Multivalued Treatments and Decomposition Analysis: An application to the WIA Program*

Wallice Ao† Sebastian Calonico‡ Ying-Ying Lee§

January 2017

Abstract

We analyze how different levels of active labor program participation affect participants' labor market outcomes. To do so, we propose two-step semiparametric efficient estimators that can be used to decompose differences in earnings distributions into (i) wage structure effects, arising due to the conditional outcome distributions associated with different levels of participation; and (ii) composition effect, arising due to differences in the distribution of observable characteristics. These counterfactual differences reveal causal relationships under a conditional independence assumption. Moreover, we calculate the semiparametric efficiency bound for the multivalued treatment effects, generalizing previous results for binary treatment effects. We employ these procedures to study the effects of the Workforce Investment Act (WIA), a large US job service program. Our estimation results show that heterogeneity in levels of participation is an important dimension to evaluate the WIA and other social programs in which participation varies. The results of this paper, both theoretically and empirically, provide rigorous assessment of intervention programs and relevant suggestions to improve the performance and cost-effectiveness of these programs.

JEL classification: I38, H53, C31, C14

Keywords: Program evaluation, decomposition analysis, treatment effects, propensity score, counterfactual distribution, multi-valued treatments, semiparametric efficiency.

*The authors are thankful to Matias Busso, Matias Cattaneo, Steven Durlauf, Rob Garlick, Jesse Gregory, Robert Haveman, Oscar Mitnik, Jack Porter, Jeffrey Smith, Christopher Taber and James Walker for their valuable advice and suggestions. The author also thanks participants at the 2013 Midwest Economic Association Annual Meeting, 2013 Public Workshop in UW-Madison, 2014 COMPIE Conference, 2015 University of Michigan Summer Econometrics Workshop, 2015 Econometric Society World Congress, and the Inter-American Development Bank Research Department for their useful comments.

†Institute for Defense Analyses. Email: ao@wisc.edu.

‡Department of Economics, University of Miami. Email: scalonico@bus.miami.edu

§Department of Economics, University of California - Irvine. Email: yingying.lee@uci.edu

1 Introduction

There is an extensive literature on the evaluation of active labor market programs.[1] In most of these works, program performance is measured by comparing employment outcomes of those who participate in the programs and those do not. However, a prominent feature of these programs is that participation varies, for example in terms of length and/or intensity. As one may expect, different levels of program participation may yield very different outcomes. For example, while active labor market programs can improve participants' labor market outcomes by providing the training and assistance they need to obtain higher earnings and find more suitable jobs, participants may face "lock-in" effects as longer participation may yield unsatisfying labor market outcomes due to lost labor market experience, which employers may view as a negative signal. Additionally, from a policy point of view, while these programs can improve participants' welfare and reduce unemployment, they can also be very costly over time.

In this paper, we propose a procedure to estimate how different levels of program participation affect participants' labor market outcomes. In addition, we seek to identify the explanatory factors driving the observed difference in outcomes. First, we decompose differences in earnings distribution for participants of different levels of program participation to those arising due to (i) wage structure effects and (ii) composition effects. A wage structure (or schedule) effect is a mapping from workers' characteristics to a certain earnings distribution. It captures how participants are distinguished in the labor market by their levels of program participation. Analyzing this effect allow us identify where the program is most effective. On the other hand, composition effect captures how much the difference in the distribution of observable characteristics for participants of different levels of participation affect the observed difference in earnings distributions. Such decomposition can be carried out easily for distributional features of the counterfactual earnings distributions, such as mean, quantiles, and inequality measures.

We propose efficient propensity score weighting estimators to implement the decomposition analysis. When the levels of participation is considered as a multi-valued treatment variable chosen randomly by participants (conditional on their characteristics), the counterfactual effects in the decomposition analysis reveal causal relationships, such as treatment effects for the treated. Furthermore, the composition effect can be decomposed to isolate the effect of different explanatory factors within participants' characteristics. We call this the *detailed composition effect*. It reveals how specific characteristics attributes to the observed earnings outcomes. We apply our procedure to study how different levels of participation in the Workforce Investment Act (WIA) program affect participants' post-program earnings outcomes. The WIA is the largest public-financed career service program in the US. Our data contains detailed demographic information and employment history for a representative sample of WIA program participants entering the program from July 2003 to June 2005. In this sample, program participation ranges from less than one month to more than four years. Our estimation results suggest that the wage structure faced by participants of different lengths of participation contribute much to the difference in earnings outcomes

[1] See handbook chapter by Heckman, LaLonde, and Smith (1999). Heckman and Vytlacil (2007) and Imbens and Wooldridge (2009) provide comprehensive review and discussions on the program evaluation literature.

among them. On the other hand, the characteristics distributions of participants do not contribute as much to these differences. Earnings outcomes are all positive and significantly different from zero for all levels of participation. This is consistent with finding in the previous studies on WIA (Heinrich, Mueser, and Troske (2009) and Hollenbeck, Schroeder, King, and Huang (2005)), which compare the post-program earnings of participants and non-participants using a binary treatment effect approach. The magnitude of the average earnings gain, however, are different across gender groups: we observe that the average gains are higher for female participants. This result is consistent with prior studies on labor market programs such as the JTPA (See Nightingale and Elaine (2011) for example).

Our key identifying assumption for estimating the causal treatment effects is unconfoundedness (selection on observables, or conditional independence); that is, conditional on a set of observable covariates, selection into different levels of treatment is random. We justify this assumption by the following: first, we have a rich set of covariates including demographic information, unemployment insurance status, pre-, during, and post-participation labor market experience (earnings and industry worked) for up to 16 quarters, to account for the unobservables. It is likely that unobservable characteristics such as motivation, self-esteem, or family condition are captured by the individual labor market history. Second, the outcome variable we consider is the earnings differential between after leaving the program and prior to entering the program. This difference-in-difference specification, along with the above justification for unconfoundedness, is also used in Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) and Kluve, Schneider, Uhlendorff, and Zhao (2012), allowing us to account for time-invariant factors that may have influenced selection. Finally, by focusing only on participants, we expect to avoid the likely more difficult issue of selection *into* the program (as discussed, for example, in Andersson, Holzer, Lane, Rosenblum, and Smith (2016)).

This paper contributes to several strands of literature. First, we contribute to a growing program evaluation literature on assessing the impact of different levels of participation in social programs. In contrast to the vast binary treatment effect literature which captures the effect of participating in a program, econometrics methods for treatment intensity effects are less developed. To the best of our knowledge, this paper is the first one to to study levels of participation using decomposition analysis. In a related work, Behrman, Cheng, and Todd (2004) considers duration in a Bolivian preschool program as a continuous treatment, using a matching-type estimator. Hirano and Imbens (2004) propose a regression approach based on the generalized propensity score for continuous treatment effects, which has been used to evaluate welfare programs, such as Progresa/Oportunidades in Mexico (Ibarraran and Villa, 2010), the South African Child Support Grant (Agüero et al., 2010), and job training programs such as Job Corps in Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) and a German adult job training program in Kluve, Schneider, Uhlendorff, and Zhao (2012). Our second contribution is to provide efficient, nonparametric estimators for distributional, multi-valued treatment effects *for the treated* and *decomposition analysis*. Our propensity score weighting estimators build on and extend Cattaneo (2010) who studies multi-valued treatment effects to study average and quantiles effect for the population. We extend the estimation and inference procedure to multi-valued average and quantile treatment effects for the treated and general inequality measures. We propose nonparametric estimators that are robust to misspecification and easy to implement. Additionally, our estimators are efficient: their asymptotic variances reach the semiparametric efficiency bound derived in Lee (2015a). For estimation, we propose a GMM procedure that allows for joint estimation and inference on the effects across treatment levels and among the distributional features. For example, we could test if the median equals the mean. The estimators are easy to implement and have many potential applications. For example, they can also be used to evaluate the effects of the multiple treatments offered by programs such as the JTPA in Plesca and Smith (2007) and the National Evaluation of Welfare-to-Work Strategies study in Flores and Mitnik (2014).[2]

This paper also contributes to the decomposition analysis literature recently reviewed by Fortin, Lemieux, and Firpo (2011). In related projects, Chernozhukov, Fernandez-Val, and Melly (2013) develop a uniform inference procedure for a semiparametric regression estimation, and Firpo and Pinto (2016) address the idea of counterfactual distribution and propose a propensity score weighting estimator for the distributional effect of a binary treatment. We contribute to this literature by providing a uniform inference result for efficient multi-valued nonparametric propensity score weighting estimators. Most notably, we add to the decomposition analysis literature on the detailed composition effects by isolating the impacts of some specific explanatory factors, introduced by DiNardo, Fortin, and Lemieux (1996) and Fortin, Lemieux, and Firpo (2011). That is, the counterfactual is based on the conditional distribution of one factor given other explanatory covariates while we keep constant the distribution of the other explanatory variables. The detailed decomposition effect provides policymakers the information on how specific characteristic affects the outcomes of different treated groups and is novel to the decomposition literature. We calculate the semiparametric efficiency bound and propose an efficient estimator to analyze the detailed decomposition effects.

The paper is organized as follows: Section 2 describes the econometric methods for decomposition analysis and discusses the relationship between our decomposition analysis and the treatment effects literature, including an efficient propensity score weighting estimator for the detailed composition effects. Section 3 collects the econometric theory for our proposed propensity score weighting estimator. Section 4 describes institutional background of the WIA programs, the data, and presents our estimation results. In the Appendix we include the proofs to our theorems and supplementary tables and figures.

2 Decomposition Analysis and Treatment Effects

In this section, we develop a decomposition analysis framework to identify the explanatory features which drive variation in earnings distributions across groups with different levels of program participation. This

[2] Plesca and Smith (2007) use a matching estimator to evaluate the JTPA that offers multiple treatments or different services to participants. They illustrate disaggregating multi-treatment programs could provide useful insights into program operation. Flores and Mitnik (2014) consider the problem of using data from multiple programs implemented at different locations. A local government considers to implement one of several possible job training programs. In contrast to the population effects in Flores and Mitnik (2014), our treatment effects for the treated can evaluate the effect of implementing one program at a specified location.

decomposition analysis relates to the seminal work of Oaxaca (1973), Blinder (1973), and DiNardo, Fortin, and Lemieux (1996). More recent, related contributions are Fortin, Lemieux, and Firpo (2011), and Chernozhukov, Fernandez-Val, and Melly (2013).

2.1 Counterfactual effects

The population of agents is categorized into mutually exclusive sub-populations indexed by $t \in \mathcal{T}$, where $\mathcal{T} = \{0, 1, 2, ..., J\}$ is a finite discrete set with some fixed positive integer J. The index t could generally indicate a policy intervention or economic environments, such as unionization or time periods. In our empirical application, t represents the level of participation in the WIA program. We label the sub-population belonging to, choosing, or assigned t as "group-t." In each group-t, an independent and identically distributed data set $\{Y_{ti}, X_{ti}\}_{i=1,...,n_t}$ is drawn from the joint distribution of $(Y_t, X_t^\top)^\top \in \mathcal{Y} \times \mathcal{X} \subset \mathbb{R}^{1+d_x}$.

Given observability, we can identify the outcome distribution F_{Y_t}, the covariate distribution F_{X_t}, and the conditional distribution $F_{Y_t|X_t}$. The actual outcome distribution of group-t can be written as

$$F_{Y_t}(y) = \int_{\mathcal{X}} F_{Y_t|X_t}(y|x) dF_{X_t}(x)$$

by the law of iterated expectations. The conditional outcome distribution given covariates x, $\{x \mapsto F_{Y_t|X_t}(y|x) : y \in \mathcal{Y}\}$, describes a *wage structure*, the stochastic assignment of outcome to program participants with characteristics x. $F_{X_t}(x)$ is the distribution of observable characteristics of group-t.

We define the *counterfactual distribution* as:

$$F_{Y\langle t|t'\rangle}(y) \equiv \int_{\mathcal{X}} F_{Y_t|X_t}(y|x) dF_{X_{t'}}(x) \tag{2.1}$$

where we replaced $F_{X_t}(x)$ by $F_{X_{t'}}(x)$, the distribution of observable characteristics for group-t'. This is a well-defined statistical object by the following two assumptions.

Assumption 1 (Common Support) *The support of X_t, $\mathcal{X} \subset \mathbb{R}^{d_x}$, is the same for all $t \in \mathcal{T}$.*

Assumption 2 (Invariance of Conditional Distribution) *The conditional distribution $F_{Y_t|X_t}(y|x)$ applies or can be extrapolated for $x \in \mathcal{X}$, or it remains valid when the marginal distribution $F_{X_{t'}}$ replaces F_{X_t}.*

The common support assumption ensures that we observe agents with the same characteristics participate in all levels/groups of the program. The *counterfactual outcome* $Y\langle t|t'\rangle$ can be viewed as a random variable generated by the distribution function $F_{Y\langle t|t'\rangle}$. This counterfactual distribution has two interpretations. First, as the counterfactual outcome distribution that would have prevailed for group-t' if they have faced group-t's wage structure $\{x \mapsto F_{Y_t|X_t}(y|x) : y \in \mathcal{Y}\}$. Second, as the counterfactual outcome distribution of group-t if they had group-t''s characteristics distribution $F_{X_{t'}}(x)$.

Following the definition of the counterfactual distribution in (2.1), the difference in outcome distributions between group-t and group-t' $F_{Y_t}(y) - F_{Y_{t'}}(y)$ can be decomposed as:

$$\begin{aligned} F_{Y_t}(y) - F_{Y_{t'}}(y) &= F_{Y\langle t|t\rangle}(y) - F_{Y\langle t'|t'\rangle}(y) \\ &= \int_{\mathcal{X}} \Big(F_{Y_t|X_t}(y|x) - \boldsymbol{F}_{Y_{t'}|X_{t'}}(y|x) \Big)\, dF_{X_t}(x) \qquad (2.2) \\ &\quad + \int_{\mathcal{X}} F_{Y_{t'}|X_{t'}}(y|x)\, d\Big(\boldsymbol{F_{X_t}(x)} - F_{X_{t'}}(x) \Big). \qquad (2.3) \end{aligned}$$

The first component (2.2) represents the *wage structure effect*, arising due to different earnings schedules across groups. The second component (2.3) is the *composition effect*, arising from the different characteristics distributions among participants across groups.

We can similarly decompose functionals of the outcome distributions, such as the mean or the quantile functions. For example, the average outcomes between group-t and group-t' can be decomposed as follows:

$$\begin{aligned} \mathbb{E}[Y_t] - \mathbb{E}[Y_{t'}] &= \mathbb{E}[Y\langle t|t\rangle] - \mathbb{E}[Y\langle t'|t'\rangle] \\ &= \int_{\mathcal{X}} \int_{\mathcal{Y}} y\, d\Big(F_{Y_t|X_t}(y|x) - F_{Y_{t'}|X_{t'}}(y|x) \Big) dF_{X_t}(x) \qquad (2.4) \\ &\quad + \int_{\mathcal{X}} \mathbb{E}[Y_{t'}|X_{t'} = x]\, d\Big(F_{X_t}(x) - F_{X_{t'}}(x) \Big). \qquad (2.5) \end{aligned}$$

Then, (2.4) is $\mathbb{E}[Y\langle t'|t\rangle] - \mathbb{E}[Y\langle t|t\rangle]$, and it represents the *average wage structure effect*, the change in average earnings for group-t, if they faced the wage schedule of group-t' participants, $F_{Y_{t'}|X_{t'}}(y|x)$. On the other hand, (2.5), which is equivalent to $\mathbb{E}[Y\langle t'|t\rangle] - \mathbb{E}[Y\langle t'|t'\rangle]$, represents the *average composition effect* and reveals the change in average earnings for group-t' participants, if they had the same covariates distribution as group-t participants.

By the same reasoning, we define the *overall counterfactual distribution*

$$F_{Y\langle t\rangle}(y) \equiv \sum_{t' \in \mathcal{T}} F_{Y\langle t|t'\rangle}(y)\mathbb{P}(T = t') = \int_{\mathcal{X}} F_{Y_t|X_t}(y|x) dF_X(x).$$

The overall counterfactual earnings outcome $Y\langle t\rangle$ is a random variable generated by the marginal distribution $F_{Y\langle t\rangle}(y)$. The overall counterfactual distribution $F_{Y\langle t\rangle}(y)$ can be interpreted as the counterfactual distribution for all program participants if they all had the same wage schedule as group-t. Therefore, the *overall average wage structure effect* $\mathbb{E}[Y\langle t\rangle] - \mathbb{E}[Y\langle t'\rangle]$ is the difference between the average earnings if *everyone* got paid according to the wage schedule of group-t and the average earnings if *everyone* got paid according to the wage schedule of group-t'.

Parameters of interest. We define the *decomposition parameter*

$$\gamma_{t|t'} = \mathbb{E}[Y\langle t|t'\rangle]$$

to be the mean of the counterfactual distribution for $t, t' \in \mathcal{T}$. Next, we define the *overall parameter*

$$\beta_t = \mathbb{E}[Y\langle t\rangle]$$

to be the mean of the overall counterfactual distribution for $t \in \mathcal{T}$.

