** In each case, it is enough to show the equivalence of the null hypotheses. For simplicity, we assume that both Y^0 and Y^1 have unbounded support $\mathbb{R}$. We first show the equivalence of $H^*_{0,a}$ and $H_{0,a}$.

• **Equivalence of $H^*_{0,a}$ and $H_{0,a}$**

(i) Sufficiency: Let $H^*_{0,a}$ be true. Suppose that $H_{0,a}$ does not hold. Then, there exist some $(\tilde{x}, \tilde{y}) \in \mathcal{X} \times \mathbb{R}$ such that $\Delta_a(\tilde{x}, \tilde{y}) = U^0(\tilde{x}, \tilde{y}) - L^1(\tilde{x}, \tilde{y}) > 0$. We need to show that there exist F^0 and F^1 contradicting to $H^*_{0,a}$. For such $\tilde{x}$ and any $\epsilon > 0$, define $F^0(\cdot|\tilde{x})$ and $F^1(\cdot|\tilde{x})$ as (a) $F^0(y|\tilde{x}) = L^0(\tilde{x}, y)$ for $y < \tilde{y}$; (b) $F^1(y|\tilde{x}) = U^1(\tilde{x}, y)$ for $y \geq \tilde{y} + \epsilon$; (c) $F^0(y|\tilde{x}) = U^0(\tilde{x}, y)$ and $F^1(y|\tilde{x}) = L^1(\tilde{x}, y)$ otherwise. By construction, they are distribution functions satisfying the inequalities in Lemma 1, but $F^0(\tilde{y}|\tilde{x}) > F^1(\tilde{y}|\tilde{x})$ contradicts to $H^*_{0,a}$.

(ii) Necessity: Let $H_{0,a}$ be true. Then, for all $(x,y) \in \mathcal{X} \times \mathbb{R}$, $\Delta_a(x,y) = U^0(x,y) - L^1(x,y) \leq 0$. For any F^0 and F^1 satisfying the inequalities in Lemma 1, this implies

$$F^0(x,y) \leq U^0(x,y) \leq L^1(x,y) \leq F^1(x,y)$$

for all $(x,y) \in \mathcal{X} \times \mathbb{R}$. Thus, $H^*_{1,a}$ holds.

We next show the equivalence of $H^*_{0,b}$ and $H_{0,b}$.

- **Equivalence of $H^*_{0,b}$ and $H_{0,b}$**

(i) Sufficiency: Let $H^*_{0,b}$ be true. Then, there exist distribution functions F^0 and F^1 such that, for any $(x,y) \in \mathcal{X} \times \mathbb{R}$, they satisfy the inequalities in Lemma 1 and $F^0(y|x) \leq F^1(y|x)$. Fix such F^0 and F^1 and suppose that $H_{0,b}$ is not true. Then, there exists $(\tilde{x},\tilde{y}) \in \mathcal{X} \times \mathbb{R}$ such that $\Delta_b(\tilde{x},\tilde{y}) = L^0(\tilde{x},\tilde{y}) - U^1(\tilde{x},\tilde{y}) > 0$. Since F^0 and F^1 satisfy the inequalities in Lemma 1, this implies

$$F^1(\tilde{y}|\tilde{x}) \leq U^1(\tilde{x},\tilde{y}) < L^0(\tilde{x},\tilde{y}) \leq F^0(\tilde{y}|\tilde{x}).$$

Therefore, $F^1(\tilde{y}|\tilde{x}) < F^0(\tilde{y}|\tilde{x})$, which contradicts to $H^*_{0,b}$.

(ii) Necessity: Let $H_{0,b}$ be true. We prove this by constructing distribution functions F^0 and F^1 satisfying the inequalities in Lemma 1 and $F^0(y|x) \leq F^1(y|x)$ for all $(x,y) \in \mathcal{X} \times \mathbb{R}$. For constants $C_1 < C_2$ and $X = x$ given, define F^0 and F^1 as follows: (a) $F^1(y|x) = \max\{L^0(x,y), L^1(x,y)\}$ for $y < C_1$; (b) $F^0(y|x) = \min\{U^0(x,y), U^1(x,y)\}$ for $y \geq C_2$; (c) $F^0(y|x) = L^0(x,y)$ and $F^1(y|x) = U^1(x,y)$ otherwise. Then, $F^0(y|x)$ and $F^1(y|x)$ satisfy the inequalities in Lemma 1 and $F^0(y|x) \leq F^1(y|x)$ for all $(x,y) \in \mathcal{X} \times \mathbb{R}$ by construction. Note also that F_0 and F^1 are distribution functions since they are CADLAG and go to 1 and 0 as $y \to \pm\infty$, respectively. □

