Syracuse University

SURFACE

Center for Policy Research

Maxwell School of Citizenship and Public Affairs

Summer 7-2014

Treatment Effects with Unobserved Heterogeneity: A Set Identification Approach

Sung Jae Jun
The Pennsylvania State University, sjun@psu.edu

Yoonseok Lee
Syracuse University, ylee41@maxwell.syr.edu

Youngki Shin
University of Western Ontario, yshin29@uwo.ca

Follow this and additional works at: https://surface.syr.edu/cpr

Part of the Health Policy Commons, Public Economics Commons, and the Social Welfare Commons

Recommended Citation

CENTER FOR POLICY RESEARCH

THE MAXWELL SCHOOL

WORKING PAPER SERIES

Treatment Effects with Unobserved Heterogeneity: A Set Identification Approach

Sung Jae Jun, Yoonseok Lee and Youngki Shin

CENTER FOR POLICY RESEARCH –Summer 2014

Leonard M. Lopoo, Director
Associate Professor of Public Administration and International Affairs (PAIA)
_____

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

John Yinger
Trustee Professor of Economics and PAIA
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi H. Baltagi.. Economics
Robert Bifulco ....PAIA
Thomas Dennison ....PAIA
Alfonso Flores-Lagunes .... Economics
Sarah Hamersma ....PAIA
William C. Horrace .... Economics
Yilin Hou ....PAIA
Duke Kao.. Economics
Sharon Kioko .... PAIA
Jeffrey Kubik.. Economics
Yoonseok Lee .... Economics
Amy Lutz... Sociology
Yingyi Ma.. Sociology

Jerry Miner ....Economics
Cynthia Morrow ....PAIA
Jan Ondrich...Economics
John Palmer...PAIA
David Popp ....PAIA
Stuart Rosenthal ....Economics
Ross Rubenstein..PAIA
Rebecca Schewe ....Sociology
Amy Ellen Schwartz .... PAIA/Economics
Perry Singleton…………………………......Economics
Abbey Steele...PAIA
Michael Wasylenko... ………………………Economics
Peter Wilcoxen...……PAIA

GRADUATE ASSOCIATES

Emily Cardon .... PAIA
Hannah Dalager .... PAIA
Maidel De La Cruz... PAIA
Carlos Diaz ....Economics
Vantiel Elizabeth Duncan .... PAIA
Alex Falevich ....Economics
Lincoln Groves .... PAIA
Ruby Jennings.. PAIA
Yusun Kim .... PAIA
Bridget Lenkiewicz .... PAIA
Michelle Lofton .... PAIA
Roberto Martinez.. PAIA

Qing Miao ....PAIA
Nuno Abreu Faro E Mota........................... Economics
Judson Murchie ....PAIA
Sun Jung Oh .... Social Science
Brian Ohl...PAIA
Laura Rodriquez-Ortiz ....PAIA
Timothy Smilnak ....PAIA
Kelly Stevens ....PAIA
Rebecca Wang .... Sociology
Pengju Zhang .... Economics
Xirui Zhang .... Economics

STAFF

Kelly Bogart......…………..……Administrative Specialist
Karen Cimilluca....…..……….….Office Coordinator
Kathleen Nasto.......................Administrative Assistant

Candi Patterson.…………..……Computer Consultant
Mary Santy.……….………….....Administrative Assistant
Katrina Wingle......….………....Administrative Assistant

Abstract

We propose the sharp identifiable bounds of the distribution functions of potential outcomes using a panel with fixed T. We allow for the possibility that the statistical randomization of treatment assignments is not achieved until unobserved heterogeneity is properly controlled for. We use certain stationarity assumptions to obtain the bounds. Dynamics in the treatment decisions is allowed as long as the stationarity assumptions are satisfied. In particular, we present an example where our assumptions are satisfied and the treatment decision of the present time may depend on the treatments and the observed outcomes of the past. As an empirical illustration we study the effect of smoking during pregnancy on infant birth weights. We found that for the group of switchers the birth weight with smoking is first order stochastically dominated by that with non-smoking.

Sung Jae Jun-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.

Yoonseok Lee-Department of Economics and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244, ylee41@maxwell.syr.edu

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

1. INTRODUCTION

We study the informational content of repeated treatments for the distributions of potential outcomes. In particular, we propose the sharp identifiable bounds of the distribution functions of potential outcomes using a panel with fixed T. Treatment assignments are not random because of the presence of *unobserved* heterogeneity. Our approach allows for dynamic treatment decisions, where e.g. the treatment decisions at the present time may depend on the treatments and the observed outcomes of the past.

Evaluating policy or treatment effects has been an important topic in diverse disciplines. In cases where randomized experiments are not available, it is often assumed that statistical randomization is possible: i.e. treatment assignments are independent of potential outcomes conditional on covariates, namely *observed* heterogeneity (e.g. Dehejia and Wahba (2002), Firpo (2007), Crump, Hotz, Imbens, and Mitnik (2008), Lee (2009), and Lee and Whang (2009)). However, this traditional assumption of *unconfoundedness* can be violated in many applications and treatment variables are often endogenous due to omitted *unobserved* heterogeneity or self–selection.

As is well summarized in Imbens and Wooldridge (2009) and Heckman and Vytlacil (2007a,b), various econometric methods have been developed to address this issue and examples include approaches based on instrumental variables (e.g. Angrist, Imbens, and Rubin (1996)) and panel data (e.g. Wooldridge (2005), Abrevaya (2006), and Arellano and Bonhomme (2012)).

When treatment effects are heterogeneous, instrumental variable (IV) estimators often lead to the analysis of the subpopulation of compliers, which is never identified from data (e.g. Heckman and Smith (1997), Djebbari and Smith (2008), Fan and Park (2010)). Panel data provide an alternative approach. Below we discuss a few examples of studies on (endogenous) treatment effects with panel data.

A classical example is estimating the wage premium of union membership, in which correlations between the union membership and unobserved ability is controlled by fixed

effects (e.g., Jones (1982), Blakemore, Hunt, and Kiker (1986), Robinson (1989), Lemieux (1998), Budd and Na (2000), and Beck and Fitzenberger (2004)). In health economics, Abrevaya (2006) studies the effect of smoking during pregnancy on the infant birthweight in a fixed–effect regression setup using pseudo panel data. Jung (2010) studies the effect of voluntary information disclosure of health insurance plans on the performance measurements of the insurance plans. Ding and Lehrer (2010) study the impact of class size reduction using multi–period experiments. We note here that these studies are all parametric and unobserved heterogeneity is handled by strong functional form assumptions.

In this paper, we take a panel data approach but we do not make parametric assumptions. Without imposing any parametric structure, we analyze the identification of the (marginal) distributions of the potential outcomes, say Y_{it}^1 and Y_{it}^0, and obtain their sharp identifiable bounds.[1] Also, we articulate the subpopulation for which the distribution functions of the potential outcomes are point identified. Unlike the group of compliers, this subpopulation is identifiable from the data.

