Treatment Effects with Unobserved Heterogeneity: A Set Identification Approach

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

March 2015

Abstract

We propose the sharp identifiable bounds of the potential outcome distributions using panel data. We allow for the possibility that statistical randomization of treatment assignments is not achieved until *unobserved* heterogeneity is properly controlled for. We use certain stationarity assumptions to obtain the sharp bounds. Our approach allows for dynamic treatment decisions, where the current treatment decisions may depend on the past treatments or the past observed outcomes. As an empirical illustration, we study the effect of smoking during pregnancy on infant birthweight. We find that for the group of switchers the infant birthweight of a smoking mother is first order stochastically dominated by that of a non–smoking mother.

JEL Classification: C12, C21, C23.

Keywords: treatment effects, dynamic treatment decisions, 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 and Center for Policy Research, Syracuse University, 426 Eggers Hall, Syracuse, NY 13244, ylee41@maxwell.syr.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, Yanqin Fan, Jin Hahn, Stefan Hoderlein, Han Hong, Kyoo il Kim, Simon Lee, Maria Ponomareva, Joris Pinkse, Christoph Rothe, Jeff Smith, and Iván Fernández–Val. Also we would like to thank seminar participants at various universities and conferences. Jun gratefully acknowledges support from NSF grant SES0922127. Jun also 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

The main purpose of this paper is to identify the potential outcome distributions in a nonparametric setup when the standard assumption of selection–on–observables does not hold. In particular, we consider a situation where treatment assignments are not necessarily exogenous because of the presence of *unobserved* heterogeneity. We study the informational contents of repeated treatments (i.e. panel data) to propose the sharp identifiable bounds of the distribution functions of potential outcomes. Our approach allows for dynamic treatment decisions, where the current treatment decisions may depend on the past treatments or the past observed outcomes.

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 selection–on–observables can be violated in many applications because of the presence of *unobserved* heterogeneity. 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. Examples include approaches based on instrumental variables (e.g. Angrist, Imbens, and Rubin (1996)) or on panel data (e.g. Wooldridge (2005), Abrevaya (2006), and Arellano and Bonhomme (2012)). In this paper we focus on the panel data approach.

There are many empirical studies using panel data models in this context. For example, Jones (1982), Blakemore, Hunt, and Kiker (1986), Robinson (1989), Lemieux (1998), Budd and Na (2000), and Beck and Fitzenberger (2004)) study the effect of union membership on the wage. Abrevaya (2006) studies smoking during pregnancy and infants' birthweight. Jung (2010) looks into the impacts of voluntary information disclosures of health insurance plans.

Ding and Lehrer (2010) study the impact of the reduction of the class size. These studies rely on parametric assumptions.

We do not use parametric assumptions at all. Therefore, unobserved heterogeneity is allowed to affect both potential outcomes and treatment decisions in a nonseparable way. A certain type of time homogeneity is the only restriction we impose on unobserved heterogeneity, which is always satisfied when it is time invariant. In this general setup, we study the identification of the potential outcome distributions and obtain their sharp identifiable bounds. Our results also 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 (cf. Heckman and Smith (1997), Djebbari and Smith (2008), Fan and Park (2010)).

Our key assumptions for identification can be described by selection–on–unobservables and time homogeneity. Unlike the traditional assumption of selection–on–observables, we allow for the possibility that the independence of the treatment history (up to the present time) and the current potential outcomes is not achieved until we condition on unobserved heterogeneity as well as exogenous covariates. When unobserved heterogeneity is time invariant (e.g. fixed effects), our assumption of time homogeneity simply requires that the marginal distributions of the potential outcomes do not vary over time. These two assumptions are the key vehicles to improve the cross–section bounds of Manski (1990) by using panel data. Please see also Manski and Pepper (2012, 2013) and Khan, Ponomareva, and Tamer (2013) for related approaches.

A similar assumption of time homogeneity is used in Chernozhukov, Fernández-Val, Hahn, and Newey (2013). However, our assumptions are not formulated in a regression setup and we do not treat the treatment variable as a control variable to obtain time homogeneity. Advantages of our approach are twofold. First, we only use the time homogeneity of the marginals, not the joint, of the potential outcomes. Second, we can be more explicit about the dynamics in treatment decisions. In Section 2.2, we discuss some structural examples that satisfy our assumptions, where the current treatment decision depends on the past treatment decisions or the past observed outcomes.

Once we derive the main results using time–invariant unobserved heterogeneity, we briefly consider an alternative setup, where unobserved heterogeneity is potentially time varying but its distribution does not change conditional on the treatment history. The sharp identifiable bounds turn out to be the same as the ones derived under time–invariant heterogeneity.

The remainder of the paper is organized as follows. In Section 2, we introduce the basic framework and discuss our assumptions, where we focus on time–invariant heterogeneity. We also discuss structural examples of dynamic treatment decisions. In Section 3, we present the sharp identifiable bounds of the potential outcome distributions and discuss inferential issues by illustrating some hypotheses of potential interest. Section 4 briefly considers the possibility of time–varying heterogeneity. In Section 5, we study the effects of a mother's smoking on infant birthweight using the pseudo panel data constructed by Abrevaya (2006).