A.4. **Proof of Theorem 1.** Note that under $H_{0,h}$, for all $(x,y) \in \mathcal{X} \times \mathbb{R}$,

$$\sqrt{n}\widehat{\Delta}_h(x,y) \leq \sqrt{n}\left(\widehat{\Delta}_h(x,y) - \Delta_h(x,y)\right),$$
$$\max\left\{\sqrt{n}\widehat{\Delta}_h(x,y),\ 0\right\} \leq \max\left\{\sqrt{n}\left(\widehat{\Delta}_h(x,y) - \Delta_h(x,y)\right),\ 0\right\},$$

where the inequalities become equalities when $\Delta_h(x,y) = 0$ for all $(x,y) \in \mathcal{X} \times \mathbb{R}$. Therefore, by the continuous mapping theorem, it suffices to show that $\sqrt{n}(\widehat{\Delta}_h - \Delta_h)$ weakly converges to $\mathbb{G}_h$ in $\ell^\infty(\mathcal{X} \times \mathbb{R})$. By Assumption 5 and the law of large numbers, we know that $1/\widehat{\mathbb{P}}[X_{it} = x]$ uniformly converges to $1/\mathbb{P}[X_{it} = x]$. Combining this with the central limit theorem shows that the weak convergence of $\sqrt{n}(\widehat{\Delta}_h - \Delta_h)$ will follow from that of $\sqrt{n}(\widehat{\Delta}^*_h - \Delta_h)$, where $\widehat{\Delta}^*_h$ is the infeasible version of $\widehat{\Delta}_h$ with $\mathbb{P}[X_{it} = x]$ in lieu of $\widehat{\mathbb{P}}[X_{it} = x]$. The weak convergence of $\sqrt{n}(\widehat{\Delta}_h^* - \Delta_h)$ follows from the fact that $\mathcal{X}$ is discrete and that the collection of functions $\mathbb{I}\{Y_{it_1} \leq y, D_{it_1} = j, X_{it_1} = x\}$ and $\mathbb{I}\{Y_{it_k} \leq y, D_{it_1} = 1-j, \cdots, D_{it_{k-1}} = 1-j, D_{it_k} = j, X_{it_k} = x\}$ is a Vapnik–Çervonenkis class. □

A.5. **Proof of Theorem 2.** Note that

$$\sqrt{n}\left|\sup \widehat{\Delta}_h(x,y) - \sup \Delta_h(x,y)\right| \leq \sqrt{n}\sup\left|\widehat{\Delta}_h(x,y) - \Delta_h(x,y)\right| = O_p(1),$$

where sup is taken over $(x,y) \in \mathcal{X} \times \mathbb{R}$. Therefore,

$$\begin{aligned}\boldsymbol{T}_h^{KS} &= \sqrt{n}\left(\sup \widehat{\Delta}_h(x,y) - \sup \Delta_h(x,y)\right) + \sqrt{n}\sup \Delta_h(x,y) \\ &= O_p(1) + \sqrt{n}\sup \Delta_h(x,y),\end{aligned}$$

where the second term goes to $+\infty$ under H_{1h}. For $\boldsymbol{T}_h^{CV}$, similarly

$$\begin{aligned}&\left|\sum_{x\in\mathcal{X}}\int_{\mathbb{R}} \max\{\sqrt{n}\widehat{\Delta}_h(x,y),\ 0\}w(x,y)dy - \sum_{x\in\mathcal{X}}\int_{\mathbb{R}} \max\{\sqrt{n}\Delta_h(x,y),\ 0\}w(x,y)dy\right| \\ &\qquad \leq \left|\sum_{x\in\mathcal{X}}\int_{\mathbb{R}} \max\left\{\sqrt{n}\left(\widehat{\Delta}_h(x,y) - \Delta_h(x,y)\right),\ 0\right\}w(x,y)dy\right| = O_p(1),\end{aligned}$$

