Heterogeneous Treatment Effects Analysis through Distribution Regression based Changes-in-Changes

Matthew Hong*

Nov. 5, 2024

Click Here for the Newest Version

Abstract

The changes-in-changes method, developed by Athey and Imbens (2006), is a powerful tool for identifying the distributional effects of a policy intervention, allowing for endogenous treatment assignment and full counterfactual distribution identification. However, challenges with incorporating control variables to address concerns akin to differential parallel trends in the difference-in-differences literature persist. In this paper, I propose a semiparametric approach to changes-in-changes based on distribution regression that can flexibly account for observed confounders. This approach can be applied to continuous and/or discrete outcome variables. I derive functional central limit theorems for the distribution regression based changes-in-changes estimator and for functionals thereof. These include unconditional distributional and quantile treatment effects, average treatment effects, and decompositional treatment effects for the treated group. Bootstrap validity result is also provided for conducting inference in practice. Lastly, I apply the approach to study the heterogeneous effects of Earned Income Tax Credit on infant weights and find that the policy had higher concentrated benefits for lower birth weights and more muted effects across the birth weight distribution than previously reported.

*Email: hongmatt@bu.edu, Ph.D. Candidate, Department of Economics, Boston University.

I am grateful to my advisors, Ivan Fernandez-Val, Hiroaki Kaido and Jean-Jacques Forneron, for their continued guidance and support. I would like to thank Zhongjun Qu, Kevin Lang, Shakeeb Khan, Stella Hong, Zhanyuan Tian, and the Econometrics seminar at Boston University for their useful comments. All errors and omissions are my own.

1 Introduction

Much interest in economics research is directed towards quantifying heterogeneous impacts of regulatory policies on outcomes of interest. As social and economic welfare policies often aim to improve the outcomes for those at the lower end of the income distribution (e.g. minimum wage policies) or for those with poor socioeconomic or health outcomes (e.g. food stamps, tax subsidies), correctly identifying and estimating distributional causal effects is crucial for informing future policy designs and decisions. For various regulatory policies, however, since randomized control trials are often infeasible, the use of quasi-experimental methods such as difference-in-differences ("DID") have become prevalent, and several extensions of the DID method to the distributional setting have been proposed (Athey and Imbens, 2006; Bonhomme and Sauder, 2011; Callaway and Li, 2019; Havnes and Mogstad, 2015; Kim and Wooldridge, 2023).

Among the alternatives, CIC stands out with many attractive features. Namely, CIC identifies the full counterfactual outcome distribution for the treated group in the absence of treatment, is scale-invariant to monotonic transformations of the outcome variable, and allows for endogeneous treatment assignment. Alternative approaches might yield counterfactual distributions that lie outside the unit interval (especially for discrete outcomes), or might rely on random treatment assignment assumptions which can be less plausible for welfare policies (Ghanem et al., 2023).

Despite the numerous attractive features of CIC, difficulty in incorporating control variables has stymied wider application of the method (Lechner, 2011; Melly and Santangelo, 2015). There are two dimensions to the importance of accounting for potential observed confounders in such a setting. First, in many economic applications, distribution of observed characteristics maybe different across groups and the outcomes of individuals with different characteristics such as age, race, and education levels may evolve differently over time. As such, implementing CIC without explicitly accounting for potential confounders may misattribute effects due to policy change. Secondly, said concerns akin to differential parallel trends in the difference-in-differences literature can be exacerbated in a distributional setting, especially if the effect of controls on the outcome variable can be heterogeneous across the outcome distribution. In a health economics setting, for instance, varying levels of basic prenatal care access or quality across U.S. states could

differentially impact mothers who would be giving birth to healthier babies compared to those who would be giving birth to less healthier, lower weight babies.

Existing remedies to account for controls are mainly threefold: (1) Nonparametric methods which suffer from the curse of dimensionality as the number of support points for the controls increase. (2) Partialling-out approach, where CIC is applied to the residuals from an ordinary least squares estimation of regressing the outcome on control variables, but one that restricts the heterogeneity of controls' effect on the outcome distribution.[1] Lastly, (3) quantile regression-based approach, which is similar to distribution regression for continuous outcomes, but may provide a poor approximation for the conditional outcome distribution when the outcome is discrete or has mass points (Chernozhukov et al., 2013, 2019). These cases arise frequently in relevant empirical applications such as censoring, bunching or count data.

To overcome these challenges, I propose a distribution regression based approach for CIC. The approach consists of modeling and estimating the conditional outcome distributions using distribution regression for each group and time pair, then plugging in the estimates to the conditional version of changes-in-changes. Distribution regression, as developed in Foresi and Peracchi (1995) and Chernozhukov et al. (2013), can flexibly model the conditional outcome distribution while allowing for the effect of potential observed confounders to vary across the outcome distribution. In practice, this ensures that individuals are compared not only with those who have similar observed characteristics, but also those whose outcomes are affected similarly by the given characteristics depending on their location in the outcome distribution. Furthermore, the distribution regression can accommodate both continuous and discrete (or mixed continuous-discrete) outcomes without adjustments to the regression estimator. Computationally, this approach is feasible and straightforward to implement, as it involves estimating a series of logit or probit regressions using off-the-shelf software packages over a grid of the outcome support.

Various economically meaningful functionals of the conditional CIC distribution function can also be recovered. The marginal distribution (or unconditional quantile) function can be obtained by integrating (and then inverting) the conditional outcome distribution over the empirical distribution of the controls in the treated group. Unconditional quantile treatment effects can be recov-

[1] In the empirical section at the end of this paper, I present qualitative and quantitative differences between residualized CIC and CIC based on distribution regression resulting from a restriction in the heterogeneity. Available software packages such as `qte` in R also offer only this option to incorporate covariates.

ered by taking the differences in the factual quantile function of the post-treatment outcomes in the treated group (identified directly from data) and the counterfactual quantile function under no treatment (identified by CIC). Additional functionals can be identified, including Gini coefficient, Lorenz curves, average treatment effects and treatment effects on the components of a distributional decomposition analysis.[2] Decomposition methods are useful for explaining outcome gaps (e.g. income, mortgage decisions, and test score gaps across race or gender) and determining relative importance of observable components such as education or experience versus unobservable components (often dubbed "discriminatory" factors). In the quasi-experimental setting, one can use this to decompose the treatment effects to assess the relative importance of each component and can provide insight into the mechanism through which outcome gaps change or persist.

This paper derives large sample theory for the distribution regression based CIC process. I show that the process converges weakly to a tight mean-zero Gaussian process under standard regularity conditions that ensure functional central limit theorem for the distribution regression based estimator. I also show that functionals of the estimator are uniformly consistent and asymptotically Gaussian using functional delta method and Hadamard differentiability of the relevant transformation maps. This paper builds on the theoretical results from (Chernozhukov et al., 2010, 2013) which establish functional central limit theorem for distribution regression based estimator of conditional distribution functions, and the Hadamard differentiability of the counterfactual operator and the monotone rearrangement operator.

I also extend the analysis for non-continuous outcome variables. As mentioned, although the quantile regression-based CIC (Melly and Santangelo, 2015) and the distribution regression-based CIC are similar in the continuous outcome case, the approach proposed in this paper can accommodate mixed continuous/discrete outcome variables without adjusting the estimation step for given group-time pairs, and can thus be applied to a wider range of important applications. For mixed continuous/discrete outcomes, the CIC-identified counterfactual outcome distribution maybe partially identified for sub-regions of the outcome support. I propose bounds on the partially identified distribution function and characterize the settings under which the counterfactual distribution remains point identified. These conditions are verifiable through data directly and can provide guidance to the applied researcher as to when and where they may obtain point

[2]See Fortin et al. (2021) for a detailed review of decomposition methods in economics.

estimates versus partially identified bounds.

Lastly, I apply the CIC based on distribution regression on an empirical application revisiting Hoynes et al. (2015) to assess the heterogeneous effect of Earned Income Tax Credit (EITC) on infant weights. I also present new results comparing the estimates from the more flexible CIC-DR approach compared to the residualized CIC approach, which is the status quo option for implementing CIC with control variables. Estimation results suggest that once heterogeneity of control variables' impact across the outcome distribution is allowed, EITC had higher concentrated benefits for lower birth weights and more muted effects across the birth weight distribution than previously reported (30g per $1,000 of tax refund spending at the bottom of the birthweight distribution, which is economically meaningful).

2 Model & Identification

2.1 Notation

Consider the canonical 2×2 quasi-experimental setting with data on individual i's outcomes, $Y_i \in \mathbb{R}$, and characteristics, $X_i \in \mathbb{R}^{d_x}$ where d_x denotes the dimension of X. Keeping in line with the notation from Athey and Imbens (2006), a unit i belongs to either a control group ($G_i = 0$) or a treatment group ($G_i = 1$), and pre- ($T_i = 0$) or post-treatment period ($T_i = 1$). Denote the treatment status as $I_i = G_i \times T_i$. Data on cross-sectional realizations of (Y_i, X_i, G_i, T_i) for $i = 1, \ldots, N_{gt}$ for each $(g, t) \in \{(0,0), (0,1), (1,0), (1,1)\}$ pair are observed. Let Y_i^N be the potential outcome for individual, i, under no treatment, and Y_i^I be the potential outcome under treatment. Also denote $\mathcal{I}_{gt} = \{i : G_i = g, T_i = t\}$ to be the set of individuals that belong to group g and time t. Then using the potential outcome framework, the observed outcome in a given time period is,

$$Y_i = (1 - I_i) \cdot Y_i^N + I_i \cdot Y_i^I$$

This notation implicitly assumes that there is no anticipation of the treatment effects, i.e. that the treatment has no effect prior to its implementation. The subscript, i, will be omitted unless otherwise required for clarity.

Let $Y_{gt} \sim Y|G = g, T = t$ be the conditional outcome distribution given the group and time

indicator variables, and $U_{gx} \sim U|G = g, X = x$ be the conditional distribution of the unobserved component of the outcome variable. I denote $F_{Y_{gt}|X}(\cdot|x)$ as the conditional distribution function of $Y|G = g, T = t, X = x$ and denote the support of X as $\mathcal{X}$. Define $\mathcal{Y}_{gt} \subseteq \mathbb{R}$ as the region of interest, contained in the support of Y_{gt} and let $\mathcal{T}$ be the set of quantile indices. For a non-decreasing function, $F : \mathcal{Y} \to \mathcal{T} \subseteq [0, 1]$, define $F^{-1} : \mathcal{T} \to \overline{\mathcal{Y}}$, where $\overline{\mathcal{Y}}$ is the closure of $\mathcal{Y}$, as $F^{-1}(u) = \inf\{y \in \mathcal{Y} \mid F(y) \geq u\}$ with the convention that $\inf\{\emptyset\} = \sup \overline{\mathcal{Y}}$ and $\inf \mathbb{R} = \inf \overline{\mathcal{Y}}$. For a well-defined cumulative distribution function, F, this generalized (left) inverse is the quantile function. Lastly, denote independence relation between random variables with $\perp$.

2.2 Quantities of Interest: Functionals of CIC Distribution

I first highlight economically meaningful functionals that are identifiable through CIC conditional on control variables. Once $F_{Y_{11}^N|X}$, the conditional distribution of untreated potential outcome for the treated group in $T = 1$ is identified through conditional CIC, parameters such as conditional and unconditional quantile/distributional treatment effects on the treated (UQTT/UDTT), conditional average treatment effects (CATT), and average treatment effects on the treated (ATT) are identified.

To study differential policy effects across the outcome distribution, quantile treatment effects are often employed. In policy discussions, moreover, unconditional quantile treatment effects might be preferred for ease of communication. UQTT can be identified by,

$$\delta_{UQTT}(\tau) = \underbrace{F_{Y_{11}}^{I,-1}(\tau)}_{\text{Identified from data}} - \underbrace{F_{Y_{11}}^{N,-1}(\tau)}_{\text{Identified from CIC}} = F_{Y_{11}}^{I,-1}(\tau) - \left(\int_{\mathcal{X}_{11}} F_{Y_{11}|X}^{N}(y|x) dF_{X_{11}}(x) \right)^{-1}$$

Note that the quantile function of Y_{11}^I is identified directly from the data, whereas the latter quantity in the second equation is identified by inverting the conditional CIC distribution after integrating over the control variable distribution. Relatedly, distributional treatment effects provide an alternative means to present heterogeneous effects across the outcome distribution.

$$\delta_{DTT}(\tau) = F_{Y_{11}}^{I}(y) - F_{Y_{11}}^{N}(y) = F_{Y_{11}}^{I}(y) - \int_{\mathcal{X}_{11}} F_{Y_{11}|X}^{N}(y|x) dF_{X_{11}}(x)$$

Since the entire counterfactual distribution for Y_{11}^N is identified, ATT is also identified. ATT

can be identified by,

$$\delta_{ATT} = \int_{-\infty}^{\infty} \left[F_{Y_{11}^N}(y) - F_{Y_{11}^I}(y) \right] dy$$

which makes use of the relationship between expectation and CDFs through integration by parts,

$$\mathbb{E}[Y_{11}^N] = \int_0^{\infty} \left[1 - F_{Y_{11}^N}(y) \right] dy - \int_{-\infty}^{0} F_{Y_{11}^N}(y) dy$$

For conditional ATT (CATT), one could replace the marginal distribution of Y_{11}^I and Y_{11}^N with that of the conditional distribution of $Y_{11}^I|X$ and $Y_{11}^N|X$.

Remark 1 *Note that this is a different expression than the one proposed in Athey and Imbens (2006) for the average treatment effect, which makes use of the inverse of the CIC transformation map, $F_{Y_{01}}^{-1} \circ F_{Y_{00}}(\cdot)$. Their expression is given by,*

$$\delta_{ATT} = \mathbb{E}\left[Y_{11}^I\right] - \mathbb{E}\left[F_{Y_{01}}^{-1}\left(F_{Y_{00}}(Y_{10})\right)\right]$$

This transformation yields the post-treatment period outcome for an individual with outcome y in the pre-treatment period (based on outcome changes observed in the control group over time). Although the two are different ways of expressing the same quantity and thus are not consequential for identification, the first expression is more amenable to straightforward estimation based on distribution regression.

Remark 2 *The CIC implied ATT and CATT, in general, differ from those implied by standard DID methods as recognized in Athey and Imbens (2006) and Roth and Sant'Anna (2023). Because the identifying assumptions of CIC and DDID are non-nested, these can result in different values. Note that although the expression for $\mathbb{E}[Y_{11}^I]$ above is expressed as a functional of $F_{Y_{11}|X}^I(y|x)$ for consistent notation with the counterfactual distribution, it is identified directly from the data. The details around the difference between estimating the mean of Y_{11}^I nonparametrically and semiparametrically is discussed during estimation in Section 3.*

2.3 Conditional CIC

I now discuss the CIC framework with control variables along with the identifying assumptions. The identification for the CIC conditional on controls follows a straightforward adaptation from Athey and Imbens (2006) and has been considered in Melly and Santangelo (2015). I provide a brief overview of the model and identification conditions below. For the main results of the current paper, I focus on the case where Y is continuous and extend the analysis to non-continuous outcomes in Section 5. Assumption 1 below introduces the model for the potential outcome under no treatment and the restrictions required to identify the counterfactual outcome distribution for the treated group in the post-treatment period, $F_{Y_{11}^N}$.

Assumption 1 *1. (Potential Outcome Under No Treatment): The outcome under no treatment is given by* $Y^N = h(U, T, X)$*, where* U *is a continuous scalar random variable and* $h : \mathcal{U} \times \{0, 1\} \times \mathcal{X} \mapsto \mathbb{R}$.

