Bounds on Average and Quantile Treatment Effects on Duration Outcomes under Censoring, Selection, and Noncompliance

German Blanco[*], Xuan Chen[†], Carlos A. Flores[‡] and Alfonso Flores-Lagunes[§]

March 28, 2019

Abstract

We consider the problem of assessing the effects of a treatment on duration outcomes using data from a randomized evaluation with noncompliance. For such settings, we derive nonparametric sharp bounds for average and quantile treatment effects addressing three pervasive problems simultaneously: self-selection into the spell of interest, endogenous censoring of the duration outcome, and noncompliance with the assigned treatment. Ignoring any of these issues could yield biased estimates of the effects. Notably, the proposed bounds do not impose the independent censoring assumption—which is commonly used to address censoring but is likely to fail in important settings—or exclusion restrictions to address endogeneity of censoring and selection. Instead, they employ monotonicity and stochastic dominance assumptions. To illustrate the use of these bounds we assess the effects of the Job Corps (JC) training program on its participants' last complete employment spell duration. Our estimated bounds suggest that JC participation may increase the average duration of the last complete employment spell before week 208 after randomization by at least 5.6 log points (5.8 percent) for individuals who comply with their treatment assignment and experience a complete employment spell whether or not they enrolled in JC. The estimated quantile treatment effects suggest the impacts may be heterogeneous, and strengthen our conclusions based on the estimated average effects.

[1]Detailed comments from the associate editor and two anonymous referees greatly improved the paper and are gratefully acknowledged. We are also grateful for comments from Adina Ardelean, Domenico Depalo, Bringham Frandsen, John Ham, Shan Li, Fabrizia Mealli, Thomas Russell, Pedro Sant'Anna, Youngki Shin, Derya Uysal, Haiqing Xu, Li Yu, and seminar/conference participants at the 2015 Southern Economic Association Meetings, the 2016 New York Camp Econometrics, the 2016 Meetings of the Midwest Econometrics Group, the 2017 Annual Meetings of the Society of Labor Economists, the 2017 Asian Meeting of the Econometric Society, the 2017 China Meeting of the Econometric Society, the 2017 International Association of Applied Econometrics Conference, the 2017 International Workshop on Causal Inference, Program Evaluation, and External Validity at LISER (Luxembourg), Central University of Finance and Economics (China), the 13th IZA Conference: Labor Market Policy Evaluation, the 2018 North America Summer Meetings of the Econometric Society, the 2018 Western Economics Association International Annual Conference, the 2018 Meetings of the Canadian Econometrics Study Group, and the 2018 Meetings of the Midwest Econometrics Group. Chen acknowledges support from the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China, 14XNF022. Flores acknowledges summer research support from the Orfalea College of Business at California Polytechnic State University. All the usual disclaimers apply.

[*]Department of Economics, Illinois State University; and GLO (gblanco@ilstu.edu).

[†]School of Labor and Human Resources, Renmin University of China (xchen11@ruc.edu.cn).

[‡]Department of Economics, California Polytechnic State University at San Luis Obispo (cflore32@calpoly.edu).

[§]Department of Economics and Center for Policy Research, Syracuse University; and IZA & GLO (aflo-resl@maxwell.syr.edu).

1 Introduction

From an economic and policy making perspective, evaluating the effectiveness of interventions (treatments) is of utmost interest. In numerous empirical settings, the relevant outcome variable represents the duration of an event. Estimating causal effects on duration variables, however, is challenging. Even when employing data from an experimental evaluation, assessing the treatment's impact on duration outcomes is often complicated by at least three identification problems: censoring, sample selection, and noncompliance. Censoring arises because the complete duration spell is fully observed only if this spell is completed before the end of the observation period, with the extent of censoring potentially affected by the treatment. Sample selection occurs because, even when individuals are randomized into the treatment, their decision to experience a spell post-randomization is endogenous and potentially affected by the treatment. Finally, noncompliance occurs because some individuals assigned to undergo the treatment do not take it up, while some assigned not to undergo the treatment take it up. In this paper, we derive nonparametric sharp bounds on average and quantile treatment effects on duration outcomes in the presence of these three pervasive identification problems. For quantile treatment effects, our bounds are on the effects on pre-specified quantiles (i.e., pointwise), and not on the corresponding quantile function (e.g., Firpo and Ridder, 2008). The proposed bounds do not require exclusion restrictions or conditional independence assumptions to address the endogeneity of selection and censoring, and do not impose distributional assumptions. Combinations of those assumptions, which are potentially strong in relevant empirical settings, are employed by most existing methods that focus on point identification. Instead, our bounds are based on weak monotonicity and stochastic dominance assumptions that are potentially weaker and are common in the growing partial identification literature (e.g., Manski, 1997; Manski and Pepper, 2000; Zhang and Rubin, 2003; Imai, 2007, 2008; Blundell et al., 2007; Lee, 2009; Lechner and Melly, 2010; Bhattacharya et al., 2012; Flores and Flores-Lagunes, 2010, 2013; Chen and Flores, 2015; Frandsen and Lefgren, 2016). We address noncompliance by employing the randomized treatment assignment as an instrumental variable for treatment receipt.

To motivate and illustrate the use of our bounds, we consider the evaluation of the effects of an important training program for disadvantaged youth in the U.S.—Job Corps or JC—on the duration of its participants' last complete employment spell, using data from a randomized evaluation that was subject to noncompliance. The methodological and empirical work in the existing literature on the evaluation of active labor market programs mainly focuses on estimating average effects of an intervention on outcomes such as earnings and employment (for a review see Heckman et al., 1999 and Imbens and Wooldridge, 2009). While estimating the effects of labor market programs on employment rates is important, it is also crucial from a policy point of view to analyze their effects on labor market duration outcomes. For example, as discussed by Ham and LaLonde (1996), such programs may improve employment rates by helping unemployed individuals find jobs faster (e.g., by improving their job search skills) or by helping employed individuals keep their jobs longer (e.g., by improving their work habits). Hence, estimating program effects on the duration of employment spells, for example, can shed light into how the program affects

employment rates, providing valuable information for policy makers. The proposed bounds are not specific to the analysis of the effects of training programs on employment durations. More generally, they can be employed in quasi-experimental studies and randomized experiments with noncompliance in which researchers are interested in analyzing the effects of a treatment on an outcome that is censored and is realized (or observed) only for a non-random subpopulation. This setting often arises in important applications. For example, when analyzing the effects of a welfare program on the duration of future welfare participation spells; when analyzing the effect of an intervention on future income, with the income variable being affected by censoring and non-response; or when analyzing the effects of a new drug or medical procedure on the duration of a given event that is not experienced by every individual (e.g., the effect of influenza vaccination on the length of flu-related hospitalizations). In addition, researchers can employ the proposed bounds to inform the sensitivity of results obtained with existing methods that rely on potentially stronger assumptions to achieve point identification.

To derive our bounds, we employ principal stratification (Frangakis and Rubin, 2002). Principal stratification—which has its roots in the instrumental variables analysis by Angrist et al. (1996)—provides a framework for analyzing treatment effects when controlling for variables that have been affected by treatment assignment: in our case, the indicators for treatment receipt, selection, and censoring. The key insight of principal stratification is that causal effects can be obtained by comparing individuals who share the same potential values of the post-treatment variable(s) one wants to adjust for under both values of the assigned treatment indicator, since membership to these groups (called principal strata) is not affected by treatment assignment. Thus, focusing on principal strata provides a way to control for unobserved confounders that correlate with the treatment, post-treatment variables, and the outcome. Bounds have been previously used in the literature to analyze the effects of training programs within this framework. Zhang et al. (2008) and Lee (2009) adjusted for sample selection when analyzing training effects on average wages (as wages are observed only for employed individuals). Lee (2009) and Blanco et al. (2013a, 2013b) employed these bounds to analyze the wage effects of JC, with the latter papers also employing Imai's (2008) extension of these bounds to quantile treatment effects. These papers, however, did not consider noncompliance. Imai (2007) and Chen and Flores (2015) derived bounds accounting for both noncompliance and selection. The latter paper used these bounds to analyze the wage effects of JC. Both Imai (2007) and Imai (2008) also considered quantile treatment effects. Importantly, none of the above papers considered duration or censored outcomes. This paper extends those bounds to the important case when the outcome is censored. Recently, Shepherd et al. (2007) employed principal stratification to propose bounds on the effects of a treatment on a censored outcome in the presence of sample selection. However, they do not consider noncompliance and impose the independent censoring assumption (discussed below), as well as parametric assumptions.

There is a large literature dealing with censoring, some of which also addresses endogeneity (in our case, noncompliance). For example, building on the work of Powell (1986) and Chernozhukov and Hong (2002)—among others—on censored quantile regression, Hong and Tamer (2003), Blundell and Powell (2007), Chernozhukov et al. (2015), Frandsen (2015) and Arellano and Bonhomme (2017) also deal with an endogenous regressor by employing instrumental variables. In a similar

setting with censoring and endogeneity, Sant'Anna (2016) proposes an alternative two-step instrumental variable estimator based on Kaplan-Meier integrals. A common assumption in these papers is that the censoring variable is conditionally independent of the latent outcome (a notable exception is Fan and Liu (2018), who model censoring using an Archimedean copula). While common, this assumption is likely to fail in important settings, where it leads to bias and loss of point identification (Khan and Tamer, 2009; Frandsen, 2018). One setting in which independent censoring is likely to fail is when the censoring points in the sample vary across observations (Frandsen, 2018). In our empirical illustration, while there is an administratively given end of the follow-up period that determines censoring (week 208 after randomization), individuals start their employment spells endogenously at different times, making it possible that unobserved characteristics of individuals confound both the latent duration outcome and the censoring indicator, rendering the independent censoring assumption dubious. Frandsen (2018) proposed a test of this assumption and applied it to unemployment duration data from three commonly-used data sets, finding that the assumption frequently fails. The same test rejects the independent censoring assumption in the data of our empirical illustration. An important contribution of our proposed bounds is to allow for endogenous censoring, in addition to endogenous sample selection, thus providing an additional tool for applied research.

Another relevant related literature is that of duration models (e.g., Heckman and Singer, 1984; Ham and LaLonde, 1996; Eberwein et al., 1997, 2002; Abbring and Van den Berg, 2003; Van den Berg et al., 2016, Ba et al., 2017). This literature usually addresses censoring, selection, and sometimes noncompliance/endogeneity, by imposing parametric, separability, or proportionality assumptions. Some standard duration models also rely on the conditional independent censoring assumption. Recent work by Vikström et al. (2018) constructs bounds on transition probabilities using a different framework from ours, although they do not allow for noncompliance. The methods developed in this paper complement this literature by constructing nonparametric bounds for treatment effects on average durations and quantiles of the duration distribution using a different framework and set of assumptions. Our approach, however, does not provide bounds on hazard or transition probabilities, which are often the objects of interest in duration models and are more difficult to estimate because of dynamic selection. Our focus on effects on the duration of spells avoids the dynamic selection problem (Vikström et al., 2018). Nevertheless, even when focusing on the duration of spells, there may be post-randomization variables (confounders) affected by randomization and correlated with compliance, selection, censoring, and the outcome. These are the confounders we address through principal stratification.

To illustrate the use of the proposed bounds, we evaluate the effects of JC on the duration of its participants' last complete employment spell. The JC is America's largest and most comprehensive education and job training program enrolling disadvantaged youth, ages 16 to 24, at no out-of-pocket cost to them. During 1995-96, the US Department of Labor funded the National Job Corps Study (NJCS), whose main feature was the randomization of eligible participants into treatment and control groups. Using data from the NJCS, various studies have examined the effects of being randomly assigned to enroll in JC (rather than actual enrollment) mostly on non-duration outcomes such as earnings, employment, and wages (e.g., Schochet et al., 2001; Schochet et al.,

2008; Lee, 2009; Zhang et al., 2009; Flores-Lagunes et al., 2010; Flores and Flores-Lagunes, 2010, 2013; Blanco et al., 2013a, 2013b). These estimated intention-to-treat effects do not capture the actual effect of receiving training when there is noncompliance with the assigned treatment, as it was the case in the NJCS. Commonly, to address noncompliance the treatment assignment indicator is employed as an instrumental variable for actual participation, which identifies actual training effects for individuals who comply with their treatment assignment, or compliers (Imbens and Angrist, 1994; Angrist et al., 1996). Studies analyzing the effects of actual JC participation on earnings, employment, and wages include Schochet et al. (2001), Schochet et al. (2008), Frumento et al. (2012), Eren and Ozbeklik (2014), Chen and Flores (2015), and Chen et al. (2018). However, to the best of our knowledge, despite the importance of analyzing the effects of training programs on duration outcomes (as previously discussed), no study has assessed the effects of this major program on such outcomes.

Our estimated bounds suggest that JC training has a plausible positive average effect on the duration (in weeks) of the last complete employment spell before week 208 after randomization for individuals who comply with their treatment assignment and who would experience a complete employment spell whether or not they enrolled in JC. The causal effect for this latent subpopulation (stratum) is the one bounded by our approach. It represents at least 21.3 percent of our sample–which excludes Hispanics–and at least 31 percent of all compliers. Following the principal stratification literature, this is the only stratum for which causal effects can be estimated without imposing further assumptions, since it is the only stratum for which complete employment durations are observed for individuals who participated and did not participate in JC. Bounding effects for other subpopulations would require adding likely stronger assumptions. Under our preferred set of assumptions, this effect is bounded between 5.6 and 55 log points (5.8 and 73.3 percent). Despite our consistent finding of positive estimated lower bounds for this mean effect across different demographic groups, the corresponding confidence intervals include zero at traditional significance levels. In addition, the results suggest that the effect may be heterogeneous across the quantiles analyzed of the employment duration distribution. In some of the lower quantiles the effect is positive and statistically significant. In contrast, while positive effects are observed at the upper part of the employment spell distribution, the confidence intervals include zero for most quantiles above the median (although they rule out plausible positive and negative values).