In our empirical application, we evaluate the WIA program by estimating the following counterfactual effects: the average wage structure effect $\mathbb{E}[Y\langle t|t'\rangle] - \mathbb{E}[Y\langle t'|t'\rangle]$, the average composition effect $\mathbb{E}[Y\langle t|t\rangle] - \mathbb{E}[Y\langle t|t'\rangle]$, and the overall average wage structure effect $\mathbb{E}[Y\langle t\rangle] - \mathbb{E}[Y\langle t'\rangle]$ for different levels of participation t, t'. In the next section, we show that the average wage structure effect $\mathbb{E}[Y\langle t|t'\rangle] - \mathbb{E}[Y\langle t'|t'\rangle]$ is interpreted as the average treatment effect for the treated and the overall average wage structure effect $\mathbb{E}[Y\langle t\rangle] - \mathbb{E}[Y\langle t'\rangle]$ is the average treatment effect under the a conditional independence assumption. The overall parameter $\beta_t = \mathbb{E}[Y(t)]$ is also known as the dose response function in the statistical literature or the average structural function in the econometrics literature.

2.2 Treatment Effects

This section relates the decomposition analysis with the treatment effect model. We use a standard treatment effects framework (Fortin, Lemieux, and Firpo (2011), Chernozhukov, Fernandez-Val, and Melly (2013)): the treatment status t is the realized value of a random treatment variable $T \in \mathcal{T}$, where the multi-valued treatment variable T is the level of participation received by the participants. The outcome of group-t Y_t is assumed to be the potential outcome for the *entire* population of interest, $Y(t)$. Using the potential outcomes framework, there is a sequence of potential outcomes $\{Y(t)\}_{t\in\mathcal{T}}$ for the population. For each individual i, we observe $Y_i = Y_i(t)$ if he receives treatment $T_i = t$. His other potential outcomes $Y_i(j)$ for $j \neq t$ are not observed. The observed outcome variable for the population can be expressed as $Y = \sum_{t=0}^{J} Y_t D_t$, where $D_t \equiv \mathbf{1}\{T = t\}$ is an indicator function of the multi-valued group status with $\sum_{t=0}^{J} D_t = 1$. The covariates for the population are denoted by $X = \sum_{t=0}^{J} X_t D_t$. Then we can write the actual conditional distribution function $F_{Y_t|X_t}(y|x) = F_{Y|T,X}(y|t,x) = F_{Y(t)|T,X}(y|t,x)$ and $F_{X_t}(x) = F_{X|T}(x|t)$ for each group $t \in \mathcal{T}$.

The counterfactual effects are well-defined statistical parameters by the common support Assumption 1. Additionally, they can also carry a causal interpretation under the following assumption.

Assumption 3 (Conditional Independence Distribution (CIA)) *$Y(t) \perp D(t)|X\ \forall t$.*

Note that, even when CIA does not hold, the above decomposition analysis is still valid. If CIA is assumed, the descriptive decomposition analysis will carry a *causal* interpretation. In the conventional treatment effect literature, the treatment effect for the treated $\mathbb{E}[Y(t) - Y(t')|T = t']$ is the effect of hypothetically assigning a different treatment level t to the subpopulation who have chosen t'.

$$F_{Y\langle t|t'\rangle}(y) = \int_{\mathcal{X}} F_{Y|X,T}(y|x,t)\ dF_{X|T}(x|t') = \int_{\mathcal{X}} F_{Y(t)|X,T}(y|x,t)\ dF_{X|T}(x|t')$$
$$= \int_{\mathcal{X}} F_{Y(t)|X,T}(y|x,t')\ dF_{X|T}(x|t') = F_{Y(t)|T}(y|t'),$$

where the third equality comes from the CIA assumption. That is to assume conditional on a set of covariates X, each individual chooses his or her treatment level T randomly over the whole choice set $\mathcal{T}$. The invariant conditional distribution Assumption 3 is replaced by the stronger CIA assumption, also known as selection on observables, ignorability, unconfoundedness or missing at random, which has been often used in the treatment effect literature. The common support Assumption 1 is known as the overlapping assumption, i.e. the propensity score $\mathbb{P}(T = t|X)$ is bounded away from zero almost surely.

For example, let t denote high-level participation and t' denote low-level participation where switching a participant from low-level participation to high-level is considered to be the "treatment." The overall counterfactual distribution $F_{Y\langle t\rangle}(y) = \int_{\mathcal{X}} F_{Y|XT}(y|x,t)dF_X(x) = F_{Y(t)}(y)$ is the marginal distribution of the potential outcome $Y(t)$. For example, the mean of the overall counterfactual distribution $\mathbb{E}[Y\langle t\rangle]$ reveals the *average treatment effect* $\mathbb{E}[Y(t)] - \mathbb{E}[Y(t')]$, the change in average earnings, if all participants switched from low-level participation to high-level participation. More importantly, the wage structure effect in (2.2) $\boldsymbol{F}_{Y\langle t'|t\rangle}(y) - F_{Y\langle t|t\rangle}(y)$, which, in our decomposition analysis, represents the counterfactual effect when the wage schedule changes, is equal to the *treatment effect for the treated* $F_{Y(t')|T}(y|t) - F_{Y(t)|T}(y|t)$, which is the impact on the earnings distribution if high-level participants (t) had participated in the program for the low-level. On the other hand, the composition effect in (2.3) $F_{Y\langle t'|t\rangle}(y) - F_{Y\langle t'|t'\rangle}(y)$, which represents the counterfactual effect of changing the covariate distribution from that of the short-term participation to that of the long-term participants, is equal to $F_{Y(t')|T}(y|t) - F_{Y(t')|T}(y|t')$, which is the difference in the earnings distribution between low-level participants (t') and high-level participants (t,) if they were both treated for a low-level.

2.3 Detailed composition effects

In this section, we further decompose the composition effect for different explanatory factors and provides an alternative method to perform the counterfactual analysis in DiNardo, Fortin, and Lemieux (1996) and Fortin, Lemieux, and Firpo (2011). For example, this detailed decomposition provides policymaker the information of how specific characteristic affects the outcome of different treated groups.

Recall that in Section 2.1, the composition effect interprets $F_{Y\langle t|t'\rangle}$ as the counterfactual outcome distribution of group-t if they had group-t''s characteristics distribution $F_{X_{t'}}(x)$. In this section, we further decompose the aggregate composition effect to isolate the contribution of different factors in the characteristics $X_t = (X_{t1}^\top, X_{t2}^\top)^\top$. More specifically, we counterfactually assign the conditional distribution of X_1 given X_2 of group-t' ($F_{X_{t'1}|X_{t'2}}$) to group-t, but X_2 remains its distribution of group-t ($F_{X_{t2}}$). That is, we perform a counterfactual experiment by changing the *conditional*, as opposed to the marginal, distribution of X_1. We focus on deposition based on sequential condoning arguments, that is, we consider the question: what would have happened to the outcome distribution if the distribution of X_1, but none of the other covariates, had changed from t to t'. Then we formally define the corresponding *factor counterfactual distribution* by

$$F_{Y\langle t|X_1 t'\rangle}(y) \equiv \int_{\mathcal{X}_2}\int_{\mathcal{X}_1} F_{Y_t|X_t}(y|x_1,x_2)\; d\boldsymbol{F}_{X_{t'1}|X_{t'2}}(x_1|x_2)\; dF_{X_{t2}}(x_2).$$

Again, the *factor counterfactual outcome* $Y\langle t|X_1 t'\rangle$ is a random variable generated by the cdf $F_{Y\langle t|X_1 t'\rangle}(y)$. So we can decompose the composition effect

$$F_{Y\langle t|t'\rangle}(y) - F_{Y\langle t|t\rangle}(y) = F_{Y\langle t|t'\rangle}(y) - F_{Y\langle t|X_1 t'\rangle}(y) \tag{2.6}$$

$$+ F_{Y\langle t|X_1 t'\rangle}(y) - F_{Y\langle t|t\rangle}(y). \tag{2.7}$$

The *factor effect* in (2.7) analyzes the change of the distribution of X_1 to that of group-t' but X_2 remains its distribution for group-t. The *remaining factor effect* in (2.6) accounts for the role of remaining attributes X_2.

Take the example in DiNardo, Fortin, and Lemieux (1996) who analyze the effects of institutional and labor market factors on the U.S. distribution of earnings on the period 1979 to 1988. The outcome of interest is earnings. The multi-valued treatment variable T indicates time: t is for year 1988 and t' is for year 1979. The factor X_1 is a dummy variable for the union status and X_2 is a vector of other attributes. So $F_{Y\langle t|X_1 t'\rangle}$ in (2.8) is the distribution of earnings that would have prevailed in 1988 if unionization, but none of the other attributes, had remained at its 1977 level. The factor composition effect in (2.7) extracts the impact of unionization, a factor of labor market institutions, on the earnings distributions between 1979 to 1988. The semiparametric procedure in DiNardo, Fortin, and Lemieux (1996) provides a visually clear representation of where in the density of earnings these various factors exert the greatest impact. Our nonparametric estimation procedure allows us to easily recover various distributional features.

We rearrange the factor counterfactual distribution as follows:

$$F_{Y\langle t|X_1 t'\rangle}(y) = \mathbb{E}\big[F_{Y|X_1X_2T}(y|x_1,x_2,t)W_{X_1t'}(X)\big|T=t\big], \text{ where} \tag{2.8}$$

$$W_{X_1t'}((x_1,x_2)) \equiv \frac{\mathbb{P}\big(T=t'\big|X=(x_1,x_2)\big)\mathbb{P}\big(T=t\big|X_2=x_2\big)}{\mathbb{P}\big(T=t\big|X=(x_1,x_2)\big)\mathbb{P}\big(T=t'\big|X_2=x_2\big)}.$$

Similar to the decomposition parameter $\gamma_{t|t'}$, we define the *factor parameter* $\lambda_{t|t'}$ to be the distributional feature defined by the factor counterfactual distribution $F_{Y\langle t|X_1 t'\rangle}$ in (2.8).

Definition 1 (Factor parameter) *Suppose a measurable function $m: \mathcal{Y}\times\Theta \mapsto \mathbb{R}^{d_m}$, where the parameter space $\Theta \subset \mathbb{R}^{d_\theta}$ and $d_m \geq d_\theta$. Define the* **factor parameter** *$\lambda_{t|t'} \in \Theta$ to be the causal distributional features of $Y\langle t|X_1 t'\rangle$ satisfying*

$$\mathbb{E}[m(Y\langle t|X_1 t'\rangle;\lambda_{t|t'})] = \int_{\mathcal{Y}} m(y;\lambda_{t|t'})\; dF_{Y\langle t|X_1 t'\rangle}(y) = 0$$

for any $t, t' \in \mathcal{T}$*.*

The estimation of $\lambda_{t|t'}$ is based on

$$\int_{\mathcal{Y}} m(y;\lambda) dF_{Y\langle t|X_1 t'\rangle}(y) = \mathbb{E}\left[m(Y;\lambda)\frac{D_t}{\mathbb{P}(T=t|X)} W_{X_1 t'}(X) \frac{\mathbb{P}(T=t|X)}{\mathbb{P}(T=t)} \right] = 0. \tag{2.9}$$

For the mean when $m(y;\lambda) = y - \lambda$, the factor parameter $\lambda_{t|t'} = \mathbb{E}[Y\langle t|X_1 t'\rangle] = \int\int \mathbb{E}[Y|X_1 = x_1, X_2 = x_2, T = t] dF_{X_{t'1}|X_{t'2}}(x_1|x_2)\ dF_{X_{t2}}(x_2)$ is the counterfactual mean of earnings that would have prevailed if (using the previous example) unionization, but none of the other attributes, had remained at its 1970 level.

3 Efficient Estimation and Inference

In this section we present the econometric theory behind our estimation and inference procedures. The general setup follows the multi-valued treatment effect model in Cattaneo (2010), extended to the treatment effect for the treated. Section 3.1 introduces the parameters of interest that are the distributional features of the counterfactual distributions defined in Section 2. Section 3.2 introduces the efficient estimators and presents their large sample properties. Section 3.3 presents inequality measures based on the counterfactual distributions. We illustrate our results by estimating quantile treatment effects for the treated. We derive a quantile process that weakly converges to a Gaussian process indexed by the quantile.

3.1 Parameters of interest

We define formally the distributional features of the counterfactual distribution $F_{Y\langle t|t'\rangle}$ and the overall counterfactual distribution $F_{Y\langle t\rangle}$ via a generic moment function m.

Definition 2 *Suppose a measurable function* $m : \mathcal{Y} \times \Theta \mapsto \mathbb{R}^{d_m}$*, where the parameter space* $\Theta \subset \mathbb{R}^{d_\theta}$ *and* $d_m \geq d_\theta$*. Consider any* $t, t' \in \mathcal{T}$*.*

1. *The* **decomposition parameter** $\gamma_{t|t'} \in \Theta$ *satisfies*

$$\mathbb{E}\left[m(Y\langle t|t'\rangle;\gamma_{t|t'})\right] = \int_{\mathcal{Y}} m(y;\gamma_{t|t'})\ dF_{Y\langle t|t'\rangle}(y) = 0, \textit{where}$$
$$F_{Y\langle t|t'\rangle}(y) = \int_{\mathcal{X}} F_{Y|XT}(y|x,t) dF_{X|T}(x|t').$$

2. *The* **overall parameter** $\beta_t \in \Theta$ *satisfies*

$$\mathbb{E}\left[m(Y\langle t\rangle;\beta_t)\right] = \int_{\mathcal{Y}} m(y;\beta_t)\ dF_{Y\langle t\rangle}(y) = 0, \textit{where } F_{Y\langle t\rangle}(y) = F_{Y|T}(y|t).$$

Definition 2 based on a generic moment function m covers various distributional features of interest. For the mean when $m(Y;\theta) = Y - \theta$, the decomposition parameter $\gamma_{t|t'} = \mathbb{E}[Y\langle t|t'\rangle] = \int_{\mathcal{X}} \mathbb{E}[Y|X=x, T=t]dF_{X|T}(x|t')$ is the mean of the counterfactual distribution $F_{Y\langle t|t'\rangle}$. When $m(Y;\theta) = \mathbf{1}\{Y \leq \theta\} - \tau$, $\gamma_{t|t'}$ is the τth quantile of the counterfactual distribution $F_{Y\langle t|t'\rangle}$. When $m(Y;\theta) = \mathbf{1}\{Y \leq y\} - \theta$, $\gamma_{t|t'}$ is the counterfactual distribution $F_{Y\langle t|t'\rangle}(y)$. Recall that in the treatment effect literature, $F_{Y\langle t|t'\rangle}$ is the distribution of the potential outcome $Y(t)$ for those who have been treated at t' under the unconfoundedness assumption. So $\gamma_{t|t'} - \gamma_{t'|t'}$ can be interpreted as the treatment effects for the treated. Similarly for the overall parameter, $F_{Y\langle t\rangle}$ is the distribution of the potential outcome $Y(t)$ for the population under the unconfoundedness assumption. So $\beta_t - \beta_{t'}$ is the overall treatment effect of switching from t' to t.

3.2 Efficient estimators

We introduce two estimators — the inverse probability weighting (IPW) and the efficient influence function (EIF) estimators. We follow and modify the estimation procedure proposed by Cattaneo (2010) for the overall treatment effect to estimate the decomposition parameter. The estimators are overidentified GMM estimators ($d_m \geq d_\theta$), which are convenient to conduct inference and implement hypothesis test with restrictions. Denote the object of interest to be a $(J+1) \times 1$ vector $\gamma_{t'} \equiv (\gamma_{0|t'}, ..., \gamma_{J|t'})^\top$ for the treated group-t'. We focus on one treated group-t' for simplicity. In general, we could consider all group-t' for $t' \in \mathcal{T}$ at the cost of notional complexity, i.e., $\gamma \equiv (\gamma_0^\top, ..., \gamma_J^\top)^\top$, a $(J+1)^2 \times 1$ vector.

At a preliminary step, we nonparametrically estimate the infinite-dimensional nuisance parameters — the propensity score $P_t(X) \equiv \mathbb{P}(T = t|X)$ and the conditional expectation of the moment $e_t(X) = \mathbb{E}\big[m(Y;\gamma_{t|t'})\big|T = t, X\big]$. The proposed estimators are also semiparametric doubly robust in the sense that the misspecification of either $P_t(X)$ or $e_t(X)$ does not affect the consistency (Graham, 2011).