2. The Framework

2.1. **The Setup and Basic Assumptions.** We consider the panel data $\{(Y_{it}, D_{it}, X_{it}) : i = 1, 2, \cdots, N, t = 1, 2, \cdots, T\}$, where Y_{it} is an outcome, D_{it} is a binary treatment, and X_{it} is a vector of strictly exogenous covariates. D_{it} is potentially endogenous in that the standard assumption of selection–on–observables may not be satisfied. We assume that T is small and fixed. We also assume that all random variables are identically distributed across i. Since we do not rely on the selection–on–observables assumption, the covariates X_{it} do not play a major role in our analysis. Therefore, throughout the paper, we assume that all of our analysis is conditional on $\overrightarrow{X}_{iT} = (X_{i1}, X_{i2}, \cdots, X_{iT})$ but we suppress "conditioning on $\overrightarrow{X}_{iT}$" for a succinct presentation.[1]

[1]For a sequence of generic random variables $A_{i1}, A_{i2}, \cdots$, we denote the history up to t by $\overrightarrow{A}_{it}$, i.e. $\overrightarrow{A}_{it} = (A_{i1}, A_{i2}, \cdots, A_{it})$.

Following the potential outcome framework, we suppose that the observed outcome Y_{it} is given by

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

and hence we observe only one of the potential outcomes Y_{it}^1 and Y_{it}^0, depending on the treatment status D_{it}. For example, let D_{it} denote the smoking status of mother i during her t^{th} pregnancy and let (Y_{it}^0, Y_{it}^1) be the potential birthweight of the baby depending on the mother's smoking status. When the mother has actually smoked (i.e. $D_{it} = 1$), we observe $Y_{it} = Y_{it}^1$ but we do not observe the counterfactual birthweight Y_{it}^0.

The main objective of this paper is nonparametrically identifying the marginal distribution functions of the potential outcomes, i.e. for $j = 0, 1$ and for all $y \in \mathbb{R}$,

$$F_t^j(y) = \mathbb{P}(Y_{it}^j \leq y).^{2} \tag{1}$$

When the treatment assignment is random in that D_{it} is independent of (Y_{it}^0, Y_{it}^1), it is well known that $F_t^j(y)$ is point identified. However, the random treatment assignment is often too strong an assumption in economics. For example, both a mother's smoking decision during pregnancy and the baby's potential birthweight may depend on the mother's characteristics such as the level of her health consciousness and her initial smoking status before the first pregnancy. In this case, controlling for such confounding factors would be a natural idea but the difficulty lies in the fact that they are not observed by the econometrician in most cases.

The following assumption formalizes the idea of the presence of unobserved confounding factors. We assume that such confounding factors are time–invariant. In Section 4, we will discuss an alternative setup, where confounding factors are potentially time varying.

Assumption 1 (Selection on Unobservables)**.** *There exists a vector α_i of time–invariant unobserved confounding factors such that for $j = 1, 0$ and for $t = 1, 2, \cdots, T$, Y_{it}^j is independent of $\overrightarrow{D}_{it}$ conditional on α_i.*

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

Assumption 1 does not concern (Y_{it}^0, Y_{it}^1) jointly but it only concerns each of Y_{it}^0 and Y_{it}^1 individually. The confounding factor α_i can be thought of as individual fixed effects in panel data models: no restrictions are imposed on the distribution of α_i and the dependence between α_i and $\overrightarrow{D}_{it}$ can be arbitrary. In our application, as we mentioned earlier, examples of α_i include a health consciousness level and an initial smoking status before the first pregnancy.

Under Assumption 1, $F_t^j(y)$ is generally not point identified but it is partially identified (e.g. Manski (1990)). We make the following assumption, based on which we utilize panel data to improve the simple Manski bounds.

Assumption 2 (Time Homogeneity)**.** *For $j = 1, 0$ and for $t = 1, 2, \cdots, T$, Y_{it}^j and Y_{i1}^j have the same distribution conditional on α_i.*

Under Assumption 2, the distribution function in (1) no longer depends on t so that we can simply write $F^j(y)$ for $F_t^j(y)$. Also, Assumption 2 is silent about both the serial dependence of $\{Y_{it}^j : t = 1, \cdots, T\}$ and the contemporaneous dependence of Y_{it}^0 and Y_{it}^1. In particular, the dependence between Y_{it}^0 and Y_{it}^1 need not be time homogeneous.

A similar condition to Assumption 2 is used in Chernozhukov, Fernández-Val, Hahn, and Newey (2013) in a nonparametric regression setup. They consider the regression equation of the observed outcome, and they assume time homogeneity of the regression error conditional on $\overrightarrow{D}_{it}$ and α_i, which can be translated into the time homogeneity of the joint distribution of (Y_{it}^0, Y_{it}^1) in our context. Khan, Ponomareva, and Tamer (2013) also use a similar condition in a censored regression setup. They assume time homogeneity of the sum of the regression error and the fixed effect given the history of treatments, which is again translated into the time homogeneity of the joint distribution of the potential outcomes in our context.

2.2. **Structural Examples: Dynamics in Treatment Decisions.** Before we derive the sharp bounds of $F^j(y)$ under Assumptions 1 and 2, we briefly discuss some structural examples that show how these assumptions can be satisfied. Assumptions 1 and 2 are in fact weak enough to allow for interesting dynamics in treatment decisions, which we focus in this subsection. However, it is not the purpose of this subsection to model precisely a dynamic structure of treatment decisions. For the merits and costs of modeling a dynamic structure of treatment decisions, see e.g. Heckman and Navarro (2007). Throughout this subsection, we use the example of the effect of mothers' smoking on birthweight.

First, we assume that the potential birthweight of the t^{th} baby depends on the mother's overall health consciousness level a_i, her smoking status before the first pregnancy D_{i0} (i.e., some initial condition), and an idiosyncratic shock ϵ_{it}^j for $j = 1, 0$:

$$Y_{it}^j = \phi_j(\alpha_i, \epsilon_{it}^j), \tag{2}$$

where $\alpha_i = (a_i, D_{i0})$.