We complement our analysis by estimating the effect of JC on the log wages in those employment spells. This is an important feature of the proposed approach: it allows bounding not only the effects of training on employment duration but also on the wages (or other characteristics) of those employment spells, since the bounds we employ address noncompliance and selection into employment, which are the identification problems for estimating wage effects. We find that, for the same subpopulation for which the effects on employment duration were estimated, JC statistically significantly increases their average log wages in those employment spells by between 5.9 and 21.2 log points (6.1 and 23.6 percent). This finding suggests that JC may not only help these individuals to maintain their jobs longer, but also that those jobs are better paid. In contrast to previous work that also analyzed wage effects of JC under noncompliance and sample selection for a different subpopulation from ours (Chen and Flores, 2015), we also consider quantile treatment effects. Our

results again suggest that the effects may be heterogeneous across the quantiles analyzed of the wage distribution, with quantiles above the median ruling out zero effects more frequently than lower quantiles.

2 Econometric Methods

2.1 Motivating Example: Job Corps

To motivate and discuss the assumptions underlying the bounds to be introduced, we consider inference on average and quantile treatment effects of JC participation on the duration of the last complete employment spell. The JC program was established in 1964 under the Economic Opportunity Act, and today is a key pillar of the Workforce Innovation and Opportunity Act (WIOA), signed in 2014. The program is America's largest and most comprehensive education and job training program. The goal of JC is to help disadvantaged young people, ages 16 to 24, improve the quality of their lives by enhancing their labor market opportunities and educational skill set, which is achieved through the offering of academic instruction, career technical training, residential living, health care, counseling, and job placement assistance. In a typical year, about 60,000 eligible students enroll in one of the 119 JC centers located nationwide, where participants typically reside. Eligible JC students are selected based on several criteria, for example, age 16 to 24, legal US residency, economically disadvantage status, living in a disruptive environment, the need of additional education or training, and be judged to have the capability and aspirations to participate in JC. The average length of enrollment in the program is about eight months.

During the mid-nineties, the Department of Labor funded the National Job Corps Study (NJCS) to determine the program's effectiveness. The main feature of the study was its randomization: individuals were taken from nearly all JC's outreach and admissions agencies located in the 48 contiguous states and the District of Columbia and randomly assigned to treatment and control groups. From a randomly selected research sample of 15,386 first time eligible applicants, 9,409 were assigned to the treatment group–permitted to enroll in JC–and the remainder 5,977 to the control group–embargoed from enrollment for three years. After recording their data through a baseline interview for both treatment and control groups, a series of follow up interviews were conducted at weeks 52, 130, and 208 after randomization (Schochet et al., 2001). By week 208 after randomization, about 74% of individuals in the treatment group had enrolled in JC (about 96% of them within 6 months after randomization), while about 4.7% of individuals in the control group had enrolled in JC.

To our knowledge, no study has assessed the effects of JC participation on the duration of employment or unemployment spells. Our illustration will focus on the duration of the last employment spell by week 208 after randomization as an outcome because it represents a long-term outcome and to avoid potentially serious threats to some of our assumptions. One of them relates to the fact that JC participants cannot hold employment while in the program, a phenomenon known as a locking-in effect. Another potential threat consists of the possibility that JC participants raise

their reservation wage, declining job offers that would be accepted under no participation. Given that most participants enroll shortly after randomization and train for an average of 8 months, the consequences of the locking-in effect and a higher reservation wage are likely diminished in the long run. This would not be the case if we were to focus on a short-term outcome. In this illustration, our bounds will account for noncompliance with the random assignment, the selection into the last employment spell by week 208 after randomization, and the censoring of the duration that occurs when the employment spell continues beyond week 208 after randomization.

2.2 Notation and Principal Strata

Consider a random sample of size n from a large population. Let $Z_i \in \{0, 1\}$ indicate whether unit i is randomly assigned to the treatment group ($Z_i = 1$) or the control group ($Z_i = 0$) at baseline, with $i = 1, ..., n$. Let $D_i \in \{0, 1\}$ be the treatment receipt indicator, i.e., whether unit i actually takes the treatment ($D_i = 1$) or not ($D_i = 0$). In our application, Z_i denotes whether i is randomly assigned to participate or not in JC, while D_i is an indicator for actual JC participation. Let Y_i^* be the latent duration of the employment spell of interest. In our application, Y_i^* denotes the latent duration of the last employment spell by the end of the follow-up observation period, which is week 208 after randomization. Let $S_i \in \{0, 1\}$ indicate whether unit i selects into the spell of interest ($S_i = 1$) or not ($S_i = 0$), and let C_i denote the time from the start of the spell of interest until censoring takes place for individual i. Not all individuals may experience the spell of interest; for example, in our application some individuals may remain unemployed from randomization until the end of the observation period. If individual i experiences the spell of interest, i.e., $S_i = 1$, then the observed spell duration is $Y_i = \min(Y_i^*, C_i)$. Let W_i be the censoring indicator variable, so $W_i = 1$ if $Y_i = C_i$, and $W_i = 0$ if $Y_i = Y_i^*$; or $W_i = 1\{Y_i = C_i\}$ with $1\{\cdot\}$ the indicator function. If $S_i = 0$, then Y_i^*, C_i, and thus Y_i and W_i are not defined since no related spell exists (the analysis below is unaffected if we were to allow these variables to be defined since the causal parameters of interest consider only individuals who would experience the spell of interest under both random assignment arms; see the related discussion in Chen and Flores, 2015, p. 525). We follow the principal stratification literature and assign them the value $*$ (e.g., Zhang and Rubin, 2003; Shepherd et al., 2007; Imai, 2007, 2008; Zhang et al., 2008, 2009; Frumento et al., 2012).

Using the potential outcomes framework, we define the following potential variables: $D_i(z)$, $S_i(z)$, $Y_i^*(z)$, $C_i(z)$, $Y_i(z)$, and $W_i(z)$ for $z = 0, 1$. Thus, for instance, $D_i(1)$ and $D_i(0)$ denote the treatment individual i would actually receive under assignment to treatment ($z = 1$) or no treatment ($z = 0$), respectively. Note that $Y_i^*(z)$, $C_i(z)$, $Y_i(z)$, and $W_i(z)$ are defined only if $S_i(z) = 1$ for $z = 0, 1$. If $S_i(z) = 0$, we let $Y_i^*(z) = C_i(z) = Y_i(z) = W_i(z) = *$. Since both treatment assignment and treatment receipt can potentially affect the post-treatment variables we consider, we also define the composite potential variables $S_i(z, d)$, $Y_i^*(z, d)$, $C_i(z, d)$, $Y_i(z, d)$, and $W_i(z, d)$ for $z, d = 0, 1$, where the last four are defined only if $S_i(z, d) = 1$, while $Y_i^*(z, d) = C_i(z, d) = Y_i(z, d) = W_i(z, d) = *$ if $S_i(z, d) = 0$. Note that, since D is potentially affected by Z, we have $Y_i^*(z) = Y_i^*(z, D_i(z))$, and similarly for the other four composite potential variables. We observe $\{Z_i, D_i, S_i\}$ for all individuals, where $D_i = D_i(Z_i)$, and $S_i = S_i(Z_i, D_i)$. We also observe

$\{Y_i, W_i\}$ if $S_i = 1$, where $Y_i = \min\{Y_i^*(Z_i, D_i), C_i(Z_i, D_i)\}$ and $W_i = 1\{Y_i = C_i(Z_i, D_i)\}$. If $S_i = 0$, we have $Y_i = *$ and $W_i = *$. In this context, causal effects of treatment receipt on the duration of the spells of interest are comparisons of the latent potential outcomes $Y_i^*(z, 0)$ and $Y_i^*(z, 1)$. In line with the potential outcomes framework, this notation implicitly assumes the stable unit treatment value assumption (SUTVA) (Rubin, 1978, 1980, 1990).

We study identification of the effects of interest in the presence of three post-randomization complications: censoring, selection, and noncompliance. To this end, we employ principal stratification (Frangakis and Rubin, 2002), which provides a useful framework for analyzing causal effects when adjusting for post-randomization variables that have been affected by the randomized treatment assignment (Z), as in our setting. The key insight is that causal effects can be obtained by comparing individuals within principal strata, which are subpopulations whose members share the same potential values of the post-randomization variables one wants to adjust for. The reason is that membership to a particular principal stratum is not affected by treatment assignment, implying that it can be treated as any other baseline covariate: just as one can obtain causal effects by comparing individuals with a given value of a baseline covariate (e.g., gender), one can obtain causal effects within principal strata. A familiar example of principal stratification is the use of random assignment as an instrumental variable for an endogenous treatment as in Imbens and Angrist (1994) and Angrist et al. (1996). In that situation with one post-randomization complication (endogeneity of treatment or noncompliance), the principal strata are the so-called "compliance types" and causal effects are obtained by comparing outcomes of individuals belonging to the same stratum. The effect that is point identified is the causal treatment effect for the subpopulation of compliers. As we add two more post-randomization complications (selection and censoring), the causal effect that will be partially identified will be for a more specific subpopulation.

Given the three identification issues we address, we partition the population based on the joint potential values of the treatment receipt (D), selection (S), and censoring (W) indicator variables under the two values of treatment assignment ($z = 0, 1$): $\{D_i(0), D_i(1)\} \times \{S_i(0), S_i(1)\} \times \{W_i(0), W_i(1)\}$. The stratification with respect to $\{D_i(0), D_i(1)\}$, done to address noncompliance using treatment assignment as an instrument for treatment receipt, is the same as that in Angrist et al. (1996). Define the following four subpopulations: $a = \{i : D_i(0) = D_i(1) = 1\}$, individuals who always take the treatment regardless of treatment assignment, or always takers; $n = \{i : D_i(0) = D_i(1) = 0\}$, individuals who never take the treatment regardless of treatment assignment, or never takers; $c = \{i : D_i(0) = 0, D_i(1) = 1\}$, individuals who take the treatment only if assigned to take it, or compliers; and $d = \{i : D_i(0) = 1, D_i(1) = 0\}$, individuals who take the treatment only if assigned not to take it, or defiers.

The stratification with respect to $\{S_i(0), S_i(1)\}$ is analogous to that used to address sample selection in Zhang and Rubin (2003), Zhang et al. (2008), Lee (2009) and Imai (2008), among others. In our application, the particular selection we address is into the employment spell of interest. Define the following four subpopulations based on the joint potential values of $\{S_i(0), S_i(1)\}$: $EE = \{i : S_i(0) = S_i(1) = 1\}$, individuals who always select into the employment spell of interest regardless of treatment assignment, or always selected; $NN = \{i : S_i(0) = S_i(1) = 0\}$, individuals who never select into the employment spell of interest regardless of treatment assignment, or never

selected; $NE = \{i : S_i(0) = 0, S_i(1) = 1\}$, individuals who select into the employment spell of interest only if assigned to treatment; $EN = \{i : S_i(0) = 1, S_i(1) = 0\}$, individuals who select into the employment spell of interest only if not assigned to treatment.

To address endogenous censoring, we propose to also stratify with respect to the joint potential values of $\{W_i(0), W_i(1)\}$. Contrary to the previous cases, this stratification has the peculiarity that $W_i(z)$ is defined only if $S_i(z) = 1$, for $z = 0, 1$: if $S_i(z) = 1$ then $W_i(z)$ can take the values of 0 or 1, while if $S_i(z) = 0$ then $W_i(z) = *$. Hence, which subpopulations can be defined as a function of the vector $\{W_i(0), W_i(1)\}$ depends on the values of $S_i(z)$. For the always selected subpopulation (EE), we can define the following four subpopulations: $CC = \{i : W_i(0) = W_i(1) = 1\}$, individuals whose spells of interest are always censored regardless of treatment assignment, or always censored; $UU = \{i : W_i(0) = W_i(1) = 0\}$, individuals whose spells of interest are never censored regardless of treatment assignment, or never censored (or "always uncensored"); $UC = \{i : W_i(0) = 0, W_i(1) = 1\}$, individuals whose spells are censored only if assigned to treatment; and $CU = \{i : W_i(0) = 1, W_i(1) = 0\}$, individuals whose spells are censored only if not assigned to treatment. For the NE, EN, and NN groups, $W_i(z)$ is not defined under at least one treatment assignment arm, so we define the following subpopulations. For NE: $*C = \{i : W_i(0) = *, W_i(1) = 1\}$, censored under assignment to treatment; and $*U = \{i : W_i(0) = *, W_i(1) = 0\}$, uncensored under assignment to treatment. For EN: $C* = \{i : W_i(0) = 1, W_i(1) = *\}$, censored under assignment to control; and $U* = \{i : W_i(0) = 0, W_i(1) = *\}$, uncensored under assignment to control. For NN, $W_i(z)$ is undefined under both treatment assignment states, so we assign them the value $*$.

After combining the four compliance types with the subpopulations previously defined in terms of their potential response to selection and censoring, we have that the principal stratification defined by the joint potential values of $\{D_i(0), D_i(1)\} \times \{S_i(0), S_i(1)\} \times \{W_i(0), W_i(1)\}$ consists of the following 36 principal strata: $hEECC$, $hEEUU$, $hEEUC$, $hEECU$, $hNE*C$, $hNE*U$, $hENC*$, $hENU*$, and $hNN**$ for $h = a, n, c, d$. Hence, for instance, the principal stratum $aEECC$ is given by the always takers who always select into the spell of interest and whose outcomes are always censored regardless of treatment assignment.

Since the values of the potential variables $D_i(z)$, $S_i(z)$, and $W_i(z)$ under both values of z for the same individual are unobservable, principal strata are latent subpopulations. Thus, while it is possible to obtain causal effects by comparing individuals within the same principal strata with different values of the randomized treatment assignment (Z), it is not generally possible to observe the stratum each individual belongs to. Hence, in order to identify causal effects, assumptions need to be made, some of which imply that certain principal strata are empty.

2.3 Basic Instrumental Variable Assumptions and Causal Parameters

This and the next subsection present the main assumptions used to partially identify the effects of interest. To simplify notation, in what follows we omit the subscript i unless deemed necessary for clarity. We address noncompliance by using the randomized treatment assignment indicator as an instrumental variable (IV) for treatment receipt. Following Imbens and Angrist (1994) and

Angrist et al. (1996), we impose the following assumptions:

Assumption 1 (Randomized Instrument). $\{D(z), S(z,d), Y^*(z,d), C(z,d)\}$ is independent of Z for all $z, d = 0, 1$.

Assumption 2 (Non-Zero Average Effect of Z on D). $E[D(1) - D(0)] \neq 0$.

Assumption 3 (Exclusion Restriction of Z). $Y_i^*(0,d) = Y_i^*(1,d)$, $S_i(0,d) = S_i(1,d)$, and $C_i(0,d) = C_i(1,d)$ for all i and $d = 0, 1$.