Our model is fully nonparametric and nonseparable, where unobserved heterogeneity is allowed. As in the correlated random effect literature (e.g. Mundlak (1978) and Chamberlain (1982)), we make a certain stationarity assumption on the distribution of the heterogeneity given the treatment history. However, our approach is more general than the correlated random effect models in that our stationarity assumption is automatic when the heterogeneity is time invariant, in which case its correlation with the treatment assignment can be arbitrary. Therefore, our approach keeps the idea of "fixed effects." We will present an example of an economic structure that is allowed in our setup, where we discuss dynamics in the treatment decisions.

In addition to the aforementioned stationarity assumption, we also assume selection–on–*unobservables* and time homogeneity. To be more specific about our assumptions, consider the potential outcomes Y_{it}^1 and Y_{it}^0 of individual i at time t. Let β_{it} be the heterogeneity of

---

[1]We do not attempt to identify the distribution function of $Y_{it}^1 - Y_{it}^0$. For discussions on welfare implications of comparisons of the potential outcome distributions, see e.g. Barrett and Donald (2003).

individual i at time t, which is not necessarily observed: it is potentially multi–dimensional and time varying. The relevance of β_{it} is twofold: it matters to achieve the randomization of the treatment assignment D_{it} (selection–on–unobservables) and the distribution of Y_{it}^{j} does not change as long as β_{it} does not change (time homogeneity). Using these assumptions, we utilize penal data to improve the cross–section bounds of Manski (1990b). See Manski and Pepper (2012, 2013) and Khan, Ponomareva, and Tamer (2013) for related approaches.

A similar assumption of time homogeneity was used in Chernozhukov, Fernández-Val, Hahn, and Newey (2013, CFHN hereafter). However, our assumptions are not formulated in a regression setup and we do not treat D_t as a a control variable to obtain time homogeneity. An advantage of our approach is that we can explicitly consider dynamics in the treatment decisions. In Subsection 2.4, we present an example where all of our assumptions are satisfied and dynamic treatment decisions are allowed. However, note please that precisely modeling a dynamic structure of treatment decisions is not the focus of this paper. For merits and costs of modeling a dynamic structure of treatment decisions, see Heckman and Navarro (2007).

The remainder of the paper is organized as follows. In Section 2 we introduce the basic framework and discuss our assumptions in a few subsections, where we also discuss an example of dynamic treatment decisions. In Section 3 we present the sharp identifiable bounds of the potential outcome distributions and discuss inferential issues of them by illustrating some hypotheses of potential interest. Section 4 presents an empirical illustration that studies the effect of smoking on birthweight using the pseudo panel data constructed by Abrevaya (2006).

2. The Framework

2.1. **Potential Outcome Distributions.** We consider the panel data $\{(Y_{it}, D_{it}) : i = 1, \cdots, N, \ t = 1, \cdots, T\}$, where D_{it} is a binary treatment variable, and Y_{it} is an outcome variable of interest. In our identification analysis we assume that T is fixed and we ignore

(strictly) exogenous covariates. The observed outcome Y_{it} is given as

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

where potential outcomes Y_{it}^1 and Y_{it}^0 are never observed together and we observe only one of them depending on the treatment status D_{it}. For example, let D_{it} be an indicator of mother i's smoking status during the pregnancy of the tth 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, and we only observe either Y_{it}^1 or Y_{it}^0, depending on whether the mother actually smoked or not during the tth pregnancy.

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 assignments conditional on observed heterogeneity (e.g., Dehejia and Wahba (2002), Firpo (2007), Crump, Hotz, Imbens, and Mitnik (2008), Lee (2009), and Lee and Whang (2009)). However, as we emphasized in the introduction, one of our main goals is to avoid this standard assumption and allow the possibility that complete randomization may not be achieved until *unobserved* heterogeneity is controlled for.

Our objects of interest are the (marginal) distribution functions of the potential outcomes:

$$F_{it}^j(y) = \mathbb{P}(Y_{it}^j \le y) \ \text{ for } j = 0, 1 \text{ and for } y \in \mathbb{R}.$$

When the assumption of randomized treatment assignments fails to hold, point identification of F_{it}^j is generally not available. However, using the idea of Manski (1990a), their (pointwisely) identifiable bounds can be obtained as

$$\mathbb{P}(Y_{it} \le y, D_{it} = j) \le F_{it}^j(y) \le \mathbb{P}(Y_{it} \le y, D_{it} = j) + \mathbb{P}(D_{it} = 1 - j), \quad (1)$$

which is sharp in the cross–section context. In what follows we show that these Manski–type bounds can be improved when data on repeated treatments are available.

Recall that β_{it} represents individual heterogeneity, which is relevant to achieve the randomization of the treatment assignment and the distributional homogeneity of the potential outcomes. In principle, β_{it} can be vector–valued, where some elements may be excluded from either the D_{it} equation or the Y_{it}^j equation. It will be helpful to have the following equations in mind: for $j = 0, 1$,

$$Y_{it}^j = g^j(\gamma_{it}, \epsilon_{it}^j), \tag{2}$$

$$D_{it} = h(\delta_{it}, \eta_{it}), \tag{3}$$

where β_{it} includes the relevant elements of γ_{it} and δ_{it}. When $\gamma_{it} = \delta_{it} = \alpha_i$ so that β_{it} is simply α_i, it is a panel data model with static treatment decisions and "fixed effects."

Below we discuss our assumptions in detail. Since we focus on an identification analysis, we will suppress the subindex i throughout the paper. For a sequence of generic random variables $\{A_1, A_2, \cdots\}$, we let $\overrightarrow{A}_t = (A_1, A_2, \cdots, A_t)$ denote the history of A up to time t. We will omit the usual qualifier "almost surely" when there is no possibility of confusion.

2.2. **Selection–on–Unobservables.** The first assumption is that the omission of the unobserved heterogeneity is the only source of endogeneity and that complete randomization can be achieved when β_t is controlled for. More precisely, at time t, we assume that each of the potential outcomes is independent of the history of the treatment assignments up to t if the history of heterogeneity up to t is controlled for.

Assumption 1. *For $j = 0, 1$ and for all $t = 1, 2, \cdots, T$, Y_t^j is independent of $\overrightarrow{D}_t$ conditional on $\overrightarrow{\beta}_t$.*

Assumption 1 looks similar to but much more flexible than the standard assumption of unconfoundedness (also known as the selection–on–observables assumption), because the source of confounding is allowed to be unobserved. Also, note that Assumption 1 does not impose restrictions on the joint distribution of (Y_t^1, Y_t^0) but it does on the marginals of them.

Assumption 1 requires that β_t be the only source of confounding. To be more concrete, consider equations (2) and (3). If $\beta_t = \gamma_t = \delta_t = \alpha$, then Assumption 1 is simply saying that ϵ_t^j is independent of $\overrightarrow{\eta}_t$. Note however that Assumption 1 does not rule out dynamics in D_t. For example, if $\delta_t = (D_{t-1}, \alpha)$ and $\gamma_t = \alpha$, then letting $\beta_t = \alpha$ shows that Assumption 1 again requires the independence of ϵ_t^j and $\overrightarrow{\eta}_t$.

Therefore, Assumption 1 is an assumption of no confounding conditional on heterogeneity, which is common in the treatment effect literature. The only difference from the standard randomization assumption is that the relevant heterogeneity need not be observed.