2. (Strict Monotonicity): $u \mapsto h(u, t, x)$ *is strictly monotonic in* u *for* $t = 0, 1$ *and* $\forall x \in \mathcal{X}$.

3. (Conditional Time-Invariance): $U \perp T \mid G, X$

4. (Support Overlap): $\mathcal{U}_{G=1,X=x} \subseteq \mathcal{U}_{G=0,X=x}$, $\forall x \in \mathcal{X}_{11}$

Assumption 1.1 and 1.3 are key conditions for CIC that play the role of a common trends assumption in DID. With a production function that is allowed to evolve over time but is common across groups, unobserved distribution is assumed to be independent of time. This does not require rank invariance among individuals across time, which is a stronger condition than the one presented here. Assumption 1.3 is also referred to as the rank similarity condition as in Chernozhukov and Hansen (2005).

Also note that Assumption 1.3 corresponds to a repeated cross-sectional sampling. When there is access to panel data, where individual observations are tracked over time, such that $Y_{it}^N = h(U_{it}, t, x_i)$ one can adjust the conditional time-invariance assumption to be $U_{i0}|G_i, X_i \stackrel{d}{=} U_{i1}|G_i, X_i$. Assumption 1.2, which states that individuals with higher realizations of the unobserved component, U will also have higher values of Y, enables the comparison of individuals across groups with similar realizations of the outcome.

Lemma 1 *Under Assumption 1, the conditional counterfactual distribution for the treated group in post-treatment period is identified by,*

$$F^N_{Y_{11}|X}(y|x) = F_{Y_{10}|X}(F^{-1}_{Y_{00}|X}(F_{Y_{01}|X}(y|x))) \quad \forall (y,x) \in \mathcal{Y}_{11}\mathcal{X}_{11}$$

The conditional CIC transformation takes the quantile change observed in the control group between $T = 0$ and $T = 1$ for individuals with given characteristics, x, i.e. $F^{-1}_{Y_{00}|X} \circ F_{Y_{01}|X}(\cdot|x)$, and applies this inter-temporal quantile change to the corresponding sub-population (with the same characteristics, x) in the treated group in $T = 0$. This is the counterfactual change in quantiles that the treated group would have observed had the treatment not been implemented. The composition mapping described above for the control group has an optimal transport interpretation and have been explored in related contexts (Arkhangelsky et al., 2019; Gunsilius, 2023). If DID takes arithmetic differences in control group's mean outcome over time, CIC takes the Wasserstein distance between the control group's outcome distribution over time and applies the distance to the treated group.

Note that the CIC-identified untreated potential outcome distribution is monotonic by construction and is invariant to monotonic transformations of the outcome variable. Because the CIC transformation mapping is a composition of monotonic transformations, the resulting counterfactual distribution is a proper cumulative distribution function (CDF). DDID, in contrast, can yield a counterfactual distribution that lie outside of the unit interval, or one that is not monotonic in y (Athey and Imbens, 2006; Ghanem et al., 2023) depending on the application.[3] Because the identifying assumptions between CIC and DDID are non-nested, these will in general result in different counterfactual distributions.

2.4 Decomposition of Heterogeneous Causal Effects

In conducting comparative sub-group analysis, there is an extensive literature on decomposition methods both for mean and distributional outcomes (Fortin et al., 2011). These methods partition the outcome gap into portions explainable by observed compositional differences, such as

[3]In fact, Ghanem et al. (2023) illustrates an example of DDID-identified counterfactual distribution violating the logical monotonicity in their empirical application based on Cengiz et al. (2019).

differences in distribution of education or experience and those not explainable by observable differences (often termed "discriminatory" or wage structure effects) in the spirit of Oaxaca (1973) and Blinder (1973).

In the current setting, one can decompose the policy's causal effect into each of these terms to measure the extent of change in the relative importance in explaining the causal effect of the policy. This could be of particular interest to a social planner that has in mind not only the change in outcome distribution as a result of a regulatory policy, but also the channel through which this occurred. For instance, the over-representation of black and/or female workers in low and medium wage industry has been recognized as a justification for increasing minimum wage to reduce wage disparities (Derenoncourt and Montialoux, 2021; Wursten and Reich, 2023; Blau et al., 2023). If minimum wage increases induce a reduction in racial or gender wage gaps, this maybe viewed as a positive outcome. If outcomes were improved through observable factors such as education or experience (for which detailed decomposition is available in the literature), then this decomposition of treatment effects can provide guidance for policymakers to look to invest in more job training or schooling. On the other hand, if the change in the unobserved component explained more of the treatment effects, then this could suggest a different policy or investment strategy moving forward.

To formalize ideas, consider the counterfactual operator as proposed in Chernozhukov et al. (2013) in a cross-sectional setting with two sub-groups. The decomposition for outcome distribution gaps between, say white and black workers can be expressed as,

$$F_{Y<W|W>} - F_{Y<B|B>} = \underbrace{(F_{Y<W|W>} - F_{Y<W|B>})}_{\triangle_{comp}} + \underbrace{(F_{Y<W|B>} - F_{Y<B|B>})}_{\triangle_{returns}}$$

where

$$F_{Y<W|W>}(y) = \int_{\mathcal{X}_W} F_{Y_W|X_W}(y|x) dF_{X_W}(x)$$

is the stochastic assignment of outcomes to individuals with characteristics, x for white workers (analogously defined for black workers), and

$$F_{Y<W|B>}(y) = \int_{\mathcal{X}_B} F_{Y_W|X_W}(y|x) dF_{X_B}(x)$$

is the counterfactual outcome distribution for white workers had they possessed characteristics observed in black workers. This is well defined under a support overlap condition, $\mathcal{X}_B \subseteq \mathcal{X}_W$.

Next, consider the following treatment effect on the outcome distribution gap e.g. wage gap,

$$(\underbrace{F^I_{Y<W,11|W,11>} - F^I_{Y<B,11|B,11>}}_{\text{Factual Wage Gap}}) - (\underbrace{F^N_{Y<W,11|W,11>} - F^N_{Y<B,11|B,11>}}_{\text{Counterfactual Wage Gap}})$$

which can also be interpreted as difference in sub-group specific treatment effects. Then, adapting the notation to the current 2×2 setting, (g,t)-specific counterfactual outcome distribution for sub-groups of interest, j and k, can be expressed as,

$$F_{Y<j,gt|k,gt>}(y) = \int F_{Y_{j,gt}|X_{j,gt}}(y|x) dF_{X_k,gt}(x)$$

Decomposing each factual and counterfactual wage gap as above and rearranging terms,

$$\begin{aligned}
&= \{(F^I_{Y<W,11|W,11>} - F^I_{Y<W,11|B,11>}) - (F^N_{Y<W,11|W,11>} - F^N_{Y<W,11|B,11>})\} \\
&\quad + \{(F^I_{Y<W,11|B,11>} - F^I_{Y<B,11|B,11>}) - (F^N_{Y<W,11|B,11>} - F^N_{Y<B,11|B,11>})\} \\
&= (\underbrace{\triangle^I_{comp} - \triangle^N_{comp}}_{\delta_{DTE,comp}}) + (\underbrace{\triangle^I_{returns} - \triangle^N_{returns}}_{\delta_{DTE,returns}})
\end{aligned}$$

one can see that the treatment effect on the outcome distribution gap can be decomposed into the change in components attributable to compositional differences and change in components attributable to unobserved differences.

3 Estimation

A plug-in estimation approach based on distribution regression is proposed in this section. Distribution regression, as previously introduced, is a flexible and computationally feasible approach to modeling conditional outcome distributions given control variables. It is a semiparametric[4] model where

$$F_{Y|X}(y|x) = \Lambda(P(x)'\beta(y)) \quad \forall x \in \mathcal{X} \text{ and } y \in \mathcal{Y} \tag{1}$$

[4] $\beta(y)$ is a function over $y \in \mathcal{Y}$.

for some link function $\Lambda(\cdot)$ e.g. logit or probit and $y \mapsto \beta(y)$ is a function-valued parameter. For any link function, Λ, $F_{Y|X}(y|x)$ can be approximated arbitrarily well for rich enough transformations, $P(X)$, and it coincides with the empirical cumulative distribution function of Y if $P(x)$ is set as a constant. Furthermore, one can add more structure on the coefficient function, $\beta(y)$ depending on the empirical application.[5]

I apply equation 1 in the quasi-experimental 2×2 setting to model the (g,t)-specific conditional distribution functions under the following assumption. Although for ease of exposition, x enters linearly inside the link function, more flexible specifications such as interaction terms or higher order terms can be used under $P(x)$. The following modeling assumption is the semiparametric structure that I introduce to be able to flexibly incorporate control variables.

Assumption 2

1. *Samples* $\{(Y_{gt,i}, X_{gt,i}) : i = 1, \ldots, n_{gt}\}$ *are* $i.i.d.$ *copies of* (Y_{gt}, X_{gt}) *for all* (g,t) *pairs that has probability law,* P_{gt}*.*

2. *The conditional distribution function takes the form,*

$$F_{Y_{gt}|X}(y|x) = \Lambda(x'\beta_{gt}(y))$$

for all $(g,t,y,x) \in \{0,1\}^2 \times \mathcal{Y} \times \mathcal{X}$ *where* $\Lambda(\cdot)$ *is a logit/probit link function.*

The second part of Assumption 2 allows for the effect of controls to vary across the outcome distribution as well as across groups and time periods. In the CIC framework, the unconditional time-invariance assumption in Athey and Imbens (2006) is made more plausible by considering the distribution of unobservables, net of the effects by potential observed confounders, to be independent of time.

3.1 Estimation Algorithm

The estimation procedure consists of estimating the conditional outcome distribution for every group and time period, then applying the CIC transformation conditional on each control variable

[5]For instance, Fortin et al. (2021) uses linear-in-y restrictions to identify and estimate spillover effects of minimum wage.

value. Once the conditional estimates have been obtained, one can further recover unconditional estimates by weighted averages. Given Assumption 2, the estimation proceeds in the following steps:

1. Specify a grid over the outcome support (e.g. 100 equidistant points across in the support over which the distribution function is evaluated; these could be the quantiles of the outcome variable or equidistant grid between the min and the max of the a sub-interval of the outcome support) that the researcher is interested in estimating, $\mathcal{S} \subseteq \mathcal{Y}_{01}$.

2. For each $y \in \mathcal{S}$, estimate conditional distribution functions for each (g,t) pair through distribution regression using conditional Maximum Likelihood,

$$\hat{F}_{Y_{gt}|X}(y|x) = \Lambda(x'\hat{\beta}_{gt}(y)), \quad \forall x \in \mathcal{X}$$

$$\hat{\beta}_{gt}(y) \in \underset{b\in\mathbb{R}^p}{\arg\max} \sum_{i\in\mathcal{I}_{gt}} \left[\mathbf{1}\{Y_i \leq y\} \ln \Lambda(X_i'b) + \mathbf{1}\{Y_i > y\} \ln\left(1 - \Lambda(X_i'b)\right)\right]$$

3. Rearrange potentially non-monotonic $y \mapsto \hat{F}_{Y_{00}|X}(y|x)$ using monotonic rearrangement operator of Chernozhukov et al. (2010) and invert $\hat{F}_{Y_{00}|X}(y|x)$ to obtain $\tilde{F}^{-1}_{Y_{00}|X}(\tau|x)$. Analytically, the inverse of the rearranged distribution function has the following representation,

$$\tilde{F}^{-1}_{Y_{00}|X}(\tau|x) = \int_0^\infty \mathbf{1}\left\{\hat{F}_{Y_{00}|X}(y|x) < \tau\right\} dy - \int_{-\infty}^0 \mathbf{1}\left\{\hat{F}_{Y_{00}|X}(y|x) > \tau\right\} dy$$

4. Plug in distribution function estimates into CIC transformation to obtain counterfactual probabilities at $y \in \mathcal{T}$,

$$\hat{F}^N_{Y_{11}|X}(y|x) = \hat{F}_{Y_{10}|X}\left(\tilde{F}^{-1}_{Y_{00}|X}\left(\hat{F}_{Y_{01}|X}(y|x)\right)\right) \quad \forall x \in \mathcal{X}$$

   The full expression after plugging in the rearranged distribution function is given by,

$$\hat{F}^N_{Y_{11}|X}(y|x) = \Lambda\left(x'\hat{\beta}_{10}\left(\int_0^\infty \mathbf{1}\{x'\hat{\beta}_{00}(\tilde{y}) < x'\hat{\beta}_{01}(y)\}d\tilde{y} - \int_{-\infty}^0 \mathbf{1}\{x'\hat{\beta}_{00}(\tilde{y}) > x'\hat{\beta}_{01}(y)\}d\tilde{y}\right)\right)$$

5. For the unconditional distribution function, integrate over the empirical distribution of controls in the treated group.

$$\hat{F}_{Y_{11}^N}(y) = \frac{1}{n_{11}} \sum_{i=1}^{n_{11}} \hat{F}^N_{Y_{11}|X}(y|X_i)$$

Computationally, Step 2 consists of estimating a series of logit or probit regression using the same set of controls over the grid of the outcome support. Although the computational burden increases with the size of $\mathcal{S}$, this process is straightforward and feasible, since it can be implemented in parallel across the grid. It is also a flexible method compared to the existing partialling out method that restrict treatment effect heterogeneity across different realizations of X.

Step 3 concerns the monotonic rearrangement of the estimated conditional distribution function. This maybe required in practice, because in finite samples, even under correct specification of the distribution function, an estimated conditional distribution function may not be monotonic in y for some $x \in \mathcal{X}$ in some regions of $\mathcal{T}$. Computationally, this amounts to sorting the initially predicted values of the conditional distribution function to obey logical monotonicity restrictions of a distribution function. Note that this is an estimation adjustment and not an identification restriction, unlike the non-monotonicity that may arise in the DDID framework due to the way in which counterfactual distributions are identified and constructed. In this case, one can apply the monotone rearrangement procedure proposed in Chernozhukov et al. (2010) to recover a monotonic distribution function. Under Assumption 2, the first-order asymptotic behavior of the rearranged distribution function is equivalent to that of the estimated conditional CiC distribution function. Therefore, the large sample theory pertaining to the distribution regression based CIC estimator will hold for the rearranged distribution function as well.

Once the counterfactual conditional distribution function has been estimated, functionals of this counterfactual distribution function can also be estimated using a plug-in approach. Unconditional quantile treatment effects, for instance, can be obtained by integrate over the empirical

distribution of controls in the treated group and invert the unconditional distribution functions.

$$\begin{aligned}\hat{\delta}_{UQTE}(\tau) &= \hat{F}^{I,-1}_{Y_{11}|X}(\tau) - \hat{F}^{N,-1}_{Y_{11}|X}(\tau) \\ &= \left(\frac{1}{n_{11}}\sum_{i=1}^{n_{11}} \hat{F}^{I}_{Y_{11}|X}(y|X_i)\right)^{-1} - \left(\frac{1}{n_{11}}\sum_{i=1}^{n_{11}} \hat{F}^{N}_{Y_{11}|X}(y|X_i)\right)^{-1}\end{aligned}$$

Note that the unconditional distribution function of Y^I_{11} can be estimated in two ways. Because Y^I_{11} is observed directly in the data, one can take the empirical CDF or could use distribution regression in the same way as rest of the (g,t) pairs and then aggregate it over the distribution of the control variables.

Remark 3 The two approaches are equivalent if Λ is equal to the logit link function.[6] Under other link functions, the two might not be equal, but the resulting estimates of the conditional distribution function can be smoother than the nonparametric CDF in many applications. This can also be achieved if parametric restriction on the $\beta(y)$ function itself is imposed, for instance to be linear in y. In this case, there would be a robustness-precision tradeoff.