Assumption 4 (Individual-Level Monotonicity of D in Z). $D_i(1) \geq D_i(0)$ for all i.

Assumption 1 states that the IV (Z) is as good as randomly assigned. It holds by design in our empirical illustration, as assignment to JC was randomized. Assumption 2 requires the IV—treatment assignment—to have a non-zero average effect on treatment receipt (D). This is also the case in our application. Assumption 3 requires that any effect of the IV on $S(z,d)$, $Y^*(z,d)$ and $C(z,d)$ (and thus $Y(z,d)$ and $W(z,d)$) is only through treatment receipt. Note that this exclusion restriction is required to hold for three post-randomization variables: the selection indicator, the time from start of the spell of interest until censoring, and the latent duration outcome. In the application, Assumption 3 implies that randomization into JC can only affect the latent employment duration, the indicator for experiencing the employment spell of interest, and the censoring indicator, through its effect on JC participation. Exclusion restriction assumptions have been used in the JC literature to analyze various outcomes, such as employment, earnings, wages, health, crime and welfare receipt (e.g., Schochet et al., 2001; Schochet et al., 2008; Frumento et al., 2012; Chen and Flores, 2015; Chen et al., 2018). The assumption seems also plausible in our context. However, it could fail, for instance, if being embargoed from JC affects individual job search behavior relative to what it would be under having the option to enroll (although this possibility is decreased by our focus on the last employment spell, a longer-term outcome). In general, note that the exclusion restriction may not be innocuous in some applications, even in the context of randomized encouragement designs and social experiments (e.g., Hirano et al., 2000; Flores and Flores-Lagunes, 2013; Chen et al., 2019). Indeed, in the context of natural experiments, this assumption is less likely to hold. Finally, Assumption 4 states that the IV cannot have a negative effect on treatment receipt for any individual. This assumption has also been widely used in the JC literature (e.g., in the papers mentioned above), as it seems improbable that an individual would participate in JC only if denied access to it.

Assumptions 3 and 4 imply that some strata are empty. Given that Z cannot affect S or W through channels other than D, Assumption 3 sets to zero the population proportion of the strata of always takers or never takers that have different potential selection or censoring status under the two treatment assignment arms because for them there is a direct effect of Z on S or W. The twelve strata that are eliminated by Assumption 3 are: $aEEUC$, $aEECU$, $aNE*C$, $aNE*U$, $aENC*$, $aENU*$, $nEEUC$, $nEECU$, $nNE*C$, $nNE*U$, $nENC*$, and $nENU*$. Assumption 4 sets to zero the proportion of defiers, thus eliminating nine of the remaining strata. Thus, under Assumptions 1 to 4, the number of remaining strata is 15.

In settings without sample selection and censoring, Imbens and Angrist (1994) and Angrist et

al. (1996) show that, under Assumptions 1 to 4, IV estimators point identify the average treatment effect for the subpopulation of compliers. In the presence of sample selection and censoring, however, those assumptions are not enough to point identify an effect of D on Y^*. This is the case we analyze next. We begin by defining our parameters of interest. Table 1 shows the relationship between the remaining 15 strata under Assumptions 1 to 4 and the observed values of $\{Z_i, D_i, S_i, W_i\}$. This table makes it clear that the only stratum for which the outcome of interest Y^* is observed under both treatment receipt ($D = 1$) and no treatment receipt ($D = 0$), is the $cEEUU$ stratum (in bold). This implies that the data does not contain information about the causal effects for other principal strata. The $cEEUU$ stratum is comprised of individuals who, regardless of treatment assignment, comply with their assignment and always select into an employment spell of interest that is uncensored.

In the spirit of the principal stratification literature (e.g., Angrist et al., 1996; Frangakis and Rubin, 2002; Zhang et al., 2008, 2009; Imai, 2007, 2008; Lee, 2009; Frumento et al., 2012; Chen and Flores, 2015), we focus on (partial) identification of effects for the $cEEUU$ stratum. The reason is that for other principal strata the latent outcome Y^* is censored or is not well-defined (i.e., $S = 0$) in at least one of the treatment-receipt arms. Even if one were willing to define the outcome when $S = 0$, additional assumptions than those needed for identification of effects for the $cEEUU$ stratum are required to deal with outcomes that are censored, or outcomes that are never observed and are thus "entirely hypothetical" (Rubin, 1990) (e.g., the latent duration of $nNN**$ individuals if we could "force" them to enroll in JC and experience an employment spell). Examples of such assumptions include distributional or independent censoring assumptions to deal with censored outcomes (e.g., Shephered et al., 2007), or assuming the outcome has bounded support to deal with never-observed outcomes (e.g., Huber and Mellace, 2015; Chen et al., 2018). Thus, in this paper we focus on partial identification of causal effects for the stratum for which latent outcomes are well-defined and observed under both treatment-receipt arms (the $cEEUU$ stratum). These results can serve as a building block to derive bounds on the effects for other subpopulations by introducing additional assumptions.

Table 1: Principal Strata under Assumptions 1 to 4 (Employment Duration Outcome)

	$Z=0$				$Z=1$		
	$S=0$	$S=1$			$S=0$	$S=1$	
	$W=*$	$W=0$	$W=1$		$W=*$	$W=0$	$W=1$
$D=0$	$nNN**$ $cNN**$ $cNE*C$ $cNE*U$	$nEEUU$ **cEEUU** $cEEUC$ $cENU*^{\dagger}$	$nEECC$ $cEECC$ $cEECU^{\ddagger}$ $cENC*^{\dagger}$	$D=0$	$nNN**$	$nEEUU$	$nEECC$
$D=1$	$aNN**$	$aEEUU$	$aEECC$	$D=1$	$aNN**$ $cNN**$ $cENC*^{\dagger}$ $cENU*^{\dagger}$	$aEEUU$ **cEEUU** $cEECU^{\ddagger}$ $cNE*U$	$aEECC$ $cEECC$ $cEEUC$ $cNE*C$

†: Strata $cENC*$ and $cENU*$ are ruled out by Monotonic Selection (Assumption 5 in Section 2.4.1).
‡: Stratum $cEECU$ is ruled out by Monotonic Censoring (Assumption 6 in Section 2.4.2).

To formally define our parameters of interest, let $f_{k|z}$ denote the probability distribution function of the *latent* potential outcome $Y^*(z)$ for stratum k, and let its $\alpha-$quantile be given by $q_{k|z}(\alpha) \equiv \inf\{y : F_{k|z}(y) \geq \alpha\}$ for $0 < \alpha < 1$, where $F_{k|z}(y)$ denotes the cumulative distribution function

(CDF) of $Y^*(z)$ for stratum k evaluated at y. The parameters we consider are given by the following stratum-specific average and quantile treatment effects of D on Y^*:

$$SATE_{cEEUU} = E[Y^*(1)|cEEUU] - E[Y^*(0)|cEEUU] \quad (1)$$

$$SQTE^{\alpha}_{cEEUU} = q_{cEEUU|1}(\alpha) - q_{cEEUU|0}(\alpha). \quad (2)$$

In our application, $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ give, respectively, the average and quantile treatment effects of participation in JC on the duration of the last complete employment spell by week 208 after randomization (our spell of interest) for the $cEEUU$ stratum. This stratum is given by those individuals who complied with their treatment assignment and experienced an uncensored last employment spell by the end of week 208 after randomization, regardless of treatment *receipt* (since for compliers $Z = D$). Hence, $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ can also be interpreted, respectively, as the average and quantile treatment effects of participation in JC on the duration of the last *complete* employment spell by week 208 after randomization for the $cEEUU$ stratum. We note that in the definition of $SQTE^{\alpha}_{cEEUU}$ we compare, for the $cEEUU$ stratum, the unconditional $\alpha-$quantile of the outcome's marginal distributions under each of the treatment-receipt arms.

2.4 Partial Identification Results

It is possible to derive bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ under only Assumptions 1 to 4; however, these bounds are typically wide and uninformative in practice. This is also the case when constructing bounds in other settings; for example, when addressing sample selection using only the random assignment assumption (e.g., Zhang and Rubin, 2003; Lee, 2009; Blanco et al., 2013a; Huber and Mellace, 2015), or when addressing sample selection and noncompliance using only the standard IV assumptions in Angrist et al. (1996) (e.g., Imai, 2007). Therefore, the partial identification results below make use of additional monotonicity and stochastic dominance assumptions to obtain tighter bounds beyond using the information contained in the data under only Assumptions 1 to 4.

2.4.1 Bounds under Monotonic Selection

In this and the following subsection, we construct bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ under two additional monotonicity assumptions similar to Assumption 4. Individual-level monotonicity assumptions are commonly used in the partial identification literature (e.g., Manski and Pepper, 2000; Zhang et al., 2008; Imai, 2007, 2008; Lee, 2009; Chen and Flores, 2015). However, they are not innocuous, and their plausibility in a particular application has to be judged carefully. The first additional monotonicity assumption we consider is:

Assumption 5 (Individual-Level Monotonicity of S in D for c). $S_i(1) \geq S_i(0)$ for all compliers.

Assumption 5 states that there is no individual complier who has a negative effect of treatment receipt (D) on the selection into the spell of interest (S). A testable implication of adding As-

sumption 5 is that $E[S(1) - S(0)|c] \geq 0$, i.e., the IV estimate of D on S using Z as an IV for D is non-negative. In the context of addressing sample selection, this assumption is similar to that used in Chen and Flores (2015), but differs from a similar assumption used by Imai (2007) and others (e.g., Zhang et al., 2008; Lee, 2009) in that it is imposed only for compliers (as opposed to all individuals) and on treatment receipt (as opposed to treatment assignment). In the context of our application, it states that there is a non-negative effect of participation in JC on the probability of experiencing an employment spell by week 208 after randomization for every complier. In other words, it assumes no individual complier would be continually unemployed from baseline to week 208 after randomization if she enrolled in JC, and would experience an employment spell in that same period if she did not enroll. One potential threat to this assumption relates to the locking-in effect described in Section 2.1, which precludes initiating employment spells for those receiving training. Another potential threat consists of the possibility that JC participants raise their reservation wage, declining job offers that would start employment spells. However, we believe that the relatively long period between randomization and the end of the follow-up observation in our application (208 weeks or 4 years), lessens the plausibility of those potential threats. Similar assumptions have been used in the JC literature (e.g., Lee, 2009; Blanco et al., 2013a, 2013b; Chen and Flores, 2015).

Assumption 5 sets the population proportions of the strata $cENC*$ and $cENU*$ to zero. The distribution of the remaining 13 principal strata under Assumptions 1 to 5 in the observed data $\{D_i, S_i, W_i, Z_i\}$ can be seen in Table 1 (see note to table). We start by analyzing identification of the stratum proportions, in particular, that of our stratum of interest ($cEEUU$), which will be partially identified. Let π_k denote the proportion of stratum k in the population. Also, let $p_{dsw|z} \equiv \Pr(D = d, S = s, W = w|Z = z)$, where $d, s, z \in \{0, 1\}$ and $w \in \{0, 1, *\}$, and let $f_{dsw|z}$ denote the conditional distribution of the *observed* outcome Y in the cell $\{D = d, S = s, W = w, Z = z\}$. Note that $p_{dsw|z}$ and $f_{dsw|z}$ are observed in the data. Under Assumptions 1 to 5, we can point identify the six stratum proportions for always takers and never takers, as well as the stratum proportion for $cNN**$, as (e.g., see Table 1):

$$\pi_{aNN**} = p_{10*|0} \qquad \pi_{aEEUU} = p_{110|0} \qquad \pi_{aEECC} = p_{111|0} \tag{3}$$

$$\pi_{nNN**} = p_{00*|1} \qquad \pi_{nEEUU} = p_{010|1} \qquad \pi_{nEECC} = p_{011|1} \tag{4}$$

$$\pi_{cNN**} = p_{10*|1} - p_{10*|0} \tag{5}$$

In addition, under the current assumptions, we can obtain five equations relating the proportions of the other strata of compliers to observed conditional probabilities and other point-identified stratum proportions:

$$\pi_{cNE*C} + \pi_{cNE*U} = p_{00*|0} - p_{00*|1} - p_{10*|1} + p_{10*|0} \tag{6}$$

$$\pi_{cEEUU} + \pi_{cEEUC} = p_{010|0} - p_{010|1} \tag{7}$$

$$\pi_{cEECC} + \pi_{cEECU} = p_{011|0} - p_{011|1} \tag{8}$$

$$\pi_{cEEUU} + \pi_{cEECU} + \pi_{cNE*U} = p_{110|1} - p_{110|0} \tag{9}$$

$$\pi_{cEECC} + \pi_{cEEUC} + \pi_{cNE*C} = p_{111|1} - p_{111|0} \tag{10}$$

Although the remaining six stratum proportions are not point identified, it is possible to construct bounds on the proportion of our stratum of interest, π_{cEEUU}. The equations above imply that the possible range of π_{cEEUU} under Assumptions 1 to 5 is: $\pi_{cEEUU} \in \Pi_1 \equiv [\max\{0, p_{111|0} + p_{010|0} - p_{010|1} - p_{111|1}\}, \min\{p_{010|0} - p_{010|1}, p_{110|1} - p_{110|0}\}]$. Given that all seven strata of compliers can be non-empty, equations (5) to (10) provide six testable implications: that all of their proportions should be non-negative, with at least one being strictly positive by Assumption 2.