Suppose that U_{it}^1 and U_{it}^0 are the utilities that mother i enjoys during the t^{th} pregnancy from smoking and not smoking, respectively. Since smoking is *addictive*, the simplest dynamic structure would be the one where both U_{it}^1 and U_{it}^0 depend on D_{it-1}. In this case, for $j = 1, 0$,

$$U_{it}^j = g_j(D_{it-1}, \alpha_i, \eta_{it}^j) \tag{3}$$

for some random element η_{it}^j. The addictive nature of smoking can be described by the slopes of g_1 and g_0 in D_{it-1}. Now, the smoking decision D_{it} can be modeled by

$$D_{it} = \mathbb{I}\{g_1(D_{it-1}, \alpha_i, \eta_{it}^1) > g_0(D_{it-1}, \alpha_i, \eta_{it}^0)\} = h(D_{it-1}, \alpha_i, \eta_{it}) = H(\alpha_i, \overrightarrow{\eta}_{it}) \tag{4}$$

for some function $H(\cdot)$, where $\mathbb{I}\{\cdot\}$ is the indicator function and $\eta_{it} = (\eta_{it}^0, \eta_{it}^1)$.

Therefore, Assumption 1 will be satisfied if ϵ_{it}^j is independent of $\overrightarrow{\eta}_{it}$ conditional on α_i. Assumption 2 will also be satisfied if ϵ_{it}^j has a time–homogeneous distribution conditional on α_i. Note, however, that there is no restriction on the dependence structure between ϵ_{it}^0 and ϵ_{it}^1. For example, $(\epsilon_{it}^0, \epsilon_{it}^1)$ can follow a bivariate normal distribution with a time–varying covariance while its mean and variance stay constant over time.

We have not imposed any restrictions on the serial dependence of $\{\epsilon_{it}\}$ so far. If we assume that $\{\epsilon_{it}^j\}$ is serially independent conditional on α_i, then we can allow for more interesting dynamics in the treatment decisions such as *feedback effects*. For example, we can extend (3) to allow for *learning* from the observed outcome in the previous period. In fact, it is quite conceivable that a mother who smoked during the first pregnancy may change her behavior in the second pregnancy if the first child was born unhealthy. It is because she might think that the outcome could have been different had she not smoked. In order to allow for this type of learning, we may consider

$$U_{it}^j = \tilde{G}_j \left(D_{it-1}, \mathbb{E}(Y_{it-1}^1 - Y_{it-1}^0 | D_{it-1}, Y_{it-1}, \alpha_i), \alpha_i, \eta_{it}^j \right) = \tilde{g}_j(D_{it-1}, Y_{it-1}, \alpha_i, \eta_{it}^j), \tag{5}$$

which is more general than equation (3). In this case, we have

$$\begin{aligned} D_{it} &= \mathbb{I}\{\tilde{g}_1(D_{it-1}, Y_{it-1}, \alpha_i, \eta_{it}^1) > \tilde{g}_0(D_{it-1}, Y_{it-1}, \alpha_i, \eta_{it}^0)\} = \tilde{h}(D_{it-1}, Y_{it-1}, \alpha_i, \eta_{it}) \\ &= \tilde{h}\left(D_{it-1}, D_{it-1}\phi_1(\alpha_i, \epsilon_{it-1}^1) + (1 - D_{it-1})\phi_0(\alpha_i, \epsilon_{it-1}^0), \alpha_i, \eta_{it} \right) = \tilde{H}(\alpha_i, \overrightarrow{\epsilon}_{it-1}, \overrightarrow{\eta}_{it}) \end{aligned} \tag{6}$$

for some function $\tilde{H}(\cdot)$, where $\epsilon_{it} = (\epsilon_{it}^0, \epsilon_{it}^1)$. Apparently, Equation (6) does not affect the conditional distribution of Y_{it}^j given α_i, so Assumption 2 is still satisfied as long as ϵ_{it}^j has a time–homogeneous distribution given α_i. However, Assumption 1 now requires restrictions on the serial dependence of $\{\epsilon_{it}\}$: i.e. we need to assume that ϵ_{it}^j is independent of $(\overrightarrow{\epsilon}_{it-1}, \overrightarrow{\eta}_{it})$ conditional on α_i.

The utility function in (5) can be further generalized to allow for a forward–looking behavior by including $\mathbb{E}(Y_{it+k}^1 - Y_{it+k}^0 | D_{it-1}, Y_{it-1}, \alpha_i)$ with $k = 0, 1, \cdots, T - t$. However, $\mathbb{E}(Y_{it+k}^1 - Y_{it+k}^0 | D_{it-1}, Y_{it-1}, \alpha_i)$ is still a function of $(D_{it-1}, Y_{it-1}, \alpha_i)$, and hence this leads to the same reduced–form equation as (6).

3. The Sharp Bounds of the Counterfactual Distribution Functions

3.1. **Bounds Identification.** Based on Assumptions 1 and 2, we now obtain the sharp bounds of the potential outcome distribution functions. Since we assume that the observations are identically distributed across i, hereafter we suppress the subindex i.

Point identification of $F^j(y)$ is generally not available when the assumption of randomized treatment assignments fails to hold. However, using the idea of Manski (1990), their (pointwise) identifiable bounds can be obtained by

$$\mathbb{P}(Y_t \le y, D_t = j) \le F^j(y) \le \mathbb{P}(Y_t \le y, D_t = j) + \mathbb{P}(D_t = 1 - j), \tag{7}$$

which are sharp for each t (i.e. in the cross–section context). These Manski–type bounds can be improved when data on repeated treatments are available. More precisely, for each $j = 1, 0$ and $t = 1, 2, \cdots, T$, we define

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

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

and let

$$L^j(y) = L_T^j(y) = \sum_{s=1}^{T} p_s^j(y), \tag{8}$$

$$U^j(y) = U_T^j(y) = L_T^j(y) + \mathbb{P}(D_1 = \cdots = D_T = 1 - j). \tag{9}$$

The following theorem shows that $L^j(y)$ and $U^j(y)$ are the pointwise sharp bounds of the potential outcome distribution $F^j(y)$ for each $j = 1, 0$ under Assumptions 1 and 2.