4 Asymptotic Theory

This section derives the asymptotic distribution of the distribution regression based CIC estimator (henceforth "CIC-DR") and shows that the functionals of the counterfactual distribution function are asymptotically Gaussian. The results are derived under the standard regularity conditions from Chernozhukov et al. (2013) that ensure the functional central limit theorem (FCLT) holds for the DR-based (g,t)-conditional distribution estimators, $\hat{F}_{Y_{gt}|X}$ for all (g,t) pairs. Using the FCLT of the plug-in ingredients along with the Hadamard differentiability of the CIC transformation map, $\phi(F_1, F_2, F_3) = F_1 \circ F_2^{-1} \circ F_3$, one can apply the functional delta method to derive the asymptotic Gaussianity of the CIC-DR process. Furthermore, the asymptotic distribution of smooth functionals such as unconditional quantile and distributional treatment effects can obtained by using the functional delta method based on the FCLT of the CIC-DR process and the Hadamard differentiability of the counterfactual operator from Chernozhukov et al. (2013), which

[6] This can be verified by examining the first order condition of the log-likelihood function and using the fact that $\lambda(z) = \Lambda(z)(1-\Lambda(z))$.

is used for integrating over the distribution of the control together with the chain rule. Following the derivation of the asymptotic distribution for the CIC-DR and the unconditional effects, I also establish the validity of exchangeable bootstrap for the CIC-DR process, which include commonly used bootstrap schemes such as the empirical, weighted and m out of n bootstrap, using the functional delta method for the bootstrap.

4.1 Asymptotic Distribution

To begin, I introduce necessary definitions for presentation of the main results and regularity conditions that ensure that the DR-based conditional distribution functions obey the FCLT. Let $\ell^\infty(\mathcal{Y}_{gt}\mathcal{X})$ be the set of bounded and measurable functions, $g : \mathcal{Y}_{gt} \times \mathcal{X} \to \mathbb{R}$ for $g, t, \in \{0, 1\}$.

Assumption 3

1. *(DR-i) For every (g, t) pair, $\mathcal{Y}_{gt}$ is a compact interval in $\mathbb{R}$ and $\mathcal{X}$ is a compact subset of $\mathbb{R}^{d_x}$. The conditional density function, $f_{Y_{gt}|X}(y|x)$ exists, is positive and uniformly bounded, and is uniformly continuous in (y, x) on $\mathcal{Y}_{gt}\mathcal{X}$.*

2. *(DR-ii) $\mathbb{E}\left[||X||^2\right] < \infty$ and the minimum eigenvalue of the Jacobian,*

$$J_{gt}(y) := \mathbb{E}\left[\frac{\lambda(X'\beta_{gt}(y))^2}{\Lambda(X'\beta_{gt}(y))\left(1 - \Lambda(X'\beta_{gt}(y)\right)} XX'\right]$$

is bounded away from zero uniformly in $y \in \mathcal{Y}_{gt}$ for every (g, t) pair, where λ is the derivative of Λ.

Under Assumptions 2 and 3, Corollary 5.4 of Chernozhukov et al. (2013) implies that as $N_{gt} \to \infty$, the following holds for the DR-based conditional distribution function, the conditional quantile function and the inverse of the monotone rearranged conditional distribution function.

$$\sqrt{N_{gt}}\left(\hat{F}_{Y_{gt}|X}(y|x) - F_{Y_{gt}|X}(y|x)\right) \rightsquigarrow \mathbb{Z}_{gt}(y, x) \quad \text{in } \ell^\infty(\mathcal{Y}_{gt}\mathcal{X}) \tag{2}$$

$$\sqrt{N_{gt}}\left(\hat{F}^{-1}_{Y_{gt}|X}(u|x) - F^{-1}_{Y_{gt}|X}(u|x)\right) \rightsquigarrow \mathbb{Z}^{F^{-1}}_{gt}(u, x) \quad \text{in } \ell^\infty(\mathcal{U}_{gt}\mathcal{X}) \tag{3}$$

$$\sqrt{N_{gt}}\left(\hat{F}^{-1,R}_{Y_{gt}|X}(u|x) - F^{-1,R}_{Y_{gt}|X}(u|x)\right) \rightsquigarrow \mathbb{Z}^{F^{-1}}_{gt}(u, x) \quad \text{in } \ell^\infty(\mathcal{U}_{gt}\mathcal{X}) \tag{4}$$

where $\mathbb{Z}_{gt}(y,x) = \mathbb{G}_{gt}(\ell_{gt,y,x})$ is a tight mean-zero Gaussian process indexed by $(y,x) \in \mathcal{Y}_{gt}\mathcal{X}$ with the covariance function[7] given by,

$$\mathbb{E}\left[\mathbb{G}_{gt}(\ell_{gt,y,x})\mathbb{G}_{gt}(\ell_{gt,\tilde{y},\tilde{x}})\right] = \mathbb{E}\left[\ell_{gt,y,x}\ell_{gt,\tilde{y},\tilde{x}}\right] - \mathbb{E}\left[\ell_{gt,y,x}\right]\mathbb{E}\left[\ell_{gt,\tilde{y},\tilde{x}}\right]$$

with,

$$\ell_{gt,y,x}(Y_{gt},X) := -\lambda(x'\beta_{gt}(y))J_{gt}^{-1}(y)x'\frac{\Lambda(X'\beta_{gt}(y)) - \mathbf{1}\{Y_{gt} \leq y\}}{\Lambda(X'\beta_{gt}(y))(1-\Lambda(X'\beta_{gt}(y))}\lambda(X'\beta_{gt}(y))X$$

and where $\mathbb{Z}_{gt}^{F,-1}(u,x)$ is a tight mean-zero Gaussian process with the covariance function given by,

$$\mathbb{E}\left[\mathbb{G}_{gt}(\bar{\ell}_{gt,u,x})\mathbb{G}_{gt}(\bar{\ell}_{gt,\tilde{u},\tilde{x}})\right] = \mathbb{E}\left[\bar{\ell}_{gt,u,x}\bar{\ell}_{gt,\tilde{u},\tilde{x}}\right] - \mathbb{E}\left[\bar{\ell}_{gt,u,x}\right]\mathbb{E}\left[\bar{\ell}_{gt,\tilde{u},\tilde{x}}\right]$$

with,

$$\bar{\ell}_{gt,u,x}(Y_{gt},X) := -\frac{1}{f_{Y_{gt}|X}(F^{-1}_{Y_{gt}|X}(u|x))}\ell_{gt,F^{-1}_{Y_{gt}|X}(u|x),x}(Y_{gt},X)$$

The result above for the inverse of the monotone rearranged conditional distribution function demonstrates that the empirical process (3) above is asymptotically equivalent to the process (2). This is a useful result, since it implies that the monotone rearrangement operator does not affect the asymptotic distribution of the conditional CIC-DR process under the same conditions that ensure the FCLT holds for the DR-based conditional distribution and quantile process.

I now take the FCLT results above as inputs for deriving the asymptotic distribution for the conditional CIC-DR process. This is achieved by applying the functional delta method (Lemma 3.10.4 in Van Der Vaart and Wellner (1996)) based on the Hadamard differentiability of the CIC transformation map,

$$\phi(F_1,F_2,F_3) = F_1 \circ F_2^{-1} \circ F_3$$

The explicit expression for the derivative map of ϕ along with the relevant function space is con-

[7] In the case of Λ being the logit link function, the expression for the covariance function is given by,

$$\lambda(x'\beta_{gt}(y))\lambda(\tilde{x}'\beta_{gt}(\tilde{y}))x'J_{gt}^{-1}(y) \times \mathbb{E}\left[\left(\Lambda(X'\beta_{gt}(y\wedge\tilde{y})) - \Lambda(X'\beta_{gt}(y))\Lambda(X'\beta_{gt}(\tilde{y}))\right)XX'\right]J_{gt}^{-1}(\tilde{y})\tilde{x}$$

and when $\Lambda = \Phi$ for probit link, a comparable expression can be derived with slight adjustments.

tained in Appendix A. The following theorem presents the formal result for the CIC-DR process.

Theorem 1 *Suppose that Assumptions 1, 2 and 3 hold. Then as $N_{gt} \to \infty$ and as $N \to \infty$, where $N = \sum_{g,t} N_{gt}$ such that $N/N_{gt} \to s_{gt} \in [0, \infty)$ for all $(g,t) \in \{0,1\}^2$ for $g, t, = 0, 1$, the following result holds in $\ell^\infty(\mathcal{Y}_{11}\mathcal{X})$:*

$$\sqrt{N_{11}}\left(\hat{F}_{Y_{11}^I|X}(y|x) - F_{Y_{11}^I|X}(y|x)\right) \rightsquigarrow \mathbb{Z}_{11}^I(y,x)$$
$$\sqrt{N}\left(\hat{F}_{Y_{11}^N|X}(y|x) - F_{Y_{11}^N|X}(y|x)\right) \rightsquigarrow \mathbb{Z}_{11}^N(y,x)$$

where $\mathbb{Z}_{11}^I(y,x)$ and $\mathbb{Z}_{11}^N(y,x)$ are independent tight mean-zero Gaussian processes indexed by $(y,x) \in \mathcal{Y}_{11}\mathcal{X}$. The covariance function for $\mathbb{Z}_{11}^I(y,x)$ is given by $\mathbb{E}\left[\ell_{11,y,x}\ell_{11,\tilde{y},\tilde{x}}\right]$ and the expression for $\mathbb{Z}_{11}^N(y,x)$ can be found in Appendix A.

Theorem 1 states that the centered and re-scaled DR-based conditional CIC process is asymptotically Gaussian. With this in hand, the asymptotic properties of smooth functionals of the conditional CIC-DR process can also be derived. Before proceeding onto the more general result on the asymptotic Gaussianity of smooth functionals of the CIC-DR process, I showcase results on two treatment effects that researchers commonly study. The first functional to consider is the UQTT. The asymptotic distribution can be obtained by applying the functional delta method in combination with Hadamard differentiability of the counterfactual operator as defined in Chernozhukov et al. (2013) as,

$$\int_{\mathcal{X}_{g't'}} F_{Y_{gt}|X}(y|x) dF_{X_{g't'}}(x)$$

where possibly $g' \neq g$ and $t' \neq t$. In the case of marginal distribution functions or quantile functions, one integrates over $F_{X_{11}}$, where $g = g' = t = t' = 1$. In the case of decompositional treatment effects, however, one could consider integrating over the $F_{X_{g',t'}}$ that belongs to another sub-group of interest, under the appropriate common support condition as mentioned in Section 2. This leads to the following corollaries.

Corollary 1.1 *Suppose that conditions for Theorem 1 hold. Then, as $N \to \infty$, the UQTT and UDTT*

processes converge weakly to the following limit processes:

$$\sqrt{N}\left(\hat{\delta}_{UQTT}(\tau) - \delta_{UQTT}(\tau)\right) \rightsquigarrow \mathbb{Z}_{UQTT}(\tau) \quad \textit{in } \ell^{\infty}(\mathcal{T})$$
$$\sqrt{N}\left(\hat{\delta}_{UDTT}(y) - \delta_{UDTT}(y)\right) \rightsquigarrow \mathbb{Z}_{UDTT}(y) \quad \textit{in } \ell^{\infty}(\mathcal{Y}_{11})$$

where $\delta_{UQTT}(\tau) = F^{-1}_{Y^I_{11}}(\tau) - F^{-1}_{Y_{11^N}}(\tau)$ *and* $\delta_{UDTT}(y) = F_{Y^I_{11}}(y) - F_{Y^N_{11}}(y)$. $\mathbb{Z}^{UQTT}_{11}(\tau)$ *and* $\mathbb{Z}^{UDTT}_{11}(y)$ *are tight mean-zero Gaussian processes indexed by* $\tau \in \mathcal{T}$ *and* $y \in \mathcal{Y}_{11}$ *, with covariance functions defined in Appendix A.*

Next, I present a more general result for smooth functionals of the potential outcome distributions such as Lorenz curves and Gini coefficients.

Theorem 2 *Suppose that conditions for Theorem 1 hold and that the map indexed by* z *given by,* $\phi\left(F_{Y^N_{11}|X}, F_{Y^I_{11}|X}, F_X\right)(z)$ *is Hadamard differentiable with derivative maps,* $\phi'_{F_{Y^N_{11}|X},F_X}$ *and* $\phi'_{F_{Y^I_{11}|X},F_X}$*. Then the following holds.*

$$\sqrt{N}\left(\phi\left(\hat{F}_{Y^I_{11}|X}, \hat{F}_{Y^N_{11}|X}, \hat{F}_X\right)(z) - \phi\left(F_{Y^I_{11}|X}, F_{Y^N_{11}|X}, F_X\right)(z)\right)$$
$$\rightsquigarrow \left(\phi'_{F_{Y^I_{11}|X},F_X}\mathbb{Z}_{F_{Y^I_{11}}}\right)(z) + \left(\phi'_{F_{Y^N_{11}|X},F_X}\mathbb{Z}_{F_{Y^N_{11}}}\right)(z)$$

Here, $\phi'_{F_{Y_{gt}|X},F_X}$ can be defined as a composition of Hadamard differentiable maps depending on the quantity of interest. For example, if unconditional distribution or quantile functions are involved, the derivative map can be a composition of the outer map and of the counterfactual operator (see proof of Corollary 1.1 as an example). If conditional effects are of interest, one can set the map so that the conditional distribution function is left not integrated over the control variable distribution.

Because the above results hold uniformly across indexes considered as an empirical process, this implies that asymptotic normality holds pointwise for a given point in the index set. As noted in Melly and Santangelo (2015), the asymptotic analysis in this paper is done uniformly for the entire conditional distribution and quantile process based on distribution regression, thereby nesting the pointwise case of Athey and Imbens (2006) as a special case where X is a constant.

4.2 Inference: Bootstrap Validity

In practice, the asymptotic variance function of the limiting Gaussian process maybe difficult to estimate analytically. Therefore, this section establishes conditions under which bootstrap validity holds and suggests an asymptotically valid bootstrap inference procedure. As have been considered in related contexts (Chernozhukov et al., 2013; Melly and Santangelo, 2015; Wüthrich, 2019), this paper proposes exchangeable bootstrap, which includes bootstrap schemes such as the empirical bootstrap, weighted bootstrap and m-out-of-n bootstrap as special cases. The weighted bootstrap, for example, is useful when the sample size is small and control variables are categorical (which can cause "small cell" problems) since no observation is discarded in the resampling process. When the sample is large, one can utilize subsampling for computational tractability. The asymptotic distribution can be consistently estimated using exchangeable bootstrap by functional delta method for bootstrap.

The following assumption around bootstrap weights establishes conditions under which bootstrap validity is achieved (appears under Condition EB in Chernozhukov et al. (2013)).

Assumption 4 *For each (g,t), $(w_{gt,1}, w_{gt,2}, \ldots, w_{gt,N_{gt}})$ is an exchangeable, non-negative random vector independent from the data, $\{(Y_{gt,i}, X_i)\}_{i=1}^{N_{gt}}$ such that for some $\varepsilon > 0$,*

$$\sup_{N_{gt}} \mathbb{E}\left[w_{gt,1}^{2+\varepsilon}\right] < \infty, \quad \frac{1}{N_{gt}} \sum_{i=1}^{N_{gt}} (w_{gt,i} - \bar{w}_{gt})^2 \to_p 1, \text{ and } \bar{w}_{gt} \to_p 1$$

where $\bar{w}_{gt} = N_{gt}^{-1} \sum_{i=1}^{N_{gt}} w_{gt,i}$ is independent across (g,t).