We now proceed to construct bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ in (1) and (2), respectively. Our bounds are derived following the approach in Horowitz and Manski (1995) and Imai (2007, 2008). We first consider partial identification of the distribution of $Y^*(1)$ and $Y^*(0)$ for the $cEEUU$ stratum ($f_{cEEUU|1}$ and $f_{cEEUU|0}$, respectively), from which $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ can be bounded. From Table 1, under Assumptions 1 to 5 the distributions of the observed outcomes Y in the cells $\{D = 1, S = 1, W = 0, Z = 1\}$ and $\{D = 0, S = 1, W = 0, Z = 0\}$ ($f_{110|1}$ and $f_{010|0}$, respectively) are weighted averages of the distributions $f_{cEEUU|1}$ and $f_{cEEUU|0}$, respectively, and the distributions of the latent potential outcomes of other strata. Specifically:

$$
\begin{aligned}
(11) \qquad f_{110|1} &= \frac{\pi_{aEEUU}}{p_{110|1}} f_{aEEUU|1} + \frac{\pi_{cEEUU}}{p_{110|1}} f_{cEEUU|1} + \frac{\pi_{cEECU}}{p_{110|1}} f_{cEECU|1} \\
&\quad + \frac{\pi_{cNE*U}}{p_{110|1}} f_{cNE*U|1} \\
(12) \qquad f_{010|0} &= \frac{\pi_{nEEUU}}{p_{010|0}} f_{nEEUU|0} + \frac{\pi_{cEEUU}}{p_{010|0}} f_{cEEUU|0} + \frac{\pi_{cEEUC}}{p_{010|0}} f_{cEEUC|0},
\end{aligned}
$$

where $p_{110|1} = \pi_{aEEUU} + \pi_{cEEUU} + \pi_{cEECU} + \pi_{cNE*U}$ and $p_{010|0} = \pi_{nEEUU} + \pi_{cEEUU} + \pi_{cEEUC}$. The distributions of the observed outcomes above are mixtures of more strata than those corresponding to the cells of interest in cases with only sample selection (e.g., Zhang and Rubin, 2003; Lee, 2009; Imai, 2008) or sample selection and noncompliance (e.g., Imai, 2007; Chen and Flores, 2015), as would be expected given the additional complication we address (censoring). From the right hand side of equations (11) and (12), only the distributions of latent potential outcomes of $aEEUU$ and $nEEUU$ are point identified: $f_{aEEUU|1} = f_{110|0}$ and $f_{nEEUU|0} = f_{010|1}$. Therefore, without further assumptions, it is not possible to point identify $f_{cEEUU|1}$ and $f_{cEEUU|0}$, and thus any of the terms in $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$. However, it is possible to bound them.

Using the point-identified $f_{aEEUU|1} = f_{110|0}$ and $f_{nEEUU|0} = f_{010|1}$, along with equations (3), (4), (7) and (10), which use themselves the point-identified stratum proportions, we can rewrite equations (11) and (12) as:

$$
\begin{aligned}
(13) \qquad f_1 &= \frac{\pi_{cEEUU}}{p_{110|1} - p_{110|0}} f_{cEEUU|1} + \frac{\pi_{cEECU}}{p_{110|1} - p_{110|0}} f_{cEECU|1} + \frac{\pi_{cNE*U}}{p_{110|1} - p_{110|0}} f_{cNE*U|1} \\
(14) \qquad f_0 &= \frac{\pi_{cEEUU}}{p_{010|0} - p_{010|1}} f_{cEEUU|0} + \frac{\pi_{cEEUC}}{p_{010|0} - p_{010|1}} f_{cEEUC|0},
\end{aligned}
$$

where

$$f_1 \equiv \frac{f_{110|1}p_{110|1} - f_{110|0}p_{110|0}}{p_{110|1} - p_{110|0}} \tag{15}$$

$$f_0 \equiv \frac{f_{010|0}p_{010|0} - f_{010|1}p_{010|1}}{p_{010|0} - p_{010|1}}. \tag{16}$$

Note that equations (13) and (14) provide the additional testable implications that f_1 and f_0 must be non-negative and integrate to one, as they are valid probability distributions.

To provide the intuition for the trimming bounds we derive, consider partially identifying $f_{cEEUU|1}$ from (13). If π_{cEEUU} were point identified, the upper (lower) bound for $f_{cEEUU|1}$ would be given by the best-case (worst-case) scenario in which the largest (smallest) $\pi_{cEEUU}/(p_{110|1} - p_{110|0})$ fraction of values of Y in the probability distribution f_1 belong to the $cEEUU$ stratum. Analogous bounds for $f_{cEEUU|0}$ could be derived using equation (14). After this, the upper (lower) bound on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ would be obtained by using the bounds on $f_{cEEUU|1}$ to construct the upper (lower) bound for the first term of $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ in equations (1) and (2), and subtracting the lower (upper) bound for the second term in equations (1) and (2) constructed using the bounds on $f_{cEEUU|0}$.

Since π_{cEEUU} is not point identified, however, we use the bounds on π_{cEEUU} obtained before, $\Pi_1 = [\max\{0, p_{111|0} + p_{010|0} - p_{010|1} - p_{111|1}\}, \min\{p_{010|0} - p_{010|1}, p_{110|1} - p_{110|0}\}]$. To obtain the upper (lower) bound for $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$, we need to write the upper (lower) trimming bound previously described as a function of π_{cEEUU}, and then maximize (minimize) it with respect to π_{cEEUU} over Π_1. In a setting with only sample selection, Huber and Mellace (2015) considered trimming bounds similar to those above, and showed that numerical optimization is not necessary because the bounds are monotonic functions of the trimming proportions. Their conclusion also applies to the trimming bounds we derive here: the trimming bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ are obtained when π_{cEEUU} reaches its minimal positive value in Π_1, $p_{111|0} + p_{010|0} - p_{010|1} - p_{111|1}$, as shown in the Internet Appendix.

To present the bounds, let the $\alpha-$quantile of f_1 and f_0 in equations (13) and (14) be given by $r_z(\alpha) \equiv \inf\{y : F_z(y) \geq \alpha\}$ for $0 < \alpha < 1$ and $z = 0, 1$. Also, define the distributions $L_{\gamma|z}$ and $U_{\gamma|z}$ as follows:

$$L_{\gamma|z}[-\infty, y] \equiv \begin{cases} \frac{F_z(y)}{\gamma}, \text{ if } y < r_z(\gamma) \\ 1 \qquad\qquad , \text{ if } y \geq r_z(\gamma) \end{cases}$$

$$U_{\gamma|z}[-\infty, y] \equiv \begin{cases} 0 \qquad\qquad\qquad , \text{ if } y < r_z(1-\gamma) \\ \frac{F_z(y)+\gamma-1}{\gamma}, \text{ if } y \geq r_z(1-\gamma) \end{cases},$$

where $\gamma \in (0, 1)$ is the trimming proportion used to construct bounds on $f_{cEEUU|1}$ or $f_{cEEUU|0}$, as described above.

Proposition 1 Suppose $\tau = p_{111|0} + p_{010|0} - p_{010|1} - p_{111|1} > 0$. Under Assumptions 1 through 5, the bounds $LB^{1A} \leq SATE_{cEEUU} \leq UB^{1A}$, and $LB^{1Q}_{\alpha} \leq SQTE^{\alpha}_{cEEUU} \leq UB^{1Q}_{\alpha}$ are sharp,

where:

$$
\begin{aligned}
LB^{1A} &= \int y dL_{\tau/(p_{110|1}-p_{110|0})|1} - \int y dU_{\tau/(p_{010|0}-p_{010|1})|0} \\
UB^{1A} &= \int y dU_{\tau/(p_{110|1}-p_{110|0})|1} - \int y dL_{\tau/(p_{010|0}-p_{010|1})|0} \\
LB^{1Q}_{\alpha} &= r_1 \left(\frac{\alpha\tau}{p_{110|1}-p_{110|0}} \right) - r_0 \left(1 - \frac{(1-\alpha)\tau}{p_{010|0}-p_{010|1}} \right) \\
UB^{1Q}_{\alpha} &= r_1 \left(1 - \frac{(1-\alpha)\tau}{p_{110|1}-p_{110|0}} \right) - r_0 \left(\frac{\alpha\tau}{p_{010|0}-p_{010|1}} \right).
\end{aligned}
$$

Proof. See Internet Appendix.

2.4.2 Bounds under Monotonic Censoring

In this subsection we add the following assumption.

Assumption 6 (Individual-Level Monotonicity of W in D for cEE). $W_i(1) \geq W_i(0)$ for all $i \in cEE$.

This assumption states that there is a non-negative individual effect of treatment receipt on censoring for the cEE subpopulation. That is, no individual who is an always-selected complier would have a censored latent outcome under no treatment receipt and an uncensored latent outcome under treatment receipt. Hence, Assumption 6 implies that the $cEECU$ stratum is empty. A testable implication of Assumption 6 is that $E[W(1) - W(0)|cEE] \geq 0$. While this conditional expectation is not point identified because of the presence of sample selection and non-compliance, bounds can be constructed on it under Assumptions 1 to 5 as in Imai (2007) or Chen and Flores (2015) to falsify Assumption 6. If such estimated bounds imply statistically negative effects of D on W for the cEE, this would be strong evidence against Assumption 6.

In our application, this assumption states that there are no always-selected compliers who would have a censored last employment spell by week 208 after randomization if they did not participate in JC and would have an uncensored last employment spell by week 208 after randomization if they participated in JC. Recall that always-selected compliers are those individuals who, regardless of treatment assignment, always comply with their assignment and experience an employment spell by week 208 after randomization. A justification for this assumption in our application comes from the insight that, for always-selected compliers, whether or not the duration of their last employment spell by week 208 after randomization is censored equals their employment status at week 208 after randomization. If they are employed at week 208 after randomization, then the duration of their last employment spell is censored, whereas the duration of the last employment spell is uncensored if they are unemployed at week 208 after randomization. Thus, Assumption 6 implies a positive effect of participation in JC on employment at week 208 after randomization for the always-selected compliers. Consequently, the plausibility of Assumption 6 in this application is threatened by at least the same factors discussed in the context of Assumption 5: the lock-in effect and the possibility of increased reservation wage by participants. As before, we believe that

the relatively long follow-up period of 208 week works towards decreasing the concerns of these particular factors. Assumptions imposing a non-negative individual effect of JC assignment or participation on employment at week 208 after randomization have been previously used when estimating bounds for the wage effects of JC (Lee, 2009; Blanco et al., 2013a, 2013b; Chen and Flores, 2015). While those assumptions are required to hold for all individuals or for compliers, Assumption 6 is required to hold only for a subset of the compliers (the cEE).

Under Assumptions 1 to 6, we have 12 strata left. Their relationship to the observed values of $\{D_i, S_i, W_i, Z_i\}$ can be seen in Table 1 (see note to table). Under these assumptions, we can further point identify $\pi_{cEECC} = p_{011|0} - p_{011|1}$, and equations (9) and (10) become:

$$\pi_{cEEUU} + \pi_{cNE*U} = p_{110|1} - p_{110|0} \tag{17}$$

$$\pi_{cEEUC} + \pi_{cNE*C} = p_{111|1} - p_{111|0} - p_{011|0} + p_{011|1} \tag{18}$$

Adding Assumption 6 provides one more testable implication: $p_{111|1} - p_{111|0} - p_{011|0} + p_{011|1} \geq 0$, from equation (18). The bounds on the stratum proportion π_{cEEUU} are given by: $\pi_{cEEUU} \in \Pi_2 \equiv [\max\{0, p_{111|0} + p_{010|0} + p_{011|0} - p_{010|1} - p_{111|1} - p_{011|1}\}, \min\{p_{010|0} - p_{010|1}, p_{110|1} - p_{110|0}\}]$.

The derivation of the bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ follows the same approach as in Section 2.4.1. Under Assumption 6, equations (11) and (13) are simplified because $\pi_{cEECU} = 0$. Equation (13), relating $f_{cEEUU|1}$ to the data and point-identified objects, now becomes:

$$f_1 = \frac{\pi_{cEEUU}}{p_{110|1} - p_{110|0}} f_{cEEUU|1} + \frac{\pi_{cNE*U}}{p_{110|1} - p_{110|0}} f_{cNE*U|1}. \tag{19}$$

As before, trimming bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ are obtained when π_{cEEUU} reaches its minimal positive value in Π_2.

Proposition 2 Suppose $\tau = p_{111|0} + p_{010|0} + p_{011|0} - p_{010|1} - p_{111|1} - p_{011|1} > 0$. Under Assumptions 1 through 6, the bounds $LB^{2A} \leq SATE_{cEEUU} \leq UB^{2A}$, and $LB^{2Q}_{\alpha} \leq SQTE^{\alpha}_{cEEUU} \leq UB^{2Q}_{\alpha}$ are sharp, where:

$$\begin{aligned}
LB^{2A} &= \int y dL_{\tau/(p_{110|1} - p_{110|0})|1} - \int y dU_{\tau/(p_{010|0} - p_{010|1})|0} \\
UB^{2A} &= \int y dU_{\tau/(p_{110|1} - p_{110|0})|1} - \int y dL_{\tau/(p_{010|0} - p_{010|1})|0} \\
LB^{2Q}_{\alpha} &= r_1 \left(\frac{\alpha\tau}{p_{110|1} - p_{110|0}} \right) - r_0 \left(1 - \frac{(1-\alpha)\tau}{p_{010|0} - p_{010|1}} \right) \\
UB^{2Q}_{\alpha} &= r_1 \left(1 - \frac{(1-\alpha)\tau}{p_{110|1} - p_{110|0}} \right) - r_0 \left(\frac{\alpha\tau}{p_{010|0} - p_{010|1}} \right).
\end{aligned}$$

Proof. See Internet Appendix.

The expressions of the bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ in Proposition 2 are similar to those in Proposition 1, with the difference being the trimming proportion. Adding Assumption 6 provides identifying information by increasing the (non-zero) lower bounding function of π_{cEEUU} from $(p_{111|0} + p_{010|0} - p_{010|1} - p_{111|1})$ to $(p_{111|0} + p_{010|0} + p_{011|0} - p_{010|1} - p_{111|1} - p_{011|1})$ (the difference

equals $\pi_{cEECC} = p_{011|0} - p_{011|1} \geq 0$). Since the bounds in Proposition 2 are obtained when π_{cEEUU} reaches a larger minimal positive value, they are tighter than those in Proposition 1.

It is important to note that, contrary to the cases previously considered in the literature under sample selection (e.g., Zhang and Rubin, 2003; Lee, 2009) and under sample selection and non-compliance (e.g., Imai, 2007; Chen and Flores, 2015), under the additional censoring complication the stratum proportions are not point identified even after imposing individual-level monotonicity of treatment receipt on the three corresponding post-randomization variables. It is known that trimming bounds constructed in settings where the stratum proportions are not point identified are generally wider than those in which all the stratum proportions are point identified, sometimes resulting in uninformative bounds in practice (e.g., Blanco et al., 2013a; Huber and Mellace, 2015). In our setting, point identification of all stratum proportions would require making an additional stratum proportion zero (out of the 12 strata under Assumptions 1 to 6). In the context of the empirical illustration, however, such an assumption would seem strong and thus we do not pursue this strategy. As an alternative, we consider a stochastic dominance assumption.