Theorem 1. *Under Assumptions 1 and 2, for each $y \in \mathbb{R}$ and $j = 1, 0$, we have*

$$0 \le L^j(y) \le F^j(y) \le U^j(y) \le 1, \tag{10}$$

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

From (8) and (9), it can be readily seen that the bounds become tighter as T increases. In fact, $F^j(y)$ is point identified 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. Unlike the group of compliers in the instrumental variables literature (e.g. Angrist, Imbens, and Rubin (1996)), this subpopulation is identifiable from the data.

The bounds in Theorem 1 are only pointwise (i.e. for each $y \in \mathbb{R}$) sharp. The idea of functional sharp bounds has been discussed in the literature as well (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 consider in the empirical illustration in Section 5.

3.2. **Inferences.** When $F^j(\cdot)$ is partially identified, it is not straightforward to consider any direct statistical inferences on $F^j(\cdot)$. However, from Theorem 1, we have the identified bounds of $F^j(\cdot)$, 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 distribution theory is well developed for such estimators (e.g. Van der Vaart (1998)). Therefore, if we can formulate hypotheses for which test statistics only depend on the bound estimators, then it is easier to test them than to conduct direct inference of partially identified $F^j(\cdot)$. In this subsection, without detailing implementation issues, we briefly discuss some hypotheses of potential interest that can be formulated based on the identified bounds in (10).

One possibility is directly comparing the potential outcome distributions over the entire support, such as stochastic dominance relations. More precisely, we reparametrize (10) in terms of the parameter of interest $\Delta(y) = F^1(y) - F^0(y)$ as

$$\begin{pmatrix} L^0(y) \le F^1(y) - \Delta(y) \le U^0(y) \\ L^1(y) \le F^1(y) \le U^1(y) \end{pmatrix} \Longleftrightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \\ -1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F^1(y) \\ \Delta(y) \end{pmatrix} \le \begin{pmatrix} -L^0(y) \\ U^0(y) \\ -L^1(y) \\ U^1(y) \end{pmatrix}, \tag{11}$$

which is a set of linear moment inequalities. For given y, let $\theta = (F^1, \Delta)'$ and let Θ be the identified set such that all elements in Θ satisfy (11). Then, we can construct a confidence region for θ, at least pointwise for each y, following the recent development in the partial identification literature (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 for first order stochastic dominance relations: i.e. $\Delta(y) \geq 0$ 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:

$$\begin{cases} H^*_{0,a} : \forall \theta \in \Theta,\ \Delta(y) \geq 0 & \text{for all } y \\ H^*_{1,a} : \exists \theta \in \Theta,\ \Delta(y) < 0 & \text{for some } y, \end{cases} \tag{12}$$

or

$$\begin{cases} H^*_{0,b} :\ \exists \theta \in \Theta,\ \Delta(y) \geq 0 & \text{for all } y \\ H^*_{1,b} :\ \forall \theta \in \Theta,\ \Delta(y) < 0 & \text{for some } y. \end{cases} \tag{13}$$

It turns out that tests for (12) or (13) can be done easily because the parametric analysis of Hahn and Ridder (2009) can be extended to show that (12) and (13) are *equivalent* to

$$\begin{cases} H_{0,a} : U^0(y) \leq L^1(y) & \text{for all } y \\ H_{1,a} : U^0(y) > L^1(y) & \text{for some } y, \end{cases} \qquad \begin{cases} H_{0,b} : L^0(y) \leq U^1(y) & \text{for all } y \\ H_{1,b} : L^0(y) > U^1(y) & \text{for some } y, \end{cases} \tag{14}$$

respectively. The proof of the equivalence is provided in the Appendix. Testing the hypotheses in (14) is a standard task in the literature (e.g., Linton, Song, and Whang (2010)). Testing for higher order stochastic dominance can be similarly done.

[3]In testing for 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 5 we consider both possibilities. For an alternative approach to this issue, see Gupta and Panchapakesan (1979). We thank an anonymous referee for this reference.

The bounds in (10) can be inverted to obtain bounds on the quantiles of the potential outcomes:

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

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 for stochastic dominance as above. Note, however, that the bounds in (15) do not always provide informative bounds on the expectations of the potential outcomes, because the bounds in (15) are not generally integrable. This problem can be resolved when the support of Y_t^j is known to be bounded. More precisely, 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

$$\max\{m^j, Q_{U^j}(\tau)\} \leq Q_{F^j}(\tau) \leq \min\{M^j, Q_{L^j}(\tau)\} \quad \text{for } \tau \in (0,1), \tag{16}$$

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

4. Discussions on Time Varying Heterogeneity

Assumptions 1 and 2 are formulated by assuming that the unobserved heterogeneity α_i is time invariant, which can be a strong assumption in some applications. For instance, in the smoking–birthweight example, a mother's overall stress level during each pregnancy can vary over t and it may affect both the baby's health and the mother's smoking decision. In this case, it is important to control for potentially time–varying unobserved confounders (β_{it}) as well as time–invariant ones (α_i).

It looks challenging to allow for a general form of time–varying unobserved confounders in this framework. However, the sharp bounds we obtained in the previous section are robust to a limited violation of time invariant unobserved heterogeneity. In particular, we show in this section that absolute time invariance of unobserved heterogeneity is not needed but conditional time homogeneity is sufficient for the bounds in Theorem 1 to be pointwise sharp.