For applications with "small cells', one can use the weighted bootstrap with weights satisfying, $\mathbb{E}[w_{gt,1}] = Var(w_{gt,1}) = 1$ such as the standard exponential distribution.

4.2.1 Bootstrap Algorithm

Any exchangeable bootstrap scheme obeying Assumption 4 will be valid, and a bootstrap procedure for conducting inference for UQTT is provided below.

1. For each (g,t)-conditional sample $\{(Y_{gt,i}, X_i)\}_{i=1}^{N_{gt}}$, draw bootstrap weights, $(w_{gt,1}, \ldots, w_{gt,N_{gt}})$ that satisfy Assumption 4, and normalize weights to sum to 1 for every (g,t)-conditional

sample.

2. Estimate the DR-based weighted conditional distribution function based on $(w_{gt,1}, \ldots, w_{gt,N_{gt}})$,

$$\hat{F}^*_{Y_{gt}|X}(y|x) = \Lambda(x'\hat{\beta}^*_{gt}(y)), \quad \forall (y,x) \in \mathcal{Y}_{gt}\mathcal{X}, \quad g,t, \in \{0,1\}$$
$$\hat{\beta}^*_{gt}(y) = \arg\max_b \sum_{i \in \mathcal{I}_{gt}} w_{gt,i} \left[\mathbf{1}\{y_{gt,i} \leq y\} \ln \Lambda(x_i'b) + \mathbf{1}\{y_{gt,i} > y\} \ln\left(1 - \Lambda(x_i'b)\right)\right]$$

3. Obtain bootstrap CIC-DR counterfactual distribution,

$$\begin{aligned}\hat{F}^*_{Y^N_{11}|X}(y|x) &= \hat{F}^*_{Y_{10}|X}\left(\hat{F}^{*,-1,R}_{Y_{00}|X}\left(\hat{F}^*_{Y_{01}|X}(y|x)\right)\right) \\ &= \Lambda\left(x'\hat{\beta}^*_{10}\left(\int_0^\infty \mathbf{1}\{x'\hat{\beta}^*_{00}(\tilde{y}) < x'\hat{\beta}^*_{01}(y)\}d\tilde{y} - \int_{-\infty}^0 \mathbf{1}\{x'\hat{\beta}^*_{00}(\tilde{y}) > x'\hat{\beta}^*_{01}(y)\}d\tilde{y}\right)\right)\end{aligned}$$

4. Estimate the treatment effect (or a functional of interest) by plugging in the relevant estimates, e.g. for UQTT,

$$\begin{aligned}\hat{\delta}^*_{UQTT}(\tau) &= \hat{F}^{*,-1}_{Y^I_{11}|X}(\tau) - \hat{F}^{*,-1}_{Y^N_{11}|X}(\tau) \\ &= \left(\sum_{i \in \mathcal{I}_{11}} w_{11,i}\hat{F}^*_{Y^I_{11}|X}(y|X_i)\right)^{-1} - \left(\sum_{i \in \mathcal{I}_{11}} w_{11,i}\hat{F}^*_{Y^N_{11}|X}(y|X_i)\right)^{-1}\end{aligned}$$

5. Repeat steps 1–4 for $b = 1, \ldots, B$ times and obtain bootstrap estimates, $\left\{\hat{\delta}^{*,(b)}_{UQTT}(\tau)\right\}_{\tau \in \mathcal{T}, b=1,\ldots,B}$.

6. Construct simultaneous $(1-\alpha)$-confidence bands based on maximal t-statistics across $\tau \in \mathcal{T}$ as the end-point functions defined by,

$$CI_{1-\alpha}(\tau) = \hat{\delta}_{UQTT}(\tau) \pm \hat{t}_{1-\alpha} \cdot \hat{\Sigma}_{UQTT}(\tau)^{1/2}/\sqrt{N}$$

where $\hat{t}_{1-\alpha}$ a consistent estimator of the $(1-\alpha)$-th quantile of the Kolmogorov-Smirnov maximal t-statistic, of $\left\{\sup_{\tau \in \mathcal{T}} \sqrt{N}\hat{\Sigma}_{UQTT}(\tau)^{-1/2}|\hat{\delta}^{(b)}_{UQTT}(\tau) - \hat{\delta}_{UQTT}(\tau)| : b = 1, \ldots, B\right\}$, and $\hat{\Sigma}_{UQTT}(\tau)$ is a uniformly consistent estimator of $\Sigma_{UQTT}(\tau)$, the asymptotic variance function of $\hat{Z}^{(b)}_{UQTT}(\tau) := \sqrt{N}\left(\hat{\delta}^{(b)}_{UQTT}(\tau) - \hat{\delta}_{UQTT}(\tau)\right)$. For instance, this could be the bootstrap inter-quartile range re-scaled with the standard normal distribution such as, $\hat{\Sigma}_{UQTT}(\tau) =$

$(q_{0.75}(\tau)-q_{0.25}(\tau))/(z_{0.75}-z_{0.25})$, where $q_p(\tau)$ is the p-th quantile of $\left\{\hat{Z}^{(b)}_{UQTT}(\tau) : b = 1, \ldots, B\right\}$ and z_p is the p-th quantile of $\mathcal{N}(0,1)$.

If the main target of inference is the CIC-DR process itself, then the asymptotic simultaneous confidence bands as constructed in Step 6 can be adjusted to be centered around $\hat{F}_{Y^N_{11}|X}(y|x)$ instead and with analogous adjustments to $\hat{t}_{1-\alpha}$ and $\hat{\Sigma}(y)$. In this paper, I consider uniform inference methods that cover standard pointwise methods for real-valued parameters as a special case. Specifically, the asymptotic simultaneous confidence band, $CI_{1-\alpha}$ above, satisfies the coverage guarantee given by,

$$\lim_{N\to\infty} \mathbb{P}\left\{\hat{\delta}_{UQTT}(\tau) \in CI_{1-\alpha}(\tau), \ \ \forall \tau \in \mathcal{T}\right\} = 1-\alpha.$$

This also makes it possible for considering inference on functionals of the CIC-DR distribution function, such as the Lorenz curve, Gini coefficient and decompositional treatment effects. For large sample sizes, one could consider computationally more tractable means such as subsampling or multiplier bootstrap that simulate the limit process of $\sqrt{N}\left(\hat{F}_{Y^I_{11}|X}(y|x) - \hat{F}_{Y^N_{11}|X}(y|x)\right)$, or the relevant centered and re-scaled empirical process (e.g. functional of the CIC-DR process) as in Barrett and Donald (2003). Formally, the following theorem formulates the validity of the exchangeable bootstrap.

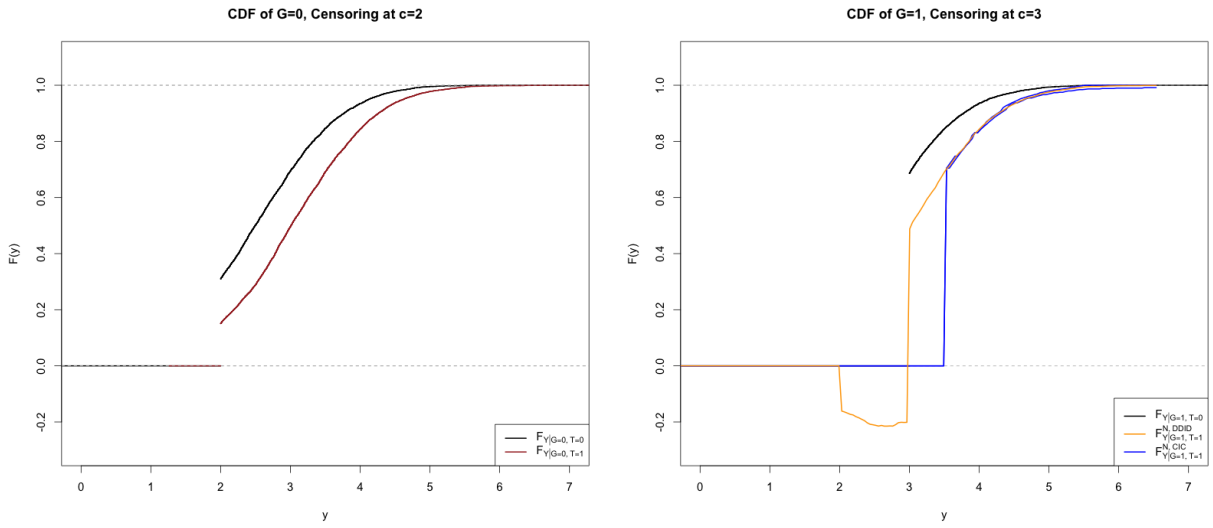
Theorem 3 *Suppose Assumptions 1, 2, 3 and 4 hold. Then the exchangeable bootstrap consistently estimates the limit laws for the processes in Theorem 1 and in Corollaries 1.1.*

5 Extension: CIC for Mixed Continuous-Discrete Outcomes

In this section, I extend the preceding framework developed for continuous outcomes to mixed continuous and discrete outcomes where the potential outcome distributions can exhibit mass points (e.g. censored outcomes or bunching). Censored outcomes, for example, can be highly relevant in analyzing the distribution of wages or hours worked in minimum wage policy settings. Such a setting highlights the general applicability of a DR-based approach for CIC. First, DR can accommodate mixed continuous and discrete outcome variables without special adjustment to the regression estimator, whereas QR does not. Second, although $F_{Y^N_{11}}$ maybe partially identified

under the CIC framework, the bounds on $F_{Y_{11}^N}$ will obey natural monotonicity restrictions and will yield predictions within the closed unit interval. This is in contrast to DDID which can yield counterfactual probabilities outside the unit interval in the lower or upper tails, precisely where possibly heterogeneous effects are of interest.

Toy Example: Illustrating with a simple example can be instructive. Consider a setting with left-censored outcomes, no control variables and censoring points ($c_{G=0} = 2$ and $c_{G=1} = 3$) that are constant across T but different between G. This simulates a common setting when analyzing the effect of a minimum wage increase on the distribution of wages between two neighboring states. Panel (b) of Figure 1 shows that the counterfactual distribution identified by DDID (in



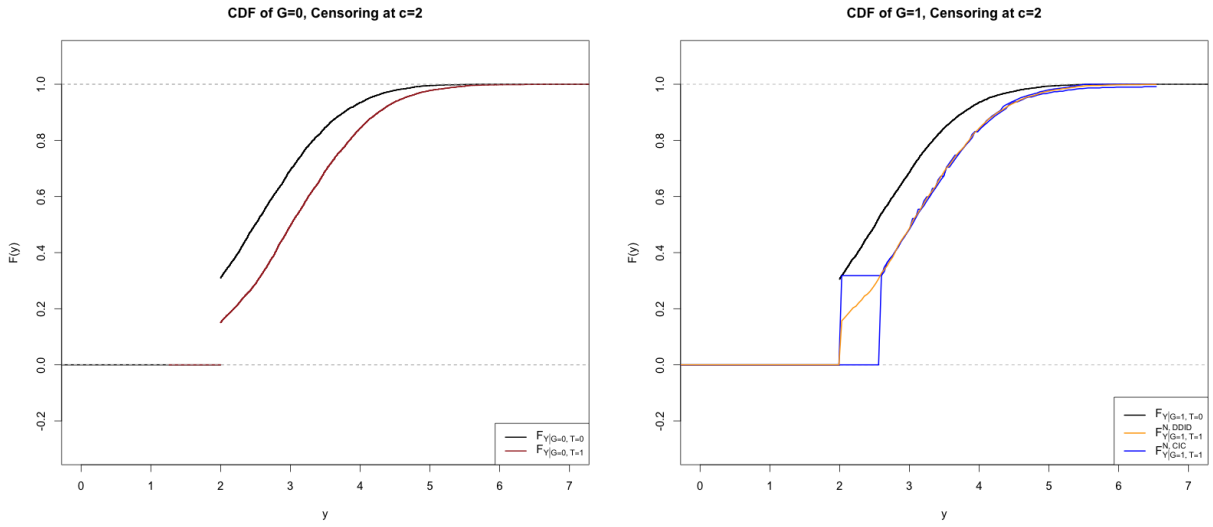
(a) Control Group (b) Treated Group

Figure 1: Example of censored outcomes at $c_0 = 2, c_1 = 3$. Point identified $F_{Y_{11}^N|X}$.

orange),

$$F_{Y_{11}^N}^{DDID}(y) = F_{Y_{10}}(y) + F_{Y_{01}}(y) - F_{Y_{00}}(y)$$

has negative probabilities for $y \in [c_0, c_1]$, because $F_{Y_{10}}(y) = 0$ while $F_{Y_{01}}(y) < F_{Y_{00}}(y)$ for $y \in [c_0, c_1]$. This stands in contrast with the CIC identified $F_{Y_{11}^N}$ (in blue), which is point identified, is non-decreasing and has range inside the unit interval. In Figure 2, the censoring point is constant across G and T. Here, $F_{Y_{11}^N}$ is partially identified for $y \in [2, 3)$ and the bounds for that region are shown in blue. The main reason for this is due to the non-uniqueness of the inverse function



(a) Control Group

(b) Treated Group

Figure 2: Example of censored outcomes at $c_0 = c_1 = 2$. Partially identified $F_{Y_{11}^N|X}$.

of $F_{Y_{00}|X}(\cdot|x)$ at the censoring points. As will be made explicit in the subsequent section, for censored outcomes, and more generally outcomes with finite number of discontinuities in the conditional outcome distribution function, the generalized left and the right inverses no longer agree with each other as in the continuous case.

Here, the upper bound is derived in the same manner as the point estimate for the counterfactual distribution in the continuous outcome case, since the generalized left-inverse expression is the same in the continuous outcome case and in this case. This upper bound is the blue line corresponding to the higher horizontal line for $[2, 3)$. The lower bound is derived by the expression to be explicitly stated in the following section, but essentially is taken by the generalized right-inverse and corresponds to the lower horizontal line in Figure 2 panel (b) and has a jump in the distribution function at $y = 3$.

The examples demonstrates that CIC does not necessarily lose point identification in a mixed continuous-discrete outcome variable setting. As will be provided below, whether or not the counterfactual distribution is point identified and to what degree it is partially identified depends on the relationship between the ranges of the conditional distribution functions of $Y_{01}|X$, $Y_{00}|X$ and $\mathbb{Y}_{10}$, the support of Y_{10}.

5.1 Identification for Conditional CIC with Mixed Continuous-DIscrete Outcomes

To formalize concepts, I introduce additional notation. Namely, for a non-decreasing function, $F : \mathcal{Y} \to \mathcal{T} \subseteq [0,1]$, define the generalized (right) inverse, $F^{\dagger} : \mathcal{T} \to \overline{\mathcal{Y}}$, as $F^{\dagger}(u) = \sup\{y \in \mathcal{Y} \cup \{-\infty\} \mid F(y) \leq u\}$ maintaining the convention that $\sup\{\emptyset\} = \inf \overline{\mathcal{Y}}$ and $\sup \mathbb{R} = \sup \overline{\mathcal{Y}}$. For continuous outcomes, the two (left and right) inverses agree, but they disagree in general. For a function $f : \mathbb{R} \to \mathbb{R}$, moreover, let $f(x-) = \sup_{z<x} f(z)$. If f is left-continuous, then $f(x-) = f(x)$. Lastly, let $Ran(F) = \{F(y) : y \in \mathbb{R}\}$ be the range of the function.

For mixed continuous-discrete outcomes, certain adjustments to the identifying assumption can be considered. The assumption of strict monotonicity as introduced in Assumption 1 can be overly restrictive and may not be practically well justified. Therefore, I relax Assumption 1 pertaining to the relationship between U and Y to be weakly monotonic instead.