2.4.3 Bounds under Stochastic Dominance

Assumptions imposing weak inequality of means or CDFs of potential outcomes across different subpopulations are common in the partial identification literature (e.g., Manski and Pepper, 2000; Zhang and Rubin, 2003; Blundell et al., 2007; Bhattacharya et al., 2012; Flores and Flores-Lagunes, 2010, 2013; Frandsen and Lefgren, 2016; Chen et al., 2018). The intuition behind such assumptions is that some subpopulations are likely to have characteristics that make them more likely to have higher potential outcomes than others. We consider such type of assumptions where, following the principal stratification literature, those subpopulations are given by strata (e.g., Zhang and Rubin, 2003; Imai, 2007, 2008; Zhang et al., 2008; Blanco et al., 2013a, 2013b; Chen and Flores, 2015).

Assumption 7 (Stochastic Dominance). (a) $F_{cEEUU|0}(y) \geq F_{cEEUC|0}(y)$ for all $y \in \mathcal{Y}$; (b) $F_{cEEUU|1}(y) \leq F_{cNE*U|1}(y)$ for all $y \in \mathcal{Y}$.

Assumption 7(a) states that the distribution of the potential latent outcome $Y^*(0)$ for the $cEEUC$ stratum weakly stochastically dominates that of the $cEEUU$ stratum, whereas Assumption 7(b) requires that the distribution of the potential latent outcome $Y^*(1)$ for the $cEEUU$ stratum weakly stochastically dominates that of the $cNE*U$ stratum. In other words, Assumption 7(a) (7(b), respectively) implies that the latent potential outcomes $Y^*(0)$ ($Y^*(1)$) of the $cEEUU$ stratum at any rank of their distribution is less (greater) than or equal to those of the $cEEUC$ ($cNE*U$) stratum. For partial identification of $SATE_{cEEUU}$ only a less-restrictive weak mean dominance version of Assumption 7 is needed: (a) $E[Y^*(0)|cEEUU] \leq E[Y^*(0)|cEEUC]$; (b) $E[Y^*(1)|cEEUU] \geq E[Y^*(1)|cNE*U]$. Since we are also considering identification of $SQTE^{\alpha}_{cEEUU}$, in what follows, we keep the stronger version for simplicity.

In our application, Assumption 7(a) states that the distribution of the potential latent employment spell durations under no participation in JC ($Y^*(0)$) of the always-selected compliers whose last employment spell by week 208 after randomization is censored only under JC participation

($cEEUC$ stratum) weakly stochastically dominates that of the always-selected-and-uncensored compliers ($cEEUU$). Informally, it implies that, under no participation in JC, $cEEUU$ members generally have shorter last employment spells by week 208 after randomization than $cEEUC$ members. The intuition behind this assumption is based on the equality of the censoring indicator and the employment indicator at week 208 after randomization for these strata. Since $cEEUU$ individuals are unemployed at week 208 after randomization regardless of JC participation, while $cEEUC$ individuals are employed at week 208 under JC participation, we postulate that the latter have characteristics that would lead them to have better labor market outcomes than the former, including longer employment duration spells. Similarly, Assumption 7(b) requires that the distribution of the potential latent employment duration under participation in JC ($Y^*(1)$) of the always-selected-and-uncensored compliers ($cEEUU$) weakly stochastically dominates that of the compliers who, under no participation in JC, do not select into a last employment spell and, under participation in JC, select into an uncensored last employment spell ($cNE * U$). Informally, it implies that, under participation in JC, $cEEUU$ members generally have longer last employment spells than $cNE * U$ members. Under treatment, both strata experience a last employment spell between randomization and week 208, and are unemployed at week 208 (since their outcome is uncensored). However, under no treatment, $cNE * U$ individuals are continuously unemployed during the same time frame, while $cEEUU$ experience at least one employment spell. Hence, we postulate that the characteristics of the $cEEUU$ stratum are such that they are more likely to have better labor market outcomes and thus higher latent potential outcomes under participation in JC than the $cNE * U$ stratum.

Based on their implications, both parts of Assumption 7 lead to a positive correlation between the employment status at week 208 after randomization and the duration of the last employment spell. Such a positive correlation between employment and its duration is supported by standard economic models of the labor market (e.g., Hall, 1979; Caliendo et al., 2013). Still, it can be seen based on the previous intuition that similar potential threats (i.e., the locking-in effect and reservation wages) to those casting shadow on Assumption 5 and Assumption 6 can affect Assumption 7. In a similar way, though, the relatively long time between randomization and week 208 works towards potentially eliminating those threats. Also, related assumptions have been considered in the JC literature. For example, in the context of analyzing the wage effects of JC, Blanco et al. (2013a, 2013b) assume that the potential wages under treatment of the always-employed individuals at week 208 after randomization stochastically dominate those of the individuals who would be employed only under assignment to JC, while Chen and Flores (2015) impose an analogous mean dominance assumption within the complier subpopulation.

Assumptions like Assumption 7 are not directly testable since the distributions are not point identified. To gather indirect evidence about their plausibility, the literature has suggested analyzing baseline characteristics of the strata that are closely related to the outcome of interest such as pre-treatment outcomes (e.g., Flores and Flores-Lagunes, 2010, 2013; Blanco et al., 2013a, 2013b; Chen and Flores, 2015). The idea is to assess whether some strata indeed appear to have more favorable characteristics than others, so that it is plausible that those strata have better outcomes. In our current setting, however, this analysis is not possible because the corresponding

stratum proportions are not point identified, and thus neither are their baseline characteristics. Alternatively, note that the implicit ranking of the three strata underlying Assumption 7 presumes that the $cEEUC$ stratum has the most favorable labor market outcomes and average baseline characteristics, whereas the $cNE*U$ stratum has the worst and the $cEEUU$ stratum lies in the middle. This implies that if Assumption 7 holds, we would expect the baseline characteristics of the subpopulation given by the set $\{cEEUU, cEEUC\}$ to be more favorable than those of the subpopulation given by the set $\{cEEUU, cNE*U\}$. Those characteristics are point identified under Assumptions 1 to 6 from equations analogous to those in (14) and (19). Assumption 7 would be less likely to hold if the baseline characteristics of the $\{cEEUU, cNE*U\}$ subpopulation are more favorable than those of the $\{cEEUU, cEEUC\}$ subpopulation.

Proposition 3 presents bounds on $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ under Assumptions 1 to 7.

Proposition 3 Suppose $\tau = p_{111|0} + p_{010|0} + p_{011|0} - p_{010|1} - p_{111|1} - p_{011|1} > 0$, and let $\overline{Y}^{zdsw} = E[Y|Z = z, D = d, S = s, W = w]$. Under Assumptions 1 through 7, the bounds $LB^{3A} \leq SATE_{cEEUU} \leq UB^{3A}$, and $LB^{3Q}_{\alpha} \leq SQTE^{\alpha}_{cEEUU} \leq UB^{3Q}_{\alpha}$ are sharp, where:

$$
\begin{aligned}
LB^{3A} &= \frac{p_{110|1}\overline{Y}^{1110} - p_{110|0}\overline{Y}^{0110}}{p_{110|1} - p_{110|0}} - \frac{p_{010|0}\overline{Y}^{0010} - p_{010|1}\overline{Y}^{1010}}{p_{010|0} - p_{010|1}} \\
UB^{3A} &= \int y dU_{\tau/(p_{110|1}-p_{110|0})|1} - \int y dL_{\tau/(p_{010|0}-p_{010|1})|0} \\
LB^{3Q}_{\alpha} &= r_1(\alpha) - r_0(\alpha) \\
UB^{3Q}_{\alpha} &= r_1\left(1 - \frac{(1-\alpha)\tau}{p_{110|1} - p_{110|0}}\right) - r_0\left(\frac{\alpha\tau}{p_{010|0} - p_{010|1}}\right),
\end{aligned}
$$

Proof. See Internet Appendix.

Compared to the bounds in Proposition 2, Assumption 7 provides identifying information that tightens the lower bounds of $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$. Consider bounding $SQTE^{\alpha}_{cEEUU}$. By integrating both sides of equations (14) and (19), it is clear that both equations hold with the probability distribution functions being replaced by their corresponding CDFs. Using the CDF version of equation (14) along with Assumption 7(a), we obtain $F_{cEEUU|0}(y) \geq F_0(y)$, which implies an upper bound for the $\alpha-$quantile of $Y^*(0)$ for the $cEEUU$ stratum. Similarly, the CDF version of equation (19) combined with Assumption 7(b) yields $F_{cEEUU|1}(y) \leq F_1(y)$, which implies a lower bound for the $\alpha-$quantile of $Y^*(1)$ for the $cEEUU$ stratum. A lower bound for $SQTE^{\alpha}_{cEEUU}$ is obtained by combining those two bounds; it equals the $\alpha-$quantile of the distribution f_1 in (15) minus the $\alpha-$quantile of the distribution f_0 in (16). The bounds for $SATE_{cEEUU}$ are obtained in an analogous way, with the resulting lower bound equal to the mean of the distribution f_1 minus the mean of the distribution f_0. In this case, the lower bounds for $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ do not depend on τ.

2.4.4 Estimation and Inference

We estimate the bounds derived in Propositions 1 to 3 using their sample analogs. Specifically, we employ the following procedure to construct estimates of our bounds, after checking that τ from those propositions is strictly positive. First, we use the indicator function $1(Y_i \leq \tilde{y})$, and M different values of $\tilde{y}$ spanning the support of the observed outcome, to estimate the conditional CDF of the observed outcome Y_i evaluated at the different values of $\tilde{y}$, $\hat{F}_{dsw|z}(\tilde{y})$. In particular, we are interested in the sample analog CDF versions of $f_{110|1}, f_{110|0}, f_{010|0}$, and $f_{010|1}$ appearing in f_1 in (15) and f_0 in (16). Second, we estimate the conditional probabilities $p_{dsw|z}$ using their sample analogs, $\hat{p}_{dsw|z}$, e.g., $\hat{p}_{110|1} = [\sum_{i=1}^{n} D_i \cdot S_i \cdot (1 - W_i) \cdot Z_i] \,/\, [\sum_{i=1}^{n}(Z_i)]$. Third, the estimates of the conditional CDFs ($\hat{F}_{dsw|z}(\tilde{y})$) and conditional probabilities ($\hat{p}_{dsw|z}$) are plugged into the CDF version of equations (15) and (16) to construct estimates $\hat{F}_1(\tilde{y})$ and $\hat{F}_0(\tilde{y})$, respectively. Next, the estimates $\hat{F}_1(\tilde{y})$, $\hat{F}_0(\tilde{y})$, and $\hat{p}_{dsw|z}$ are used to compute the expected values and α-quantiles (obtained by inverting the corresponding CDFs) needed to estimate the bounds in Propositions 1 to 3.

To undertake inference, standard errors are obtained via bootstrapping, while confidence intervals that contain the true value of the parameter with a given probability can be constructed following Imbens and Manski (2004; IM hereafter). Sample analog estimators of the type of trimming bounds we consider are known to be consistent and asymptotically normally distributed, as they are smooth functions of conditional probabilities, means, quantiles, and trimmed means and quantiles. Based on standard results in Newey and McFadden (1994), Lee (2009) provided asymptotic normality results for sample analog estimators of trimming bounds that deal with sample selection under similar monotonicity assumptions. The asymptotic normality of our bounds, combined with the fact that, by construction, our upper bounds are always larger than our lower bounds, justifies the use of Imbens and Manski's (2004) confidence intervals by Lemma 3 of Stoye (2009). Lastly, we note that the proposed inference based on the estimated bounds on quantile treatment effects is pointwise, that is, for the corresponding effect on the α-quantile specified. As a consequence, statistical statements about the quantile treatment effect function cannot be directly made due to the problem of testing multiple hypothesis. Alternatively, uniform inference on the entire quantile treatment effect function could be undertaken, for example, by extending the approach of Chernozhukov et al. (2013), which is not pursued here.

2.5 Remarks

Remark 1. As a by-product of our approach, it is possible to bound the effect of treatment receipt on other characteristics of the spells of interest (besides duration) in the presence of selection and noncompliance, for the *same* subpopulation we estimated the effects on duration (the *cEEUU* stratum). This is done by simply replacing the outcome variable in the bounds in Propositions 1 to 3. As an example, in the next section we also consider the effect of JC on the log wages paid in those same employment spells for the *cEEUU* stratum, which provides a more complete picture of the effects of JC on those employment spells.

Remark 2. A setting that recently has received attention in the literature is one with a censored outcome, an endogenous binary treatment, and a binary instrumental variable to address treatment endogeneity (Frandsen, 2015; Sant'Anna 2016). This setting is the same as ours in the absence of selection into the spells of interest, i.e., when $S_i(0) = S_i(1) = 1$ for all i. We present bounds for this case in the Internet Appendix. In the setting with only censoring and noncompliance (or treatment endogeneity), Frandsen (2015) and Sant'Anna (2016) consider point identification of treatment effects employing conditional independent censoring assumptions. These assumptions, however, may not hold in relevant applications (as illustrated by Frandsen, 2018). Hence, our bounds for this setting offer a possible alternative for applications in which those assumptions fail and point identification is lost. Other important differences are as follows. The parameter estimated in those papers differs from the one we consider. Frandsen (2015) and Sant'Anna (2016) consider treatment effects for all compliers, while the bounds presented in the Internet Appendix for this setting bound treatment effects for the always-uncensored compliers (i.e., those compliers whose outcomes are uncensored regardless of the value of the instrument). Also, Frandsen's (2015) approach imposes support restrictions that prevent estimation of average treatment effects and quantile treatment effects at all quantiles of the outcome distribution for compliers. Thus, Frandsen's (2015) approach estimates quantile treatment effects for all compliers for a subset of all quantiles, while our bounds are for quantile treatment effects on all quantiles (as well as average effects) for a subset of all compliers. Still, as discussed in Section 2.3, it is possible to construct bounds for the average and quantile treatment effects for all compliers (as well as other subpopulations) if one is willing to define "entirely hypothetical" latent durations for some individuals and impose further assumptions.