2.3. **Time Homogeneity.** The second set of assumptions is time homogeneity.

Assumption 2. *For $j = 0, 1$, for all b and y and for all t and s, we have*

$$\mathbb{P}(Y_t^j \leq y | \beta_t = b) = \mathbb{P}(Y_s^j \leq y | \beta_s = b).$$

Moreover, when $s \leq t$, Y_t^j is independent of β_s conditional on β_t.

The first part of Assumption 2 is the time homogeneity of Y_t^j given β_t. The second part of Assumption 2 says that dynamics or trends, if any, should be captured by β_t. For instance, if γ_t in (2) includes Y_{t-1}^j, then serial dependence in $\{\epsilon_t^j\}$ is not allowed.

A similar assumption to Assumption 2 is used in CFHN in a nonparametric regression setup and in Khan, Ponomareva, and Tamer (2013) in a more parametric setup. It is worth comparing Assumption 2 with CFHN here.

CFHN assume time homogeneity on $\epsilon_t = (\epsilon_t^1, \epsilon_t^0)$, where ϵ_t^j is the error term in equation (2). For example, when $t = 2$ and $s = 1$, focusing on time invariant heterogeneity α, they require that the distribution of ϵ_2 given D_1, D_2, and α be the same as the distribution of ϵ_1 given D_1 and α. Within our counterfactual outcome framework with potentially time varying heterogeneity and in view of Assumption 1, a naive modification of CFHN might seem to

be the distributional equivalence of Y_2^j given β_1, β_2 and Y_1^j given β_1, but this is not so.[2] Assumption 2 seems to be the only natural assumption of time homogeneity in our setup.

When there is a time trend in Y_t^j, Assumption 2 can still be satisfied as long as β_t includes the time trend. However, we do not recommend this approach, because our next assumption is necessarily violated in that case. Addressing issues of time trends generally requires additional assumptions. For instance, we may assume that Y_t^1 and Y_t^0 have a common trend that is additively separable. Then, the observed outcome Y_t will have the same (additively separable) trend, in which case we can attempt to estimate and subtract the trend to get the data satisfying time homogeneity.

We now make one more assumption of time homogeneity.

Assumption 3. *For all $s \leq t$, β_t and β_s have the same distribution conditional on $\overrightarrow{D}_s$.*

Assumption 3 is similar to the correlated random effect assumption of e.g. Mundlak (1978) and Chamberlain (1982) in that it directly imposes restrictions on the distribution of heterogeneity given the history of the treatment assignments. However, Assumption 3 is more general than the correlated random effect assumption. For instance, if β_t is time invariant, then Assumption 3 is automatically satisfied and there are no restrictions on the dependence of the treatment history and β_t.

However, Assumption 3 is different from assuming that β_t is time invariant. For instance, when $T = 2$, if (D_1, β_1) and (D_2, β_2) has the same distribution, then Assumption 3 is satisfied when the distribution of D_2 given (D_1, β_1) is the same as that of D_1 given (D_2, β_2). This is a time reversibility condition. When $T > 2$, Assumption 3 becomes more restrictive.

Assumptions 1–3 also impose restrictions on the joint dynamics of Y_t^0, Y_t^1 and D_t. To be more concrete, consider equations (2) and (3) again. Suppose that γ_t includes D_{t-1} so that in view of Assumption 1, β_t includes D_{t-1}. Then, Assumption 3 will be violated unless β_t depends on the entire history of the treatment assignments. Therefore, our assumptions do

---

[2]I.e. $\mathbb{P}(Y_2^j \leq \cdot | \beta_1 = a, \beta_2 = b) = \mathbb{P}(Y_1^j \leq \cdot | \beta_1 = a)$ for all a, b is an unnatural assumption unless β_t is time invariant.

not allow for the dynamics of Y_t^j such that Y_t^j depends on D_{t-1}. We rule out the possibility that the treatment decisions of the economic agent may change the natural evolution of the potential outcomes.

However, our assumptions do not rule out the dynamics of D_t such that D_t may depend on D_{t-1} and $Y_{t-1} = D_{t-1}Y_{t-1}^1 + (1 - D_{t-1})Y_{t-1}^0$. In the next subsection we present a specific example where our assumptions are satisfied and dynamic treatment decisions are allowed.

2.4. **A Canonical Example: Dynamic Treatment Decisions.** We now discuss a structural example with sufficient conditions for our assumptions. Consider the following version of (2) and (3): for $j = 0, 1$,

$$Y_t^j = g^j(\mu_t, \alpha, \epsilon_t^j), \tag{4}$$

$$D_t = h(Y_{t-1}, D_{t-1}, \alpha, \eta_t), \tag{5}$$

where Y_t and D_t are observables and α represents a time invariant type. μ_t is individual heterogeneity that is excluded from D_t. ϵ_t^j and η_t are idiosyncratic errors, which are independent of $(\overrightarrow{\mu}_t, \alpha)$. Further, we assume that $\{(\epsilon_t^0, \epsilon_t^1, \eta_t)\}$ is independent and identically distributed (i.i.d.), where $(\epsilon_t^0, \epsilon_t^1)$, and η_t are independent of each other.

To verify our assumptions it is convenient to represent (5) as follows:

$$D_t = H(Y_0, D_0, \alpha, \overrightarrow{\mu}_{t-1}, \overrightarrow{\epsilon}_{t-1}^0, \overrightarrow{\epsilon}_{t-1}^1, \overrightarrow{\eta}_t), \tag{6}$$

where Y_0, D_0 are initial values: we will interpret Assumptions 1–3 conditional on the initial values of Y_0 and D_0. Let $\beta_t = (\mu_t, \alpha)$. Then, by inspection of (4) and (6), Assumptions 1 and 2 are clearly satisfied. For Assumption 3, note that if $\{\mu_t\}$ is serially independent given α, then μ_t is independent of $\overrightarrow{D}_s$ for all $s \leq t$ given α. Therefore, Assumption 3 is satisfied when the distribution of μ_t given α does not vary over time.

Therefore, equations (4) and (5) show an example of a structural model that is allowed under our assumptions. Equation (5) is of our interest. At time t when the economic agent makes a

decision about D_t she has information for Y_{t-1} and D_{t-1} (along with α) but she does not have information for $Y_{t-1}^1 - Y_{t-1}^0$ or $Y_t^1 - Y_t^0$. In this situation the agent may choose D_t by simply considering $\mathbb{E}(Y_{t-1+k}^1 - Y_{t-1+k}^0 | Y_{t-1}, D_{t-1}, \alpha)$ and/or $\mathrm{Corr}(Y_{t-1}^1 - Y_{t-1}^0, Y_{t+k}^1 - Y_{t+k}^0 | Y_{t-1}, D_{t-1}, \alpha)$ with $k = 0, 1, 2, \cdots$. Equation (5) allows for this possibility of dynamic treatment decisions.

In the birthweight example of Section 4, assuming that the potential birthweights depend only on the mother's time invariant type, the following possibility is allowed: when a smoking mother observes a less healthy baby (i.e. a low Y_{t-1}), she suspects that smoking may have had a bad influence on the baby's health this time (i.e. a high $\mathbb{E}(Y_{t-1}^1 - Y_{t-1}^0 | Y_{t-1}, D_{t-1}, \alpha)$), which will be reflected in the smoking decision during the next pregnancy.