Assumption 1′ *(Weak Monotonicity):* $u \mapsto h(u,t,x)$ *is weakly monotonic in* u *for* $t = 0,1$ *and* $\forall x \in \mathcal{X}$*. Also, the remaining Assumption 1.1, 1.3, 1.4 hold.*

With this modified assumption, I present a modified identification result. The identification of CIC with mixed continuous-discrete outcomes and no control variables have recently been studied in Ghanem et al. (2023), rationalized using a copula stability condition on the selection mechanism rather than a conditional time invariance assumption as in Athey and Imbens (2006)[8]. Adapting their result to my case of conditional CIC with control variables, I restate the result as a lemma below.

Lemma 2 *Suppose Assumption 1′ holds. Then,* $F_{Y_{11}^N|X}(\cdot)$ *is partially identified with bounds,*

$$\begin{aligned}
&\limsup_{\tilde{y} \downarrow y} \{F^{LB}(y|x) : y \leq \tilde{y} \;\&\; y \in \mathcal{Y}_{01} \cup \{-\infty\}\} \\
\leq &F_{Y_{11}^N|X}(y|x) \\
\leq &\limsup_{\tilde{y} \downarrow y} \{F^{UB}(y|x) : y \leq \tilde{y} \;\&\; y \in \mathcal{Y}_{01} \cup \{-\infty\}\}
\end{aligned}$$

[8]Ghanem et al. (2023) show that the two conditions are equivalent for continuous outcomes, but are non-nested in general for mixed continuous-discrete or discrete outcome settings.

$\forall (x,y) \in \mathcal{Y}_{11}\mathcal{X}_{11}$, *where*

$$F^{LB}(y|x) := F_{Y_{10}|X}\left(F^{\dagger}_{Y_{00}|X}\left(F_{Y_{01}|X}(y|x)\right) - \right)$$
$$F^{UB}(y|x) := F_{Y_{10}|X}\left(F^{-1}_{Y_{00}|X}\left(F_{Y_{01}|X}(y|x)\right)\right)$$

The $\limsup$ expression that wraps around $F^{LB}(y|x)$ and $F^{UB}(y|x)$ are made to ensure that the bounds are themselves right-continuous since they are intended to cover distribution functions. These bounds are valid for continuous, discrete and mixed continuous-discrete outcome variables. Note that for continuous outcomes, the bounds collapse to a single point, specifically the upper bound. Ghanem et al. (2023) also show that these bounds are sharp when $\overline{Ran}(F_{Y_{00}})$ is closed and they numerically coincide with those proposed by Athey and Imbens (2006) when outcomes are purely discrete.

Next, I specialize Lemma 2 to censored outcomes and provide conditions under which the counterfactual distribution is point identified and where it is partially identified, along with their corresponding regions. Note that these conditions are directly verifiable through data, providing a useful check for the applied researcher.

Proposition 1 *Suppose Assumption 1′ holds and $F_{Y_{gt}|X}$ has finite number of discrete components. Then, $F_{Y^N_{11}|X}$ is point identified for the set,*

$$\{(y,x) \in (\mathbb{Y}_{01} \cap \mathbb{Y}_{10}) \times \mathcal{X} : \exists y' \in \mathcal{Y}_{00} \; s.t. \; F_{Y_{10}|X}(y|x) = F_{Y_{00}|X}(y'|x)\}$$

and $F_{Y^N_{11}|X}$ is partially identified for the set,

$$\{(y,x) \in (\mathbb{Y}_{01} \cap \mathbb{Y}_{10}) \times \mathcal{X} : \nexists y' \in \mathcal{Y}_{00} \; s.t. \; F_{Y_{10}|X}(y|x) = F_{Y_{00}|X}(y'|x)\}$$

and the bounds are given by Lemma 2.

The proposition implies that even if $Ran(F_{Y_{10}|X}) \nsubseteq Ran(F_{Y_{00}|X})$, which would result in partial identification in the purely discrete outcome case, if the set $\Gamma_{yx} := \{(y,x) \in \mathbb{Y}_{01} \times \mathcal{X} : Ran(F_{Y_{10}|X}) \nsubseteq Ran(F_{Y_{00}|X})\}$ is empty, then $F_{Y^N_{11}|X}$ remains point identified.

Remark 4 The examples in Figures 1 and 2 highlight that based on the conditions regarding the

range of the distribution functions and the support of the outcome variables, it is possible for the researcher to directly determine from the data whether they will be obtaining bounds on the counterfactual distribution or will have point identification.

Example (Minimum Wage) To solidify intuition and to outline the usefulness of the result for an applied researcher, consider an example of minimum wage in the 2×2 setting again, but with no control variables. Denote the censoring points as w_{gt} for $g, t = 0, 1$ where $w_{gt} = w_g \ \forall t$ for simplicity. Suppose that the control group's baseline minimum wage is set at w_0 and that of the treated group is set at w_1. If a researcher observes in the control group that the fraction of minimum wage workers increases over time i.e. $F_{Y_{01}}(w_0) > F_{Y_{00}}(w_0)$, then the researcher will obtain point identified counterfactual distribution, $F_{Y_{11}^N}$, for the entire region of interest $\mathcal{Y}_{11}$.

Suppose, instead, that the researcher observes that the fraction of minimum wage workers decreases over time in the control group i.e. $F_{Y_{01}}(w_0) < F_{Y_{00}}(w_0)$. Then, if minimum wage is constant across groups, i.e. $w_1 = w_0$ as in Figure 2 panel (b), then $F_{Y_{11}^N}$ is partially identified for $w_0 \leq y \leq F_{Y_{00}}^{-1}\left(F_{Y_{01}}(w_0)\right)$ and the counterfactual distribution defined on the set $\mathcal{Y}_{11}$ excluding the interval $[w_0, F_{Y_{00}}^{-1}\left(F_{Y_{01}}(w_0)\right)]$ will be point identified. If $w_1 > w_0$, then it is less likely for $F_{Y_{11}^N}$ to be partially identified in general. In the running example, given that the distribution function is discontinuous only at the censoring point, the counterfactual distribution will be point identified, since one will be able to find a wage level, w' possibly not equal to w_1, such that $F_{Y_{00}}(w') = F_{Y_{01}}(w_1)$ holds and the inverse of $F_{Y_{00}}$ at $F_{Y_{01}}(w_1)$ is uniquely determined.

Generalization: Time-varying or Conditional Censoring Points One could also consider more general settings, where for example, (1) a control group's censoring point, c_{0t} is allowed to change over time or (2) allowed to vary across individual characteristics, c_{gtx} more broadly. (1) can allow for analyzing minimum wage policy effects even when the available control group also experiences a minimum wage change during the period. This enables the identification of the effect of minimum wage change in the treatment group beyond what would have occurred had they experienced the same minimum wage path as the control group. Because the analysis presented above holds conditionally on observables, the set-up also accommodates settings where the censoring point depends on such observable characteristics. Such a feature allows for the

possibility that minimum wages can vary across industries as in Derenoncourt and Montialoux (2021) or that tax schedules may change differently over time for people with some eligibility status (e.g. disability insurance or having children).

Estimation, Asymptotic Distribution and Inference Estimation consists of a plug-in strategy similar to the one proposed for the continuous outcome case. Because the difference in the bounds arise from non-uniqueness of inverses given finite number of jumps in the otherwise continuous distribution function of the conditional outcome variable, the plug-in estimators for the conditional distribution functions do not require any adjustments. The generalized left- and right-inverses taken for $F_{Y_{00}|X}$ is the only adjustment required in the implementation of conditional CIC that is necessary. For QR-based CIC, the conditional distribution function estimator itself would need to be modified in addition to the different definitions of generalized inverses.

Next, I outline how the asymptotic distribution for the plug-in estimates of the bounds based on DR estimates of the conditional distribution function can be derived. The result can be established similar to how FCLT was derived for the conditional CIC estimator in the continuous outcome case. Note that the DR-based estimator of the upper bound is equivalent to the point identified function in the continuous case, so no adjustment to the DR estimator is required for continuous and mixed continuous-discrete outcomes. Therefore, only the case for the lower bound remains to be shown.

For thew lower bound, one can use results from Cárcamo et al. (2020) to show that the right inverse operator is Hadamard directionally differentiable, a notion that is different from (full) Hadamard differentiability. Then the asymptotic distribution for the lower bound can be derived by applying the functional delta method for Hadamard directionally differentiable functionals. Note, however, that **?** implies that the asymptotic distribution of the lower bound for conditional CIC is no longer a Gaussian process since we do not have full Hadamard differentiability. Because the asymptotic distribution is even more involved than the Gaussian process for the upper bound, bootstrap inference based on **?** can be implemented.

6 Effect of EITC on Birthweights

In this section, I apply the CIC-DR estimator developed in Section 3 to revisit the study of Hoynes et al. (2015). I re-analyze the effect of Earned Income Tax Credit ("EITC") on low birthweights ($\leq$ 2,500g or 5.5 lbs), which have been well recognized in the health economics literature to be a consequential indicator of short and long-term health and economic outcomes (Almond et al., 2005; Currie, 2011; Almond et al., 2011). Given the concentrated risk at the low end of the birth weight (Y) distribution, estimating heterogeneous effects of a welfare policy or a treatment across the outcome distribution, while controlling for potential confounders (X) in a flexible manner is of critical importance.

The importance of accounting for control variables for CIC can be highlighted by two issues in this empirical context. First, birthweight trends may vary across geographic region that invalidates the unconditional time-invariance condition as outlined in Athey and Imbens (2006) if control variables are not accounted for. Socio-economic and health characteristics correlated with race, age and education levels of mothers may also induce differential trends for birthweights over time. Second, U.S. state of residence may matter more for mothers who would be giving birth to low weight infants than healthier mothers, given that access to prenatal care and quality of care, more generally, may vary geographically. Accordingly, assuming that race, age, education and/or residence location matter equally across the potential outcome distribution may yield misleading results. CIC-DR in this regard can provide reliable treatment effects estimates and confidence bands compared to existing restrictive methods for informing future policy decisions.

6.1 Background & Data

EITC is a welfare policy targeting working low-income individuals in the form of additional tax refunds. The maximum allowable amount of these credits increases in the number of pre-existing children for the mother/family. Hoynes et al. (2015) study the impact of the largest EITC expansion (Omnibus Reconciliation Act of 1993; OBRA93) using a DID design to identify the impact of income gains (through EITC) on birth weights. They find significant effects of EITC on reducing incidences of low birth weight and on increasing average birth weights among the treated. Because OBRA93 generated large tax credit gains for those with two or more children

versus for those with one child, they define the control group as those with one pre-existing child and the treated group as those with two or more children.[9]

Microdata from U.S. Vital Statistics Natality Data between 1989-1999 is used for analysis. Data contains infant characteristics, such as birth weight, gender, birth order (parity) and birth date, and mother's demographic characteristics, such as age, race, education, state of residence and marital status. To be in line with the original analysis, the sample is limited to single-mothers with at least 18 years of age and at most 12 years of education, which yields 2-3 million observations per (g, t) pair.[10] The implementation in this section follows suit with the two identification conditions, "cash-in-hand" and "sensitive developmental stage" proposed by Hoynes et al. (2015). The first condition assumes that infant health is affected by the EITC through the cash payment in the form of tax refunds and that this cash is spent in the subsequent 12 months of receipt. The second condition assumes that the last three months of pregnancy before birth is important for infant's birth weight production. The paper also assumes full take-up of the treatment based on high take-up evidence from the IRS around the time frame of analysis and I maintain this assumption.[11]

Due to potential heterogeneity in treatment effects over time as well as the plausible pre-trends based on the previous OBRA expansions, I consider treatment effects for pairs of samples in 1993 compared to those in 1994, 1993 compared to those in 1995, etc. I also compare my results to the partialling out approach (via OLS) as proposed by Athey and Imbens (2006) and implemented by Sasaki and Wang (2024) for the same empirical study.[12] This approach significantly restricts the heterogeneity in the controls' effect on the outcome distribution. Note that Sasaki and Wang (2024) focus on the extreme quantiles of the birthweight distribution, and as such they impose such parametric restrictions on the effect of covariates on the outcome distribution. I extend their analysis to the intermediate quantiles to showcase how a researcher would implement CIC with

[9] They also consider the comparison between individuals with no child vs. those with one or more children, but due to pre-existing trends in the gap between no children and one child, I focus on the one child vs. two or more children case.

[10] Sample selection in such a manner focuses the analysis on "high-impact sample" as described in Hoynes et al. (2015).

[11] See Footnote 19 of Hoynes et al. (2015) for details.

[12] Sasaki and Wang (2024) considers a 2×2 setting by comparing observations between 1989–1993 to those in 1994–1999. The results for the pooled sample as well as the pairwise samples differ qualitatively, but the relative differences between CIC-DR and the partialling out approach persist throughout the two cases.

control variables given available tools in the literature.

The restriction of the control variables' effect on the outcome distribution can be seen by letting $\tilde{Y}_{gt} = Y - X'\beta_{gt}$ and observing that for a given (g,t) pair,

$$F_{\tilde{Y}_{gt}|X}(y) = P\left(\tilde{Y}_{gt} \leq y | X = x\right) = \Lambda\left(y + x'\beta_{gt}\right)$$

where β_{gt} is assumed to be homogeneous across $y \in \mathcal{Y}$ and only the intercept is allowed to vary. Contrasting this with the more general framework used under CIC-DR,

$$F_{Y_{gt}|X}(y|x) = \Lambda\left(x'\beta_{gt}(y)\right)$$

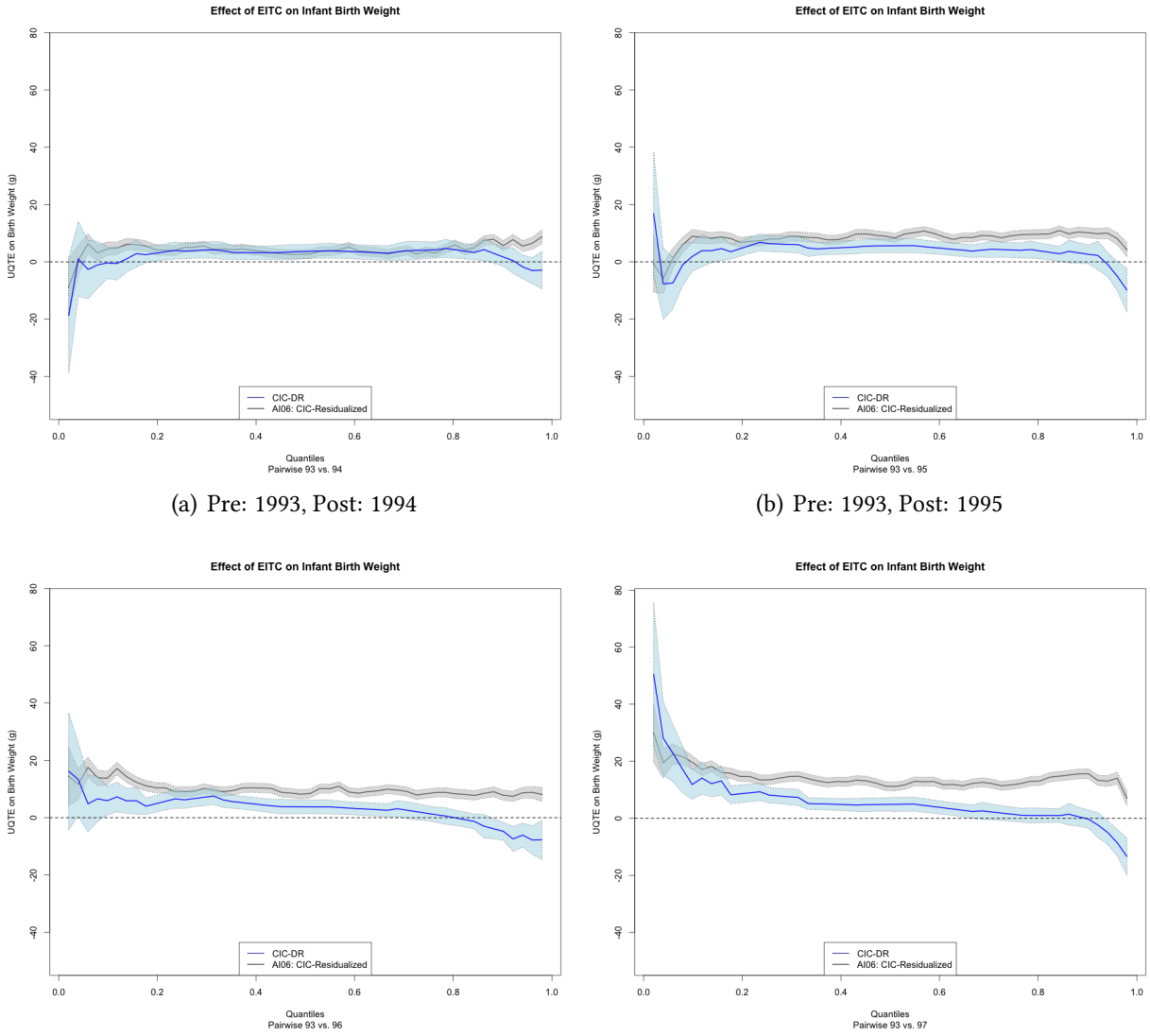
it is easy to see that the latter expression is much more flexible and includes the former specification as a special case. Indeed one can test $H_0 : \beta_{gt}(y) = 0 \;\; \forall y \in \mathcal{Y}_{gt}$ and I show an example in Appendix B Figure 6, illustrating that the data rejects the more restrictive effect of control variables on the outcome distribution. In particular, the null hypothesis for the coefficient function corresponding to mother's race being black is rejected at the 5% significance level for low birth weights, whereas the control variable's effect is not significant for higher birth weights.