Remark 3. As more post-treatment complications are added, estimation and inference is based on finer cells defined by the observed values of treatment assignment (Z) and the corresponding post-treatment variables (e.g., D, S, W). This negatively affects the precision of the estimates. For instance, when only selection is present, trimming bounds for the mean of $Y^*(1)$ for always-selected individuals are constructed based on the cell $\{Z = 1, S = 1\}$; in our case with three complications, trimming bounds for the mean of $Y^*(1)$ for always-selected-and-uncensored compliers are constructed based on the finer cell $\{Z = 1, D = 1, S = 1, W = 0\}$. Similarly, note that the greater the censoring is, the smaller the number of observations in those cells. This is a reflection of the challenging nature of the problem at hand rather than of the approach used, and in some applications it may point to the need to impose more assumptions.

Remark 4. The general approach followed to construct bounds on employment durations can be followed to construct bounds on unemployment durations. The resulting bounds are different from those on employment durations since the direction of the monotonicity assumptions change in accordance to the nature of the unemployment spells. For example, the direction of the monotonicity assumption of the treatment on the censoring indicator of the unemployment spell is reversed relative to the case of employment spells. The ability to choose the direction of the monotonicity and stochastic dominance assumptions to match the requirements of a particular application is a feature of our approach. The bounds on unemployment duration, a discussion of the corresponding assumptions, and their illustration in the context of JC are contained in the Internet Appendix.

Remark 5. Trimming bounds, such as those developed here, can be narrowed using a pre-

randomization covariate that is correlated with the outcome of interest. This is done by obtaining the bounds within sub-samples defined based on the pre-randomization covariate and subsequently averaging the bounds across those sub-samples. The reason narrowing occurs relates to a property of means of truncated distributions. For details, including how this can be implemented, see Lee (2009) and Blanco et al. (2013a).

Remark 6. In some instances, it may be possible to interpret the lower bound for $SATE_{cEEUU}$ and $SQTE^{\alpha}_{cEEUU}$ in Proposition 3 as the lower bound for a different parameter in the absence of a combination of Assumptions 5, 6 and 7, following arguments similar to those in Remark 2 in Chen and Flores (2015). For example, consider bounding $SATE_{cEEUU}$. In this case, in the absence of Assumptions 5 and 7, the lower bound for $SATE_{cEEUU}$ in Proposition 3 can be interpreted as the lower bound for $\beta = E[Y^*(1) - Y^*(0)|cEEUU, cNE * U]$—the average effect of D on Y^* for the $cEEUU$ and $cNE * U$ strata, or the compliers who are selected and uncensored under treatment—under Assumptions 1 to 4, Assumption 6, plus the following assumption: $E[Y^*(0)|cEEUU, cEEUC,\ cENU*] \geq E[Y^*(0)|cEEUU, cNE * U]$. In our application, this assumption exploits the positive correlation between having an employment spell in the observation period and employment duration, since the stratum $cNE * U$ is continually unemployed during the observation period under no participation in JC. See the Internet Appendix for additional intuition about this point, and for an indirect assessment of the plausibility of this alternative mean-dominance assumption.

3 Employment Duration Effects of Job Corps

We illustrate the use of the bounds presented in Section 2 by assessing the effect of JC on the duration of employment. Section 3.1 discusses some details on the JC program and the data, and provides a preliminary analysis of the effects of JC on the observed duration outcomes. Section 3.2 presents the estimated bounds, and Section 3.3 offers a discussion of the results.

3.1 National Job Corps Study Data and Preliminary Analysis

The data for our empirical illustration comes from the NJCS, briefly described in Section 2.1. Following the previous JC literature, and since part of our analysis also bounds the wage effects of JC on the employment spells of interest, our sample is restricted to individuals who have non-missing values for weekly earnings and hours worked for every week after randomization. These are the same sample restrictions employed by Lee (2009) and Blanco et al. (2013a). In addition, the sample is restricted to individuals with information on actual enrollment, captured by a binary indicator of whether the individual was ever enrolled in JC during the 208 weeks after randomization. Chen and Flores (2015) employed the same set of restrictions to analyze the effects of enrolling in JC on wages. As those papers do, we also assume that the information is missing completely at random (in principle, the bounds in Section 2 could be extended to deal with missing data). Similar to Blanco et al. (2013a) and Chen and Flores (2015), we analyze a sample of 7,531

individuals that excludes Hispanics to increase the plausibility of Assumption 5 and Assumption 6. Hispanics were the only demographic group in the NJCS for which negative but statistically insignificant effects of JC on employment and earnings were found (Schochet et al., 2001), thus making those assumptions harder to justify for them. The Internet Appendix presents separate analyses for Whites, Blacks, (Non-Hispanic) Males, and (Non-Hispanic) Females. We employ the NJCS design weights throughout the analysis, since different subgroups in the population had different probabilities of being included in the research sample (for details on the NJCS design weights see Schochet, 2001).

Our duration outcome of interest is the log of the observed number of weeks in employment for the last spell by week 208 after randomization. We focus on the last employment spell in order to consider a longer-term outcome and to increase the plausibility of our assumptions (see section 2.1). Note that not every individual may experience an employment spell during the follow-up period—a consequence of the selection into who may experience a spell of interest. In addition, the employment duration is censored for individuals who are employed at week 208 after randomization. Since the timing of the start of an employment spell of interest (if it takes place) is endogenous, opening the door for unobserved characteristics of individuals to confound both the latent duration outcome and the censoring time, the independent censoring assumption is likely invalid. Indeed, applying Frandsen's (2018) test to these data soundly rejects this assumption (p-value= 0.00 for the non-Hispanic sample). We also analyze the average log wages received during the corresponding employment spells to complement the analysis of durations.

Note that with the outcome we consider in our application, the potential outcome of each individual under treatment may correspond to an employment spell number (e.g., the k-th spell after randomization) that is different from the one corresponding to her potential outcome under control (e.g., the $(k+m)$-th spell, for $m \neq 0$). One way to think of "employment spell number" in this setting is as an intermediate or mechanism variable, since it is a variable potentially affected by treatment assignment (Z) that in turn can affect the potential outcomes of interest. In this case, differences in the spell number at which the potential outcomes are evaluated would not cause bias in our estimated bounds. The reason is that our focus is on the reduced-form effect on the latent duration of the last employment spell by week 208 after randomization (rather than on transition probabilities or another structural effect), and that, conditional on having an employment spell (the selection variable S we control for), the full observability of the outcome is not determined by the number of spells experienced (it is determined by W, which we control for). In this sense, the number of spells is just a possible mechanism: the values of such variables may differ between treatment and control states for each individual, which in turn may affect her potential outcomes, but this does not cause bias in the estimated bounds on the reduced-form effect we consider. Specifically, as previously discussed in Section 2.2, randomization of Z makes the groups with $Z = 0$ and $Z = 1$ within the $cEEUU$ stratum comparable, which allows us to bound the causal reduced-form effect of treatment assignment Z (which equals treatment receipt D for compliers) on the outcome of interest without bias. It may also be of interest to analyze the effect of JC on the latent duration of the last employment spell by week 208 after randomization when measured at a fixed spell number (e.g., at the first employment spell after randomization). For that, our

general approach could be extended to construct bounds on those effects by also stratifying on that dimension. That analysis, however, is beyond the scope of this paper.

Table 2 presents some preliminary estimates (summary statistics are given in Table E1 of the Internet Appendix). The main sample of Non-Hispanics contains 7,531 individuals: 4,554 and 2,977 assigned to treatment and control groups, respectively. Roughly 74 and 4.7 percent of the individuals in the treatment and control group enrolled in JC within the 208 weeks after randomization, respectively. Randomization into JC increases the probability of employment at week 208 by a statistically significant 4.9 percentage points, and it also increases the length of the observed employment spells by a statistically significant 6.5 log points (ITT). Estimates of the local average treatment effect for the compliers ($LATE$) that employ random assignment as an IV for treatment receipt is larger (9.5 log points) and also statistically significant. While the latter estimate controls for noncompliance, both the ITT and $LATE$ ignore selection and censoring and they are thus likely biased. Estimates of quantile treatment effects (QTE) that compare points of the employment duration distributions by treatment assignment (Koenker and Bassett, 1978) show statistically significant effects of randomization on the 25th and 50th quantiles. These estimates are similar in magnitude, at around 8.7 log points. Analogous to the $LATE$ estimates, we compute estimates from the instrumental variable quantile treatment effect ($IVQTE$) of Abadie et al. (2002). The estimated effects on the 25th and 50th quantiles of the employment duration for compliers are 16.7 and 11.8 log points, respectively, and statistically significant. Estimates based on QTE and $IVQTE$ are also likely biased since they ignore selection and censoring. We note here that, when comparing the point estimates above (i.e., ITT, $LATE$, QTE, and $IVQTE$) to the estimated bounds reported in the following section, it is important to keep in mind that each of them identifies different parameters. As a result, for instance, it cannot be concluded that a given point estimate is unbiased because it falls within the estimated bounds, as this can occur because the bias in the point estimate is offset by the difference in the true values of the parameters that each method identifies.

3.2 Estimated Nonparametric Bounds

We now present our estimated nonparametric bounds that address noncompliance, selection, and censoring. We start by verifying the testable implications of our assumptions and assessing indirect evidence on their plausibility. Then, we present estimates of the bounds to do inference on the impact of JC on the duration of the last complete employment spells and the average log wages earned by individuals during those spells.

3.2.1 Assessment of the Plausibility of the Assumptions

The plausibility of the basic IV assumptions was discussed in Section 2.3, so we do not consider them further. As for the monotonicity and stochastic dominance assumptions in Assumptions 5, 6, and 7 are not innocuous and have to be judiciously assessed in any application. When introducing those assumptions in section 2.4, we advanced arguments for why we regard them plausible in the

context of analyzing the effects of JC on employment duration and also discussed potential threats to their validity. In this subsection, we empirically assess the plausibility of Assumptions 5, 6, and 7 by analyzing their testable implications and the indirect evidence that can be gathered, discussed in Section 2.4.

A testable implication of Assumption 5 is that the IV estimate of D on S using Z as an instrument for D is non-negative (see Section 2.4.1). We find a positive and highly statistically significant estimate of 0.017 (standard error of 0.007). As shown in Table 3, under Assumptions 1 to 5, estimates for the principal stratum proportions for all the always takers and the never takers are positive and statistically significant. Another testable implication is that one of the stratum proportions involving compliers in equations (5) to (10) is strictly positive: the stratum proportion π_{cNN**} is statistically positive at 0.030 ($= 0.032 - 0.002$, with a standard error of 0.003), while the estimates of equations (6) to (10) are all statistically positive. A testable implication of Assumption 6 is that the parameter $E[W(1) - W(0)|cEE]$ is non-negative (see Section 2.4.2). Employing Imai's (2007) bounds under Assumptions 1 to 5, the estimated lower and upper bounds for that parameter are .058 and .077, respectively, with the 90 percent Imbens and Manski (2004) confidence interval excluding zero. Also, under Assumptions 1 to 6, we find that the estimated stratum proportion π_{cEECC} is statistically positive: $0.547 - 0.160 = 0.386$ (standard error of 0.011). Finally, the implied testable inequality in equation (18) has a positive and statistically significant estimate of 0.049 (standard error of 0.011). In sum, the empirical evidence on the testable implications of Assumptions 1 to 6 does not falsify those assumptions.

One way to assess the plausibility of Assumption 7 (stochastic dominance) is by providing indirect evidence for the assumption's implied weak ranking of the strata's average labor market outcomes, whereby $cEEUC$ has the most favorable average outcomes, followed by $cEEUU$ and $cNE * U$ (see Section 2.4.3). The assessment compares average baseline characteristics of the subpopulations $\{cEEUU, cEEUC\}$ and $\{cEEUU, cNE * U\}$, since averages for each stratum are not point identified. The estimates of $E[X|cEEUU, cEEUC]$ and $E[X|cEEUU, cNE*U]$, where X denotes relevant baseline characteristics, are consistent with the implied weak ranking (numerical results are presented in Table F6 in the Internet Appendix). On average, the former group is older, more educated and with better labor market experiences at baseline (e.g., months employed, yearly and weekly earnings). The estimated differences are not statistically significant, which is still consistent with the implied weak ranking. Hence, this exercise does not provide indirect evidence against Assumption 7, with the sign of the point estimates being consistent with the assumption.

3.2.2 Estimated Bounds on $SATE_{cEEUU}$

We now analyze the average effect of JC on the duration of the last complete employment spell before week 208 after randomization for compliers who experience a complete employment spell whether or not they participate in JC ($SATE_{cEEUU}$). Estimates under Proposition 1 are not presented because the lower bound for the stratum proportion π_{cEEUU} is zero (i.e., $\tau \leq 0$). Column 1 in Table 4 presents the estimated bounds in Propositions 2 and 3 for $SATE_{cEEUU}$, along with their IM confidence intervals, for the Non-Hispanic sample. Under Proposition 2 (Assumptions 1

to 6), the first row shows an estimated minimal positive value in the domain of π_{cEEUU} (τ) of 0.213. Given that the fraction of compliers is 0.69, this implies the $cEEUU$ stratum is at least 31 percent of all compliers. The estimated lower and upper bounds on $SATE_{cEEUU}$ are -0.426 and 0.550, respectively. Based on these estimated bounds, it is plausible that the effects for the stratum of interest are considerably larger or smaller than those based on the ITT and $LATE$ estimates, implying that these point estimates fall inside the estimated bounds. Inference based on the bounds, however, is likely to be more reliable because, contrary to the ITT and $LATE$ estimates, they address selection and censoring.

Table 4, column 1 also presents the estimated bounds for $SATE_{cEEUU}$ under Proposition 3 (adding stochastic dominance in Assumption 7), which are considerably narrower due to a tighter lower bound. The estimated lower bound is now positive, suggesting an average training effect of at least a 5.6 log points increase on the duration of the last complete employment spell for compliers who experienced a complete employment spell regardless of JC participation. Based on the 90 percent IM confidence interval, however, the true effect could be zero. At the same time, with 90 percent confidence, negative effects larger than 2.8 log points of JC on the duration of the last complete employment spells for $cEEUU$ individuals are ruled out. In addition to analyzing the sample of Non-Hispanics, we provide estimates for other demographic groups in the Internet Appendix (Table H1). An interesting finding is that the lower bounds for Males and Blacks are relatively small, where the estimate for the former group is negative. Hinted by their smaller estimates, we found that after excluding Black male individuals from the sample of Non-Hispanics, the estimated lower bound increases to 0.108 log points with a 90 percent IM confidence interval that rules out zero effects.

Analysis of Wages During the Employment Spells of Interest.