The inverse probability weighting estimator uses the moment condition:

$$\mathbb{E}[m(Y\langle t|t'\rangle;\gamma_{t|t'})] = \mathbb{E}\left[m(Y;\gamma_{t|t'})\frac{D_t}{\mathbb{P}(T=t|X)}\frac{\mathbb{P}(T=t'|X)}{\mathbb{P}(T=t')}\right] = 0$$

by rewriting the definition of the decomposition parameter.[3] To define our GMM estimators, denote $|\cdot|$ to be the Euclidean norm given by $|A| = \sqrt{trace(A'A)}$ for any matrix A. Choose a $(J+1)d_\theta \times (J+1)d_m$ matrix $A_n = A + o_p(1)$ such that a weighting matrix $W = A'A$ guarantees the resulting estimator to be efficient. The inverse probability weighting (IPW) estimator is defined by

$$\hat{\gamma}^{IPW} \equiv \arg\min_{\theta \in \Theta^{J+1}} \big|A_n\ M^{IPW}_{t'n}(\theta, \hat{P}, \hat{p})\big| + o_p(n^{-1/2}), \text{ where}$$

$$M^{IPW}_{t'n}(\theta, P, p) \equiv \frac{1}{n}\sum_{i=1}^{n} m(Y_i;\theta)\left(\frac{D_0}{P_0(X_i)}, ..., \frac{D_J}{P_J(X_i)}\right)^\top \frac{P_{t'}(X_i)}{p_{t'}},$$

$$P_t(X) \equiv \mathbb{P}(T = t'|X), \text{ and } p_t \equiv \mathbb{P}(T = t).$$

[3] An alternative expression $\mathbb{E}\big[\mathbb{E}\big[m(Y;\gamma_{t|t'})\big|T=t, X\big]\big|T=t'\big] = 0$ motivates the regression estimator, for example, Hahn (1998), Chernozhukov, Fernandez-Val, and Melly (2013).

The propensity scores $P \equiv (P_0(X), ..., P_J(X))^\top$ are estimated nonparametrically by a multinormial logistic series estimator that satisfies Assumption A.NP in the Appendix. The probability of being treated at t $p_t \equiv \mathbb{P}(T = t)$ is estimated by the sample analog $\hat{p}_t = n^{-1} \sum_{i=1}^n D_{ti}$. Let $p \equiv (p_0, ..., p_J)^\top$.

The second estimator uses the efficient influence function derived in Lee (2015a). We define the main component of the efficient influence function to be $\psi_{t'}(Z; \gamma_{t'}, P, p, e(\gamma_{t'}))$, a $(J+1) \times 1$ vector whose t-th component is

$$\left(\frac{D_t}{P_t(X)} m(Y; \gamma_{t|t'}) \; + e_t(X; \gamma_{t|t'}) \left(\frac{D_{t'}}{P_{t'}(X)} - \frac{D_t}{P_t(X)} \right) \right) \frac{P_{t'}(X)}{p_{t'}}, \text{ where} \tag{3.1}$$
$$e_t(X; \gamma_{t|t'}) \equiv \mathbb{E}\big[m(Y; \gamma_{t|t'}) \big| T = t, X\big].$$

The conditional expectations of the moments $e(\gamma_{t'}) \equiv \big(e_0(\cdot; \gamma_{0|t'})^\top, ..., e_J(\cdot; \gamma_{J|t'})^\top\big)^\top$ are estimated nonparametrically by series estimators that satisfies Assumption A.NP in the Appendix. The efficient influence function (EIF) estimator is defined by

$$\hat{\gamma}^{EIF} \equiv \arg \min_{\theta \in \Theta^{J+1}} \big| A_n \; M^{EIF}_{t'n}(\theta, \hat{P}, \hat{p}, \hat{e}(\theta)) \big| + o_p(n^{-1/2}), \text{ where}$$
$$M^{EIF}_{t'n}(\theta, P, p, e(\theta)) \equiv \frac{1}{n} \sum_{i=1}^n \psi_{t'}(Z_i; \theta, P, p, e(\theta)).$$

The asymptotic behavior of the estimators $\hat{\gamma}^{IPW}$ and $\hat{\gamma}^{EIF}$ follows from the next assumption.

Assumption 4 *For all $t \in \mathcal{T}$: (a) $\mathbb{E}\big[m(Y\langle t|t'\rangle; \theta)^2\big] < \infty$ and $\mathbb{E}\big[m(Y\langle t|t'\rangle; \theta)\big]$ is differentiable in $\theta \in \Theta$ at $\gamma_{t|t'}$; and (b) Define the gradient matrix*

$$\Gamma_{*|t'} \equiv \begin{bmatrix} \Gamma_{0|t'} & \mathbf{0} & ... & \mathbf{0} \\ \mathbf{0} & \Gamma_{1|t'} & ... & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & ... & \Gamma_{J|t'} \end{bmatrix}, \textit{where } \; \Gamma_{t|t'} \equiv \frac{\partial}{\partial \theta^\top} \mathbb{E}\big[m(Y\langle t|t'\rangle; \theta)\big]\Big|_{\theta = \gamma_{t|t'}}$$

and $\mathbf{0}$ is a $d_m \times d_\theta$ matrix of zeros. The rank of $\Gamma_{*|t'}$ is $(J+1)d_\theta$.

Theorem 1 (Asymptotic Linear Representation) *Suppose Assumption 4 and all Assumptions in Appendix hold. Then,*

$$\hat{\gamma}^{IPW} - \gamma_{t'} = \hat{\gamma}^{EIF} - \gamma_{t'} + o_p(n^{-1/2}) = -\big(\Gamma^\top_{*|t'} W \Gamma_{*|t'}\big)^{-1} \Gamma^\top_{*|t'} W M^{EIF}_{t'n}(\gamma, P, p, e(\gamma_{t'})) + o_p(n^{-1/2})$$

The treatment effects and treatment effects for the treated are continuous transformations of the distributional features $\gamma_{t'}$, respectively. So a delta-method argument recovers any such collection of treatment effects.

Remark (Efficient estimators) *$\hat{\gamma}^{IPW}$ and $\hat{\gamma}^{EIF}$ are efficient for $\gamma_{t'}$ in the following two cases:*

1. *$d_m = d_\theta$ for the just-identified case, where $\hat{\gamma}^{IPW}$ solves $M^{IPW}_{t'n}(\hat{\gamma}^{IPW}, \hat{P}, \hat{p}, \hat{e}(\hat{\gamma}^{IPW})) = 0$ and $\hat{\gamma}^{EIF}$ solves $M^{EIF}_{t'n}(\hat{\gamma}^{EIF}, \hat{P}, \hat{p}, \hat{e}(\hat{\gamma}^{EIF})) = 0$.*

2. *The optimal weighting matrix $W = V^{-1}_{*|t'}$ is chosen.*

The natural plug-in estimator of $V_{*|t'}$ is given by

$$\hat{V}_{t'} = \frac{1}{n}\sum_{i=1}^{n} \psi_{t'}\big(Z_i; \hat{\gamma}, \hat{P}, \hat{p}, \hat{e}(\hat{\gamma})\big)\psi_{t'}\big(Z_i; \hat{\gamma}, \hat{P}, \hat{p}, \hat{e}(\hat{\gamma})\big)^{\top} \tag{3.2}$$

for some consistent estimator $\hat{\gamma}$ of $\gamma_{t'}$. The asymptotic covariance matrix and the optimal weighting matrix can be estimated consistently as Section 5.3 in Cattaneo (2010), so we do not repeat the proofs.

3.3 Inequality measures

We now consider inequality measures based on the counterfactual distribution $F_{Y\langle t|t'\rangle}$. We first obtain weak convergence of the counterfactual distribution estimator $\hat{F}_{Y\langle t|t'\rangle}(y)$ that is an empirical process indexed by $y \in \mathcal{Y}$. Once the weak convergence of distribution process is established, we can extend the results to the Hadamard-differentiable functionals of the distribution process. We are able to provide the limit distribution and uniform inference for estimating common inequality measures, and various distributional structural features; for example, quantile functions, the Lorenz curves, and the Gini coefficients (note that some inequality measures, such as the Gini coefficient, cannot be expressed as a parameter defined by the moment function m in the previous setup). The class of inequality measures that is Hadamard-differentiable has been studied recently in Firpo and Pinto (2016), Chernozhukov, Fernandez-Val, and Melly (2013), Donald and Hsu (2014), among others.

Consider the moment function to be a process indexed by the distributional threshold value $y \in \mathcal{Y}$, $m(\cdot; \theta, y) \equiv \big\{Y \mapsto \mathbf{1}\{Y \le y\} - \theta : y \in \mathcal{Y}\big\}$. This is a just-identified case in the previous setup. That is, for $t, t' \in \mathcal{T}$,

$$\begin{aligned}
\hat{F}^{IPW}_{Y\langle t|t'\rangle}(y) &= \frac{1}{n}\sum_{i=1}^{n} \frac{D_{ti}}{\hat{P}_t(X_i)}\mathbf{1}\{Y_i \le y\}\frac{\hat{P}_{t'}(X_i)}{\hat{p}_{t'}}, \\
\hat{F}^{EIF}_{Y\langle t|t'\rangle}(y) &= \hat{F}^{IPW}_{Y\langle t|t'\rangle}(y) + \frac{1}{n}\sum_{i=1}^{n} \hat{F}_{Y|TX}(y|t, X_i)\ \left(\frac{D_{t'i}}{\hat{P}_{t'}(X_i)} - \frac{D_{ti}}{\hat{P}_t(X_i)}\right)\frac{\hat{P}_{t'}(X_i)}{\hat{p}_{t'}}.
\end{aligned}$$

To shorten the notation, let the efficient influence function for estimating $F_{Y\langle t|t'\rangle}$ from (3.1) be

$$\psi_{t|t'}(Z; y) \equiv \left(\frac{D_t}{P_t(X)}\big(\mathbf{1}\{Y \le y\} - F_{Y\langle t|t'\rangle}(y)\big)\ + \big(F_{Y|TX}(y|t, X) - F_{Y\langle t|t'\rangle}(y)\big)\ \left(\frac{D_{t'}}{P_{t'}(X)} - \frac{D_t}{P_t(X)}\right)\right)\frac{P_{t'}(X)}{p_{t'}}.$$

Theorem 2 (Weak Convergence) *Suppose the conditions in Theorem 1 and Assumption 6 in the Appendix hold. For any $t, t' \in \mathcal{T}$, uniformly in $y \in \mathcal{Y}$,*

$$\begin{aligned}\sqrt{n}\big(\hat{F}^{IPW}_{Y\langle t|t'\rangle}(y) - F_{Y\langle t|t'\rangle}(y)\big) &= \sqrt{n}\big(\hat{F}^{EIF}_{Y\langle t|t'\rangle}(y) - F_{Y\langle t|t'\rangle}(y)\big) + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^{n}\psi_{t|t'}\big(Z_i; y\big) + o_p(1) \Longrightarrow \mathbb{G}_{t|t'}(y).\end{aligned}$$

The empirical processes converge weakly to a Gaussian process $\mathbb{G}_{t|t'}(\cdot)$ with mean zero and the covariance kernel $Cov\big(\mathbb{G}_{t|t'}(y_1), \mathbb{G}_{t|t'}(y_2)\big) = \lim_{n\to\infty}\mathbb{E}\big[\psi_{t|t'}\big(Z; y_1\big)\psi_{t|t'}\big(Z; y_2\big)\big]$ for $y_1, y_2 \in \mathcal{Y}$.

We can then implement the functional delta method on the Hadamard-differentiable functional of this distribution process. Denote $D_\theta \subset l^\infty(\mathcal{Y})$ to be a function space of bounded functions on $\mathcal{Y}$.

Corollary 1 (Functional Delta Method) *Assume the conditions in Theorem 2 hold. Consider the parameter θ as an element of a parameter space $D_\theta \subset l^\infty(\mathcal{Y})$ with D_θ containing the true value $\theta_0(y) \equiv F_{Y\langle t|t'\rangle}(y)$. Suppose a functional $\Gamma(\theta)$ mapping D_θ to $l^\infty(\mathcal{W})$ is Hadamard differentiable*[4] *in θ at θ_0 with derivative Γ'_θ. Then*

$$\begin{aligned}&\Big|\sqrt{n}\big(\Gamma(\hat{\theta}_0)(w) - \Gamma(\theta_0)(w)\big) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Gamma'_\theta(\psi_{t|t'}(Z_i; y))(w)\Big| = o_p(1) \\ &\sqrt{n}\big(\Gamma(\hat{\theta}_0)(w) - \Gamma(\theta_0)(w)\big) \Longrightarrow \Gamma'_\theta(\mathbb{G}_{t|t'})(w) \equiv G(w)\end{aligned}$$

where G is a Gaussian process indexed by $w \in \mathcal{W}$ in $l^\infty(\mathcal{W})$, with mean zero and covariance kernel defined by the limit of the second moment of $\Gamma'_\theta(\psi_{t|t'}(Z_i; y))$.

We illustrate Corollary 1 by letting Γ be the τ-quantile operator on $\theta_0(y) \equiv F_{Y\langle t|t'\rangle}(y)$, i.e., Γ is a generalized inverse $\theta_0^{-1} : (0, 1) \to \mathcal{Y}$ given by $\theta_0^{-1}(\tau) = \inf\{y : \theta_0(y) \geq \tau\}$. For the quantile treatment effects for the treated, $\theta_0^{-1}(\tau)$ is the τth-quantile function of $Y(t)$ for the treated t', denoted by $Q_\tau = Q_\tau(Y(t)|T = t') = F^{-1}_{Y(t)|T}(\tau|t')$. Hadamard-differentiability requires $F_{Y\langle t|t'\rangle}(y)$ to be continuously differentiable at the τth-quantile, with the derivative being strictly positive and bounded over a compact neighborhood. Additional assumptions might be needed for different policy functionals. For instance, Bhattacharya (2007) gives regularity conditions for Hadamard-differentiability of Lorenz and Gini functionals.

Corollary 2 (Quantile treatment effect for the treated) *Assume the conditions in Corollary 1.*

[4] See, for example, van der Vaart (2000) for dentition: let Γ be a Hadamard-differentiable functional mapping from $\mathcal{F}$ to some normed space $\mathcal{E}$, with derivative Γ'_f, a continuous linear map $\mathcal{F} \mapsto \mathcal{E}$. For every $h_n \to h$ and $f \in \mathcal{F}$,

$$\lim_{u\to 0}\frac{1}{u}\Big(\Gamma(f + uh_n) - \Gamma(f)\Big) = \Gamma'_f(h).$$

Then uniformly in $\tau \in [a,b] \subset (0,1)$,

$$\sqrt{n}\Big(\hat{Q}_\tau - Q_\tau\Big) = \frac{1}{n}\sum_{i=1}^{n}\psi^{Q}_{t|t'}(Z_i;\tau) + o_p(1)$$

that is an empirical process indexed by τ converging weakly to a Gaussian process with mean zero and covariance matrix $\mathbb{E}\big[\psi^{Q}_{t|t'}(Z;\tau_1)\psi^{Q}_{t|t'}(Z;\tau_2)\big]$ for any $\tau_1,\tau_2 \in [a,b]$. The influence function is

$$\begin{aligned}\psi^{Q}_{t|t'}(Z_i;\tau) \equiv\ & \frac{-D_{ti}}{p_{t'} f_{Y(t)|T}(Q_\tau|t')}\ \frac{P_{t'}(X_i)}{P_t(X_i)}\Big(\mathbf{1}\{Y_i \le Q_\tau\} - F_{Y|TX}(Q_\tau|t,X_i)\Big)\\ &+ \frac{-D_{t'i}}{p_{t'} f_{Y(t)T}(Q_\tau|t')}\Big(F_{Y|TX}(Q_\tau|t,X_i) - \tau\Big).\end{aligned}$$

To carry out point-wise inference, the asymptotic variance can be estimated by $n^{1}\sum_{i=1}^{n}\hat{\psi}^{Q}_{t|t'}(Z_i;\tau)^2$ as (3.2) in the previous section.

3.4 Detailed composition effect

3.4.1 Semiparametric efficiency bound

We consider a $(J+1)\times 1$ vector of interest for the detailed composition effect $\lambda_{t'} \equiv (\lambda_{0|t'},...,\lambda_{J|t'})^\top$. Define $\psi_{X_1t'}(Z;\lambda_{t'},p,e(\lambda_{t'}),P_{-1})$ to be a $(J+1)\times 1$ vector whose t-th element is

$$\begin{aligned}\equiv\Bigg[&\Bigg(\frac{D_t}{P_t(X)}m(Y;\lambda_{t|t'})\ + e_t(X;\lambda_{t|t'})\ \Bigg(\frac{D_{t'}}{P_{t'}(X)} - \frac{D_t}{P_t(X)}\Bigg)\Bigg)\frac{P_{t'}(X)}{\mathbb{P}(T=t'|X_2)}\\ &+\mathbb{E}\big[m(Y;\lambda_{t|t'})W_{X_1t'}((X_1,X_2))\big|T=t,X_2\big]\Bigg(\frac{D_t}{\mathbb{P}(T=t|X_2)} - \frac{D_{t'}}{\mathbb{P}(T=t'|X_2)}\Bigg)\Bigg]\frac{\mathbb{P}(T=t|X_2)}{p_t}\end{aligned} \tag{3.3}$$

for $t \in \mathcal{T}$. Denote the propensity score given X_2 by $P_{-1}(x_2) \equiv \big(\mathbb{P}(T=0|X_2=x_2),...,\mathbb{P}(T=J|X_2=x_2)\big)^\top$. The following assumption guarantees the existence of the efficiency bound for $\lambda_{t'}$.