6.2 Estimation Results

Estimation results comparing CIC-DR, residualized CIC and the original results from Hoynes et al. (2015) are presented in this section. Quantile and distributional treatment effects are presented first, followed by average treatment effects implied by CIC-DR for comparison with alternative methods.

Heterogeneous Treatment Effects Quantile treatment effects of EITC across the birth weight distribution (grid of 50 equidistant points in $[0.02, 0.98]$-th quantiles) are reported for pairwise samples ('93 vs. '94, '93 vs. '95, etc.). Figure 3 below shows UQTT estimates of EITC on birth weights based on CIC-DR and the OLS-residualized CIC based on Athey and Imbens (2006) and Sasaki and Wang (2024). The uniform confidence bands for CIC-DR (in blue) are based on 500 bootstrap replications using standard exponential weights and the confidence bands for residualized CIC (in red) are based on the expression for the asymptotic variance given by Athey and

Imbens (2006).



(a) Pre: 1993, Post: 1994

(b) Pre: 1993, Post: 1995

(c) Pre: 1993, Post: 1996

(d) Pre: 1993, Post: 1997

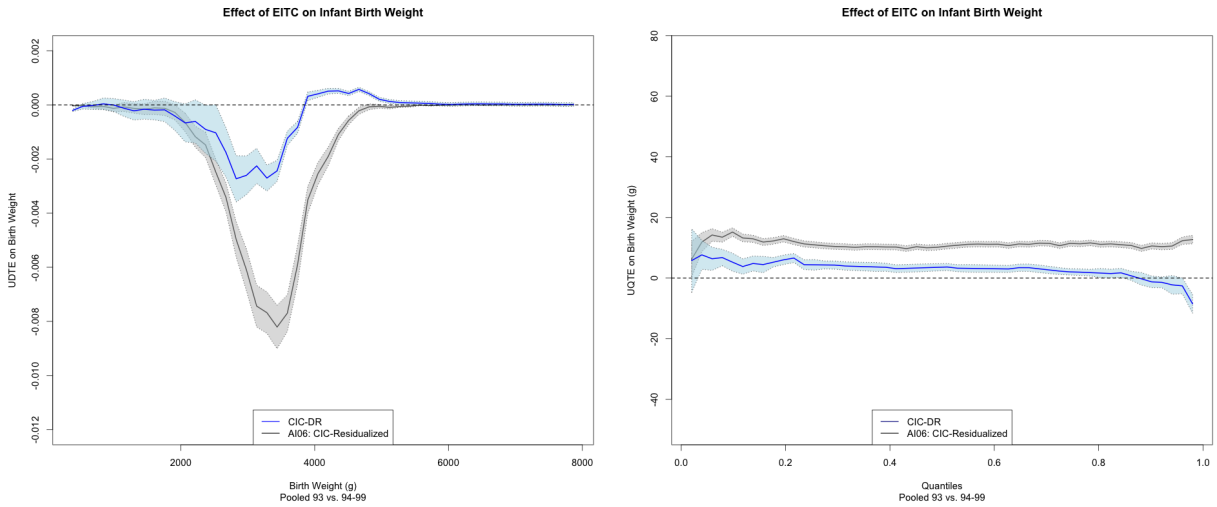
Figure 3: Pairwise UQTT of EITC on Birth Weight (g). Blue line plots estimates from CIC-DR and red line plots estimates from OLS-residualized CIC. Blue shaded area is the uniform 90% confidence band obtained from 500 weighted bootstrap replications and red shaded area is the confidence band obtained based on analytical expression of asymptotic variance from Athey and Imbens (2006).

As the years pass from 1994 to 1997, and thereby the amount of EITC allowable increases, the quantile treatment effects demonstrates an increasing pattern for the lower end of the infant weight distribution. Note that in Figure 8 panel (d) for 1997, which is when the maximum EITC allowable is the highest among the years considered, the effect of EITC on birth weight is estimated to be more positive (albeit noisy) in the bottom decile than previously reported. This

is an approximately 50 g increase in birth weight at the 0.02-th quantile for the 1997 sample. Also, although CIC-DR and the residualized CIC seems to be similar for 1994, as time passes, the gap between the two demonstrably increases leading to statistically and economically significant differences in the resulting estimates. Restricting heterogeneity of X's impact across the distribution of birth weights would, therefore, seem to overstate the impact of EITC on the intermediate and upper birth weight quantiles.

Interpretation of Heterogeneous Treatment Effects To contextualize these heterogeneous estimates in light of Hoynes et al. (2015), I also present distributional treatment effects on the treated. The aforementioned paper presents the treatment effects in terms of probability changes of an infant weighing less than 2,500 g (or 5.5 lbs). In comparison, CIC-DR delivers a complete picture across the entire distribution of birth weights with uniformly valid confidence bands as opposed to only pointwise estimates. Because the original paper considered pooled sample letting 1992-1993 effective tax years as the pre-treatment period and 1994-1999 as the post-treatment period, I also estimate UDTT using the same sample for comparison. Figure 4 shows the changes in probability of birth weight being below a certain threshold, over a grid of thresholds. Since the treatment effect is on the cumulative distribution, a negative UDTT can be interpreted as a positive health outcome implying that more children are born with higher weights compared to the control group. Similar to the pairwise results, the pooled results suggest that residualized CIC would overstate the impact of EITC on birth weights. CIC-DR estimates a UDTT of -0.10 percentage points at $y =$2,500 g. DID results for $y =$2,500 g from Hoynes et al. (2015) on Table 2 suggests -0.36 percentage points; residualized CIC suggest a UDTT of -0.22 percentage points.

Two points are worth mentioning in interpreting these results. First, the estimation results suggest that for the pooled sample, EITC had smaller-than-reported effects in reducing the incidence of low birth weight i.e. proportion of those who had babies weight under 2,500 g. In monetary terms, based on the overall after-tax income calculated in the original paper (Table 4 of Hoynes et al. (2015)), this translates to 0.06 percentage point reduction in incidences of low birth weights (in 2009 dollar terms) or a 2.75 g / $1,000 increase in birth weights at the quantile rank corresponding to 2,500 g threshold. Recall that the pairwise result for the 1997 sample would suggest a three times as large (6.42 g) gain at the same quantile compared to the pooled sample



(a) UDTT for Pooled Sample

(b) UQTT for Pooled Sample

Figure 4: Pooled UDTT and UQTT of EITC on Birth Weight (g). Pre: 1992-1993, Post: 1994-1999. Blue line plots estimates from CIC-DR and red line plots estimates from OLS-residualized CIC. Shaded areas are the uniform 90% confidence band obtained from 500 weighted bootstrap replications.

estimate. For the bottom decile, the benefits are much higher at approximately 32 g / $1,000.

Second, I argue that these results are nevertheless economically very meaningful improvements. To see this, comparing EITC's effect on low birth weight to that of smoking on low birth weight can be instructive. A conservative estimate by Currie et al. (2009) suggests that the effect of a mother being a smoker is estimated to reduce birth weight by 38.9 g, with each additional cigarette reducing birth weight further by 2.2 g. Although the estimated effects of EITC around the intermediate quantiles suggest 5-10 g of increased birthweight depending on the sample used, a near 50 g increase in birth weight at the bottom of the weight distribution in 1997 is notable. This is on par with the effect of being a smoker when translated into $1,000 terms. Furthermore, as Hoynes et al. (2015) argue, additional cash could induce a negative income effect of smoking behavior and through reduced stress and added employment effects of mothers, which can further help improve birth outcomes through multiple channels.

CIC-implied Average and Median Treatment Effects Next, I present CIC-DR implied average (ATT) and median treatment effects (MTT) estimates. Although the main spirit of the proposed method is to analyze heterogeneous beyond average effects, comparing those obtained

via CIC-DR with those found in Hoynes et al. (2015), which uses pointwise DDID[13] can be an instructive exercise. The CIC implied ATT can easily be obtained by using the expression from Section 2.2. The results presented also include the residualized CIC. Both the CIC-DR implied ATT and MTT are much smaller than those found by DDID or using residualized CIC. The ATT is no longer statistically significant at the 5% significance level. Estimates from DDID and residualized

	ATT	MTT
DDID	9.95	–
	(2.05)	–
CIC-DR	2.67	3.60
	(1.51)	(1.46)
CIC-Resid	10.49	10.35
	(1.03)	(0.94)
DDID-DR	2.46	3.38
	(1.47)	(1.52)

Table 1: Average and Median Treatment Effects on the Treated. Numbers in parentheses are bootstrapped standard errors based on 500 replications.

CIC are much closer to each other and are statistically significant at the 5% significance level. Although the DDID design from Hoynes et al. (2015) control for potential observed confounders, they do so by aggregating observations by cells induced by the combination of control variable values. Because they do this for a series of thresholds, this could be considered to be more flexible than the residualized CIC approach. The difference in the estimates between CIC-DR and DDID could be attributed to largely two factors: First, a difference in the identifying assumptions and the resulting estimate of the counterfactual distribution could lead to a divergence in the treatment effects estimates. This could be especially salient if the marginal outcome densities exhibit skewness or multi-modality and these patterns are different across group and time indicators. Upon visual inspection (demonstrated in Appendix B), the CDFs and the quantile functions seem similar in their shapes. The second factor could be attributed to distribution regression as a flexible modeling tool which could be capturing heterogeneity of control variables effect across the outcome distribution. If this is the case and the first posited factor is less relevant, then a DR-

[13]In the sense that they consider a DID design on the probability of being low birthweight but only for several thresholds, and thus not having simultaneously valid standard errors.

based DDID estimator could yield treatment effects estimates closer to those obtained using the CIC-DR estimator.

Indeed, the results on the table appear to be consistent with the aforementioned narrative. One can observe the DDID based on DR-estimates of the marginal outcome distributions (where $\hat{F}_{Y_{gt}|X}(y|x)$ yields similar ATT and median treatment effects as those implied by CIC-DR. Here, the potential outcome under no treatment in the treated group using DR-based DDID obtained by,

$$\hat{F}^{DDID}_{Y^N_{11}}(y) = \hat{F}^{DR}_{Y_{10}}(y) + \left(\hat{F}^{DR}_{Y_{01}}(y) - \hat{F}^{DR}_{Y_{00}}(y)\right)$$

where $\hat{F}^{DR}_{Y_{gt}}(y) = \int_{\mathcal{X}_{11}} \Lambda(x'\hat{\beta}_{gt}(y))d\hat{F}_{X_{11}}(x)$ is denoted to distinguish the DR-based estimates of the distribution function. DDID and CIC will not be similar in general, especially if the distributions across groups are quite different, and DDID may even yield invalid CDFs that violate the natural monotonicity condition when outcomes are discrete. In this empirical example, given that the marginal distribution functions are smooth and are quite similar in their shape, the second aforementioned factor appears to exhibit higher relevance. If the difference in the probability gaps versus the quantile gaps across group and time indicators are large, this could increase the relevance of the first factor in explaining the difference in the estimates.

Lastly, I report that I also find multiple violations of the natural monotonicity restriction of conditional distribution functions for various control variables using DDID in this empirical exercise. This implies that the data clearly rejects the identifying assumption of DDID. Although there are statistical tests available to test the violation of monotonicity as outlined in Roth and Sant'Anna (2023), this could give rise to issues with pre-testing. On the other hand, the counterfactual conditional distributions produced by CIC-DR do not violate monotonicity restrictions.

7 Conclusion

Heterogeneous treatment effects of a policy intervention are often of great interest, because policies are often motivated to improve the outcomes for those at the lower end of the outcome distribution or to reduce inter-group outcome gaps. For example, welfare policies often target those at the lower end of the income distribution or are intended to reduce labor market outcome disparities between sub-groups of interest. To reliably estimate the differential effects of a given

treatment across the outcome distribution, controlling for the effect of potential observed confounders is crucial. Doing so in a restrictive manner e.g. assuming that a given control variable's effect on the entire outcome distribution is the same, however, can yield misleading estimates.

Towards this effort, I propose a distribution regression-based changes-in-changes estimator and derive its asymptotic distributions along with bootstrap validity for inference. This paper also examines the heterogeneous effects of EITC on infant weights across the entire weight distribution and find more concentrated benefits at the low end of the distribution, while more muted effects across the rest of the distribution than previously reported. I also demonstrate that average effects may mask important heterogeneity and deem policies unhelpful for improving outcomes where they matter. For example, even though EITC served to help reduce low birth weight incidences and increase the birth weights at the bottom of the weight distribution which is the most consequential part of the distribution for downstream health outcomes, estimates of average treatment effects on the treated suggest that the policy effect was statistically indistinguishable from null.

The current paper focuses on repeated cross sectional data in the canonical 2×2 setting with low-dimensional controls. It would also be insightful to develop theory for when panel data is available, when treatments are staggered in their adoption (along with a means to aggregate the quantile or distributional treatment effects) and when high-dimensional controls are present. The last direction of research can be useful when there are a large number of potential confounders, but the researcher does not know which ones are the most relevant in determining conditional outcome trends over time.

References

Almond, D., Chay, K. Y., and Lee, D. S. (2005). The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1083.

Almond, D., Hoynes, H. W., and Schanzenbach, D. W. (2011). Inside the war on poverty: The impact of food stamps on birth outcomes. *The review of economics and statistics*, 93(2):387–403.