To complement the previous analysis, we estimate bounds on the average effect of JC on the average wage in the last complete employment spell for $cEEUU$ individuals. We employ the same bounds used to analyze employment duration, but using the outcome of log wages during the spells (see Remark 1 in Section 2.5). These results are reported in column 2 of Table 4. The estimated bounds under Proposition 2 do not identify the sign of the average log wage effect for Non-Hispanics. Estimates under Proposition 3 indicate that the average effects of JC training on the average log wage during the employment spells of interest is bounded between 5.9 and 21.2 log points, where the lower bound statistically excludes zero based on the IM confidence intervals. Taken together, these results suggest that, for the $cEEUU$ stratum, training in JC may increase the average duration of the last complete employment spells for Non-Hispanics, along with their average wages in those employment spells (and that these log wage effects can potentially be high). While we cannot rule out zero average effects on the duration of the last complete employment spells for this stratum at conventional significance levels, it is important to note that average log wages during these spells significantly increase due to training in JC. A similar conclusion applies to the samples of Whites and Males. The analysis of average log wages for Females and Blacks are suggestive of positive effects, however, their 90 percent IM confidence intervals do not exclude zero. These results can be found in Table H2 in the Internet Appendix.

3.2.3 Estimated Bounds on $SQTE^{\alpha}_{cEEUU}$

Figure 1 presents our estimated bounds on $SQTE^{\alpha}_{cEEUU}$ for the duration of the last complete employment spells. The estimated upper and lower bounds are denoted by a short dash, while their respective (pointwise) 90 percent IM confidence intervals are denoted by a long dash at the end of the dashed vertical line. The numerical results for all figures can be found in Tables J1 to J4 in the Internet Appendix. Similar to the estimated bounds for average effects, Figure 1(a) shows that the estimates of the lower and upper bounds for $SQTE^{\alpha}_{cEEUU}$ on the duration of the last complete employment spells under Proposition 2 are not informative about the sign of the effect of interest on any of the analyzed quantiles. Relative to the bounds on $SATE_{cEEUU}$, the bounds on the median are narrower: estimates are -0.32 and 0.43 for the lower and upper bound, respectively. As shown in Figure 1(b), the estimated lower bounds under Proposition 3 become tighter and positive on all the quantiles we consider. However, only the (pointwise) 90 percent IM confidence intervals on the 10th and 30th quantiles exclude zero. In contrast to the positive estimated lower bound for $SATE_{cEEUU}$, the estimated lower bound on the median is zero. Still, plausible values for the effects on the quantiles analyzed can be ruled out based on the estimated bounds. For instance, the (pointwise) 90 percent IM confidence interval on the median rule out negative effects larger than 11 log points on the duration of the spell of interest. Point estimates based on the $IVQTE$ (for the compliers), presented in Table 2, are inside the estimated bounds for $SQTE^{\alpha}_{cEEUU}$ on the 25th and 50th quantiles, but on the 75th quantile the estimated lower bound is larger in magnitude than the estimated $IVQTE$. As shown in Figure J2 in the Internet Appendix, containing the estimates for different demographic groups, we only find positive and statistically significant effects on lower quantiles analyzed for the sample of Females. This finding stands in contrast with prior literature that has documented smaller effects for females on earnings and employment relative to other groups (e.g., Flores et al., 2012; Blanco et al., 2013a; Eren and Ozbeklik, 2014).

Analysis of Wages During the Employment Spells of Interest.

The analysis of effects on selected quantiles of the distribution of average log wages for the $cEEUU$ stratum during the same employment spells is presented in Figure 2 for Non-Hispanics. Figure 2(a) shows that the estimated lower bounds under Proposition 2 are negative on all the quantiles considered. Under Proposition 3, as shown in Figure 2(b), all the estimated lower bounds become positive, and the corresponding (pointwise) 90 percent IM confidence intervals exclude zero for most quantiles above the first quartile. Furthermore, effects can be considerably large on quantiles analyzed beyond the 75th quantile. In combination with the results for the duration of employment, the log wage results suggest the insight that JC training has a likely positive impact on quantiles analyzed in the bottom of the distribution of employment duration and also a positive impact on most of the analyzed quantiles beyond the 25th quantile of the wage distribution during the employment spells of interest. A similar pattern as the one noted for Non-Hispanics is evident when analyzing the samples of Whites and Males (Figure J4 in the Internet Appendix). We also find that wage impacts are positive and significant on several quantiles analyzed below the median for the sample of Blacks, but small and not statistically significant on quantiles analyzed

in the upper part of their wage distribution. In contrast to the positive effects found for their employment durations, the sample of Females is the only one where we do not rule out zero wage effects throughout most of the analyzed quantiles of the average log wage distribution, consistent with earlier results in Blanco et al. (2013a).

3.3 Discussion of Results

The evidence presented above suggests that JC training may have a positive average effect on the duration of the last complete employment spell before week 208 after randomization for compliers who would experience a complete employment spell whether or not they participate in JC (the $cEEUU$ stratum). This stratum represents at least 21.3 percent of Non-Hispanics, and at least 31 percent of all compliers. Under our preferred set of assumptions, those used in Proposition 3, the estimated bounds suggest that the increase in the duration of the last complete employment spells is between 5.6 to 55 log points (5.8 and 73 percent). However, the 90 percent IM confidence intervals do not statistically rule out a zero effect. The analysis for other samples reveals that, while the 90 percent IM confidence intervals do not rule out zero for any sample, the groups of Males and Blacks have the smallest estimated lower bounds. Indeed, when excluding Black male individuals from our main sample, the estimated lower bound increases from 5.6 to 10.8 log points (from 5.8 to 11.4 percent) and the corresponding 90 percent IM confidence interval excludes zero. Our own interpretation is that there can exist statistically significant average employment duration effects for at least some non-Hispanics. This effect could be heterogeneous across the quantiles of the distribution of the last complete employment spells' duration, since the estimated bounds are consistent with larger effects on the upper quantiles analyzed. Nevertheless, only on some of the lower quantiles considered the (pointwise) IM confidence intervals rule out zero effects and, in general, the confidence intervals largely overlap across the analyzed quantiles. We also find that the estimated bounds of JC training on the average log wages received during those spells analyzed are consistent with positive values and their (pointwise) IM confidence intervals exclude zero.

4 Conclusions

We derive nonparametric sharp bounds for average and quantile effects of a treatment on duration outcomes. Even when employing data from a randomized evaluation, three common complications in such settings are noncompliance with the assigned treatment, selection into the spell of interest, and censoring of the duration outcome. Ignoring any of these issues could yield biased estimates of the effects. Conventional techniques require exclusion restrictions, distributional assumptions, and/or conditional independence assumptions to address endogenous selection and censoring. Instead, the proposed nonparametric bounds employ weak monotonicity and stochastic dominance assumptions, while requiring an instrumental variable to address noncompliance. These assumptions are potentially weaker in relevant empirical settings. The bounds are on the causal effects for the latent subpopulation of individuals that comply with their treatment assignment, and have an uncensored spell of interest regardless of random assignment.

The derived bounds are illustrated in an analysis of the effect of Job Corps (JC) on the log duration (in weeks) of the last complete employment spell using data from a randomized evaluation of JC. To our knowledge, this is the first analysis to shed light on the program's impact on an important duration outcome. Evidence based on the estimated bounds suggests that JC participation may increase the average duration of the last complete employment spell before week 208 after randomization by at least 5.6 log points (5.8 percent) and up to 55 log points (73 percent). However, the 90 percent confidence intervals of these estimated bounds include zero. This effect is bounded for the compliers (with their random assignment to JC) who would experience a complete employment spell by week 208 after randomization regardless of JC participation. The estimated bounds on the quantile treatment effects on the duration of the last complete employment spell for the same subpopulation are also suggestive of positive effects of JC, with the (pointwise) confidence intervals of some of the quantiles analyzed excluding zero. We complement the analysis on employment spell durations by estimating bounds for the average and quantile treatment effects on the average log wage during the same spells and for the same subpopulation, where, in general, we find statistically positive wage effects. We interpret these findings as suggesting that JC can help these individuals to maintain their jobs longer and that those jobs are better paid.

The methods developed herein can be extended in important ways. First, while commonly used in the econometrics literature (e.g., Imbens and Angrist, 1994; Manski, 1997; Heckman et al., 1999; Manski and Pepper, 2000; Vytlacil, 2002; Lee, 2009; Zhang and Rubin, 2003), individual-level monotonicity assumptions are not innocuous. As a result, this type of assumptions need to be assessed on a case-by-case basis, as they may fail to hold in specific applications (e.g., de Chaisemartin, 2018). In some cases—such as when the instrumental variable has a weak effect on the treatment—the type of bounds considered in this paper may be sensitive to violations of individual-level monotonicity assumptions (e.g., Chen and Flores, 2015). In instances where the use of those assumptions may be suspect, it may be possible to explore other assumptions to derive bounds, such as assuming the outcome has bounded support, or other stochastic dominance assumptions (see, e.g., Remark 6). Alternatively, it may be possible to perform a sensitivity-type of analysis (e.g., in the spirit of Manski and Pepper, 2018; Masten and Poirier, 2018; or the simulations in Chen and Flores, 2015) that does not rely on specific individual-level monotonicity assumptions or ruling out specific strata. For instance, instead of imposing any particular individual-level monotonicity assumptions (e.g., Assumption 6), one could derive bounds for our parameters of interest setting the proportion of each stratum that is ruled out by such assumptions equal to a given constant $k > 0$ (e.g., $\pi_{cEECU} = k$), and then analyze how the estimates of these bounds vary as those constants are changed (the assumptions employed in this paper set those constants equal to zero). Second, the proposed methodology in this paper was developed within the original principal stratification framework of Frangakis and Rubin (2002), which does not use "entirely hypothetical" outcomes (it only uses potential outcomes, which are outcomes that can be observed in the data depending of the value of the instrument; entirely hypothetical or a priori counterfactual outcomes are never observed regardless of the value of the instrument). As a result, the proposed methods share what in some applications can be a shortcoming of principal stratification: its focus on a particular stratum (e.g., Deaton, 2010; Heckman and Urzua, 2010; Pearl, 2011). Therefore, it is of interest to build on the bounds presented herein to derive bounds on the effects for other subpop-

ulations that may be of interest, such as those that include individuals with censored durations, or the subpopulations of all compliers or all treated individuals. This could follow work that has been done in other settings (e.g., Huber and Mellace, 2015; Chen et al., 2018). Lastly, it is also relevant to consider the derivation of uniform bounds (e.g., Firpo and Ridder, 2008) and methods to undertake uniform inference on the quantile function (e.g., Chernozhukov et al., 2013; Fan and Liu, 2018) to complement the pointwise approach employed here. These are all important topics for future research.

References

Abadie, A., Angrist, J., and Imbens, G. (2002), "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91-117.

Abbring J., and Van den Berg, G. (2003), "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71, 1491-1517.

Angrist, J., Imbens, G., and Rubin, D. (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the Statistical Association*, 91, 444-455.

Arellano, M., and Bonhomme, S. (2017), "Quantile Selection Models with an Application to Understanding Changes in Wage Inequality," *Econometrica*, 85, 1-28.

Ba, B., Ham, J., LaLonde, R., and Li, X. (2017), "Estimating (Easily Interpreted) Dynamic Training Effects from Experimental Data," *Journal of Labor Economics*, 35, S149-S200.

Bhattacharya, J., Shaikh, A., and Vytlacil, E. (2012), "Treatment Effect Bounds: An Application to Swan-Ganz Catheterization," *Journal of Econometrics*, 168, 223-243.

Blanco, G., Flores, C., and Flores-Lagunes, A. (2013a), "Bounds on Average and Quantile Treatment Effects of Job Corps Training on Wages," *Journal of Human Resources*, 48, 659-701.

Blanco, G., Flores, C., and Flores-Lagunes, A. (2013b), "The Effects of Job Corps Training on Wages of Adolescents and Young Adults," *American Economic Review: Papers and Proceedings*, 103, 418-422.

Blundell, R., and Powell, J. (2007), "Censored Regression Quantiles with Endogenous Regressors," *Journal of Econometrics*, 141, 65-83.

Blundell, R., Gosling, A., Ichimura, H., and Meghir, C. (2007), "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds," *Econometrica*, 75, 323-63.

Caliendo, M., Tatsiramos, K., and Uhlendorff, A. (2013), "Benefit Duration, Unemployment Duration and Job Match Quality: A Regression-Discontinuity Approach," *Journal of Applied Econometrics*, 28, 604-627.

Chen, X. and Flores, C. (2015), "Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps," *Journal of Business and Economic Statistics*,

33, 523-540.

Chen, X., Flores, C., and Flores-Lagunes, A. (2018), "Going beyond LATE: Bounding Average Treatment Effects of Job Corps Training," *Journal of Human Resources*, 53, 1050-1099.

Chen, X., Flores, C., and Flores-Lagunes, A. (2019), "Bounds on Average Treatment Effects with an Invalid Instrument, with an Application to the Oregon Health Insurance Experiment," Mimeo, Department of Economics, California Polytechnic State University at San Luis Obispo.

Chernozhukov, V., Fernandez-Val, I., and Kowalski, A. (2015), "Quantile Regression with Censoring and Endogeneity," *Journal of Econometrics*, 186, 201-221.

Chernozhukov, V., Fernandez-Val, I., and Melly, B. (2013), "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205-2268.

Chernozhukov, V., and Hong, H. (2002), "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of the American Statistical Association*, 97, 872-882.

De Chaisemartin, C. (2017), "Tolerating Defiance? Local Average Treatment Effects without Monotonicity," *Quantitative Economics*, 8, 367-396.

Deaton, A. (2010), "Instruments, Randomization, and Learning About Development." *Journal of Economic Literature*, 48, 424-55.

Eberwein, C., Ham, J., and LaLonde, R. (1997), "The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data," *Review of Economic Studies*, 64, 655-682.

Eberwein, C., Ham, J., and LaLonde, R. (2002), "Alternative Methods of Estimating Program Effects in Event History Models," *Labour Economics*, 9, 249-278.

Eren, O., and Ozbeklik, S. (2014), "Who Benefits from Job Corps? A Distributional Analysis of an Active Labor Market Program," *Journal of Applied Econometrics*, 29, 586-611.

Fan, Y., and Liu, R. (2018), "Partial Identification and Inference in Censored Quantile Regression," *Journal of Econometrics*, in press.