where we used the fact that $\max(a+b,0) \leq \max(a,0) + \max(b,0)$. Therefore,

$$\begin{aligned}\boldsymbol{T}_h^{CV} &= \sum_{x\in\mathcal{X}}\int_{\mathbb{R}} \max\{\sqrt{n}\widehat{\Delta}_h(x,y),\ 0\}w(x,y)dy - \sum_{x\in\mathcal{X}}\int_{\mathbb{R}} \max\{\sqrt{n}\Delta_h(x,y),\ 0\}w(x,y)dy \\ &\quad + \sum_{x\in\mathcal{X}}\int_{\mathbb{R}} \max\{\sqrt{n}\Delta_h(x,y),\ 0\}w(x,y)dy \\ &= O_p(1) + \sqrt{n}\int_{\mathbb{R}} \max\{\Delta_h(x,y),\ 0\}w(x,y)dy,\end{aligned}$$

where the second term again goes to $+\infty$ under H_{1h}. □

A.6. **Proof of Theorem 3.** Note that $\boldsymbol{T}_h^{KS}$ is equal to

$$\sup_{(x,y)\in\mathcal{X}\times\mathbb{R}}\left\{\sqrt{n}\left(\widehat{\Delta}_h(x,y) - \Delta_h(x,y)\right) + \delta_h(x,y)\right\} \xrightarrow{d} \sup_{(x,y)\in\mathcal{X}\times\mathbb{R}}\left\{\mathbb{G}_h(x,y) + \delta_h(x,y)\right\}. \tag{30}$$

Similarly, $\boldsymbol{T}_h^{CV}$ is equal to

$$\sum_{x\in\mathcal{X}}\int_{\mathbb{R}}\max\left\{\sqrt{n}\left(\widehat{\Delta}_h(x,y)-\Delta_h(x,y)\right)+\delta_h(x,y),\ 0\right\}w(x,y)dy$$
$$\xrightarrow{d}\sum_{x\in\mathcal{X}}\int_{\mathbb{R}}\max\left\{\mathbb{G}_h(x,y)+\delta_h(x,y),\ 0\right\}w(x,y)dy. \quad (31)$$

The expressions on the right hand side of (30) and (31) are simply right–shifted versions of the null distributions of $\boldsymbol{T}_h^{KS}$ and $\boldsymbol{T}_h^{CV}$, respectively. Since it is assumed that $\delta_h(x,y)\geq 0$, as long as $\delta_h(x,y)>0$ for some $(x,y)\in\mathcal{X}\times\mathbb{R}$ and $\sum_{x\in\mathcal{X}}\int_{\mathbb{R}}\max\{\delta_h(x,y),\ 0\}w(x,y)dy=\sum_{x\in\mathcal{X}}\int_{\mathbb{R}}\delta_h(x,y)w(x,y)dy>0$, respectively, the right–shift is strict, which implies that the rejection probability will be strictly larger than the nominal size. $\square$

References Cited

Abadie, A. (2002): "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American Statistical Association*, 97(457), 284–292.

Abrevaya, J. (2006): "Estimating the effect of smoking on birth outcomes using a matched panel data approach," *Journal of Applied Econometrics*, 21(4), 489–519.

Abrevaya, J., and C. M. Dahl (2008): "The Effects of Birth Inputs on Birthweight: Evidence From Quantile Estimation on Panel Data," *Journal of Business and Economic Statistics*, 155, 56–70.

Andrews, D., S. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete GAmes with Multiple Equilibria with an Application to Discount Chain Store," mimeo, Yale University.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of causal effects using instrumental variables.," *Journal of the American Statistical Association*, 91(434).

Arellano, M., and S. Bonhomme (2011): "Identifying distributional characteristics in random coefficients panel data models," *The Review of Economic Studies*, forthcoming.

Barrett, G., and S. Donald (2003): "Consistent tests for stochastic dominance," *Econometrica*, 71(1), 71–104.

Beck, M., and B. Fitzenberger (2004): "Changes in Union Membership Over Time: A Panel Analysis for West Germany," *LABOUR*, 18, 329–362.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Bhattacharya, J., A. Shaikh, and E. Vytlacil (2012): "Treatment effect bounds: An application to Swan–Ganz catheterization," *Journal of Econometrics*, 168(2), 223–243.

Blakemore, A., J. Hunt, and B. Kiker (1986): "Collective bargaining and union membership effects on the wages of male youths," *Journal of Labor Economics*, 4(2), 193–211.

Budd, J., and I. Na (2000): "The union membership wage premium for employees covered by collective bargaining agreements," *Journal of Labor Economics*, 18(4), 783–807.

Chernozhkukov, V., S. Lee, and A. Rosen (2009): "Intersection Bounds: Estimation and Inference," mimeo, MIT.