Assumption 5 *For all $t \in \mathcal{T}$: (a) $\mathbb{E}\big[m(Y\langle t|X_1t'\rangle;\theta)^2\big] < \infty$ and $\mathbb{E}\big[m(Y\langle t|X_1t'\rangle;\theta)\big]$ is differentiable in $\theta \in \Theta$ at $\lambda_{t|t'}$; and (b) Define the gradient matrix*

$$\Gamma_{*|X_1t'} \equiv \begin{bmatrix}\Gamma_{0|X_1t'} & \mathbf{0} & ... & \mathbf{0}\\ \mathbf{0} & \Gamma_{1|X_1t'} & ... & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{0} & \mathbf{0} & ... & \Gamma_{J|X_1t'}\end{bmatrix},\ \textit{where}\ \ \Gamma_{t|X_1t'} \equiv \frac{\partial}{\partial\theta^\top}\mathbb{E}\big[m(Y\langle t|X_1t'\rangle;\theta)\big]\Big|_{\theta=\lambda_{t|t'}}$$

and $\mathbf{0}$ is a $d_m \times d_\theta$ matrix of zeros. The rank of $\Gamma_{*|t'}$ is $(J+1)d_\theta$.

Theorem 3 (Factor parameter) *Suppose Assumptions 1 and 5 hold. Then the efficient influence function of $\lambda_{t'}$ is given by*

$$\Psi_{X_1t'} = -\big(\Gamma^\top_{*|X_1t'} V^{-1}_{*|X_1t'} \Gamma_{*|X_1t'}\big)^{-1} \Gamma^\top_{*|X_1t'} V^{-1}_{*|X_1t'} \psi_{X_1t'}$$

where $V_{*|X_1t'} = var[\psi_{X_1t'}]$. The semiparametric efficiency bound for any regular estimator of $\lambda_{t'}$ is given by $V^*_{X_1t'} = \big(\Gamma^\top_{*|X_1t'} V^{-1}_{*|X_1t'} \Gamma_{*|X_1t'}\big)^{-1}$.

3.4.2 Efficient estimators for the detailed composition effect

The factor parameter $\lambda_{t'}$ can be estimated by

$$\begin{aligned}
\hat{\lambda}^{IPW} &= \arg\min_{\theta\in\Theta^{J+1}} \big|A_n\ M^{IPW}_{X_1t'n}(\theta, \hat{P}, \hat{p})\big| + o_p(n^{-1/2}), \text{where}\\
M^{IPW}_{X_1t'n}(\theta, P, p) &= \frac{1}{n}\sum_{i=1}^n m(Y_i;\theta)\left(\frac{D_0}{P_0(X_i)}, ..., \frac{D_J}{P_J(X_i)}\right)^\top \frac{P_t(X_i)}{p_t} W_{X_1t'}(X_i);\\
\hat{\lambda}^{EIF} &= \arg\min_{\theta\in\Theta^{J+1}} \big|A_n\ M^{EIF}_{X_1t'n}(\theta, \hat{P}, \hat{p}, \hat{e}(\theta), \hat{P}_{-1})\big| + o_p(n^{-1/2}), \text{where}\\
M^{EIF}_{X_1t'n}(\theta, P, p, e(\theta), P_{-1}) &= \frac{1}{n}\sum_{i=1}^n \psi_{X_1t'}(Z_i;\theta, P, p, e(\theta), P_{-1}).
\end{aligned}$$

The IPW estimator is based on (2.9). The additional weight $\mathbb{P}(T = t|X_2)/\mathbb{P}(T = t'|X_2)$ can be estimated similarly as the propensity score $P_J(X)$ for $j \in \mathcal{T}$. The additional regression of $\mathbb{E}[m(Y;\lambda)W_{X_1t'}|T = t, X_2]$ in (3.3) can be estimated similarly as $e_t(X;\lambda)$. The asymptotic linear representation for these estimators can be derived similarly as Theorem 1. We do not repeat the proofs.

4 Empirical Application

4.1 The Workforce Investment Act (WIA) Programs

Replacing the Job Training Partnership Act (JTPA), the Workforce Investment Act (WIA) of 1998 has two main goals. First, as stated in the Act, a main goal is to "...consolidate, coordinate, and improve employment, training, literacy, and vocational rehabilitation programs in the United States..." by reforming the former public workforce programs that had become "fragmented" and "uncoordinated." Specifically, the WIA established the largest network of public-financed career service programs and unified them to be available at over 3,000 One-Stop Career Centers around the country. Second, the Act established 3 flagship programs focusing on assessment, counseling, job readiness skills, occupational skills and trainings: WIA Adult program, WIA Dislocated Workers program and WIA Youth program. The WIA programs are an integral part of Employment and Training Administration (ETA) under the U.S. Department of Labor. In 2010, the programs serviced more than 7 million Adult workers and 1.5 million dislocated workers nationwide. In this paper, we focus on the WIA Adult and Dislocated programs.

An individual is eligible for WIA Adult if he or she is age 18 and older who are unemployed at time of application or who are under-employed (in a job earning $10.10 or less) or whose family meets adult low income guidelines. Dislocated workers are officially defined by meeting one of the following criteria: (1) has been laid off or terminated, or received notice of termination or lay off and is unlikely to return to previous industry of occupation, (2) has been terminated or laid off, or has received a notice of termination or lay off, as a result of permanent closure of, or substantial layoff at a plant or facility, (3) was self-employed and now unemployed because of a natural disaster, (4) was self-employed (including farmer, rancher, or fisherman), but is unemployed as a result of general economic conditions in the community in which he or she resides or because of a natural disaster, or (5) is a displaced homemaker.

WIA career services are offered at three levels. All individuals entering the program receive the "core" services, which include staff-assisted job search and placement, labor market information, and basic counseling. After that, staff may recommend the participants to receive "intensive" services, which involves more comprehensive assessment and counseling, career planning, and possibly some short courses. Participants may then be recommended for "training" services, which may be on-the-job training with local employers and apprenticeships in different fields, or educational training programs in vocational schools and community college using vouchers authorized by WIA program staffs. These are the three levels of participations that we consider here. The participants can be linked to job opportunities in their communities. In our data, about 2/3 of training recipients receive some kind of credentials. Although participation in WIA is voluntary, access is restricted. Program staffs must admit participants and authorize any services that are provided.

Previous literature on WIA focuses on comparing the earnings outcomes of the program participants and non-participants. Heinrich, Mueser, and Troske (2009) compare the earnings of WIA participants to those of UI claimants and Employment Service (ES) program participants for 16 quarters after program entries. They find that on average, female WIA participants earns $482-$638 more per quarter than the comparison group while male WIA participants earns $320-$692 more per quarter. Hollenbeck, Schroeder, King, and Huang (2005) compare the earnings of WIA participants to those of Employment Service (ES) program participants for 8 quarters after leaving the programs. They find that on average, female WIA participants earns $887 more per quarter than the comparison group and male WIA participants earns $773 more per quarter. Overall, previous studies have shown that the WIA programs have positive effects on participants' earnings outcomes. However, as we previously argued, while active labor market programs can improve participants' labor market outcomes by providing training and assistance they need to obtain higher earnings and find more suitable jobs, participants may face "lock-in" effects as longer participation may yield unsatisfying labor market outcomes due to lost labor market experience and employers may view this as a negative signal. From a policy point of view, while active labor market programs can improve the welfare of the participants, they can be very costly. In the raw WIASRD data, program participation varies from less than 1 month to more than 4 years while the expenditure on each participant exiting the program ranges from about $1000 to $15000. Therefore, to better evaluate the program and to give suggestions to improve their performance and cost-efficiency, we focus on studying how different lengths of participation in the WIA Adult program affect participants' post-program earnings outcomes.

Unconfoundedness. The key identifying assumption for our estimates to carry a causal interpretation in the treatment effect framework is selection on observables; that is, conditional on a set of observable covariates, selection into different levels of treatment is random. This assumption can be justified by the following: first, we have a rich set of covariates which includes demographic information, unemployment insurance status, pre-, during, and post-participation labor market experience (earnings and industry worked) for up to 16 quarters, to account for the unobservables. It is likely that unobservable characteristics such as motivation, self-esteem, or family condition are captured by the individual labor market history. Second, although participation in WIA is voluntary, access is restricted. Program staffs must admit participants and authorize any services that are provided. It is likely the recommended services provided to the participants, which is closely linked to the duration of participation in the program, are contingent on their prior labor market experience such as earnings, which are accounted for in our set of covariates. Finally, the outcome variable we consider in this paper is the the earnings differential between after leaving the program and prior to entering the program. This difference-in-difference specification, along with the above justification for unconfoundedness, is also used in Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) and Kluve, Schneider, Uhlendorff, and Zhao (2012), allowing us to account for time-invariant factors that may have influenced selection. Finally, by focusing only on participants, we expect to avoid the likely more difficult issue of selection *into* the program (as discussed, for example, in Andersson, Holzer, Lane, Rosenblum, and Smith (2016)).

4.2 Data and descriptive statistics

Our dataset comes from three sources: the annual Workforce Investment Act Standardized Record Data (WIASRD), the Unemployment Insurance data, and the Unemployment Insurance Wage Record data. The WIASRD dataset was primarily collected in December 2007 by the state workforce agencies, as requested by the US Department of Labor, for evaluating federal-funded WIA activities. Agreements were reached and data were provided by twelve states: Connecticut, Indiana, Kentucky, Maryland, Minnesota, Missouri, Mississippi, Montana, New Mexico, Tennessee, Utah, and Wisconsin. This dataset includes individual-level information on the time of program entry and exit (month and year), qualification status (adult or dislocated worker), and detailed demographic characteristics such as age, race, level of education, gender, disability and veteran status of all participants entering the program from July 2003 to June 2005 in nine of the above twelve states.

More importantly, each individual in the WIASRD dataset is assigned a random identification number that can be matched with other administrative data. Thus, we cross-reference the WIASRD dataset with the Unemployment Insurance data and the Unemployment Insurance Wage Record data. The Unemployment Insurance (UI) data cover all individuals who filed an unemployment insurance claim and contain demographic information of claimants and the sum of insurance payments received. The Unemployment Insurance Wage Record (UIWR) data provide quarterly earnings from all employees in unemployment insurance-covered firms. Using the unified identification numbers, we match the UI data and the (UIWR) data with the WIASRD to compile information on detailed labor market experience of the WIA participants including earnings and employment status before and after program participation. All earnings are then adjusted for inflation in 2006 Q1 dollars.

Combining the information from the above three sources, we obtain detailed demographic information, unemployment insurance status, pre- and post-participation labor market experience for our sample of WIA Adult and Dislocated program participants. Then, observations with missing entry and exit dates are dropped. We also discard observations with length of participation less than one month (those who entered and exited the program within the same month). We restrict our sample to be participants between the ages of 18-65. In addition, in order to consider the effect of the level of program participation on labor market outcomes, information on post-program labor market experience is required. Therefore, we also discard observations with an exit date later than June 2007 because the latest earnings data we have is the second quarter of 2007. The variation of program participation ranges from less than one month to more than four years. Our final samples comprise of observations from 62,852 WIA Adult and 40,657 Dislocated program participants. Table 1 presents the basic summary statistics for our sample of WIA participants. The length of participation has a mean of 7.22 months. The participants were mostly in their 30's and had obtained a high school degree or equivalent[5]. As in previous studies of WIA programs (for example, see Heinrich, Mueser, and Troske (2009)), we separate the analysis for male and female because the labor market activities can be very different for reasons such as fertility, marriage, and household production[6]. Figure 1 shows the trajectory of quarterly earnings before and after treatment, for the entire Adult sample, and by gender. Overall, we do not observe large pre-program differences in earnings trajectories between participants at different levels.

4.3 Estimation and Results

The primary goal of this paper is to assess the heterogeneity of WIA participants' post-program earnings outcomes arising from different levels of participation. The decomposition analysis described in Section 2 allows us to explore the role of the wage structure and composition effects associated with different levels of participation in attributing to the difference in the distribution of earnings.

We define the *earnings outcome* to be the difference in earnings between 8 quarters after leaving the program and 4 quarters prior to entering the program. Given the variation of participation observed in the data, we aim to assess the heterogeneity of participants' post-program earnings outcomes arising from different levels of participation, in terms of the services received: core ($T = 1$), intensive ($T = 2$) and training ($T = 3$) services. In particular, we explore the role of wage structures and participants' characteristics distributions associated with different levels of participation in attributing to the difference in the *earnings outcome* distributions. The set of covariates includes age, gender, years of education,

[5]The current version of the paper only analyses the Adult WIA program participants.

[6]We also plan to separate our analysis for different racial group as they display substantial differences in levels of participation.

veteran status, disability status, and pre-participation earnings, 5-8 quarters prior to entering the program. We use the notation for the treatment effects model to present our objects of interest, i.e., the overall parameter $\beta_t \equiv \mathbb{E}[Y\langle t\rangle] = \mathbb{E}[Y(t)]$ for the average treatment effect (ATE) and the decomposition parameter $\gamma_{t|t'} \equiv \mathbb{E}[Y\langle t|t'\rangle] = \mathbb{E}[Y(t)|T = t']$ for the average treatment effect for the treated (ATET).

Our estimation procedure extends Cattaneo (2010) and Cattaneo, Drukker, and Holland (2013) who estimate the average treatment effect for the population to also include multi-valued treatment effect for the treated, in particular, ATET and QTET. First, we estimate nonparametrically the probability of treatment for each individual i, given their characteristics, i.e., the propensity scores $P_t(X_i) \equiv \mathbb{P}(T = t|X = X_i)$ for $t \in \{1, 2, 3\}$ and $i = 1, .., n$. We use a Multinomial Logistic Series Estimator, where the order of the polynomial is selected using the Akaike Information Criterion (AIC). Given the estimated propensity scores $\hat{P}_t(X_i)$, we select the common support region for estimation following Flores, Flores-Lagunes, Gonzalez, and Neumann (2012): for each group-t, we find the minimum and maximum estimated propensity scores:

$$p_t^{min} \equiv \min_{\{i:T_i=t\}} \hat{P}_t(X_i) \quad \text{and} \quad p_t^{max} \equiv \max_{\{i:T_i=t\}} \hat{P}_t(X_i).$$

Define the *support region* for t to be the subpopulation whose $\hat{P}_t(X_i)$ bounded between p_t^{min} and p_t^{max}: $S_t \equiv \{i : \hat{P}_t(X_i) \in [p_t^{min}, p_t^{max}]\}$. The *common support region* is the intersection of the support regions for all $t \in \mathcal{T}$: $CS \equiv \cap_{t\in\mathcal{T}} S_t$. Observations that fall outside of the common support region is dropped.

The means $\mathbb{E}[Y\langle t\rangle]$ and $\mathbb{E}[Y\langle t|t'\rangle]$ are estimated by:

$$\hat{\beta}_t = \hat{\mathbb{E}}[Y(t)] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_{ti}}{\hat{P}_t(X_i)}Y_i - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - 1\Big)\hat{e}_t(X_i)\right)$$

$$\hat{\gamma}_{t|t'} = \hat{\mathbb{E}}[Y(t)|T = t'] = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_{ti}}{\hat{P}_t(X_i)}Y_i - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - \frac{D_{t'i}}{\hat{P}_{t'}(X_i)}\Big)\hat{e}_t(X_i)\right)\frac{\hat{P}_{t'}(X_i)}{\hat{p}_{t'}},$$

where $e_t(X_i) \equiv \mathbb{E}[Y|T = t, X = X_i]$ is also estimated using polynomial-regression series estimators with AIC to select the order of the polynomial. Finally, $p_t \equiv \mathbb{P}(T = t)$ is estimated by the sample analogue $\hat{p}_t = n^{-1}\sum_{i=1}^{n} D_{ti}$.

The τth-quantiles $Q_\tau(Y\langle t\rangle)$ and $Q_\tau(Y\langle t|t'\rangle)$ are estimated by:

$$\hat{\beta}_t = \hat{Q}_\tau(Y(t)) = \arg\min_{q\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_{ti}}{\hat{P}_t(X_i)}(\mathbf{1}\{Y_i \le q\} - \tau) - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - 1\Big)\hat{e}_t(X_i; q)\right)\right|$$

$$\hat{\gamma}_{t|t'} = \hat{Q}_\tau(Y(t)|T = t') = \arg\min_{q\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_{ti}}{\hat{P}_t(X_i)}(\mathbf{1}\{Y_i \le q\} - \tau) - \Big(\frac{D_{ti}}{\hat{P}_t(X_i)} - \frac{D_{t'i}}{\hat{P}_{t'}(X_i)}\Big)\hat{e}_t(X_i; q)\right)\frac{\hat{P}_{t'}(X_i)}{\hat{p}_{t'}}\right|,$$

where $\hat{e}_t(X_i; q) = \hat{\mathbb{E}}[\mathbf{1}\{Y \le q\}|T = t, X = X_i] - \tau$.