3. The Sharp Bounds of F_t^j

3.1. **Bounds Identification.** Based on Assumptions 1–3, we now obtain the sharp identified bounds of the potential outcome distribution functions. Note first that under Assumptions 2 and 3, $F_t^j(y)$ does not depend on t. Therefore, we will simply write $F^j(y)$ without the subindex t.

For each $j = 0, 1$ and $t = 1, 2, \cdots, T$, we first define

$$p_1^j(y) = \mathbb{P}(Y_1 \leq y, D_1 = j),$$

$$p_s^j(y) = \mathbb{P}(Y_s \leq y, D_1 = \cdots = D_{s-1} = 1 - j, D_s = j), \quad \text{for } s = 2, 3, \cdots, T,$$

and

$$L^j(y) = L_T^j(y) = \sum_{s=1}^{T} p_s^j(y), \tag{7}$$

$$U^j(y) = U_T^j(y) = L_T^j(y) + \mathbb{P}(D_1 = \cdots = D_T = 1 - j). \tag{8}$$

The following theorem shows the pointwisely sharp bounds of the potential outcome distributions under Assumptions 1–3.

Theorem 1. *Under Assumptions 1–3, for each $y \in \mathbb{R}$ and $j = 0, 1$, we have*

$$0 \leq L^j(y) \leq F^j(y) \leq U^j(y) \leq 1, \tag{9}$$

where the bounds by $L^j(y)$ and $U^j(y)$ are pointwisely sharp.

Theorem 1 shows that the bounds become tighter as T increases. In fact, the upper and lower bounds coincide when $T \to \infty$. Also, for the subpopulation characterized by $\mathbb{P}(D_1 = \cdots = D_T = 1 - j) = 0$, $F^j(y)$ is point identified for every y.

The bounds in Theorem 1 are only pointwisely sharp. The idea of functional sharp bounds has also been discussed in the literature (e.g. Henry and Mourifié (2012)), but we do not discuss this issue in this paper. In fact, functional sharp bounds do not provide an advantage in testing for first order stochastic dominance, which we will consider in Section 4.

We focus on the identification of the unconditional distributions of the potential outcomes. However, one may be interested in conditional distributions of the potential outcomes, where the conditioning is on some covariates. When the conditioning variables are time–invariant covariates such as gender, race, or simply the entire history over $t = 1, \cdots, T$ of any (time-varying) covariates, Theorem 1 can be readily extended to the conditional distribution functions as long as Assumptions 1 through 3 hold conditional on those covariates. However, even with a single conditioning variable, when the conditioning variable is time-varying and T is large, conditioning on the entire history of the covariates may not be practically attractive. In this case, conditioning only on the contemporaneous covariates can be of practical interest. If we assume time–invariant (unobserved) heterogeneity and have no dynamics, then we can derive a similar result to Theorem 1 under the exchangeability assumption of the covariate distribution given unobserved heterogeneity. However, it is too special a situation in a panel data setup and we do not formally present this result here.

3.2. **Inferences.** When F^j is not point identified, it is not straightforward to consider any direct statistical inferences on F^j. However, from Theorem 1, we have the identified bounds

of F^j, where the bounds $L^j(\cdot)$ and $U^j(\cdot)$ are easy to estimate. For instance, provided that we have independent and identically distributed observations across i, both $L^j(\cdot)$ and $U^j(\cdot)$ can be estimated by empirical–distribution–like estimators and a well–developed distribution theory is readily available for such estimators (e.g. Van der Vaart (2000)). Inferences for a partially identified nonparametric object is generally challenging. However, once we formulate hypotheses of interest for which test statistics only depend on the bound estimators, we can easily conduct statistical inferences. In this section, without detailing implementation–related issues, we briefly discuss some hypotheses of potential interest that can be formulated based on the identified bounds in (9).

One possibility is directly comparing the potential outcome distributions over the entire support, such as stochastic dominance relations. More precisely, we reparametrize (9) in terms of the parameter of interest $\Delta(y) = F^1(y) - F^0(y)$ as

$$
\begin{pmatrix} L^0(y) \leq F^1(y) - \Delta(y) \leq U^0(y) \\ L^1(y) \leq F^1(y) \leq U^1(y) \end{pmatrix} \iff \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ -1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F^1(y) \\ \Delta(y) \end{pmatrix} \leq \begin{pmatrix} -L^0(y) \\ U^0(y) \\ -L^1(y) \\ U^1(y) \end{pmatrix},
\tag{10}
$$

which is a set of (nonparametric) linear moment inequalities. Let $\theta = (F^1, \Delta)'$ and let Θ_T be the identified set such that all elements in Θ_T satisfy (10). Then, we can construct a confidence region for θ, at least pointwisely for each y, following the recent development in the partial identification literature: see e.g., Andrews, Berry, and Jia (2004), Chernozhukov, Hong, and Tamer (2007), Beresteanu and Molinari (2008), Rosen (2008), Romano and Shaikh (2010), Kim (2009), Andrews and Soares (2010), and Chernozhukov, Lee, and Rosen (2013).

The only complications here are that θ is a nonparametric object and that we are interested only in the subvector of θ. However, these complications can be resolved once we clarify the hypothesis of interest. To be more specific, consider testing first order stochastic dominance

relations: e.g. $\Delta(y) \geq 0$ for all y.[3] Since Δ is only partially identified, directly testing "$\Delta(y) \geq 0$ for all y" is not possible but there are two possibilities:

$$\begin{cases} H_{0,a}^* : \forall \theta \in \Theta_T, \ \Delta(y) \geq 0 \quad \text{for all } y \\ H_{1,a}^* : \exists \theta \in \Theta_T, \ \Delta(y) < 0 \quad \text{for some } y, \end{cases} \tag{11}$$

or

$$\begin{cases} H_{0,b}^* : \ \exists \theta \in \Theta_T, \ \Delta(y) \geq 0 \quad \text{for all } y \\ H_{1,b}^* : \ \forall \theta \in \Theta_T, \ \Delta(y) < 0 \quad \text{for some } y. \end{cases} \tag{12}$$

It turns out that tests for (11) or (12) can be done easily because the parametric analysis of Hahn and Ridder (2009) can be extended to show that (11) and (12) are *equivalent* to

$$\begin{cases} H_{0,a} : U^0(y) \leq L^1(y) \quad \text{for all } y \\ H_{1,a} : U^0(y) > L^1(y) \quad \text{for some } y, \end{cases} \qquad \begin{cases} H_{0,b} : L^0(y) \leq U^1(y) \quad \text{for all } y \\ H_{1,b} : L^0(y) > U^1(y) \quad \text{for some } y, \end{cases} \tag{13}$$

respectively. The proof of the equivalence is provided in the appendix. Testing for the hypotheses in (13) is a standard task (e.g., Linton, Song, and Whang (2010)). Testing higher order stochastic dominance can be similarly done.