We first consider the following assumption. We let $\gamma_{it} = \gamma(\alpha_i, \beta_{it})$ be unobserved heterogeneity that can vary over time. Apparently, Assumption 3 below is implied by Assumptions 1 and 2 when absolute time invariance holds, i.e. $\gamma_{it} = \alpha_i$.

Assumption 3. *For $j = 1, 0$ and $t = 1, 2, \cdots, T$, we assume*

(i) *Y_{it}^j is independent of $\overrightarrow{D}_{it}$ conditional on $\overrightarrow{\gamma}_{it}$;*

(ii) *For any c and y, $\mathbb{P}(Y_{it}^j \leq y|\gamma_{it} = c) = \mathbb{P}(Y_{i1}^j \leq y|\gamma_{i1} = c)$ and Y_{it}^j is independent of $\overrightarrow{\gamma}_{it-1}$ conditional on γ_{it};*

(iii) *γ_{it} has the same distribution as γ_{is} for any $s < t$ conditional on $\overrightarrow{D}_{it}$.*

Assumption 3–(i) is a generalized version of Assumption 1. However, unlike Assumption 1, Assumption 3–(i) only assumes the existence of some unobserved confounding factor γ_{it}, which is probably so general that it is rarely violated. So, the real content of Assumption 3–(i) should be understood in the context of the rest of Assumption 3. Assumption 3-(ii) is about time homogeneity similar to Assumption 2. Assumption 3–(iii) is an extra assumption, which is trivially satisfied when $\gamma_{it} = \alpha_i$.

As in Mundlak (1978) and Chamberlain (1982), Assumption 3–(iii) directly imposes restrictions on the conditional distribution of the unobserved heterogeneity given the treatment history. This condition does allow for time–varying heterogeneity but only in a limited sense. A simple example is to assume that the time–varying element β_{it} in γ_{it} (such as the stress level during the t^{th} pregnancy) is retrospectively defined by

$$\beta_{it} = \varphi(\overrightarrow{D}_{iT}, \zeta_{it})$$

with an idiosyncratic error ζ_{it}. In this case, $\overrightarrow{D}_{iT}$ determines the overall type of mother i along with α_i, but the time–varying aspect solely comes from an idiosyncratic noise ζ_{it}. Assumption 3–(iii) holds when ζ_{it} is identically distributed over t given $\overrightarrow{D}_{iT}$.

Wooldridge (2005) considers a nonparametric version of Mundlak (1978)'s conditional mean assumption and proposes a consistent estimator for time–varying average treatment effects in a correlated–random coefficient setup. However, Wooldridge (2005)'s approach and ours are complementary since the objects of interest are different and they depend on different sets of assumptions.

The following theorem shows that the sharp bounds of $F^j(y)$ under Assumption 3 are the same as the ones in Theorem 1. This result is reminiscent of Mundlak (1978), who shows the equivalence between the within–group estimator with fixed effects and the generalized least squares estimator with correlated random effects.

Theorem 2. *Under Assumption 3, the pointwise sharp bounds of $F^j(y)$ are the same as the bounds given in Theorem 1.*

5. 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 $F^j(\cdot)$ denote the potential birthweight distribution and let D_{it} be the binary indicator denoting smoking status of mother i during her t^{th} pregnancy, with $D_{it} = 1$ indicating smoking. We obtain the bound estimates of the potential birthweight distributions for the population of *ever–smokers* and the subpopulation of *switchers*. We 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 do not include those who never smoked (*never–smokers*) in our analysis because the proportion of never–smokers is so large (i.e., 82.7% of the entire sample of three births) that we cannot 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 but never–smokers are unlikely to start smoking during pregnancy.

A mother's smoking decision is generally correlated with her health–related lifestyle factors, which need to be controlled for to isolate the causal effects of smoking on birthweight. However, many of such lifestyle factors (e.g., health consciousness and smoking behavior before the first pregnancy) are hard or nearly impossible to be observed, and hence the standard assumption of selection–on–observables is violated for the presence of omitted confounding factors. Similarly to Abrevaya (2006), we instead introduce unobserved heterogeneity and we assume selection–on–unobservables. We then assume that unobserved heterogeneity consists of general health consciousness, initial smoking status, and other time–invariant factors, in which case Assumption 1 is satisfied. In addition, knowing of a folk belief that the firstborn tends to be lighter than those born in the later order, we detrend all birthweight observations by adjusting their means for each t. After detrending, we observe the distribution of Y_{it} remains almost identical over t, and therefore we do not find evidence of violation of time homogeneity, i.e. Assumption 2. This framework allows for dynamics in a mother's smoking decisions as we discussed in Section 2.2. Lastly, there may still remain some concerns about the possibility that unobserved heterogeneity is time varying. However, as we discuss in Section 4, the bounds we use are robust to a limited violation of time invariant heterogeneity. For instance, if the relative stress level during the t^{th} pregnancy is defined as a function of the smoking history and a random noise, then Assumption 3 holds and the same bounds are achieved without changing our analysis.