4.3.1 Estimation results and discussions

Figures 2-9 in the Appendix present the estimation results for the overall parameter β_t and the decomposition parameter $\gamma_{t|t'}$ for the entire sample and by gender group.

We first focus on the *overall effects*. Figure 2 depicts the estimated overall parameters for the mean and selected quantiles. We interpret our results under the familiar treatment effects model when the unconfoundedness assumption holds. For example, $\mathbb{E}[Y(t)] - \mathbb{E}[Y(t')]$ is the *average treatment effect*, that is the change in average earnings if all participants switch from level t' to t. The average earnings are positive for all groups for the mean, median and upper quantiles, but negative for the lower ones. Then, even though, on average, all groups experienced some improvement in earnings between 4 quarters prior to program entry to 8 quarters after program exit, there is heterogeneity across the distribution of the outcome. The magnitude of the average earnings gain, however, are different across gender groups and the different levels of participation, as we can see from the different contrast displayed in Figure 3. The average earnings gains are higher for female participants compared to males. Also, the relationship between the levels of participation and earnings outcomes is monotonic. For all gender groups, the core-service participants experience the smallest earnings gain. We find that females gain the most from training-services participation. Our results suggest that there are no "lock-in" effects for extended participation in the WIA Adult program. That is, longer participation leads to higher earnings outcomes. All estimates are displayed in Table 2 in the Appendix.

We next analyze *decomposition effects*. Figure 4 presents the quantile function for each treatment group. Then, figure 5 depicts the results for the decomposition parameters $\gamma_{t|t'} = \mathbb{Q}[Y\langle t|t'\rangle] = \mathbb{Q}[Y(t)|T = t']$. Again, we can interpret our estimation results in both treatment effect framework and decomposition analysis framework. For example, $\mathbb{Q}_\tau[Y(3)|T = 1] - \mathbb{Q}_\tau[Y(1)|T = 1]$ is the *quantile treatment effect for the treated*, which is the change in earnings for the core-service participants if they extend their participation to training-service, evaluated at quantile τ. The distribution of earnings outcomes between two groups can be decomposed into *wage structure effects* and *composition effects*. Our estimates suggest that all composition effects are negligible for all groups. That is, we find that the wage structure (WS) component accounts pretty much for all of the observed differences between any treatment group.

5 Conclusion

This paper presents a decomposition analysis for multi-valued treatment effect in the program evaluation context. We apply the methodology to study how different levels of participation in WIA program affect participants' post-program earnings outcomes. In particular, we explore the role of wage structures and participants' characteristics distributions associated with different levels of participation in attributing to the difference in observed earnings distributions. When the level of participation in the WIA program is considered as a multi-valued treatment and is chosen by participants randomly conditional on their characteristics (the unconfoundedness assumption), the decomposition analysis reveals causal effects, such as treatment effects for the treated. For estimation, we propose efficient nonparametric estimators for multi-valued treatment effects. We further propose an efficient estimator for the detailed composition effects by isolating the impacts of some specific explanatory factors, similar to DiNardo, Fortin, and Lemieux (1996) and Fortin, Lemieux, and Firpo (2011). That is, the counterfactual is based on the conditional distribution of one factor given other explanatory covariates while other explanatory covariates remain the same distribution. Our results suggest that the heterogeneity in the level of participation is an important dimension to investigate for program evaluation. In future work, we plan to estimate the quantile counterfactual effects and quantile treatment effect for the treated to provide a more comprehensive assessment for the program. The results of this paper, both theoretical and empirical, provide rigorous assessment of intervention programs and relevant suggestions to improve the performance and cost-effectiveness of these programs.

Table 1: Summary Statistics

	Age	Female	Education	Disable	Veteran	UI	Black	MoP
Total	31.63	0.56	12.27	0.09	0.11	0.36	0.48	7.23
$T=1$	31.08	0.54	12.21	0.11	0.14	0.35	0.55	6.50
$T=2$	34.26	0.63	12.40	0.08	0.09	0.42	0.43	7.63
$T=3$	31.86	0.61	12.40	0.05	0.07	0.34	0.30	9.24

Table 2: Treatment Effects Contrasts

	All	s.e	Females	s.e.	Males	s.e.
T3-T1						
Q10	659.2	75.1	749.7	87.9	589.7	150.3
Q25	632.3	21.4	613.3	23.3	707.3	35.7
Median	1121.2	49.9	1228.6	59.0	987.2	92.2
Mean	1409.8	43.9	1580.2	55.5	1197.5	65.9
Q75	2175.2	63.5	2326.4	100.1	1933.6	127.7
Q90	2643.0	101.8	2924.8	113.4	2257.6	156.1
T3-T2						
Q10	440.6	108.1	535.8	117.9	337.6	201.1
Q25	526.2	44.0	415.6	55.4	680.0	74.3
Median	798.1	74.0	802.5	87.0	793.4	122.9
Mean	939.5	66.4	1003.2	74.2	832.9	113.0
Q75	1372.2	131.2	1191.1	161.5	1235.0	225.2
Q90	1599.9	142.0	1642.3	152.9	1416.2	250.4
T2-T1						
Q10	218.6	98.2	213.9	92.4	252.1	167.8
Q25	106.1	53.2	197.7	70.3	27.3	84.4
Median	323.1	41.6	426.1	48.0	193.8	61.4
Mean	470.4	52.9	577.0	59.6	364.6	92.7
Q75	803.0	102.9	1135.3	101.6	698.6	153.0
Q90	1043.1	102.0	1282.4	102.2	841.4	209.2