The bounds in (9) can be inverted to obtain bounds on the quantiles of the potential outcomes:

$$Q_{U^j}(\tau) \leq Q_{F^j}(\tau) \leq Q_{L^j}(\tau) \text{ for } \tau \in (0,1), \tag{14}$$

where $Q_F(\tau) = \inf\{y : F(y) \geq \tau\}$. Therefore, comparing particular quantiles of the potential outcomes is not any more difficult than testing stochastic dominance as above.

Note, however, that the bounds in (14) do not always provide informative bounds on the expectations of the potential outcomes, because the bounds in (14) are not generally integrable. This problem can be resolved when the support of Y_t^j is known to be bounded.

---

[3]In testing first order stochastic dominance, choosing a null between "$\Delta(y) \geq 0$ for all y" and "$\Delta(y) \leq 0$ for all y" can be an issue. In Section 4 we consider both possibilities. For an alternative approach to this issue, see Gupta and Panchapakesan (1979). We thank an anonymous referee for this reference.

Suppose that M^j and m^j are the upper and the lower bounds of the support of Y_t^j. It then follows that

$$\max\{m^j, Q_{U^j}(\tau)\} \leq Q_{F^j}(\tau) \leq \min\{M^j, Q_{L^j}(\tau)\} \ \text{ for } \tau \in (0,1), \tag{15}$$

after which integrating over $\tau \in (0,1)$ yields bounds for $\mathbb{E}(Y_t^j)$. Therefore, inference on the average treatment effects can also be similarly done.

4. An Empirical Illustration: Birth Weight and Stochastic Dominance

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)). Let $F_{it}^j(\cdot)$ denote the potential birthweight distribution of mother i's t^{th} baby and let D_{it} be the indicator of mother i's smoking status during the tth pregnancy with $D_{it} = 1$ indicating smoking. We will obtain the bound estimates of potential birthweight distributions for the populations of *ever–smokers* and *switchers*. We will then use them to formally test for the presence of first order stochastic dominance.

Our analysis is based on the (pseudo panel) 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 (2012) in the random coefficients panel model. We start with the $n = 2,137$ sample of those who had *three births* ($T = 3$) and had ever smoked during pregnancy (*ever–smokers*), i.e. $D_{it} = 1$ at least for one period t. Among these ever–smokers, 692 mothers smoked during all of the three pregnancies (*always–smokers*), which leaves 1,445 *switchers*. The focus of our analysis is on the ever–smokers and the switchers. We did not include those who never smoked (*never–smokers*) in our analysis because the sample size of never–smokers is too large (i.e., 82.7% of the entire sample of three births) 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.

A mother's smoking choice should generally be correlated with her health–related life–style factors, which are confounding variables to explain the causal effects of smoking on a baby's birthweight. Therefore, the complete randomization of treatment assignments will not be achieved until those confounding factors are properly controlled for. However, many of such life–style factors are hard or nearly impossible to observe and they are frequently time varying. Those unobserved life–style factors correspond to β_{it} in Assumption 1. Note that, as discussed in Section 2.2, Assumption 1 allows for the possibility that a smoking status during the tth pregnancy may depend on her smoking status and birth outcomes in the past. Therefore, a general form of dynamics is allowed, but we do not attempt to specifically model the smoking decision in our framework.

Figures 1 and 2 show the bound estimates of the distribution functions of potential birthweights for ever–smokers and switchers, respectively.[4] The case of $T = 3$ shows the bounds that are estimated by all the three time periods and the case of $T = 1$ is the cross–sectional Manski type bounds. The case of $T = 2$ shows the bounds based on the two period panel. Note that all the estimated bounds become tighter as T increases. In particular, with $T = 3$, $F^1(\cdot)$ is point–identified in both Figures 1 and 2, which is explained by the fact that never-smokers are excluded from our analysis. In Figure 2, both F^1 and F^0 are point–identified with $T = 3$.

Consider Figure 1 first. $\widehat{F}^1(\cdot)$ is located above the lower bound estimate of $F^0(\cdot)$ over all birthweights, which suggests that there may exist F^0 that first order stochastically dominates F^1. However, $\widehat{F}^1(\cdot)$ is clearly above the upper bound estimate of $F^0(\cdot)$ around the birthweight near $3,500$ grams, suggesting that there cannot exist $F^0(\cdot)$ that is first-order stochastically dominated by $F^1(\cdot)$. In comparison, Figure 2 depicts the bound estimates for switchers by excluding always–smokers from the sample. In Figure 2, with $T = 3$, both $F^1(\cdot)$ and $F^0(\cdot)$ are

---

[4]Dotted and dashed lines show (pointwise) 95% confidence sets for $F^j(y)$ for each y, using Imbens and Manski (2004) and Stoye (2009).

point identified, from which we can easily tell that $F^0(\cdot)$ first–order stochastically dominates $F^1(\cdot)$.

Table 1 shows the results of formal tests of stochastic dominance, which consider null and alternative hypotheses given as (11) and (12) (or as (13) in their dual forms). The p–values are obtained by using the bootstrap method of Linton, Song, and Whang (2010). Specifically, in the case of (11), we use a Cramér–von Mises type statistic based on

$$\int \max\{U^0(y) - L^1(y), 0\}^2 w(y) dy,$$

where we use the empirical density of Y_{it} on equi–spaced 100 grid points for the weight function $w(\cdot)$. We choose Linton, Song, and Whang (2010)'s weight function $q(\cdot) = 1$ and the cutoff value $c_n = 2n^{-1/2} \log \log n$, following what their simulation studies guide. Each column in Table 1 summarizes the bootstrap p–values for each hypothesis, which are obtained from $1,000$ replications. With $T = 3$ and for the ever–smoker group, we conclude that the true F^0 may first–order stochastically dominate F^1. For the switcher group there is clear evidence of first order stochastic dominance. These results re-enforce the existing empirical findings such as Abrevaya and Dahl (2008) and Arellano and Bonhomme (2012). Also notice that, different from Abadie (2002), the stochastic dominance result for the switcher group is obtained without using instrumental variables.

Bound estimates of the potential outcome distributions in Figures 1 and 2 also imply bound estimates of the quantiles and the means of the potential outcomes as (14) and (15) in Section 3.2. Table 2 summarizes the median and mean bound estimates of each potential outcome distributions, from which we can also obtain bounds of the median treatment effects or the average treatment effects.

Finally, we comment that we detrend all birthweight observations by adjusting their means for each t for this analysis, because there is a folk belief that the first born tend to be lighter than those born in the later order on average. Using the original data without detrending did not change a meaningful difference though.