Chernozhukov, V., I. Fernandez-Val, and W. Newey (2009): "Quantile and average effects in nonseparable panel models," Discussion paper, CEMMAP.

Chernozhukov, V., and C. Hansen (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132(2), 491–525.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2008): "Nonparametric Tests for Treatment Effect Heterogeneity," *The Review of Economics and Statistics*, 90(3), 389–405.

Dehejia, R., and S. Wahba (2002): "Propensity score-matching methods for nonexperimental causal studies," *Review of Economics and statistics*, 84(1), 151–161.

Djebbari, H., and J. Smith (2008): "Heterogeneous impacts in PROGRESA," *Journal of Econometrics*, 145(1-2), 64–80.

Donald, S., and Y. Hsu (2010): "Improving the power of tests of stochastic dominance," mimeo, University of Missouri at Columbia.

Evans, W., and J. S. Ringel (1999): "Can Higher Cigarette Taxes Improve Birth Outcomes?," *Journal of Public Economics*, 72, 135–154.

Fan, Y., and S. Park (2010): "Sharp bounds on the distribution of treatment effects and their statistical inference," *Econometric Theory*, 26(3), 931–951.

——— (2011): "Confidence intervals for the quantile of treatment effects in randomized experiments," *Journal of Econometrics*, 167(2), 330–344.

Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259–276.

Giné, E., and J. Zinn (1990): "Bootstrapping general empirical measures," *The Annals of Probability*, 18(2), 851–869.

Guggenberger, P., J. Hahn, and K. Kim (2008): "Speicification testing under moment inequalities," *Economics Letters*, 99, 375–378.

Hahn, J. (1998): "On the role of the propensity score in semiparametric estimation of average treatment effects," *Econometrica*, 66, 315–331.

Hahn, J., and G. Ridder (2009): "Partial identification and confidence intervals," mimeo, UCLA.

Hansen, B. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica: Journal of the Econometric Society*, pp. 413–430.

Heckman, J., and J. Smith (1997): "Making the Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economics Studies*, 64, 487–535.

Heckman, J. J., and E. J. Vytlacil (2007a): "Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B of *Handbook of Econometrics*, chap. 70, pp. 4779 – 4874. Elsevier.

——— (2007b): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6B of *Handbook of Econometrics*, chap. 71, pp. 4875–5143. Elsevier.

Hirano, K., G. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161–1189.

Imbens, G. W., and J. M. Wooldridge (2009): "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 47(1), 5–86.

Jones, E. (1982): "Union/Nonunion differentials: Membership or coverage?," *Journal of Human Resources*, 17(2), 276–285.

Jung, K. (2010): "Incentives for Voluntary Disclosure of Quality Information in HMO Markets," *Journal of Risk and Insurance*, 77(1), 183–210.

Khan, S., M. Ponomareva, and E. Tamer (2011): "Identification of Panel Data Models with Endogenous Censoring," mimeo, Duke University.

Kim, K. (2009): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," mimeo, University of Minnesota.

LEE, M. J. (2009): "Non-parametric tests for distributional treatment effect for randomly censored responses," *Journal of the Royal Statistical Society Series B*, 71(1), 243–264.

LEE, S., AND Y. J. WHANG (2009): "Nonparametric Tests of Conditional Treatment Effects," mimeo, Cowles Foundation.

LEMIEUX, T. (1998): "Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection," *Journal of Labor Economics*, 16(2), 261–291.

LINTON, O., K. SONG, AND Y. WHANG (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154(2), 186–202.

MAIER, M. (2011): "Tests for distributional treatment effects under unconfoundedness," *Economics Letters*, 110(1), 49–51.

MANSKI, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80(2), 319–323.

PERMUTT, T., AND J. HEBEL (1989): "Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight," *Biometrics*, 45(2), 619–622.

PONOMAREVA, M. (2010): "Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates," mimeo, the University of Western Ontario.

ROBINSON, C. (1989): "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models," *Journal of Political Economy*, 97(3), 639–667.

ROMANO, J. P., AND A. M. SHAIKH (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–211.

ROSEN, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146(1), 107–117.

ROSENBAUM, P., AND D. RUBIN (1983): "The central role of the propensity score in observational studies of causal effects," *Biometrika*, 70, 41–55.

ROTHE, C. (2010): "Nonparametric estimation of distributional policy effects," *Journal of Econometrics*, 155, 56–70.