Appendix

A Tables and Figures

Figure 1: Quarterly Earnings: All Sample, Females and Males

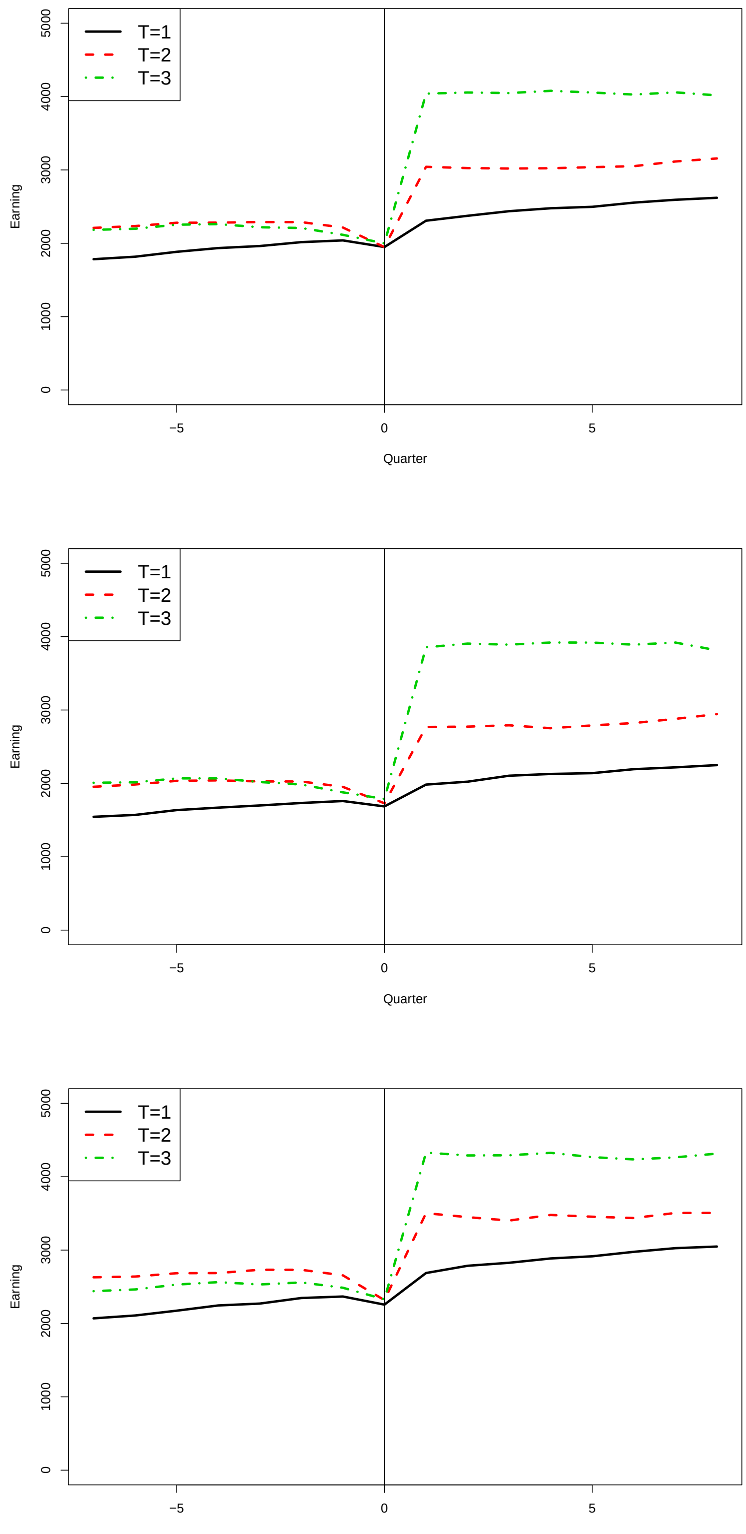


Figure 2: Treatment Effects: All Sample, Females and Males

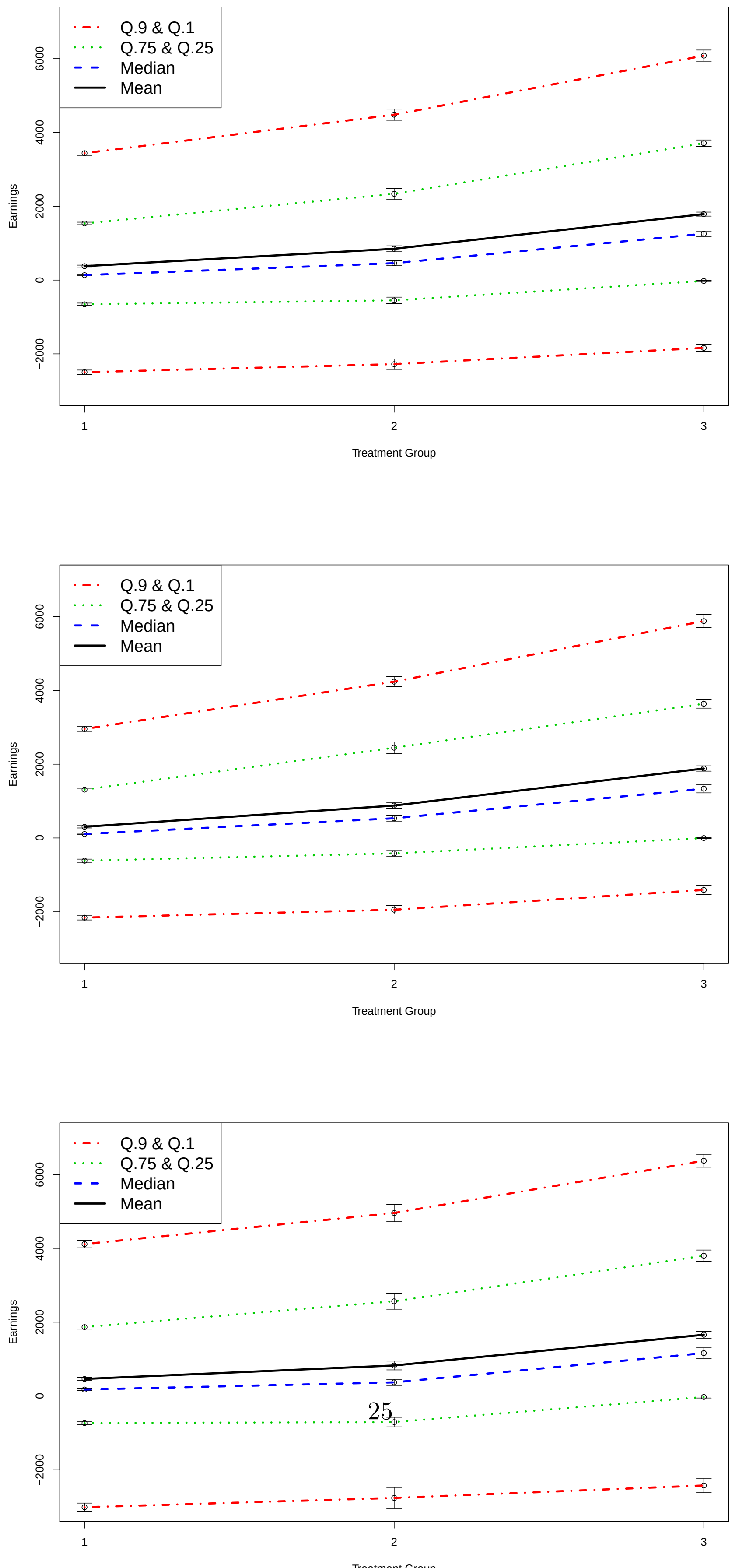


Figure 3: Treatment Effects (contrasts): All Sample, Females and Males

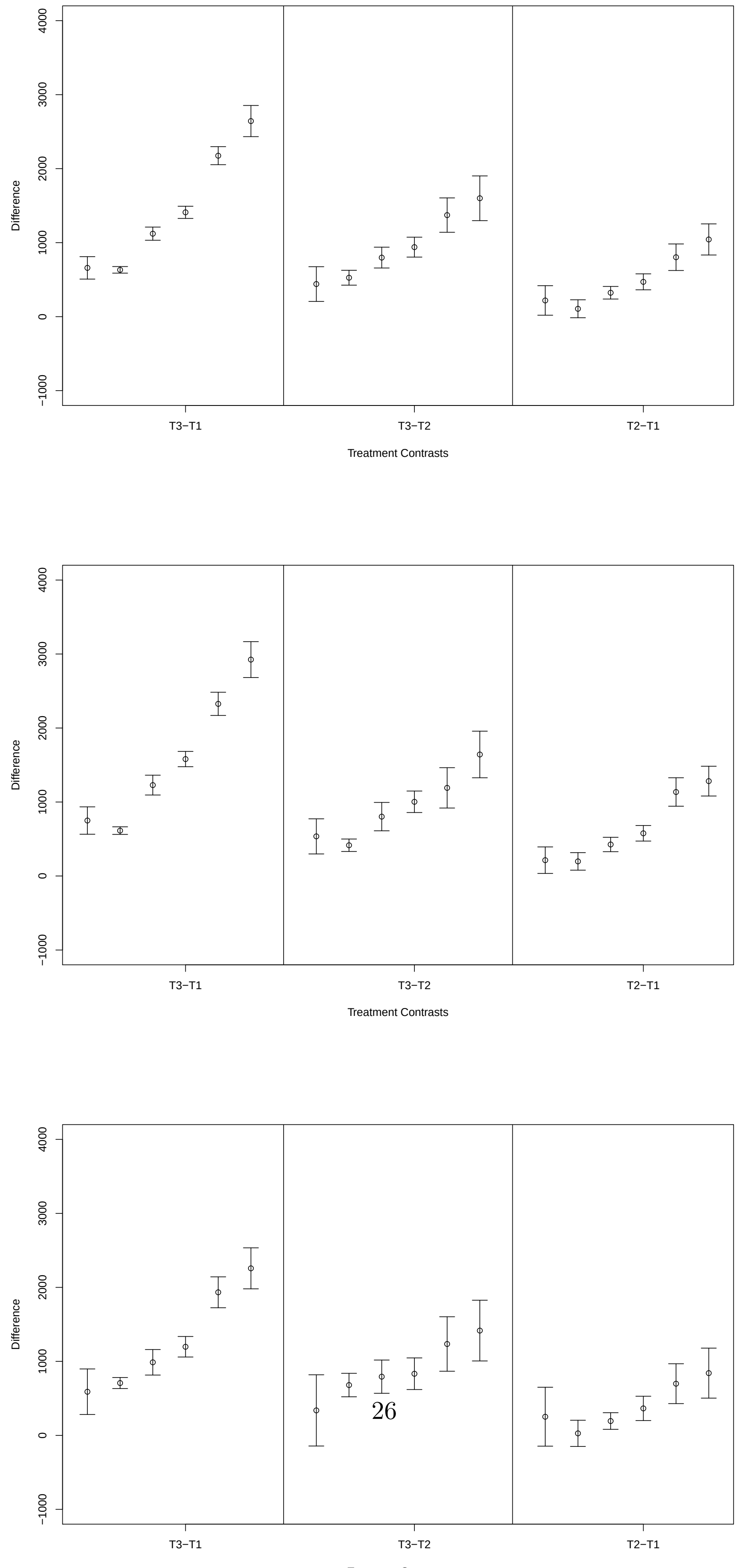


Figure 4: QTE: All Sample, Females and Males

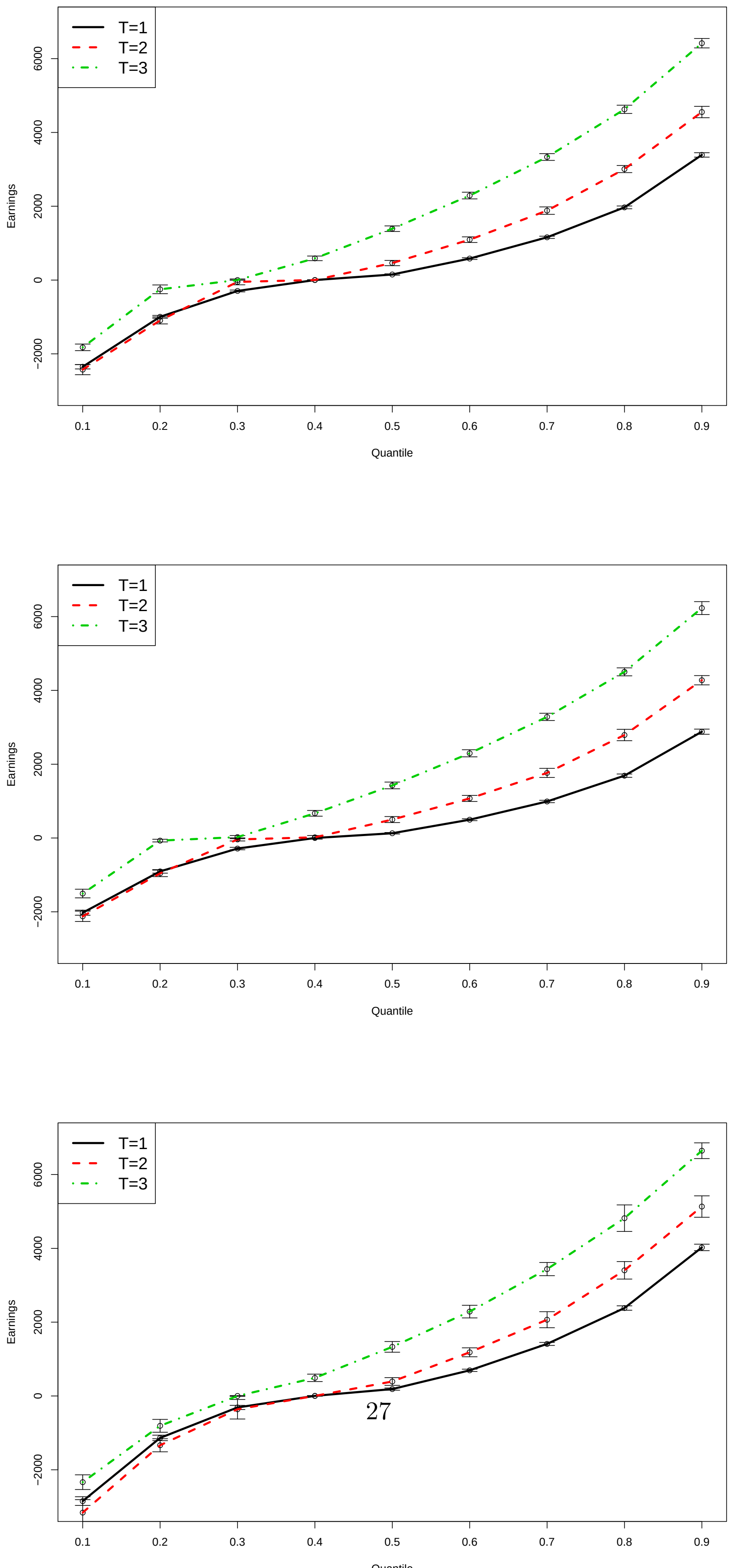
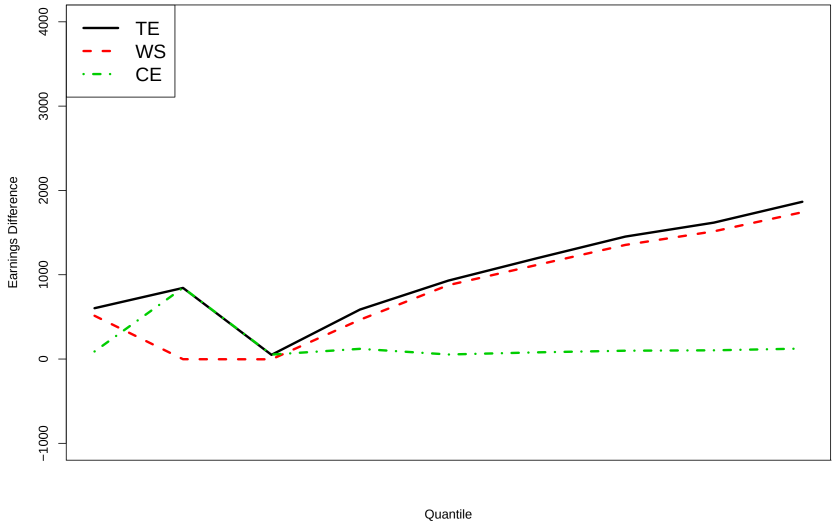
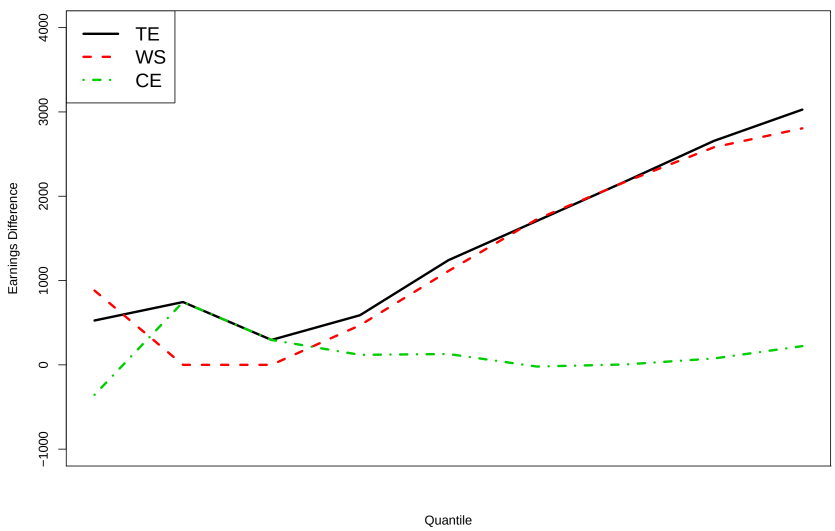
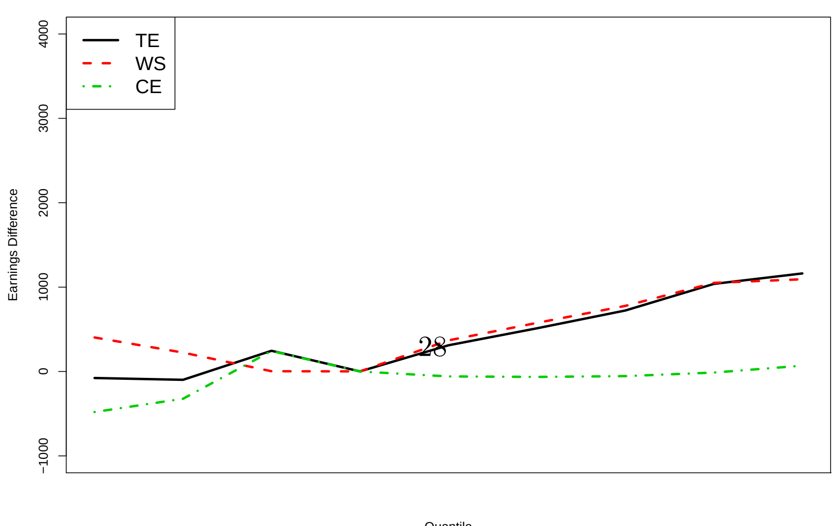


Figure 5: Decomposition Effects: All Sample, Females and Males







B Proof of Theorem 1

The proof of the asymptotic theorem for $\hat{\gamma}^{IPW}$ and $\hat{\gamma}^{EIF}$ follows from Cattaneo (2010). Denote the true parameters and functions with the superscript $*$, i.e., $e^*(X) \equiv \mathbb{E}[Y|T=t,X]$ and $p_t^*(x) = \mathbb{P}(T = t|X=x)$. We suppress the subscript of t' for simplicity, i.e., denote the true parameters $\gamma_{t'} \equiv \gamma^*$ and $\gamma_{t|t'} = \gamma_t^*$ throughout the proofs in this section. The following assumptions are from Assumptions 4 to 8 in Cattaneo (2010) by changing β_t^* to γ_t^*, for all $t \in \mathcal{T}$.

A.C (a) the class of functions $\{\theta \mapsto m(\cdot;\theta) : \theta \in \Theta\}$ is Glivenko-Cantelli; (b) $\mathbb{E}\big[\sup_{\theta\in\Theta} |m(Y(t);\theta)|\big] < \infty$; (c) $\{\theta \mapsto e_t^*(\cdot;\theta) : \theta \in \Theta\}$ is Glivenko-Cantelli.

A.IPW For some $\delta > 0$: (a) $\{\theta \mapsto m(\cdot;\theta) : |\theta - \gamma_t^*| < \delta\}$ is a Donsker class; (b) there exist constant $C > 0$ and $r \in (0,1)$ such that $\mathbb{E}\big[\sup_{|\theta-\tilde{\theta}|<\delta} |m(Y(t);\theta) - m(Y(t);\tilde{\theta})|^2\big] \le C\delta^{2r}$ for all $\tilde{\theta} \in \Theta$; (c) $\mathbb{E}\big[\sup_{|\theta-\gamma_t^*|} |m(Y(t);\theta)|^2\big] < \infty$;

A.EIF For some $\delta > 0$, and for all $x \in \mathcal{X}$ and all θ such that $|\theta - \gamma_t^*| < \delta$: (a) $e_t^*(x,\theta)$ is continuously differentiable with derivative given by $\partial_\theta e_t^*(x;\theta) \equiv (\partial/\partial\theta)e_t^*(x;\theta)$ with $\mathbb{E}[\sup_{|\theta-\gamma_t^*|<\delta} |\partial_\theta e_t^*(X;\theta)|] < \infty$; and (b) there exist $\epsilon > 0$ and a measurable function $b(x)$, with $\mathbb{E}[|b(x)|] < \infty$, such that $|\partial_\theta e_t(x;\theta) - \partial_\theta e_t^*(x;\theta)| \le b(x)\|e_t - e_t^*\|_\infty^\epsilon$ for all function $e_t(\theta) \in \mathcal{E}$ such that $\|e_t - e_t^*\|_\infty < \delta$, where $\mathcal{E}$ is a subspace of smooth functions on $\mathcal{X}$, endowed with the supremum norm.

A.NP (a) $p_t^*(\cdot)$ and $e_t^*(\cdot,\gamma_t^*)$ are s times differentiable with $s/d_x > 5\eta/2+1/2$, where $\eta = 1$ or $1/2$ depending on whether power series or splines are used as basis functions, respectively; (b) X is continuously distributed with density bounded and bounded away from zero on its compact support $\mathcal{X}$; and (c) for some $\delta > 0$, $var[m(Y(t);\theta)|X=x]$ is uniformly bounded for all $x \in \mathcal{X}$ and all θ such that $|\theta - \gamma_t^*| < \delta$.

Lemma 1 (Asymptotic Linear Representation) *Assume γ^* belongs to the interior of Θ^{J+1}. Suppose Assumptions 1, 4, A.C, A.IPW, and A.EIF hold. Assume (a) $\|\hat{p} - p^*\|_\infty = o_p(n^{-1/4})$.*

(i) Assume (b) $M_{t'n}^{IPW}(\gamma^*, \hat{P}, \hat{p}) = M_{t'n}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2})$. Then
$$\hat{\gamma}^{IPW} - \gamma^* = -\big(\Gamma'_{*|t'} W \Gamma_{*|t'}\big)^{-1} \Gamma'_{*|t'} W M_{t'n}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2}).$$

(ii) Assume (c) $\sup_{|\theta-\gamma^*|<\delta} \|\hat{e}(\theta) - e^*(\theta)\|_\infty = o_p(1)$, for some $\delta > 0$. (d) $M_{t'n}^{EIF}(\gamma^*, \hat{P}, \hat{p}, \hat{e}(\gamma^*)) = M_{t'n}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2})$. Then
$$\hat{\gamma}^{EIF} - \gamma^* = -\big(\Gamma'_{*|t'} W \Gamma_{*|t'}\big)^{-1} \Gamma'_{*|t'} W M_{t'n}^{EIF}(\gamma^*, P^*, p^*, e^*(\gamma^*)) + o_p(n^{-1/2}).$$

Lemma 2 (Nonparametric Estimation) *Suppose Assumption A.NP holds. The nonparametric estimators for p^* and e^* are described in Section 5.4 in Cattaneo (2010) with $K = n^v$, $4s/d_x - 6\eta > 1/v > 4\eta + 2$, $\eta = 1$ or $\eta = 1/2$ depending on whether power series or splines are used as basis functions. Then the conditions (a) to (d) in Lemma 1 hold.*

Proof of Lemma 1 The consistency $\hat{\gamma}^{IPW} = \gamma^* + o_p(1)$ and $\hat{\gamma}^{EIF} = \gamma^* + o_p(1)$ is directly implied by the proofs of Theorems 2 and 3 in Cattaneo (2010). The result for $\hat{\gamma}^{IPW}$ in Lemma 1 (i) is derived by the same argument in the proof of Theorem 4 in Cattaneo (2010). The result for $\hat{\gamma}^{EIF}$ in Lemma 1 (ii) is implied by the proof of Theorem 5 in Cattaneo (2010). We only note the main difference in the following proof for $\hat{\gamma}^{IPW}$ and $\hat{\gamma}^{EIF}$. Denote $m_i(\gamma_t) \equiv m(Y_i;\gamma_t)$, $P_{ti} \equiv P_t(X_i)$, and $e_{ti}(\gamma_t) \equiv e_t(X_i;\gamma_t)$.

For $\hat{\gamma}^{IPW}$, in the proof of Theorem 4 in Cattaneo (2010), $\theta = \gamma, \theta_0 = \gamma^*$, and the t-th element of $M^{IPW}_{t'n}$ is $M^{IPW}_{[t],n}$. The main difference is in

$$\Delta_{[t],n}(\gamma, P-P^*, p-p^*) = \frac{1}{n}\sum_{i=1}^{n} D_{ti} m_i(\gamma_t)\Lambda_i, \text{ where}$$

$$\Lambda_i \equiv \Lambda_n(X_i) \equiv -\frac{\left(P_{ti}-P^*_{ti}\right)}{P^{*2}_{ti}}\frac{P^*_{t'i}}{p^*_{t'}} + \frac{1}{P^*_{ti}p^*_{t'}}\left(P_{t'i}-P^*_{t'i}\right) - \frac{P^*_{t'i}}{P^*_{ti}p^*_{t'}}\left(p_{t'}-p^*_{t'}\right).$$

The last two terms are from estimating the ratio for adjusting for the treated $P_{t'}(X_i)/p_{t'}$. We modify

$$R_{3n} = \sup_{|\gamma_t-\gamma^*_t|\le\delta_n} \frac{\left|M^{IPW}_{[t],n}(\gamma^*,\hat{P},\hat{p}) - M^{IPW}_{[t],n}(\gamma^*,P^*,p^*) - \Delta_{[t],n}(\gamma^*,\hat{P}-P^*,\hat{p}-p^*)\right|}{1+C\sqrt{n}|\gamma_t-\gamma^*_t|}$$

$$R_{4n} = \sup_{|\gamma_t-\gamma^*_t|\le\delta_n} \frac{\left|\Delta_{[t],n}(\gamma^*,\hat{P}-P^*,\hat{p}-p^*) - \Delta_{[t],n}(\gamma,\hat{P}-P^*,\hat{p}-p^*)\right|}{1+C\sqrt{n}|\gamma_t-\gamma^*_t|}.$$

R_{1n} and R_{2n} are the same.