Null Hypothesis	Type-a (all F^0 and F^1 satisfies)		Type-b (there exist F^0 and F^1 s.t.)	
	F^0 FSD F^1	F^1 FSD F^0	F^0 FSD F^1	F^1 FSD F^0
Ever–Smokers $(n = 2137)$				
$T = 1$	0.000	0.000	0.905	0.899
$T = 2$	0.000	0.000	0.937	0.919
$T = 3$	0.000	0.000	0.943	0.000
Switchers $(n = 1445)$				
$T = 1$	0.000	0.000	0.888	0.896
$T = 2$	0.000	0.000	0.914	0.913
$T = 3$	0.660	0.000	0.660	0.000

TABLE 1. *p*–values of the First-order Stochastic Dominance tests (3–births)

|	F^0		F^1	
	Median	*Mean*	*Median*	*Mean*
Ever–Smokers $(n = 2137)$				
$T = 1$	[440.0, 4933.0]	[1236.6, 3974.1]	[2737.5, 3520.3]	[1961.3, 3462.5]
$T = 2$	[2737.5, 3833.4]	[1811.0, 3768.4]	[3102.8, 3311.6]	[2508.2, 3078.5]
$T = 3$	[3050.6, 3676.9]	[2073.9, 3611.9]	3207.2	2855.2
Switchers $(n = 1445)$				
$T = 1$	[2372.2, 3990.0]	[1779.6, 3681.2]	[680.0,4763.0]	[1779.6, 3681.2]
$T = 2$	[3311.6, 3468.1]	[2677.6, 3251.8]	[3050.6, 3363.7]	[2294.3, 3164.1]
$T = 3$	3363.7	3016.9	3207.2	2876.0

TABLE 2. Median and mean bounds (3–births)

We have also conducted the same analysis for a few more sub-populations with three births, but the overall results do not change, either. For instance, the subgroup of those who ever drank alcohol during at least one of the pregnancies (ever–drinkers) was considered. For this group the bounds of F^1 were tighter (with $T = 1, 2$) than the bounds from all the ever smoker observations, i.e. $\mathbb{P}(\text{never smoke}|\text{ever drink}) < \mathbb{P}(\text{never smoke})$ as we would expect.
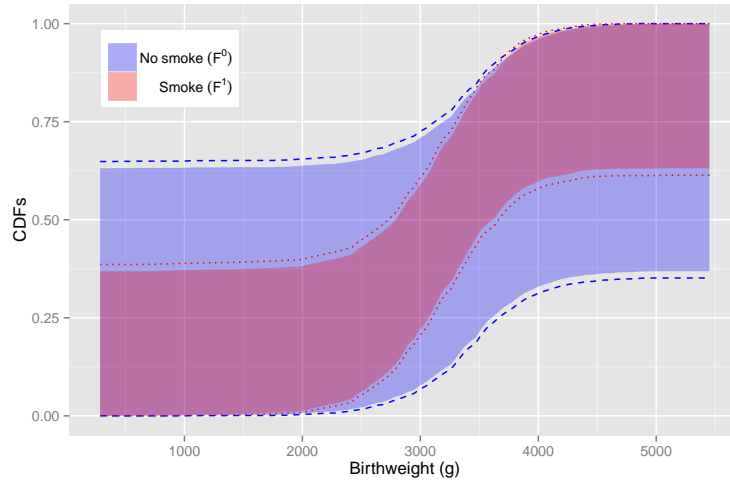
(a) $T = 1$
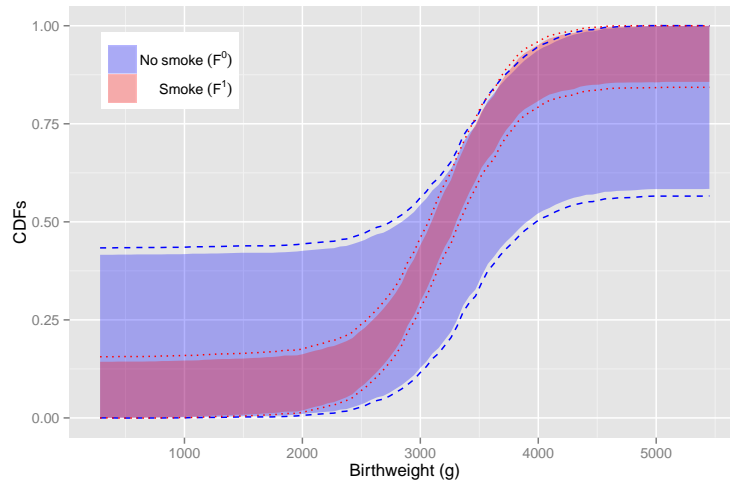
(b) $T = 2$
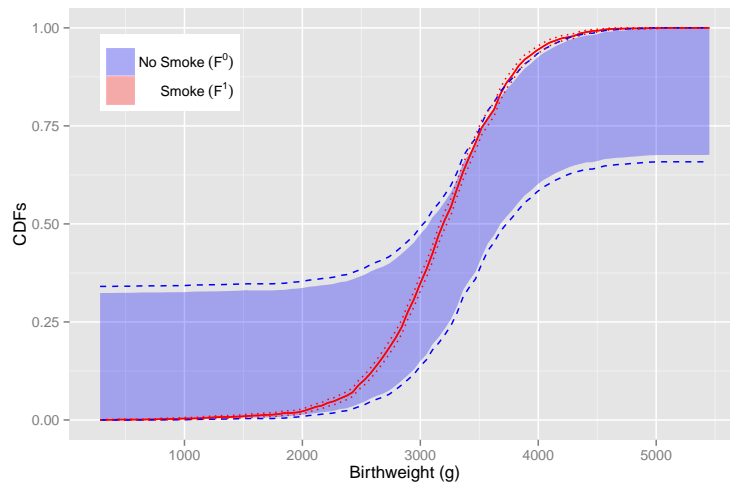
(c) $T = 3$

FIGURE 1. Bounds of the Distributions of Potential Birthweights (Ever–Smokers)
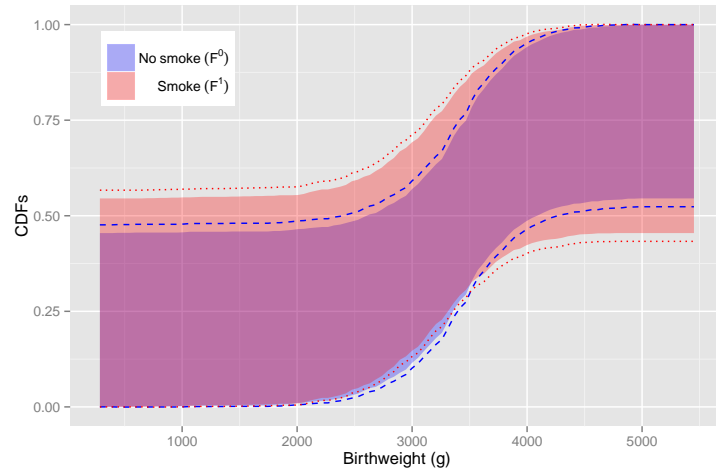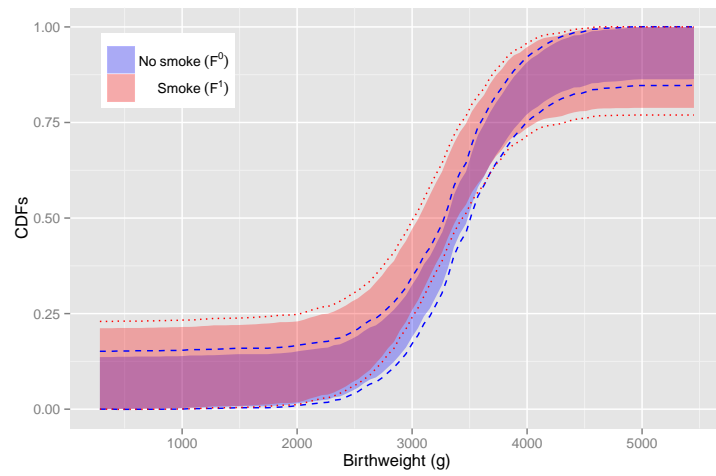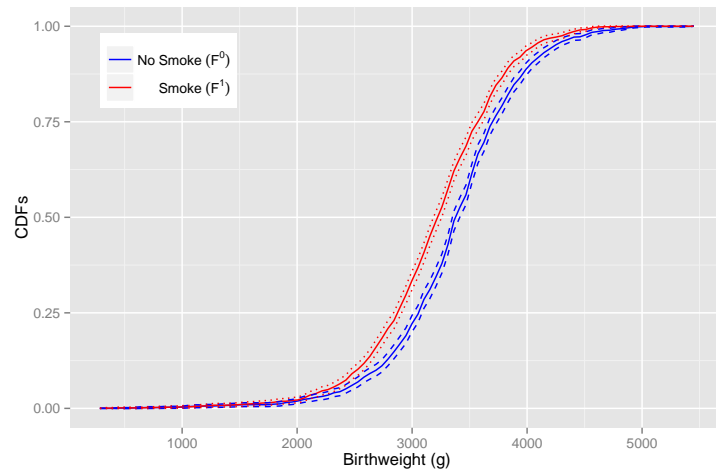
(a) $T = 1$