Figures 1 and 2 show the bound estimates of the distribution functions of potential birthweights for ever–smokers and switchers, respectively. Dotted and dashed lines show pointwise 95% confidence sets for $F^j(y)$ for each y, using Imbens and Manski (2004) and Stoye (2009). The case of $T = 3$ shows the bounds that are estimated using all the three births 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. When $T = 3$, $\widehat{F}^1(\cdot)$ 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 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, where the null and alternative hypotheses are as given (12) and (13) (or (14) 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 (12), 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 reinforce the existing empirical findings such as Abrevaya and Dahl (2008) and Arellano and Bonhomme (2012). Also notice that, unlike

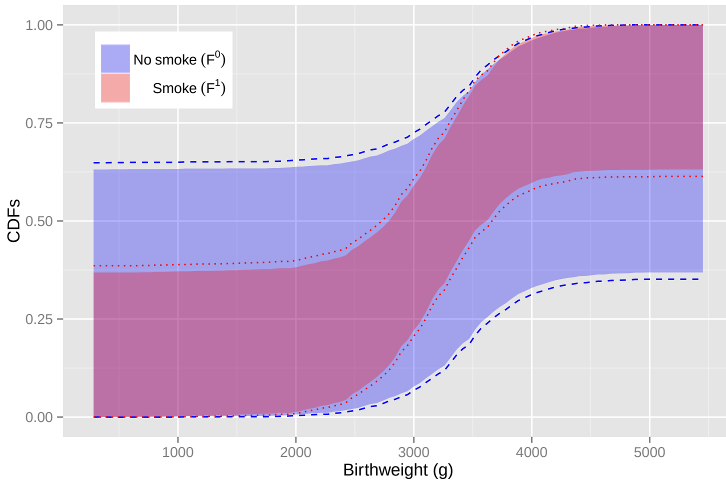
Null Hypothesis	Type-a (all F^0 and F^1 satisfies)		Type-b (there exist F^0 and F^1 s.t.)	
	F^0 FSD F^1	F^1 FSD F^0	F^0 FSD F^1	F^1 FSD F^0
Ever–Smokers ($n = 2137$)				
$T = 1$	0.000	0.000	0.905	0.899
$T = 2$	0.000	0.000	0.937	0.919
$T = 3$	0.000	0.000	0.943	0.000
Switchers ($n = 1445$)				
$T = 1$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

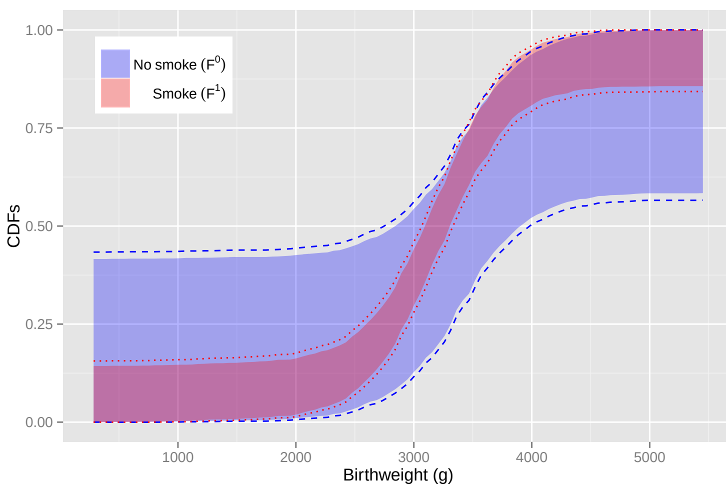
Abadie (2002), the stochastic dominance result for the switcher group is obtained without using instrumental variables.

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 (15) and (16) 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 and the average treatment effects.

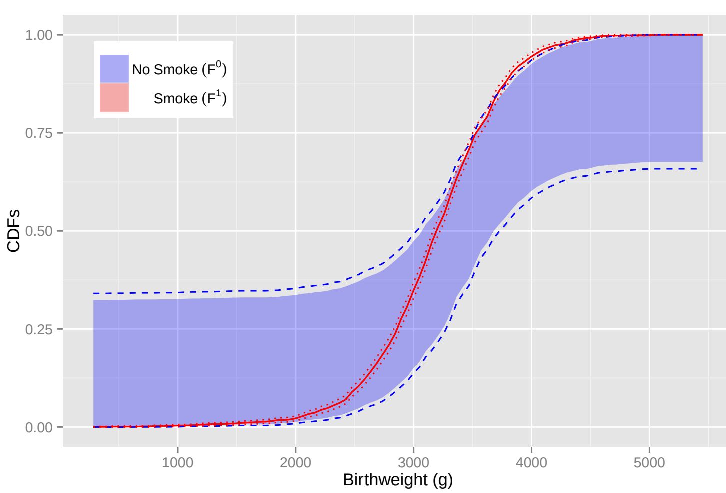
We also conduct the same analysis for a few more sub-populations with three births, but the overall results do not change. For instance, the subgroup of those who ever drank alcohol during at least one of the pregnancies (ever–drinkers) is considered. For this group, the bounds of F^1 are tighter (with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