Firpo, S., and Ridder, G. (2008), "Bounds on Functionals of the Distribution of Treatment Effects," IEPR Working Paper 08.09.

Flores, C., and Flores-Lagunes, A. (2010), "Nonparametric Partial Identification of Causal Net and Mechanism Average Treatment Effects," Mimeo, Department of Economics, California Polytechnic State University at San Luis Obispo.

Flores, C., and Flores-Lagunes, A. (2013), "Partial Identification of Local Average Treatment Effects with an Invalid Instrument," *Journal of Business and Economic Statistics*, 31, 534-545.

Flores, C., Flores-Lagunes, A., Gonzales, A., and Neumann, T. (2012), "Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps," *The Review of Economics and Statistics*, 94, 153-171.

Flores-Lagunes, A., Gonzalez, A., and Neumann, T. (2010), "Learning but not Earning? The Impact of Job Corps Training on Hispanic Youth," *Economic Inquiry*, 48, 651-67.

Frandsen, B. (2015), "Treatment Effects with Censoring and Endogeneity," *Journal of the American Statistical Association*, 110, 17451752.

Frandsen, B. (2018), "Testing Censoring Point Independence," *Journal of Business and Economic Statistics* (in press).

Frandsen, B., and Lefgren, L. (2016), "Partial Identification of the Distribution of Treatment Effects," Mimeo, Department of Economics, Brigham Young University.

Frangakis, C., and Rubin, D. (2002), "Principal Stratification in Causal Inference," *Biometrics*, 58, 21-29.

Frumento, F., Mealli, F., Pacini, B. and Rubin, D. (2012), "Evaluating the Effect of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data," *Journal of the American Statistical Association*, 107, 450-466.

Hall, R. (1979), "A Theory of the Natural Unemployment Rate and the Duration of Employment," *Journal of Monetary Economics* 5, 153-169.

Ham, J., and LaLonde, R. (1996), "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training," *Econometrica* 64, 175-205.

Heckman, J., LaLonde, R., and Smith, J. (1999), "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics*, Volume 3A, eds. O. Ashenfelter and D. Card, Amsterdam: North Holland, pp. 1865-2097.

Heckman, J., and Singer. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271-320.

Heckman, J., and Urzua, S. (2010), "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify," *Journal of Econometrics*, 156, 27-37.

Hirano, K., Imbens, G., Rubin, D. and Zhou, X. (2000), "Assessing the Effect of an Influenza Vaccine in an Encouragement Design With Covariates," *Biostatistics*, 1, 69-88.

Hong, H., and Tamer, E. (2003), "Inference in Censored Models With Endogenous Regressors, *Econometrica*, 71, 905-932.

Horowitz, J., and Manski, C. (1995), "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281-302.

Huber, M., and Mellace, G. (2015), "Sharp Bounds on Causal Effects under Sample Selection," *Oxford Bulletin of Economics and Statistics*, 77, 129-151.

Imai, K. (2007), "Identification Analysis for Randomized Experiments with Noncompliance and Truncation-by-Death," Technical Report, Department of Politics, Princeton University.

Imai, K. (2008), "Sharp Bounds on the Causal Effects in Randomized Experiments with 'Truncation-by-Death'," *Statistics and Probability Letters*, 78, 144-149.

Imbens, G., and Angrist, J. (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.

Imbens, G., and Manski, C. (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

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

Khan, S. and Tamer, E. (2009), "Inference on Endogenously Censored Regression Models Using Conditional Moment Inequalities," *Journal of Econometrics*, 152,104-119.

Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.

Lechner, M., and Melly, B. (2010), "Partial Identification of Wage Effects of Training Programs," Mimeo, Department of Economics, Brown University.

Lee, D. (2009), "Training Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071-1102.

Manski, C. (1997), "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.

Manski, C., and Pepper, J. (2000), "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997-1010.

Manski, C., and Pepper, J. (2018), "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity using Bounded-Variation Assumptions," *Review of Economics and Statistics*, 100, 232-244.

Masten, M., and Poirier, A. (2018), "Identification of Treatment Effects Under Conditional Partial Independence," *Econometrica*, 86, 317-351.

Newey, W. and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Volume 4, eds. R. Engle and D. McFadden, Amsterdam: North Holland, pp. 2111-2245.

Pearl, J. (2011), "Principal Stratification–a Goal or a Tool?," *The International Journal of Biostatistics*, 7, 1-13.

Powell, J. (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-155.

Rubin, D. (1978), "Bayesian Inference for Causal Effects," *The Annals of Statistics*, 6, 34-58.

Rubin, D. (1980), Comment on "Randomization Analysis of Experimental Data: The Fisher Randomization Test," by D. Basu, *Journal of American Statistical Association*, 75, 591-593.

Rubin, D. (1990), "Comment on 'Neyman (1923) and Causal Inference in Experiments and Observational Studies'," *Statistical Science*, 5, 472-480.

Sant'Anna, P. (2016), "Program Evaluation with Right Censored Data," Mimeo, Department of Economics, Vanderbilt University.

Schochet, P. (2001), *National Job Corps Study: Methodological Appendixes on the Impact Analysis*, Mathematica Policy Research, Inc., Princeton, NJ.

Schochet, P., Burghardt, J., and Glazerman, S. (2001), *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes*, Mathematica Policy Research, Inc., Princeton, NJ.

Schochet, P., Burghardt, J., and McConnell, S. (2008), "Does Job Corps Work? Impact Findings from the National Job Corps Study," *American Economic Review*, 98, 1864-1886.

Shepherd, B., Gilbert, P., and Lumley, T. (2007), "Sensitivity Analyses Comparing Time-to-Event Outcomes Existing Only in a Subset Selected Postrandomization," *Journal of the American Statistical Association*, 102, 573-582.

Stoye, J. (2009), "More on Confidence Intervals for Partially Identified Parameters" *Econometrica*, 77, 1299-1315.

Van den Berg, G., Bonev, P., and Mammen, E. (2016), "Nonparametric Instrumental Variable Methods for Dynamic Treatment Evaluation," IZA Discussion Paper No. 9782.

Vikström, J., Ridder, G., and Weidner, M. (2018), "Bounds on Treatment Effects on Transitions," *Journal of Econometrics*, 205, 448-469.

Vytlacil, E. (2002), "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331-341.

Zhang, J. and Rubin, D. (2003), "Estimation of Causal Effects Via Principal Stratification When Some Outcomes are Truncated by 'Death'," *Journal of Educational and Behavioral Statistics*, 28, 353-368.

Zhang, J., Rubin, D., and Mealli, F. (2008), "Evaluating the Effect of Job Training Programs on Wages Through Principal Stratification," in *Modelling and Evaluating Treatment Effects in Econometrics (Advances in Econometrics, Volume 21)*, eds. T. Fomby, R. Carter Hill, D. Millimet, J. Smith, and E. Vytlacil, Emerald Group Publishing Limited, pp.117 - 145.

Zhang, J., Rubin, D., and Mealli, F. (2009), "Likelihood-based Analysis of the Causal Effects of Job Training Programs Using Principal Stratification," *Journal of the American Statistical Association*, 104, 166-176.

Table 2: Preliminary Analysis for the Non-Hispanic Sample

	Treatment $Z_i = 1$	Control $Z_i = 0$	Difference
At week 208:			
Enrollment	0.737	0.047	0.689^{***}
			(0.008)
Employed	0.613	0.564	0.049^{***}
			(0.011)
Ever experienced:			
Employment spell	0.953	0.942	0.011^{**}
			(0.005)
Intention to Treat Effect (ITT)			
Log Employment duration			0.065^{**}
			(0.032)
Local Average Treatment Effect (LATE)			
Log Employment duration			0.095^{**}
			(0.047)
Observations	4554	2977	
		α-quantiles	
	0.25	0.50	0.75
Quantile Treatment Effect (QTE)			
Log Employment duration	0.087^{**}	0.090^{*}	0.036
	(0.039)	(0.049)	(0.037)
Intrumental Variable QTE (IVQTE)			
Log Employment duration	0.167^{**}	0.118^{*}	0.048
	(0.071)	(0.061)	(0.050)

Note: Standard errors are in parentheses. *, **, and *** denote statistical significance at a 90, 95 and 99 percent confidence level. Computations use design weights.

Table 3: Estimates of Principal Stratum Proportions under Assumptions 1 to 6, Non-Hispanic Sample

	$Z = 0$				$Z = 1$		
	$S = 0$	$S = 1$			$S = 0$	$S = 1$	
	$W = *$	$W = 0$	$W = 1$		$W = *$	$W = 0$	$W = 1$
$D = 0$	$nNN**$ $cNN**$ $cNE*C$ $cNE*U$	$nEEUU$ $cEEUU$ $cEEUC$	$nEECC$ $cEECC$	$D = 0$	$nNN**$	$nEEUU$	$nEECC$
	0.057	0.350	0.547		0.015	0.088	0.160
	(0.004)	(0.009)	(0.009)		(0.002)	(0.004)	(0.005)
$D = 1$	$aNN**$	$aEEUU$	$aEECC$	$D = 1$	$aNN**$ $cNN**$	$aEEUU$ $cEEUU$ $cNE*U$	$aEECC$ $cEECC$ $cEEUC$ $cNE*C$
	0.002	0.027	0.018		0.032	0.252	0.453
	(0.001)	(0.003)	(0.002)		(0.003)	(0.006)	(0.007)

Note: Bootstrapped standard errors, based on 1000 replicates, are reported within parentheses Computations use design weights.

Table 4: Estimated Bounds on the Average Treatment Effect of JC on the Duration of the Last Complete Employment Spell and its Average Wage for the $cEEUU$ stratum, Non-Hispanic Sample

	Duration	Average wage
Bounds under Proposition 2		
(Assumptions 1 to 6)		
Minimal positive value		
in the domain of π_{cEEUU}, τ	0.213	
(Standard error)	(0.009)	
$[LB^{2A}, UB^{2A}]$	[-0.426, 0.550]	[-0.097, 0.212]
(90% IM confidence intervals)	(-0.574, 0.709)	(-0.137, 0.258)
Bounds under Proposition 3		
(Assumptions 1 to 7)		
$[LB^{3A}, UB^{2A}]$	[0.056, 0.550]	[0.059, 0.212]
(90% IM confidence intervals)	(-0.028, 0.709)	(0.027, 0.258)
Observations	7531	

Note: Bootstrapped standard errors to compute the Imbens and Manski (IM, 2004) confidence intervals are based on 1000 replications. Computations use design weights.

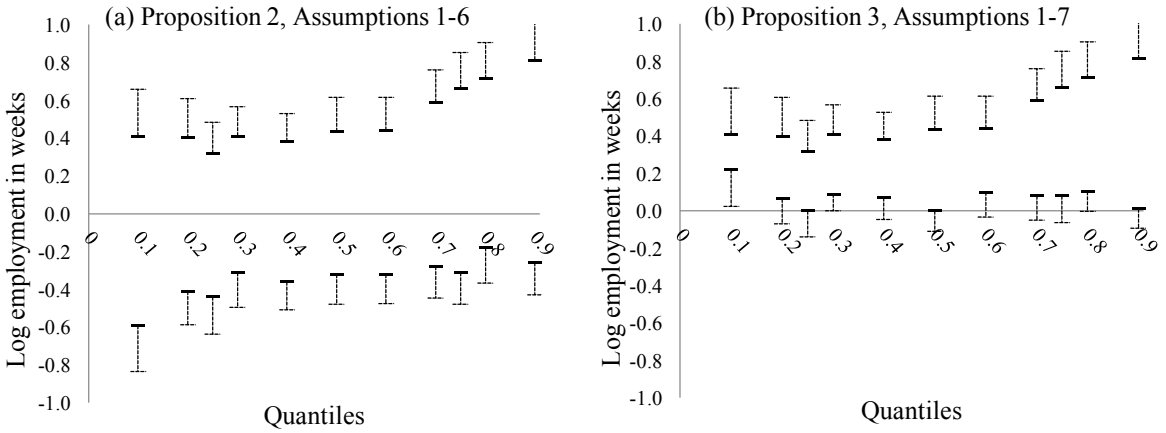

Figure 1. Estimated bounds on $SQTE_{cEEUU}$ on the duration of the last complete employment spells for the Non-Hispanic sample. Upper and lower bounds are denoted by a short dash, while the 90 percent Imbens and Manski (2004) confidence intervals are denoted by a long dash at the end of the dashed vertical line.

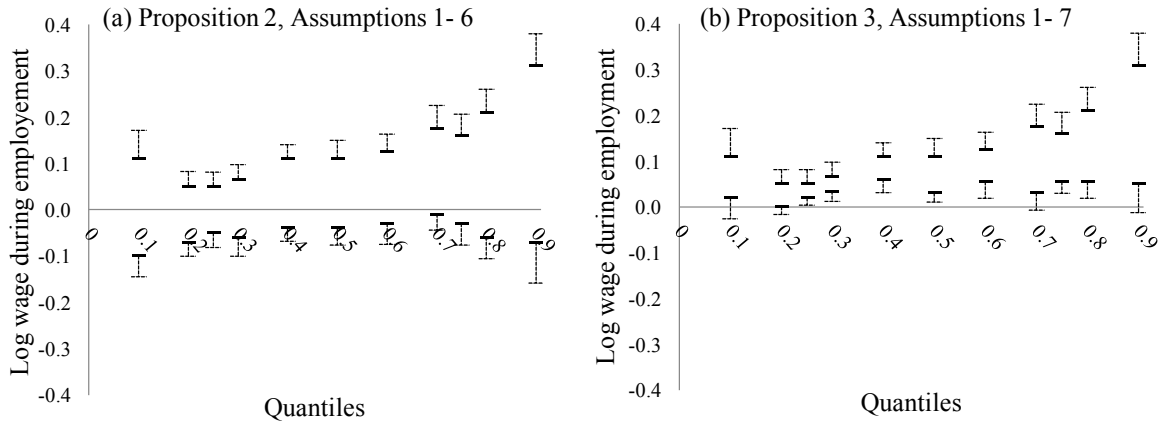

Figure 2. Estimated bounds on $SQTE_{cEEUU}$ on log wages during the last complete employment spells for the Non-Hispanic sample. Upper and lower bounds are denoted by a short dash, while the 90 percent Imbens and Manski (2004) confidence intervals are denoted by a long dash at the end of the dashed vertical line.