Arkhangelsky, D. et al. (2019). *Dealing with a technological bias: The difference-in-difference approach*. Centro de Estudios Monetarios y Financieros.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.

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

Blau, F. D., Cohen, I., Comey, M. L., Kahn, L., and Boboshko, N. (2023). The minimum wage and inequality between groups. (31725).

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

Bonhomme, S. and Sauder, U. (2011). Recovering distributions in difference-in-differences models: A comparison of selective and comprehensive schooling. *Review of Economics and Statistics*, 93(2):479–494.

Callaway, B. (2021). Bounds on distributional treatment effect parameters using panel data with an application on job displacement. *Journal of Econometrics*, 222(2):861–881.

Callaway, B. and Li, T. (2019). Quantile treatment effects in difference in differences models with panel data. *Quantitative Economics*, 10(4):1579–1618.

Cárcamo, J., Cuevas, A., and Rodríguez, L.-A. (2020). Directional differentiability for supremum-type functionals: Statistical applications. *Bernoulli*, 26(3):2143 – 2175.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

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

Chernozhukov, V., Fernandez-Val, I., Melly, B., and Wüthrich, K. (2019). Generic inference on quantile and quantile effect functions for discrete outcomes. *Journal of the American Statistical Association*.

Currie, J. (2011). Inequality at birth: Some causes and consequences. *American Economic Review*, 101(3):1–22.

Currie, J., Neidell, M., and Schmieder, J. F. (2009). Air pollution and infant health: Lessons from new jersey. *Journal of Health Economics*, 28(3):688–703.

Derenoncourt, E. and Montialoux, C. (2021). Minimum Wages and Racial Inequality*. *The Quarterly Journal of Economics*, 136(1):169–228.

Foresi, S. and Peracchi, F. (1995). The conditional distribution of excess returns: An empirical analysis. *Journal of the American Statistical Association*, 90(430):451–466.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. In *Handbook of labor economics*, volume 4, pages 1–102. Elsevier.

Fortin, N. M., Lemieux, T., and Lloyd, N. (2021). Labor market institutions and the distribution of wages: The role of spillover effects. *Journal of Labor Economics*, 39(S2):S369–S412.

Ghanem, D., Kédagni, D., and Mourifié, I. (2023). Evaluating the impact of regulatory policies on social welfare in difference-in-difference settings. *arXiv preprint arXiv:2306.04494*.

Gunsilius, F. F. (2023). Distributional synthetic controls. *Econometrica*, 91(3):1105–1117.

Havnes, T. and Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of public economics*, 127:100–114.

Hoynes, H., Miller, D., and Simon, D. (2015). Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy*, 7(1):172–211.

Kim, D. and Wooldridge, J. M. (2023). Difference-in-differences estimator of quantile treatment effect on the treated. Working Paper. Available at: https://drive.google.com/file/d/1zd-vpKBdC4JddZX2OUacyaijje78eMWv/view (Accessed: May 23rd, 2024).

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends® in Econometrics*, 4(3):165–224.

Melly, B. and Santangelo, G. (2015). The changes-in-changes model with covariates. *Universität Bern, Bern*, 1.

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

Roth, J. and Sant'Anna, P. H. (2023). When is parallel trends sensitive to functional form? *Econometrica*, 91(2):737–747.

Sasaki, Y. and Wang, Y. (2024). Extreme changes in changes. *Journal of Business & Economic Statistics*, 42(2):812–824.

Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and Empirical Processes*. Springer.

Wursten, J. and Reich, M. (2023). Racial inequality in frictional labor markets: Evidence from minimum wages. *Labour Economics*, 82:102344.

Wüthrich, K. (2019). A closed-form estimator for quantile treatment effects with endogeneity. *Journal of Econometrics*, 210(2):219–235.

A Appendix: Proofs

A.1 Notation and Statistical Definitions

To present the proofs for the theorems and corollaries in the main text, I formally introduce notation to describe the mathematical set up and the relevant function spaces. Let $\ell^\infty(\mathcal{YX})$ denote the set of all bounded and measurable mappings $\mathcal{YX} \mapsto \mathbb{R}$. $\mathcal{YX}$ is a subset of $\bar{\mathbb{R}}^{1+d_x}$ with topology induced by the standard metric, ρ, on $\bar{\mathbb{R}}^{1+d_x}$, where $\bar{\mathbb{R}}$ is the extended real line. Also let $\lambda(f, \tilde{f}) = \left[\int \left(f - \tilde{f}\right)^2 dF_X\right]^{1/2}$ be a metric on $\mathcal{F}$. Here, $\mathcal{F}$ is defined as a class of measurable functions that includes $\{F_{Y_{gt}|X}(y|\cdot) : y \in \mathcal{Y}_{gt}, (g,t) \in \{0,1\}^2\}$ as well as the indicators of all rectangles in $\bar{\mathbb{R}}^{d_x}$ such that $\mathcal{F}$ is totally bounded under λ. The preceding definition ensures that the uniform CLT holds for the estimators of the conditional distribution function and control variable distributions.

A.2 Proof of Theorem 1

I utilize the derivative map obtained in Appendix SE.3 of Callaway (2021) to apply the functional delta method. Specifically, let $\bar{\mathcal{Y}}_{gt}\mathcal{X}$ be the support of Y_{gt} conditional on X, $\mathbb{D} := \ell^\infty(\mathcal{Y}_{10}\mathcal{X}) \times \ell^\infty(\mathcal{Y}_{00}\mathcal{X}) \times \ell^\infty(\mathcal{Y}_{01}\mathcal{X})$ and consider the map, $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \ell^\infty(\bar{\mathcal{Y}}_{01}\mathcal{X})$ defined by

$$\phi(F_1, F_2, F_3) = F_1 \circ F_2^{-1} \circ F_3$$

for $F = (F_1, F_2, F_3) \in \mathbb{D}_\phi$. Here, $\mathbb{D}_\phi = \mathcal{E}^3$ where $\mathcal{E}$ denotes the set of all distribution functions with density function that is uniformly bounded above and bounded away from zero. Then, ϕ is Hadamard differentiable at $F_0 = (F_{Y_{10}|X}, F_{Y_{00}|X}, F_{Y_{01}|X}) \in \mathbb{D}$ with the derivative map given by,

$$\phi'_{F_0}(\gamma) = \gamma_1 \circ F^{-1}_{Y_{00}|X} \circ F_{Y_{01}|X} + f_{Y_{10}|X}\left(F^{-1}_{Y_{00}|X} \circ F_{Y_{01}|X}\right) \frac{\gamma_3 - \gamma_2 \circ F^{-1}_{Y_{00}|X} \circ F_{Y_{01}|X}}{f_{Y_{00}|X}\left(F^{-1}_{Y_{00}|X} \circ F_{Y_{01}|X}\right)}$$

tangentially to $\mathbb{D}_\phi$ in $\gamma = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{D}_\phi$. This can be derived from the functional chain rule and from the result that composition of Hadamard differentiable maps are themselves Hadamard

differentiable (Lemmas 3.10.24 and 3.10.28 of Van Der Vaart and Wellner (1996)). Then taking the derivative map and applying the functional delta method as in Theorem 3.10.4 of Van Der Vaart and Wellner (1996) based on the limit laws of the input conditional distribution processes (i.e. $\mathbb{Z}_{10}(y,x), \mathbb{Z}_{00}(y,x), \mathbb{Z}_{01}(y,x)$) yields the limit law for the conditional CIC-DR process,

$$\begin{aligned}\mathbb{Z}_{11}^{N}(y,x) =& \mathbb{Z}_{10}\left(F_{Y_{00}|X}^{-1}\left(F_{Y_{01}|X}(y|x)|x\right),x\right) \\ &+ f_{Y_{10}|X}\left(F_{Y_{00}|X}^{-1}\left(F_{Y_{01}|X}(y|x)|x\right)|x\right)\mathbb{Z}_{00,01}(y,x)\end{aligned}$$

where,

$$\mathbb{Z}_{00,01}(y,x) = \frac{\mathbb{Z}_{01}(y,x) - \mathbb{Z}_{00}\left(F_{Y_{00}|X}^{-1}\left(F_{Y_{01}|X}(y|x)|x\right),x\right)}{f_{Y_{10}|X}\left(F_{Y_{00}|X}^{-1}\left(F_{Y_{01}|X}(y|x)|x\right)|x\right)}$$

The result for the asymptotic distribution of $\hat{F}_{Y_{11}^I|X}$ follows directly from Corollary 5.4 of Chernozhukov et al. (2013) as shown in the expression following Assumption 3. This completes the proof of Theorem 1. ■

A.3 Proof of Corollary 1.1

To show that the DR-based UQTT process is asymptotically Gaussian, note that the UQTT process is the difference in the conditional quantile process integrated over the control variable distribution between the treated and the untreated potential outcomes.

$$\hat{\delta}_{UQTT}(\tau) = \left(\int_{\mathcal{X}_{11}} \hat{F}_{Y_{11}|X}^{I}(y|x)d\hat{F}_{X_{11}}(x)\right)^{-1} - \left(\int_{\mathcal{X}_{11}} \hat{F}_{Y_{11}|X}^{N}(y|x)d\hat{F}_{X_{11}}(x)\right)^{-1}$$

This is a smooth functional of the respective distribution functions, and therefore the functional delta method can be applied to the following map based on Theorem 1. To be concrete, let $\mathbb{D}_{\varphi}$ be the product space of measurable functions, $\Gamma : \mathcal{Y}_{gt}\mathcal{X} \mapsto [0,1]$ defined by $(y,x) \mapsto \Gamma(y,x)$ and $\Pi : \mathcal{F} \mapsto \mathbb{R}$ defined by $f \mapsto \int f d\Pi$, where Π is bounded and restricted to be a probability

measure on $\mathcal{X}$. Consider the map $\varphi : \mathbb{D}_\varphi \subset \mathbb{D} = \ell^\infty(\mathcal{Y}_{gt}\mathcal{X})^2 \times \ell^\infty(\mathcal{F}) \mapsto \mathcal{E} = \ell^\infty(\mathcal{T})$ defined by

$$\varphi(\Gamma_1, \Gamma_2, \Pi) := \left(\int \Gamma_1(\cdot, x) d\Pi(x) \right)^{-1} - \left(\int \Gamma_2(\cdot, x) d\Pi(x) \right)^{-1}.$$

Then, φ can be defined as a composition of two maps, $\varphi_1 : \mathbb{D}_\varphi \mapsto \mathbb{D}_{\varphi_2}$ and $\varphi_2 : \mathbb{D}_{\varphi_2} \mapsto \ell^\infty(\mathcal{T})$ where $\mathbb{D}_{\varphi_2} = \ell^\infty(\mathcal{Y}_{gt})$ defined by,

$$\varphi(\Gamma_1, \Gamma_2, \Pi) = \varphi_2(\varphi_1(\Gamma_1, \Pi), \varphi_1(\Gamma_2, \Pi))$$

where

$$\varphi_1(\Gamma, \Pi) := \int \Gamma(\cdot, x) d\Pi(x) \text{ and } \varphi_2(\Psi_1, \Psi_2) := \Psi_1^{-1} - \Psi_2^{-1}$$

for $\Gamma \in \mathbb{D}_{varphi}$ and By Lemma D.1. of Chernozhukov et al. (2013), ϕ_1 is Hadamard differentiable at $(F_{Y_{gt}|X}, F_X)$ tangentially to the subset, $\mathbb{D}_0 = UC(\mathcal{Y}_{gt}\mathcal{X}, \rho) \times UC(\mathcal{F}, \lambda)$ with the derivative map,

$$\phi'_{1, F_{Y_{gt}|X}, F_X}(\gamma, \pi)(y) = \int \gamma(y, x) dF_X(x) + \pi(F_{Y_{gt}|X}(y|\cdot)) \tag{5}$$

ϕ_2 is also a difference of inverse maps, which are Hadamard differentiable by Lemma 3.10.21 of Van Der Vaart and Wellner (1996) with the derivative map given by,

$$\phi'_{2, \Psi_1, \Psi_2}(k_1, k_2) = -\frac{k_1}{\psi_1} \circ \Psi_1^{-1} + \frac{k_2}{\psi_2} \circ \Psi_2^{-1}$$

Then, by Lemmas 3.10.24 and 3.10.28 of Van Der Vaart and Wellner (1996), ϕ is also Hadamard differentiable at $(F_{Y_{11^I}|X}, F_{Y_{11^N}|X}, F_X)$ with the derivative map given by,

$$
\begin{aligned}
&\phi'_{F_{Y_{11}^I|X}, F_{Y_{11}^N|X}, F_X}(\gamma_1, \gamma_2, \pi) \\
&= \phi'_{2,\phi_1(F_{Y_{11}^I|X}, F_X), \phi_1(F_{Y_{11}^N|X}, F_X)} \left(\int \gamma_1(y,x) dF_X(x) + \pi(F_{Y_{11}^I|X}(y|\cdot)), \int \gamma_2(y,x) dF_X(x) + \pi(F_{Y_{11}^N|X}(y|\cdot)) \right) \\
&= -\frac{\int \gamma_1(y,x) dF_X(x) + \pi(F_{Y_{11}^I|X}(y|\cdot))}{f_{Y_{11}^I}} \circ F^{-1}_{Y_{11}^I} + \frac{\int \gamma_2(y,x) dF_X(x) + \pi(F_{Y_{11}^N|X}(y|\cdot))}{f_{Y_{11}^N}} \circ F^{-1}_{Y_{11}^N}
\end{aligned}
$$

Next, defining the empirical processes for the conditional distribution process and the control variable distribution as,

$$
\begin{aligned}
\hat{Z}_{gt}(y,x) &:= \sqrt{N} \left(\hat{F}_{Y_{gt}|X}(y|x) - F_{Y_{gt}|X}(y|x) \right) \\
\hat{G}_{gt}(f) &:= \sqrt{N_{gt}} \int_{\mathcal{X}_{gt}} f d \left(\hat{F}_X - F_X \right)
\end{aligned}
$$

where $\hat{F}_X$ is the empirical distribution of X. Under Assumption 2 and 3, and as $N \to \infty$ with $N/N_{gt} \to s_{gt} \in [0, \infty)$, along with the definition of $\mathcal{F}$ in Appendix A.1, the following holds

$$
\left(\hat{Z}^I_{11}(y,x), \hat{Z}^N_{11}(y,x), \hat{G}_{11}(f) \right) \rightsquigarrow \left(\mathbb{Z}^I_{11}(y,x), \mathbb{Z}^N_{11}(y,x), \mathbb{G}_{11}(f) \right)
$$

as stochastic processes indexed by $(y, x, g, t, f) \in \mathcal{Y}\mathcal{X}\{0,1\}^2\mathcal{F}$. The limit process is a tight mean-zero Gaussian process, where $\mathbb{Z}_{gt}$ and $\mathbb{G}_{gt}$ almost surely have uniformly continuous paths with respect to the metric ρ and λ respectively.