(a) $T = 1$

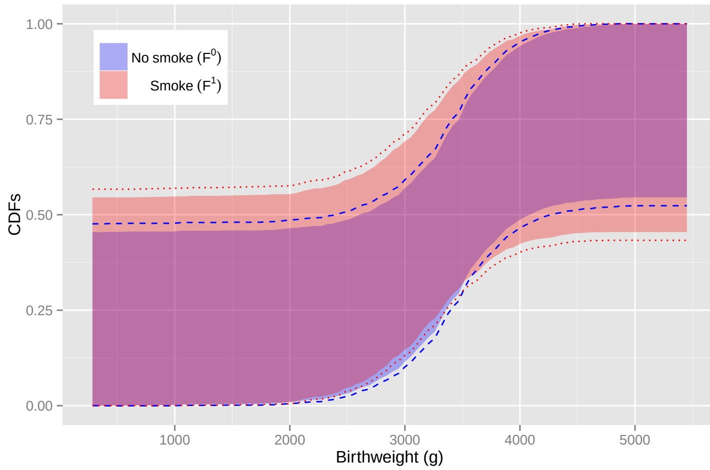


(b) $T = 2$

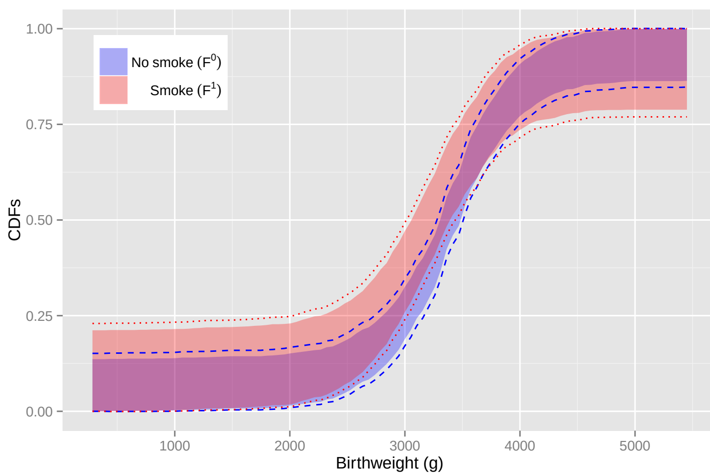


(c) $T = 3$

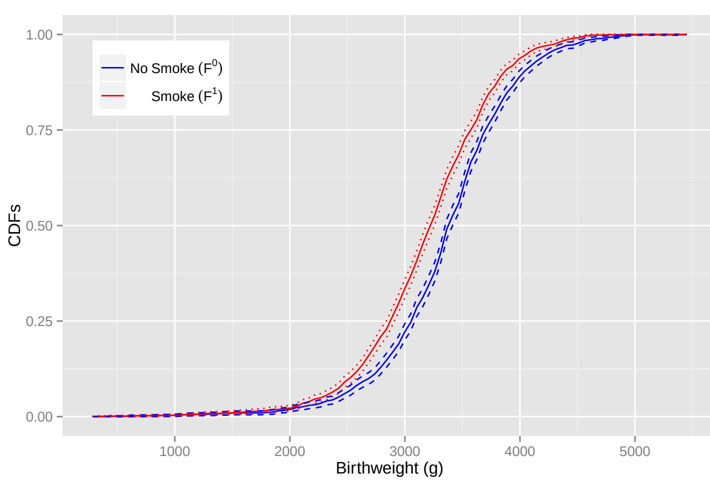
FIGURE 1. Bounds of the Distributions of Potential Birthweights (Ever–Smokers)



(a) $T = 1$



(b) $T = 2$



(c) $T = 3$

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

	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)

6. 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.

6.1. **Proof of Theorem 1.**

We only consider $j = 1$ because $j = 0$ can be shown in a similar way. We also omit the index i. For given t, note that

$$\begin{aligned}\mathbb{P}(Y_t^1 \leq y) = \mathbb{P}(Y_t^1 \leq y, D_1 = 1) + \mathbb{P}(Y_t^1 \leq y, D_1 = 0, D_2 = 1)& \\ + \cdots + \mathbb{P}(Y_t^1 \leq y, D_1 = D_2 = \cdots = D_{t-1} = 0, D_t = 1)& \\ + \mathbb{P}(Y_t^1 \leq y, D_1 = D_2 = \cdots = D_{t-1} = D_t = 0).& \quad (17)\end{aligned}$$

Note also that for any $s \leq t$,

$$\begin{aligned}\mathbb{P}(Y_t^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s) &= \int \mathbb{P}(Y_t^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s | \alpha = a) d\mu_\alpha(a) \\ &\overset{A1}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \alpha = a)\mathbb{P}(Y_t^1 \leq y | \alpha = a) d\mu_\alpha(a) \\ &\overset{A2}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \alpha = a)\mathbb{P}(Y_s^1 \leq y | \alpha = a) d\mu_\alpha(a) \\ &\overset{A1}{=} \int \mathbb{P}(Y_s^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s | \alpha = a) d\mu_\alpha(a) = \mathbb{P}(Y_s^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s). \quad (18)\end{aligned}$$

Therefore, equation (17) is equal to

$$\mathbb{P}(Y_t^1 \leq y) = \mathbb{P}(Y_1 \leq y, D_1 = 1) + \sum_{s=2}^{t} \mathbb{P}(Y_s \leq y, \overrightarrow{D}_{s-1} = 0, D_s = 1) + \mathbb{P}(Y_t^1 \leq y, \overrightarrow{D}_t = 0),$$

which yields the desired result, because the last term on the right hand side is between 0 and $\mathbb{P}(\overrightarrow{D}_t = 0)$. In fact, these bounds are sharp, because

$$\mathbb{P}(Y_t^1 \leq y, \overrightarrow{D}_t = 0) = \int \mathbb{P}(\overrightarrow{D}_t = 0 | \alpha = a)\mathbb{P}(Y_t^1 \leq y | \alpha = a) d\mu_\alpha(a), \quad (19)$$

where our assumptions do not impose any restrictions on the support of $\mathbb{P}(Y_t^1 \leq y|\alpha)$ except that it is between 0 and 1. $\square$