(b) $T = 2$

(c) $T = 3$

FIGURE 2. Bounds of the Distributions of Potential Birthweights (Switchers)

5. Appendix

Throughout the appendix, $\mu_A(\cdot)$ and $\mu_{A|B}(\cdot|b)$ represent the probability distribution of A and the conditional distribution of A given $B = b$, respectively, where A and B are generic random variables.

5.1. The proof of Theorem 1.

We only consider $j = 1$ because $j = 0$ is symmetric. For a given t, note that

$$\mathbb{P}(Y_t^1 \le y) = \mathbb{P}(Y_t^1 \le y, D_1 = 1) + \mathbb{P}(Y_t^1 \le y, D_1 = 0, D_2 = 1)$$

$$+ \cdots + \mathbb{P}(Y_t^1 \le y, D_1 = D_2 = \cdots = D_{t-1} = 0, D_t = 1)$$

$$+ \mathbb{P}(Y_t^1 \le y, D_1 = D_2 = \cdots = D_{t-1} = D_t = 0). \quad (16)$$

Now, for any $s \le t$, we have

$$\mathbb{P}(Y_t^1 \le y, \overrightarrow{D}_s = \overrightarrow{d}_s) = \int \mathbb{P}(Y_t^1 \le y, \overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\beta}_t = \overrightarrow{b}_t) d\mu_{\overrightarrow{\beta}_t}(\overrightarrow{b}_t)$$

$$\overset{A1}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\beta}_t = \overrightarrow{b}_t) \mathbb{P}(Y_t^1 \le y | \overrightarrow{\beta}_t = \overrightarrow{b}_t) d\mu_{\overrightarrow{\beta}_t}(\overrightarrow{b}_t)$$

$$\overset{A2}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\beta}_t = \overrightarrow{b}_t) \mathbb{P}(Y_t^1 \le y | \beta_t = b_t) d\mu_{\overrightarrow{\beta}_t}(\overrightarrow{b}_t)$$

$$= \int \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\beta}_{t-1} = \overrightarrow{b}_{t-1}, \beta_t = b_t) \mathbb{P}(Y_t^1 \le y | \beta_t = b_t) d\mu_{\overrightarrow{\beta}_{t-1}|\beta_t}(\overrightarrow{b}_{t-1}|b_t) d\mu_{\beta_t}(b_t)$$

$$= \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \beta_t = b_t) \mathbb{P}(Y_t^1 \le y | \beta_t = b_t) d\mu_{\beta_t}(b_t)$$

$$\overset{A2,A3}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \beta_s = b_t) \mathbb{P}(Y_s^1 \le y | \beta_s = b_t) d\mu_{\beta_s}(b_t) \quad (17)$$

$$= \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\beta}_{s-1} = \overrightarrow{b}_{s-1}, \beta_s = b_t) \mathbb{P}(Y_s^1 \le y | \beta_s = b_t) d\mu_{\overrightarrow{\beta}_{s-1},\beta_s}(\overrightarrow{b}_{s-1}, b_t)$$

$$\overset{A2}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\beta}_{s-1} = \overrightarrow{b}_{s-1}, \beta_s = b_t) \mathbb{P}(Y_s^1 \le y | \overrightarrow{\beta}_{s-1} = \overrightarrow{b}_{s-1}, \beta_s = b_t) d\mu_{\overrightarrow{\beta}_{s-1},\beta_s}(\overrightarrow{b}_{s-1}, b_t)$$

$$\overset{A1}{=} \int \mathbb{P}(Y_s^1 \le y, \overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\beta}_{s-1} = \overrightarrow{b}_{s-1}, \beta_s = b_t) d\mu_{\overrightarrow{\beta}_{s-1},\beta_s}(\overrightarrow{b}_{s-1}, b_t)$$

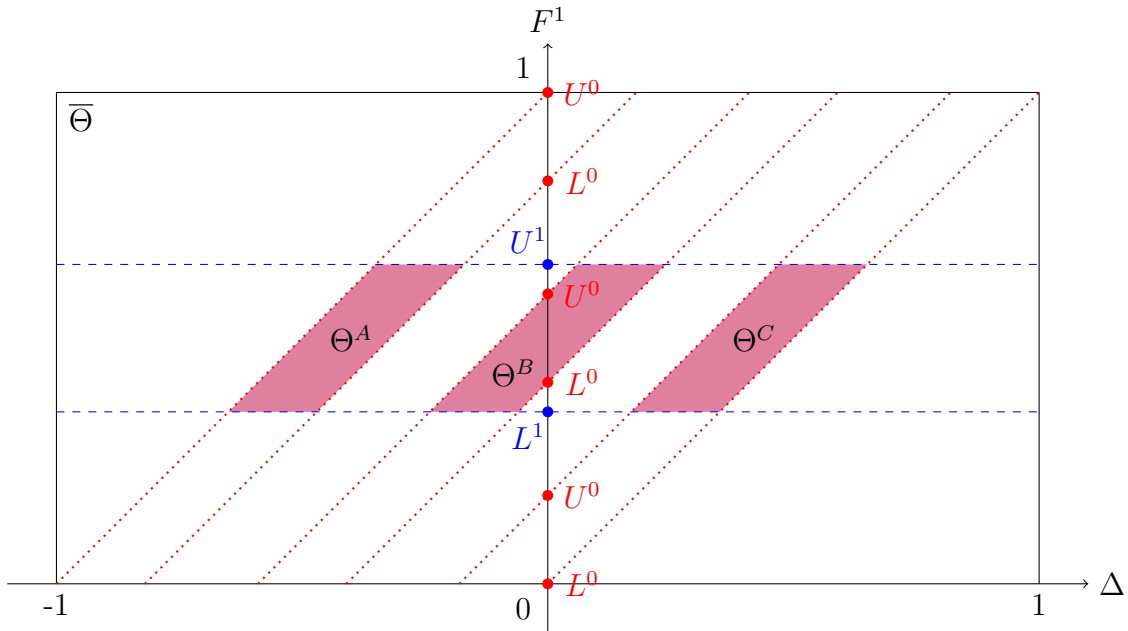$$= \mathbb{P}(Y_s^1 \le y, \overrightarrow{D}_s = \overrightarrow{d}_s).$$