For $\hat{\gamma}^{EIF}$, Equation (A.2) in Cattaneo (2010) becomes

$$\sup_{|\gamma_t-\gamma^*_t|\le\delta_n} \frac{\sqrt{n}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\hat{e}_{ti}(\gamma_t)-e_{ti}(\gamma^*_t)\right)\left(\frac{D_{ti}}{\hat{P}_{ti}}-\frac{D_{t'i}}{\hat{P}_{t'i}}\right)\frac{\hat{P}_{t'i}}{\hat{p}_{t'}}\right|}{1+C\sqrt{n}|\gamma_t-\gamma^*_t|} \le R_{1n}+R_{2n}.$$

Define $\hat{\Upsilon}_i \equiv \left(\frac{D_{ti}}{\hat{P}_{ti}} - \frac{D_{t'i}}{\hat{P}_{t'i}}\right)\frac{\hat{P}_{t'i}}{\hat{p}_{t'}}$ to be approximated by $\Upsilon_i + D_{ti}\Lambda_i - D_{t'i}\frac{p^*_{t'}-\hat{p}_{t'}}{p^{*2}_{t'}}$ and $\Upsilon_i \equiv \left(\frac{D_{ti}}{P^*_{ti}} - \frac{D_{t'i}}{P^*_{t'i}}\right)\frac{P^*_{t'i}}{p^*_{t'}}$.

For some convex linear combination between γ_t and γ_t^*, $\tilde{\gamma}_t$,

$$
\begin{aligned}
R_{1n} = & \sup_{|\gamma_t-\gamma_t^*|\le\delta_n, \|e_t-e_t^*\|_\infty\le\delta_n} \frac{\sqrt{n}\Big|\frac{1}{n}\sum_{i=1}^n \Big(\frac{\partial}{\partial\gamma}e_i(\tilde{\gamma}_t) - \frac{\partial}{\partial\gamma}e_i^*(\gamma_t^*)\Big)\Big(\gamma_t-\gamma_t^*\Big)\hat{\Upsilon}_i\Big|}{1+C\sqrt{n}|\gamma_t-\gamma_t^*|} \\
& \le C \sup_{|\gamma_t-\gamma_t^*|\le\delta_n, \|e_t-e_t^*\|_\infty\le\delta_n} \frac{1}{n}\sum_{i=1}^n \Big|\frac{\partial}{\partial\gamma}e_i(\gamma_t) - \frac{\partial}{\partial\gamma}e_i^*(\gamma_t)\Big| \\
& + C \sup_{|\gamma_t-\gamma_t^*|\le\delta_n} \Big|\frac{1}{n}\sum_{i=1}^n \Big(\frac{\partial}{\partial\gamma}e_i(\gamma_t) - \frac{\partial}{\partial\gamma}e_i^*(\gamma_t)\Big)\Upsilon_i\Big| \\
& + \frac{C}{n}\sum_{i=1}^n \sup_{|\gamma_t-\gamma_t^*|\le\delta_n}\Big|\frac{\partial}{\partial\gamma}e_i^*(\gamma_t)\Big|\big|\hat{\Upsilon}_i-\Upsilon_i\big| \\
R_{2n} = & \sup_{|\gamma_t-\gamma_t^*|\le\delta_n} \frac{\sqrt{n}\Big|\frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\gamma}e_i^*(\gamma_t^*)(\gamma_t-\gamma_t^*)\hat{\Upsilon}_i\Big|}{1+C\sqrt{n}|\gamma_t-\gamma_t^*|}.
\end{aligned}
$$

□

Proof of Lemma 2 We modify the proof of Theorem 8 in Cattaneo (2010). We verify the condition (b) for $\hat{\gamma}_t^{IPW}$ in Lemma 1 by showing the followings are $o_p(1)$:

$$
\begin{aligned}
R_{1n} &= \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{\frac{D_{ti}m_i(\gamma_t^*)}{\hat{P}_{ti}}\frac{\hat{P}_{t'i}}{\hat{p}_{t'}} - \frac{D_{ti}m_i(\gamma_t^*)}{P_{ti}^*}\frac{P_{t'i}^*}{p_{t'}^*} - D_{ti}m_i(\gamma_t^*)\Lambda_i\Big\}\Big| \\
R_{2n} &= \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{\frac{D_{ti}m_i(\gamma_t^*)}{P_{ti}^*}\Lambda_i P_{ti}^* - e_i^*(\gamma_t^*)\Lambda_i P_{ti}^*\Big\}\Big| \\
R_{3n} &= \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{e_i^*(\gamma_t^*)\Lambda_i P_{ti}^* + e_i^*(\gamma_t^*)\Big(\frac{D_{ti}}{P_{ti}^*} - \frac{D_{t'i}}{P_{t'i}^*}\Big)\frac{P_{t'i}^*}{p_{t'}^*}\Big\}\Big| \\
&\le \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{-e_i^*(\gamma_t^*)\frac{\hat{P}_{ti}-P_{ti}^*}{P_{ti}^*} + e_i^*(\gamma_t^*)\frac{D_{ti}-P_{ti}^*}{P_{ti}^*}\Big\}\frac{P_{t'i}^*}{p_{t'}^*}\Big| && \text{(B.1)} \\
&+ \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{e_i^*(\gamma_t^*)\frac{\hat{P}_{t'i}-P_{t'i}^*}{P_{t'i}^*} - e_i^*(\gamma_t^*)\frac{D_{t'i}-P_{t'i}^*}{P_{t'i}^*}\Big\}\frac{P_{t'i}^*}{p_{t'}^*}\Big| && \text{(B.2)} \\
&+ \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^*(\gamma_t^*)\frac{P_{t'i}^*}{p_{t'}^*}\big(\hat{p}_{t'}-p_{t'}^*\big)\Big| && \text{(B.3)}
\end{aligned}
$$

R_{1n}, R_{2n}, (B.1), and (B.2) are $o_p(1)$ following the same arguments in Cattaneo (2010). (B.3) is $O_p(1) \times o_p(1) = o_p(1)$ because $\mathbb{E}[e_i^*(\gamma_t^*)P_{t'i}^*/p_{t'}^*] = 0$. For the condition (d) for $\hat{\gamma}_t^{EIF}$ in Lemma 1,

$$
\begin{aligned}
R_{4n} &= \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{D_{ti}\big(m_i(\gamma_t^*) - e_i(\gamma_t^*)\big)\Lambda_i\Big\}\Big| \\
R_{5n} &= \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{-\big(\hat{e}_i(\gamma_t^*) - e_i^*(\gamma_t^*)\big)\Big(\frac{D_{ti}}{P_{ti}^*} - \frac{D_{t'i}}{P_{t'i}^*}\Big)\frac{P_{t'i}^*}{p_{t'}^*}\Big\}\Big| \\
R_{6n} &= \Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_{t'i}e_i(\gamma_t^*)\frac{\hat{p}_{t'} - p_{t'}^*}{p_{t'}^{*2}}\Big|.
\end{aligned}
$$

R_{4n} and R_{5n} are $o_p(1)$ following the same arguments as the case for $\hat{\beta}_t^{EIF}$. $R_{6n} = o_p(1)$ follows the same reasoning as (B.3) above. $\square$

C Proofs in Section 3.3

Let $(Z_1, Z_1, ..., Z_n)$ be an $i.i.d.$ sequence of random variables taking values in a probability space $(\mathcal{Z}, \mathcal{B})$ with distribution P. For some measurable function $\phi : \mathcal{Z} \to \mathbb{R}$, define $\mathbb{E}\phi = \int \phi dP$ and $G_n\phi = \sqrt{n}(n^{-1}\sum_i \phi(Z_i) - \mathbb{E}\phi)$ for the empirical process at ϕ. Let $\bar{O}_p(a_n)$ and $\bar{o}_p(a_n)$ be $O_p(a_n)$ and $o_p(a_n)$ uniformly in $y \in \mathcal{Y}$.

Definition 3 ($\mathcal{C}_M^\alpha(\mathcal{X})$, van der Vaart and Wellner (1996) (P. 154))) *$\mathcal{C}_M^\alpha(\mathcal{X})$ is defined on a bounded set $\mathcal{X}$ in $\mathbb{R}^{d_x}$ as follows: For any vector $q = (q_1, ..., q_d)$ of q_d integers, let D^q denote the differential operator $D^q = \partial^{q\cdot}/\partial x_1^{q_1}...\partial x_d^{q_d}$. Denote $q. = \sum_{l=1}^d q_l$ and $\underline{\alpha}$ to be the greatest integer strictly smaller than α. Let $\|g\|_\alpha = \max_{q.\leq\underline{\alpha}}\sup_x |D^q g(x)| + \max_{q.\leq\underline{\alpha}}\sup_{x\neq x'}|D^q g(x) - D^q g(x')|/\|s - s'\|^{\alpha-\underline{\alpha}}$ where $\max_{q.\leq\underline{\alpha}}$ denotes the maximum over $(q_1, ..., q_d)$ such that $q. \leq \underline{\alpha}$ and the suprema are taken over the interior of $\mathcal{X}$. Then $\mathcal{C}_M^\alpha(\mathcal{X})$ is the set of all continuous functions $g : \mathcal{X} \subset \mathbb{R}^{d_x} \mapsto \mathbb{R}$ with $\|g\|_\alpha \leq M$.*

Assumption 6 *For any $t, t' \in \mathcal{T}$ and $y, y_1, y_2 \in \mathcal{Y}$,*

(a) $P_t(X) \in \mathcal{C}_M^\alpha(\mathcal{X})$ and $F_{Y|TX}(y|t, X) \in \mathcal{C}_M^\alpha(\mathcal{X})$ for $\alpha > d_x/2$. $\sup_{x\in\mathcal{X}}\big|F_{Y|TX}(y_1|t, x) - F_{Y|TX}(y_2|t, x)\big| < C|y_1 - y_2|^{1/2}$ for some positive constant C.

(b) $\sup_{x\in\mathcal{X}}\|\partial_x^q\hat{P}_t(x) - \partial_x^q P_t(x)\| = o_p(1)$ and $\sup_{x\in\mathcal{X}}\big\|\partial_x^q\hat{F}_{Y|TX}(y|t, x) - \partial_x^q F_{Y|TX}(y|t, x)\big\| = o_p(1)$ for all $q < d_x/2$.

(c) (i) (EIF) $\int\big(\hat{F}_{Y|TX}(y|t, x) - F_{Y|TX}(y|t, x)\big)\big(\hat{P}_{t'}(x) - P_{t'}(x)\big)\, f_X(x)dx = \bar{o}_p(n^{-1/2})$.

(ii) (IPW) $\int F_{Y|TX}(y|t, x)\frac{P_{t'}(x)}{p_{t'}}\Big(\frac{\hat{P}_{t'}(x)}{P_{t'}(x)} - \frac{\hat{P}_t(x)}{P_t(x)}\Big)\, f_X(x)dx = n^{-1}\sum_{i=1}^n F_{Y|TX}(y|t, X_i)\frac{P_{t'}(X_i)}{p_{t'}}\Big(\frac{D_{t'i}}{P_{t'}(X_i)} - \frac{D_{ti}}{P_t(X_i)}\Big) + \bar{o}_p(n^{-1/2})$.

Assumption 6 (c)(ii) is for the IPW estimator. This is analogous to the condition (b) in Lemma 1 that the nonparametrically estimated propensity score captures the correction term in the efficient influence function. Assumption 6 (c)(i) is for the EIF estimator. The doubly robust property of the EIF estimator implies we can allow either $F_{Y|TX}$ or $P_t(X)$ to be misspecified and estimated at a root-n rate. Then we only need to consistently estimate the other unknown function without restricting the convergence rate. Denote

$$\hat{F}^{IPW}_{Y\langle t|t'\rangle}(\cdot) = \frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_1(Z_i;y,t,t'), \text{ where } \varphi_1(Z;y,t,t') \equiv \frac{D_t}{P_t(X)}\mathbf{1}\{Y \leq y\}\frac{P_{t'}(X)}{p_{t'}}$$
$$\hat{F}^{EIF}_{Y\langle t|t'\rangle}(\cdot) = \frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_2(Z_i;y,t,t'), \text{ where}$$
$$\varphi_2(Z;y,t,t') \equiv \varphi_1(Z;y,t,t') + F_{Y|TX}(y|t,X)\,\Big(\frac{D_{t'}}{P_{t'}(X)} - \frac{D_t}{P_t(X)}\Big)\frac{P_{t'}(X)}{p_{t'}}.$$

Proof of Theorem 2 We decompose the estimator as follows:

$$\sqrt{n}(\hat{F}^{EIF}_{Y\langle t|t'\rangle} - F_{Y\langle t|t'\rangle}) = \sqrt{n}\big(\frac{1}{n}\sum_{i=1}^{n}\hat{\varphi}_2 - \varphi_2\big) = G_n\big[\hat{\varphi}_2 - \varphi_2\big] + G_n\big[\varphi_2\big] + \sqrt{n}\;\mathbb{E}\big[\hat{\varphi}_2 - \varphi_2\big]. \tag{C.1}$$

The second term $G_n[\varphi_2]$ is $\bar{O}_p(1)$ by the Donsker property in Assumption 6 (a). The first term $G_n[\hat{\varphi}_2 - \varphi_2] = \bar{o}_p(1)$ by a stochastic equicontinuity argument using Lemma A.1 in Lee (2015b). Assumption 6 (a) and (b) ensure the conditions of Lemma A.1 in Lee (2015b). That is, the estimators $\hat{P}_t(X)$ and $\hat{F}_{Y|TX}(y|t,X)$ belongs to $\mathcal{C}^{\alpha}_{M}(\mathcal{X})$ and satisfies the Hölder continuity with probability approaching to one. We calculate the third term in the following.

$$\begin{aligned}
\mathbb{E}\big[\hat{\varphi}_2 - \varphi_2\big] &= \mathbb{E}\bigg[\bigg(\frac{D_t}{\hat{P}_t(X)}\mathbf{1}\{Y \leq y\} + \hat{F}_{Y|TX}(y|t,X)\,\Big(\frac{D_{t'}}{\hat{P}_{t'}(X)} - \frac{D_t}{\hat{P}_t(X)}\Big)\bigg)\frac{\hat{P}_{t'}(X)}{\hat{p}_{t'}} - \varphi_2\ \bigg] \\
&= \mathbb{E}\bigg[\bigg(\frac{P_t(X)}{\hat{P}_t(X)}F_{Y|TX}(y|t,X) + \hat{F}_{Y|TX}(y|t,X)\,\Big(\frac{P_{t'}(X)}{\hat{P}_{t'}(X)} - \frac{P_t(X)}{\hat{P}_t(X)}\Big)\bigg)\frac{\hat{P}_{t'}(X)}{\hat{p}_{t'}} - F_{Y|TX}(y|t,X)\frac{P_{t'}(X)}{p_{t'}}\ \bigg] \\
&= \mathbb{E}\bigg[\bigg(\frac{P_t(X)}{\hat{P}_t(X)}F_{Y|TX}(y|t,X) + F_{Y|TX}(y|t,X)\,\Big(\frac{P_{t'}(X)}{\hat{P}_{t'}(X)} - \frac{P_t(X)}{\hat{P}_t(X)}\Big)\bigg)\frac{\hat{P}_{t'}(X)}{\hat{p}_{t'}} - F_{Y|TX}(y|t,X)\frac{P_{t'}(X)}{p_{t'}}\ \bigg] \\
&\quad + \mathbb{E}\bigg[\Big(\hat{F}_{Y|TX}(y|t,X) - F_{Y|TX}(y|t,X)\Big)\,\Big(\frac{P_{t'}(X)}{\hat{P}_{t'}(X)} - \frac{P_t(X)}{\hat{P}_t(X)}\Big)\frac{\hat{P}_{t'}(X)}{\hat{p}_{t'}}\bigg] \\
&= \mathbb{E}\bigg[F_{Y|TX}(y|t,X)\,\Big(\frac{P_{t'}(X)}{\hat{p}_{t'}} - \frac{P_{t'}(X)}{p_{t'}}\Big)\ \bigg] + \bar{o}_p(n^{-1/2}) \\
&= \mathbb{E}\bigg[F_{Y|TX}(y|t,X)\frac{P_{t'}(X)}{p_{t'}}\bigg]\Big(1 - \frac{1}{n}\sum_{i=1}^{n}D_{t'i}/p_{t'}\Big) + \bar{o}_p(n^{-1/2})
\end{aligned}$$

where the fourth equality is implied by Assumption 6 (c)(i). Together with the second term $G_n[\varphi_2]$, $\sqrt{n}(\hat{F}^{EIF}_{Y\langle t|t'\rangle} - F_{Y\langle t|t'\rangle}) = G_n[\psi_{t|t'}(Z;y)] + \bar{o}_p(1)$.