6.2. **Proof of Theorem 2.** We show how to derive an equation like (18) under Assumption 3: the remaining arguments are the same as the proof of Theorem 1. Note first that

$$\begin{aligned}
\mathbb{P}(Y_t^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s) &= \int \mathbb{P}(Y_t^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\gamma}_t = \overrightarrow{c}_t) d\mu_{\overrightarrow{\gamma}_t}(\overrightarrow{c}_t) \\
&\overset{A3(i)}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\gamma}_t = \overrightarrow{c}_t) \mathbb{P}(Y_t^1 \leq y | \overrightarrow{\gamma}_t = \overrightarrow{c}_t) d\mu_{\overrightarrow{\gamma}_t}(\overrightarrow{c}_t) \\
&\overset{A3(ii)}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\gamma}_t = \overrightarrow{c}_t) \mathbb{P}(Y_t^1 \leq y | \gamma_t = c_t) d\mu_{\overrightarrow{\gamma}_t}(\overrightarrow{c}_t),
\end{aligned}$$

which is equal to

$$\begin{aligned}
&\int\int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\gamma}_{t-1} = \overrightarrow{c}_{t-1}, \gamma_t = c_t) d\mu_{\overrightarrow{\gamma}_{t-1}|\gamma_t}(\overrightarrow{c}_{t-1}|c_t) \mathbb{P}(Y_t^1 \leq y | \gamma_t = c_t) d\mu_{\gamma_t}(c_t) \\
&= \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \gamma_t = c_t) \mathbb{P}(Y_t^1 \leq y | \gamma_t = c_t) d\mu_{\gamma_t}(c_t) \\
&= \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \gamma_s = c_t) \mathbb{P}(Y_s^1 \leq y | \gamma_s = c_t) d\mu_{\gamma_s}(c_t), \quad (20)
\end{aligned}$$

where the last equality follows from Assumptions 3–(ii) and 3–(iii). Then, (20) is equal to

$$\begin{aligned}
&\int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\gamma}_{s-1} = \overrightarrow{c}_{s-1}, \gamma_s = c_t) \mathbb{P}(Y_s^1 \leq y | \gamma_s = c_t) d\mu_{\overrightarrow{\gamma}_{s-1},\gamma_s}(\overrightarrow{c}_{s-1}, c_t) \\
&\overset{A3(ii)}{=} \int \mathbb{P}(\overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\gamma}_{s-1} = \overrightarrow{c}_{s-1}, \gamma_s = c_t) \mathbb{P}(Y_s^1 \leq y | \overrightarrow{\gamma}_{s-1} = \overrightarrow{c}_{s-1}, \gamma_s = c_t) d\mu_{\overrightarrow{\gamma}_{s-1},\gamma_s}(\overrightarrow{c}_{s-1}, c_t) \\
&\overset{A3(i)}{=} \int \mathbb{P}(Y_s^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s | \overrightarrow{\gamma}_{s-1} = \overrightarrow{c}_{s-1}, \gamma_s = c_t) d\mu_{\overrightarrow{\gamma}_{s-1},\gamma_s}(\overrightarrow{c}_{s-1}, c_t) \\
&= \mathbb{P}(Y_s^1 \leq y, \overrightarrow{D}_s = \overrightarrow{d}_s)
\end{aligned}$$

as desired. $\square$

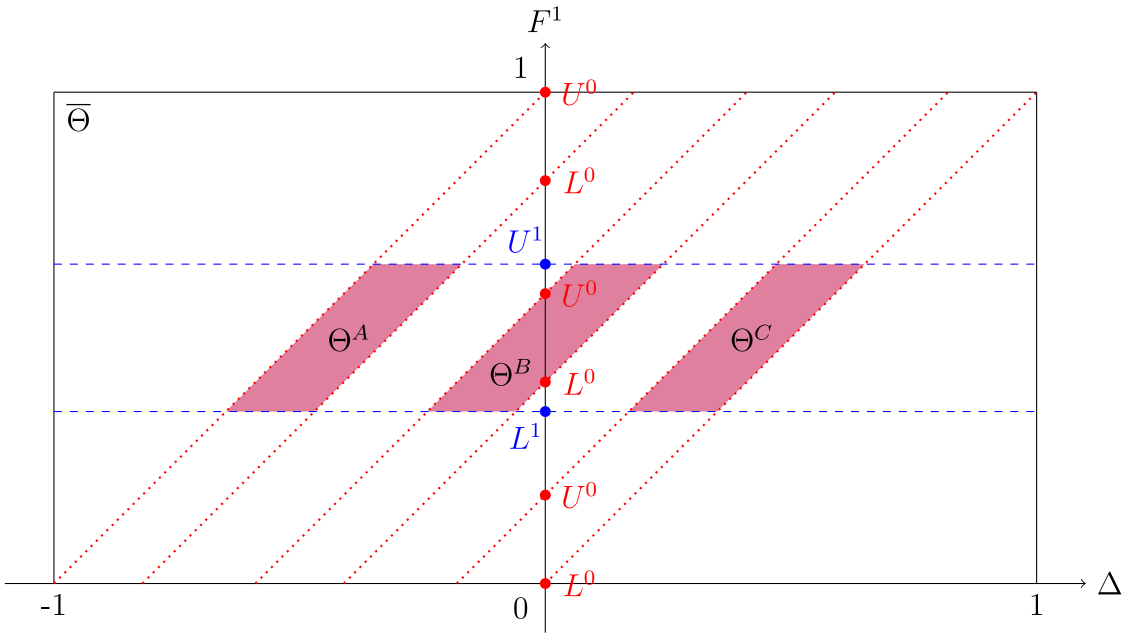


FIGURE 3. Graphical Illustration of the Duality

6.3. **Duality.**

Proposition 1. *The hypotheses in (12) and (13) are equivalent to the hypotheses in (14), 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 (14).

• **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 (10), 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 (10), 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 (14).

• **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 (10), 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 (10) 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 (10) 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. $\square$

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., and S. Panchapakesan (1979): *Multiple decision procedures: theory and methodology of selecting and ranking populations*, no. 71 in Wiley series in probability and mathematical statistics. Wiley, New York.

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, chap. 70, pp. 4779 – 4874. Elsevier, Amsterdam.

——— (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, chap. 71, pp. 4875–5143. Elsevier, Amsterdam.

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. F. (1990): "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*, 46(1), 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. (1998): *Asymptotic statistics*, vol. 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

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.