FIGURE 3. Graphical Illustration of the Duality

Therefore, equation (16) is thus equal to

$$\mathbb{P}(Y_t^1 \leq y) = \mathbb{P}(Y_1 \leq y, D_1 = 1) + \mathbb{P}(Y_2 \leq y, D_1 = 0, D_2 = 1)$$

$$+ \cdots\cdots + \mathbb{P}(Y_t \leq y, D_1 = \cdots = D_{t-1} = 0, D_t = 1) + \mathbb{P}(Y_t^1 \leq y, D_1 = \cdots = D_{t-1} = 0, D_t = 0).$$

Here, by the expression in (17), the last term is equal to

$$\int \mathbb{P}(D_1 = \cdots = D_{t-1} = D_t = 0 | \beta_t = b_t) \mathbb{P}(Y_t^1 \leq y | \beta_t = b_t) d\mu_{\beta_t}(b_t), \tag{18}$$

where our assumptions do not impose any restrictions on the support of $\mathbb{P}(Y_t^1 \leq y | \beta_t = b_t)$. Therefore, the sharp bounds of (18) will be the interval between 0 and $\mathbb{P}(D_1 = \cdots = D_t = 0)$, which yields the sharp bound of $\mathbb{P}(Y_t^1 \leq y)$ as desired. $\qquad \square$

5.2. **Duality.**

Proposition 1. *The hypotheses in (11) and (12) are equivalent to the hypotheses in (13), respectively.*

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

We now formally prove the duality result. In each case it suffices to show the equivalence of the null hypotheses. For simplicity, we assume that both Y_t^0 and Y_t^1 have unbounded support $\mathbb{R}$. We first show the equivalence of $H_{0,a}^*$ and the first null hypothesis, say $H_{0,a}$, of (13).

● **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 $\tilde{y} \in \mathbb{R}$ such that $U^0(\tilde{y}) - L^1(\tilde{y}) > 0$. We need to show that there exist F^0 and F^1 contradicting to $H_{0,a}^*$. For any $\epsilon > 0$, define $F^0(\cdot)$ and $F^1(\cdot)$ as follows: (a) $F^0(y) = L^0(y)$ for $y < \tilde{y}$; (b) $F^1(y) = U^1(y)$ for $y \geq \tilde{y}+\epsilon$; (c) $F^0(y) = U^0(y)$ and $F^1(y) = L^1(y)$ otherwise. By construction, they are distribution functions satisfying the inequalities of (9), but $F^0(\tilde{y}) > F^1(\tilde{y})$ that contradicts to $H_{0,a}^*$.

(ii) Necessity: Let $H_{0,a}$ be true. Then, for all $y \in \mathbb{R}$, $U^0(y) - L^1(y) \leq 0$. For any F^0 and F^1 satisfying the inequalities in (9), this implies $F^0(y) \leq U^0(y) \leq L^1(y) \leq F^1(y)$ for all $y \in \mathbb{R}$ and thus $H_{0,a}^*$ holds.

We next show the equivalence of $H_{0,b}^*$ and the second null hypothesis, say $H_{0,b}$ in (13).

● **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 $y \in \mathbb{R}$, they satisfy the inequalities in Theorem 1 and $F^0(y) \leq F^1(y)$. Fix such F^0 and F^1 and suppose that $H_{0,b}$ is not true. Then, there exists $\tilde{y} \in \mathbb{R}$ such that $L^0(\tilde{y}) - U^1(\tilde{y}) > 0$. Since F^0 and F^1 satisfy the inequalities in (9), this implies $F^1(\tilde{y}) \leq U^1(\tilde{y}) < L^0(\tilde{y}) \leq F^0(\tilde{y})$. Therefore, $F^1(\tilde{y}) < F^0(\tilde{y})$, 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 (9) and $F^0(y) \leq F^1(y)$ for all $y \in \mathbb{R}$. For some constants $c_1 < c_2$, define F^0 and F^1 as follows: (a) $F^1(y) = \max\{L^0(y), L^1(y)\}$ for $y < c_1$; (b) $F^0(y) = \min\{U^0(y), U^1(y)\}$ for $y \geq c_2$; (c) $F^0(y) = L^0(y)$ and $F^1(y) = U^1(y)$ otherwise. Then, $F^0(y)$ and $F^1(y)$ satisfy the inequalities in (9) and $F^0(y) \leq F^1(y)$ for all $y \in \mathbb{R}$ by construction. Note also that F^0 and F^1 are distribution functions since they are CADLAG and go to 1 and 0 as $y \to +\infty$ and $y \to -\infty$, respectively. $\qquad\square$

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 (2012): "Identifying distributional characteristics in random coefficients panel data models," *The Review of Economic Studies*, 79(3), 987–1020.

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.

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.

CHAMBERLAIN, G. (1982): "Multivariate regression models for panel data," *Journal of Econometrics*, 18(1), 5–46.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, J. HAHN, AND W. NEWEY (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81(2), 535–580.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81(2), 667–737.

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.

DING, W., AND S. F. LEHRER (2010): "Estimating treatment effects from contaminated multi-period education experiments: The dynamic impacts of class size reductions," *Review of Economics and Statistics*, 92(1), 31–42.

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

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.

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

GUPTA, S. S., AND S. PANCHAPAKESAN (1979): *Multiple decision procedures: theory and methodology of selecting and ranking populations*, vol. 44. Siam.

HAHN, J., AND G. RIDDER (2009): "Partial identification and confidence intervals," mimeo, UCLA.

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 S. NAVARRO (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136(2), 341–396.

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.

HENRY, M., AND I. MOURIFIÉ (2012): "Sharp Bounds in the Binary Roy Model," mimeo, Pennsylvania State University.

IMBENS, G. W., AND C. F. MANSKI (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72(6), 1845–1857.

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 (2013): "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.

MANSKI, C. (1990a): "Nonparametric bounds on treatment effects," *American Economic Review*, 80(2), 319–323.

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

MANSKI, C. F., AND J. V. PEPPER (2012): "Partial identification of treatment response with data on repeated cross sections," mimeo, Northwestern University.

——— (2013): "Deterrence and the death penalty: partial identification analysis using repeated cross sections," *Journal of Quantitative Criminology*, 29(1), 123–141.

MUNDLAK, Y. (1978): "On the pooling of time series and cross section data," *Econometrica: journal of the Econometric Society*, pp. 69–85.

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.

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.

STOYE, J. (2009): "More on confidence intervals for partially identified parameters," *Econometrica*, 77(4), 1299–1315.

VAN DER VAART, A. W. (2000): *Asymptotic statistics*, vol. 3. Cambridge university press.

WOOLDRIDGE, J. M. (2005): "Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models," *Review of Economics and Statistics*, 87(2), 385–390.