We decompose $\sqrt{n}(\hat{F}^{IPW}_{Y\langle t|t'\rangle} - F_{Y\langle t|t'\rangle})$ similarly as in (C.1). The third term

$$
\begin{aligned}
\mathbb{E}\big[\hat{\varphi}_1 - \varphi_1\big] &= \mathbb{E}\Big[\frac{D_t}{\hat{P}_t(X)}\mathbf{1}\{Y \le y\}\frac{\hat{P}_{t'}(X)}{\hat{p}_{t'}} - \varphi_1\ \Big] \\
&= \mathbb{E}\Big[\frac{P_t(X)}{\hat{P}_t(X)}F_{Y|TX}(y|t,X)\frac{\hat{P}_{t'}(X)}{\hat{p}_{t'}} - F_{Y|TX}(y|t,X)\frac{P_{t'}(X)}{p_{t'}}\ \Big] \\
&= \mathbb{E}\Big[F_{Y|TX}(y|t,X)\ \Big(-\frac{P_{t'}(X)}{p_{t'}}\Big(\frac{\hat{p}_{t'}}{p_{t'}} - 1\Big) - \frac{P_{t'}(X)}{p_{t'}}\Big(\frac{\hat{P}_t(X)}{P_t(X)} - 1\Big) + \frac{\hat{P}_{t'}(X) - P_{t'}(X)}{p_{t'}}\Big)\ \Big] + \bar{o}_p(n^{-1/2}) \\
&= \gamma - \gamma\frac{1}{n}\sum_{i=1}^{n}\frac{D_{t'i}}{p_{t'}} + \mathbb{E}\Big[F_{Y|TX}(y|t,X)\frac{P_{t'}(X)}{p_{t'}}\Big(\frac{\hat{P}_{t'}(X)}{P_{t'}(X)} - \frac{\hat{P}_t(X)}{P_t(X)}\Big)\Big] + \bar{o}_p(n^{-1/2})
\end{aligned}
$$

The third term in the last equation is by Assumption 6 (c)(ii). Together with the third term $G_n[\varphi_1]$, $\sqrt{n}(\hat{F}^{IPW}_{Y\langle t|t'\rangle} - F_{Y\langle t|t'\rangle}) = G_n[\psi_{t|t'}(Z;y)] + \bar{o}_p(1)$.

Define the class of measurable functions $\mathcal{H} = \{(\mathcal{Y} \times \mathcal{T} \times \mathcal{X}) \to \psi(Y,T,X;y) : y \in \mathcal{Y}\}$. By Lemma A2 in Donald and Hsu (2014) and the Assumptions in the Appendix, $\mathcal{H}$ is P-Donsker. The weak convergence is implied by Donsker's Theorem in Section 2.8.2 in van der Vaart and Wellner (1996). $\square$

Proof of Corollary 1 By the functional delta method (e.g., Theorem 3.9.4 in van der Vaart and Wellner (1996)) and the linearity of the Hadamard derivative, the weak convergence to a Gaussian process is implied. $\square$

Proof of Corollary 2 Suppose $\sqrt{n}\big(\hat{\theta}(\cdot) - \theta_0(\cdot)\big) = n^{-1/2}\sum_{i=1}^{n}\psi_{tin}(\cdot) + o_p(1) \Rightarrow \mathbb{G}_t(\cdot)$ from Theorem 2. Assume θ_t is continuously differentiable with strictly positive derivative $(\partial/\partial y)\theta_0(y)\big|_{y=Q_\tau} \equiv \theta_0'(Q_\tau)$. The Hadamard derivative is shown in Example 3.9.24 in van der Vaart and Wellner (1996). Then the influence function for estimating the quantile process is $\psi^Q_{t|t'}(Z_i;\tau) \equiv -\psi_{tin}(Q_\tau)/\theta_0'(Q_\tau)$. Therefore,

$$
\sqrt{n}\big(\hat{Q}_{\cdot} - Q_{\cdot}\ \big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi^Q_{t|t'}(Z_i;\cdot) + o_p(1) \Rightarrow -\mathbb{G}_t\big(Q_{\cdot}\big)\Big/\theta_0'(Q_{\cdot}) \equiv \mathbb{G}^Q_t(\cdot)
$$

where $\mathbb{G}^Q_t$ is a Gaussian process indexed by $\tau \in [a,b] \subset (0,1)$ in the metric space $l^\infty([a,b])$. The Gaussian process $\mathbb{G}^Q_t$ has zero mean and covariance kernel, for any $\tau_1 < \tau_2 \in [a,b]$, $Cov\big(\mathbb{G}^Q_t(\tau_1), \mathbb{G}^Q_t(\tau_2)\big) = \mathbb{E}\big[\psi^Q_{t|t'}(Z;\tau_1)\psi^Q_{t|t'}(Z;\tau_2)\big]$. $\square$

D Proof of Theorem 3

We calculate the semiparametric efficiency bound for the parameter λ_t defined by

$$\int\int\int m(y;\lambda_t) f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(X)}{p_t}\ W_{X_1t'}((x_1,x_2))\ dydx_1dx_2 = 0.$$

The pathwise derivative w.r.t. θ is

$$\int\int\int \frac{d}{d\theta} m(y;\lambda_t(\theta)) f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(x)}{p_t}\ W_{X_1t'}((x_1,x_2))\ dydx_1dx_2 \tag{D.1}$$

$$+\int\int\int m(y;\lambda_t)\frac{d}{d\theta}\Big(f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(x)}{p_t}\Big)\ W_{X_1t'}((x_1,x_2))\ dydx_1dx_2 \tag{D.2}$$

$$+\int\int\int m(y;\lambda_t) f_{YTX}(y,t,x_1,x_2)\ \frac{P_t(x)}{p_t}\ \frac{d}{d\theta} W_{X_1t'}((x_1,x_2))\ dydx_1dx_2 = 0. \tag{D.3}$$

The result for the decomposition parameter $\gamma_{t|t}$ is directly applied to (D.2) where the moment function is replaced by $m(y;\lambda_t)W_{X_1t'}((x_1,x_2))$. That is, (D.2) contributes

$$\frac{P_t(X)}{p_t}\frac{D_t}{P_t(X)} m(Y;\lambda_t) W_{X_1t'}((X_1,X_2)) = \frac{D_t}{P_t(X)} m(Y;\lambda_t) \frac{P_{t'}(X)}{\mathbb{P}(T=t'|X_2)}\frac{\mathbb{P}(T=t|X_2)}{p_t}.$$

For (D.3),

$$\begin{aligned}&\mathbb{E}\left[e_t(X;\lambda_t)\ \frac{P_t(X)}{p_t}\ \frac{d}{d\theta} W_{X_1t'}((X_1,X_2))\right]\\ &= \mathbb{E}\left[e_t(X;\lambda_t)\ \frac{P_t(X)}{p_t}\ W_{X_1t'}((X_1,X_2))\ \left(\frac{\dot{P}_{t'}(X)}{P_{t'}(X)} - \frac{\dot{p}_t(X)}{p_t(X)} + \frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)} - \frac{\dot{\mathbb{P}}(T=t'|X_2)}{\mathbb{P}(T=t'|X_2)}\right)\right].\end{aligned}$$

The proof of Theorem 2 in Lee (2015a) implies the first part containing $\dot{P}_t(X)$ and $\dot{P}_{t'}(X)$ contributes

$$e_t(X;\lambda_t)\frac{P_t(X)}{p_t}\ W_{X_1t'}((X_1,X_2))\left(\frac{D_{t'}}{P_{t'}(X)} - \frac{D_t}{P_t(X)}\right)$$

to the efficient influence function.

For the rest part containing $\dot{\mathbb{P}}(T=t|X_2)$ and $\dot{\mathbb{P}}(T=t'|X_2)$, we define the score as

$$S(y,t,x;\theta_0) = S_y(y,t,x) + S_1(x_1,x_2,t) + S_{p2}(t,x_2) + S_{x_2}(x_2)$$

where $S_1(x_1,x_2,T) \equiv \sum_{j\in\mathcal{T}} D_j s_{xj}(x_1,x_2)$, $s_{xj}(x_1,x_2) \equiv \frac{d}{d\theta}\log f_{X_1|X_2T}(x_1|x_2,j;\theta)\big|_{\theta_0}$, $S_{p2}(T,x_2) \equiv \sum_{j\in\mathcal{T}} D_j\dot{\mathbb{P}}(T=j|X_2=x_2)/\mathbb{P}(T=j|X_2=x_2)$, $\dot{\mathbb{P}}(T=j|X_2) \equiv \frac{d}{d\theta}\mathbb{P}(T=j|X_2;\theta)\big|_{\theta_0}$, and $S_{x_2}(x_2) \equiv \frac{d}{d\theta}\log f_{X_2}(x_2;\theta)\big|_{\theta_0}$. The tangent space is characterized $\mathcal{H}_y + \mathcal{H}_1 + \mathcal{H}_{p2} + \mathcal{H}_{x_2}$, where $\mathcal{H}_1 \equiv \{S_1(X_1,X_2,T) : s_{xj}(X_1,X_2) \in L_0^2(F_{X_1|X_2T}(X_1|X_2,j)), \forall j\in\mathcal{T}\}$, $\mathcal{H}_{p2} \equiv \{S_{p2}(T,X_2) : S_{p2}(T,X_2) \in L_0^2(F_{T|X_2})\}$, and $\mathcal{H}_{x_2} \equiv \{S_{x2}(X_2) : S_{x2}(X_2) \in L_0^2(F_{X_2})\}$.

Similar to Equation (12) in Lee (2015a),

$$
\begin{aligned}
&\mathbb{E}\left[\frac{D_t-\mathbb{P}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}S(Z;\theta_0)\middle|X_2\right]\\
&=\mathbb{E}\left[\frac{D_t-\mathbb{P}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}\left(\sum_{j\in\mathcal{T}}D_js_j(Y,X)+D_js_{xj}(X_1,X_2)+D_j\frac{\dot{\mathbb{P}}(T=j|X_2)}{\mathbb{P}(T=j|X_2)}+S_{x2}(X_2)\right)\middle|X_2\right]\\
&=\mathbb{E}\left[\frac{D_t}{\mathbb{P}(T=t|X_2)}\left(s_t(Y,X)+s_{xt}(X_1,X_2)+\frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}\right)\middle|X_2\right]\\
&\quad-\mathbb{E}\left[\sum_{j\in\mathcal{T}}\left(D_js_j(Y,X)+D_js_{xj}(X_1,X_2)+D_j\frac{\dot{\mathbb{P}}(T=j|X_2)}{\mathbb{P}(T=j|X_2)}\right)\middle|X_2\right]\\
&=\mathbb{E}[s_t(Y,X)|T=t,X_2]+\mathbb{E}[s_{xt}(X_1,X_2)|T=t,X_2]+\frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}\\
&\quad-\sum_{j\in\mathcal{T}}\left(\mathbb{P}(T=j|X_2)\mathbb{E}[s_j(Y,X)|T=j,X]+\dot{\mathbb{P}}(T=j|X_2)\right)=\frac{\dot{\mathbb{P}}(T=t|X_2)}{\mathbb{P}(T=t|X_2)}
\end{aligned}
$$

by the low of iterated expectations, $\mathbb{E}[s_j(Y,X)|T=j,X]=0$, and $\mathbb{E}[s_{xj}(X_1,X_2)|T=j,X_2]=0$, $\forall j\in\mathcal{T}$. We first calculate

$$
\mathbb{E}[e_t(X;\lambda_t)P_{t'}(X)|X_2]=\mathbb{E}\left[m(y;\lambda_t)\frac{P_{t'}(X)}{P_t(X)}\middle|T=t,X_2\right]\mathbb{P}(T=t|X_2).
$$

Then by the law of iterated expectations,

$$
\begin{aligned}
&\mathbb{E}\left[e_t(X;\lambda_t)\frac{P_t(X)}{p_t}W_{X_1t'}(X)\frac{\dot{\mathbb{P}}(T=t'|X_2)}{\mathbb{P}(T=t'|X_2)}\right]\\
&=\mathbb{E}\left[\mathbb{E}\left[e_t(X;\lambda_t)\frac{P_t(X)}{p_t}W_{X_1t'}(X)\middle|X_2\right]\mathbb{E}\left[\frac{D_{t'}-\mathbb{P}(T=t'|X_2)}{\mathbb{P}(T=t'|X_2)}S(Z;\theta_0)\middle|X_2\right]\right]\\
&=\mathbb{E}\left[\frac{\mathbb{P}(T=t|X_2)}{p_t}\mathbb{E}\left[m(Y;\lambda_t)W_{X_1t'}(X)\middle|T=t,X_2\right]\frac{D_{t'}-\mathbb{P}(T=t'|X_2)}{\mathbb{P}(T=t'|X_2)}S(Z;\theta_0)\right].
\end{aligned}
$$

We obtain the main component of the efficient influence function

$$
\begin{aligned}
&\psi_{X_1t'}(Z;\lambda_t,p,e(\lambda_t))\equiv\Bigg(\psi_{t'}(Z;\lambda_t,p,e(\lambda_t))\frac{p_{t'}}{\mathbb{P}(T=t'|X_2)}\\
&+\mathbb{E}\big[m(Y;\lambda_t)W_{X_1t'}((X_1,X_2))\big|T=t,X_2\big]\left(\frac{D_t}{\mathbb{P}(T=t|X_2)}-\frac{D_{t'}}{\mathbb{P}(T=t'|X_2)}\right)\Bigg)\frac{\mathbb{P}(T=t|X_2)}{p_t}.
\end{aligned}
$$

□

References

Agüero, J., M. Carter, and I. Woolard (2010). The impact of unconditional cash transfers on nutrition: The south african child support grant. working paper.

Andersson, F., H. J. Holzer, J. I. Lane, D. Rosenblum, and J. A. Smith (2016). Does Federally-Funded Job Training Work? Non-experimental Estimates of WIA Training Impacts Using Longitudinal Data on Workers and Firms. Technical report.

Behrman, J. R., Y. Cheng, and P. E. Todd (2004, February). Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach. *The Review of Economics and Statistics 86*(1), 108–132.

Bhattacharya, D. (2007). Inference on inequality from household survey data. *Journal of Econometrics 137*(2), 674–707.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources 8*(4), pp. 436–455.

Cattaneo, M., D. Drukker, and A. Holland (2013). Estimation of multivalued treatment effects under conditional independence. *Stata Journal 13*(3), 407–450(44).

Cattaneo, M. D. (2010, April). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics 155*(2), 138–154.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996, September). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–44.

Donald, S. G. and Y.-C. Hsu (2014). Estimation and inference for distribution functions and quantile functions in treatment effect models. *Journal of Econometrics 178*(3), 383–397.

Firpo, S. and C. Pinto (2016). Identification and estimation of distributional impacts of interventions using changes in inequality measures. Technical Report 3. jae.2448.

Flores, C. A., A. Flores-Lagunes, A. Gonzalez, and T. C. Neumann (2012, February). 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*(1), 153–171.

Flores, C. A. and O. Mitnik (2014). Comparing treatments across labor markets: An assessment of nonexperimental multiple-treatment strategies. *The Review of Economics and Statistics forthcoming.*

Fortin, N., T. Lemieux, and S. Firpo (2011). *Decomposition Methods in Economics*, Volume 4 of *Handbook of Labor Economics*, Chapter 1, pp. 1–102. Elsevier.

Graham, B. S. (2011). Efficiency bounds for missing data models with semiparametric restrictions. *Econometrica 79*(2), 437–452.

Hahn, J. (1998, March). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*(2), 315–332.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. *Handbook of labor economics 3*, 1865–2097.

Heckman, J. J. and E. J. Vytlacil (2007, January). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6 of *Handbook of Econometrics*, Chapter 70-71. Elsevier.

Heinrich, C. J., P. R. Mueser, and K. R. Troske (2009). Workforce investment act non-experimental net impact evaluation, final report. *U.S. Department of Labor*.

Hirano, K. and G. Imbens (2004). *The Propensity Score with Continuous Treatments*, Chapter 7. John Wiley and Sons.

Hollenbeck, K., D. Schroeder, C. T. King, and W. J. Huang (2005). Net impact estimates for services provided through the workforce investment act. Division of research and demonstration, office of policy and research, employment and training administration report, U.S. Department of Labor.

Ibarraran, P. and J. Villa (2010). Labor insertion assessment of conditional cash transfer programs: A dose-response estimate for mexicos oportunidades. Mimeo, Inter-American Development Bank.

Imbens, G. W. and J. M. Wooldridge (2009, September). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Kluve, J., H. Schneider, A. Uhlendorff, and Z. Zhao (2012). Evaluating continuous training programs using the generalized propensity score. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 175*(2), 587–617.

Lee, Y.-Y. (2015a). Efficient propensity score regression estimators of multi-valued treatment effects for the treated. *Working paper*.

Lee, Y.-Y. (2015b). Partial mean processes with generated regressors: Continuous treatment effects and nonseparable models.

Nightingale, D. S. and S. Elaine (2011). *The Availability and Use of Workforce Development Programs Among Less-Educated Youth*, Chapter 8, pp. 185–288. Black Males Left Behind. The Urban Insitute.

Oaxaca, R. (1973, October). Male-female wage differentials in urban labor markets. *International Economic Review 14*(3), 693–709.

Plesca, M. and J. Smith (2007, May). Evaluating multi-treatment programs: theory and evidence from the u.s. job training partnership act experiment. *Empirical Economics 32*(2), 491–528.

van der Vaart, A. (2000). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: with Application to Statistics.* New York: Springer-Verlag.