Finally, apply the functional delta method to the composition of Hadamard differentiable maps and obtain the following limit laws for the UQTT.

$$
\sqrt{N} \left(\hat{\delta}_{UQTT} - \delta_{UQTT} \right) \rightsquigarrow \mathbb{Z}_{UQTT}(\tau)
$$

as a stochastic process indexed by $\tau \in \mathcal{T}$, and $\mathbb{Z}_{UQTT}(\tau)$ is a tight mean-zero Gaussian process

defined by,

$$\mathbb{Z}_{UQTT}(\tau) = -\frac{\mathbb{Z}^I_{11,U}\left(F^{-1}_{Y^I_{11}}(\tau)\right)}{f_{Y^I_{11}}\left(F^{-1}_{Y^I_{11}}(\tau)\right)} + \frac{\mathbb{Z}^N_{11,U}\left(F^{-1}_{Y^N_{11}}(\tau)\right)}{f_{Y^N_{11}}\left(F^{-1}_{Y^N_{11}}(\tau)\right)}$$

where

$$\mathbb{Z}^I_{11,U}(y) := \int_{\mathcal{X}_{11}} \mathbb{Z}^I_{11}(y,x)dF_X(x) + \mathbb{G}\left(F_{Y^I_{11}|X}(y|\cdot)\right)$$
$$\mathbb{Z}^N_{11,U}(y) := \int_{\mathcal{X}_{11}} \mathbb{Z}^N_{11}(y,x)dF_X(x) + \mathbb{G}\left(F_{Y^N_{11}|X}(y|\cdot)\right)$$

Note that for UDTT, the one need only consider the first Hadamard differentiable map, ϕ_1, and apply the functional delta method directly to obtain the limit law,

$$\mathbb{Z}_{UDTT}(y) = \mathbb{Z}^I_{11,U}(y) - \mathbb{Z}^N_{11,U}(y)$$

This completes the proof of Corollary 1.1. ■

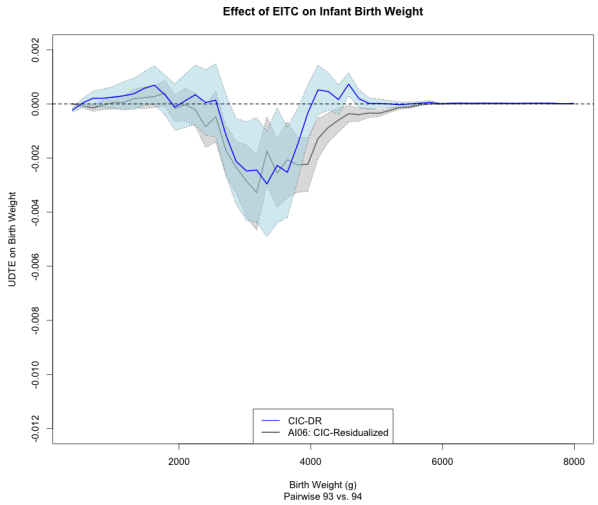
A.4 Proof of Theorem 2

This follows directly from the functional delta method. For quantities involving the integration over the control variable distribution, the same map found in expression (5) can be used. ■

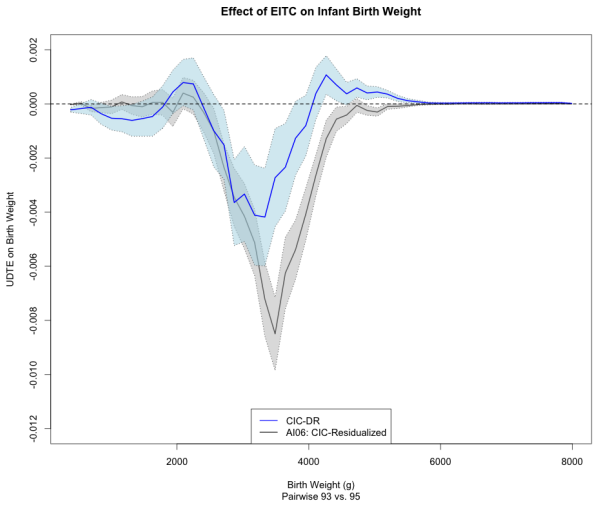
A.5 Proof of Theorem 3

Under the stated assumptions and results found in (2), by Corollary 5.4 of Chernozhukov et al. (2013), the exchangeable bootstrap is valid for the the DR-based estimators for the conditional distribution and quantile functions. Combining this with Theorem 1 and the Hadamard differentiability of the functionals involved, one can apply the functional delta method for the bootstrap based on section 3.10.3 of Van Der Vaart and Wellner (1996) and the result holds. ■

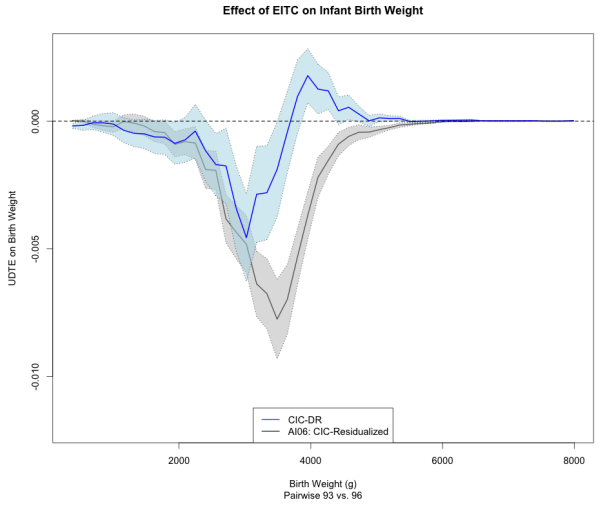
B Appendix: Figures



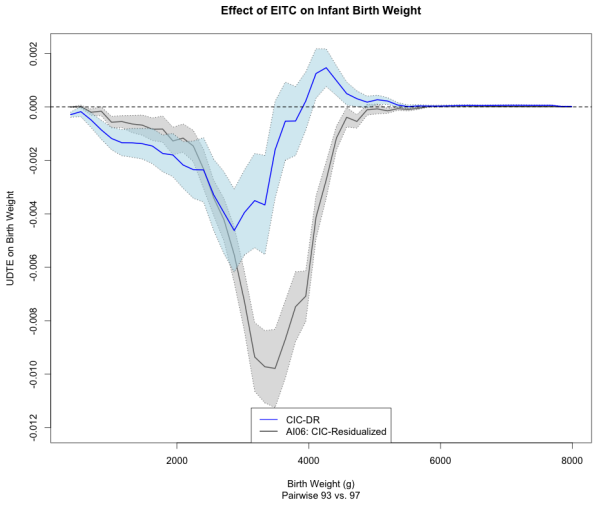
(a) Pre: 1993, Post: 1994



(b) Pre: 1993, Post: 1995

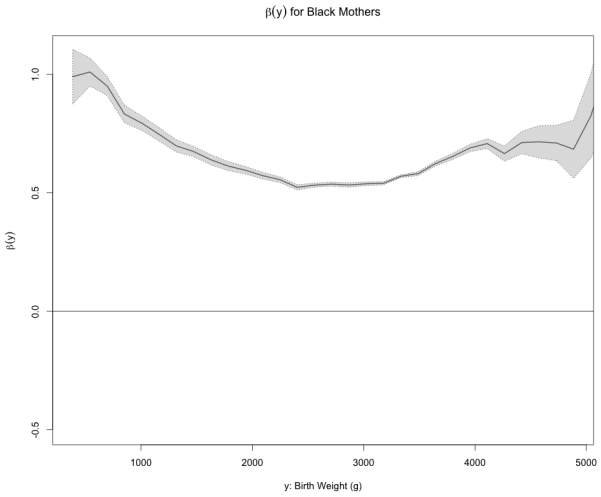


(c) Pre: 1993, Post: 1996

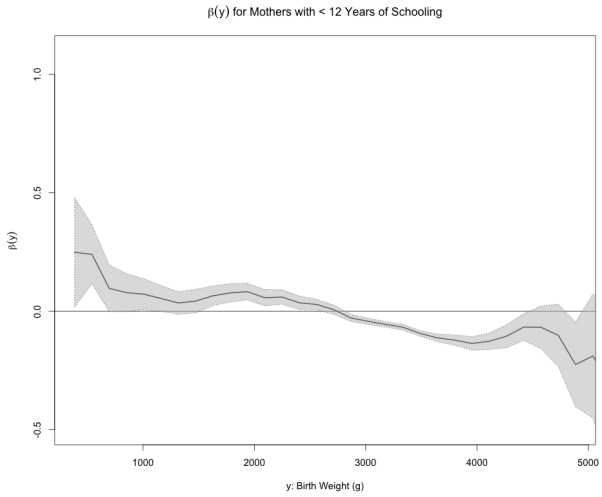


(d) Pre: 1993, Post: 1997

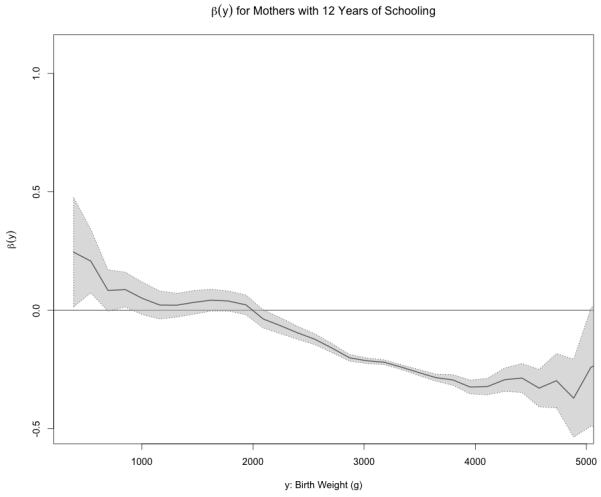
Figure 5: Pairwise UDTT of EITC on Birth Weight (g). Blue line plots estimates from CIC-DR and red line plots estimates from OLS-residualized CIC. Shaded areas are the uniform 90% confidence band obtained from 500 weighted bootstrap replications.



(a) Pre: 1993, Post: 1997

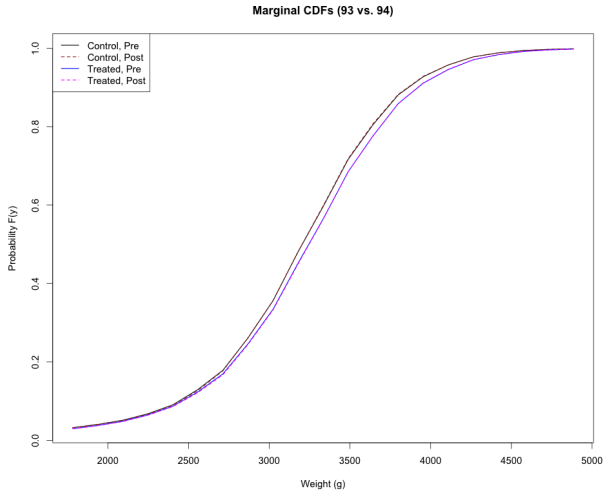


(b) Pre: 1993, Post: 1997

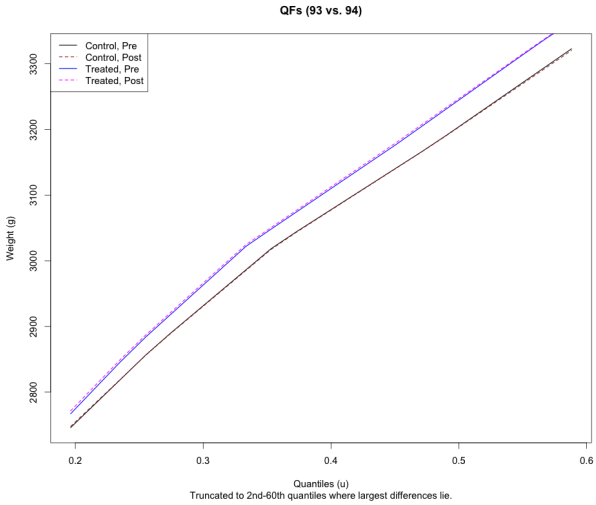


(c) Pre: 1993, Post: 1997

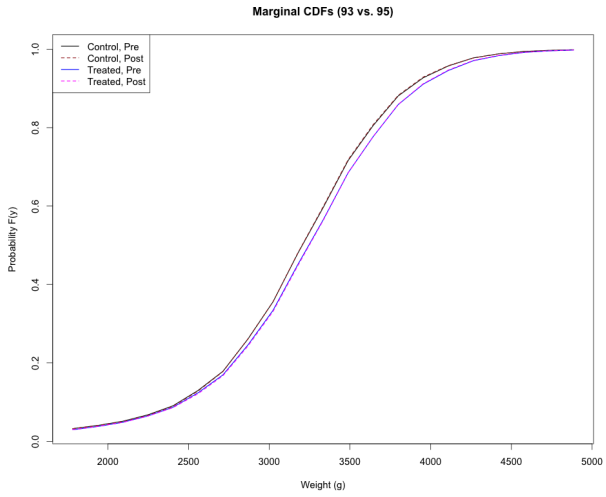
Figure 6: $\beta_{01}(y)$ for select mothers' characteristics.



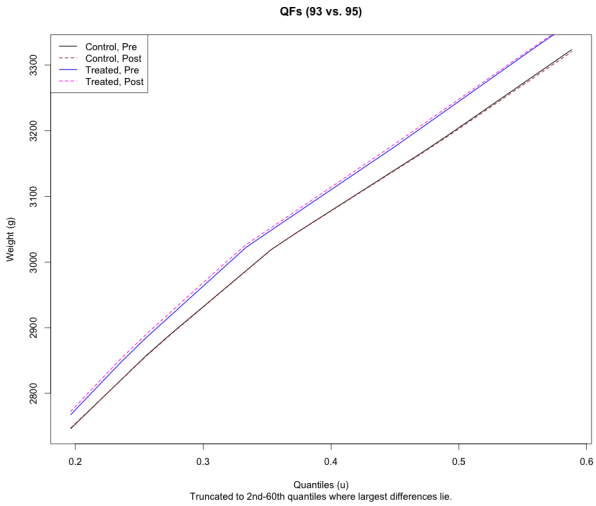
(a) Pre: 1993, Post: 1994



(b) Pre: 1993, Post: 1994

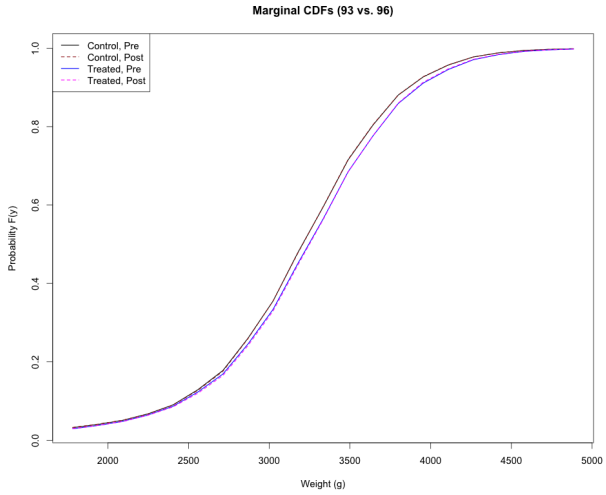


(c) Pre: 1993, Post: 1995

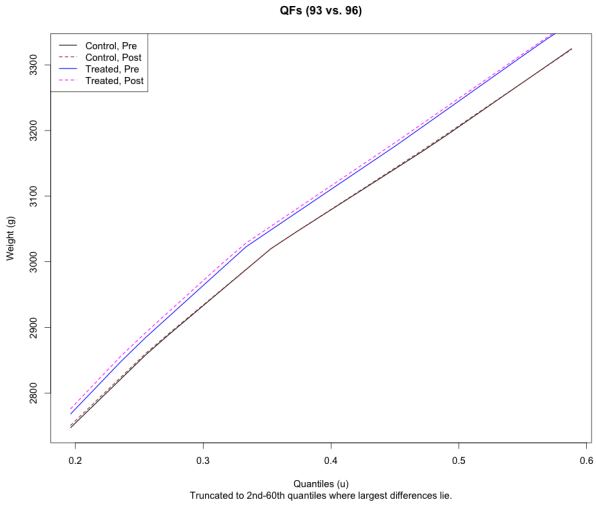


(d) Pre: 1993, Post: 1995

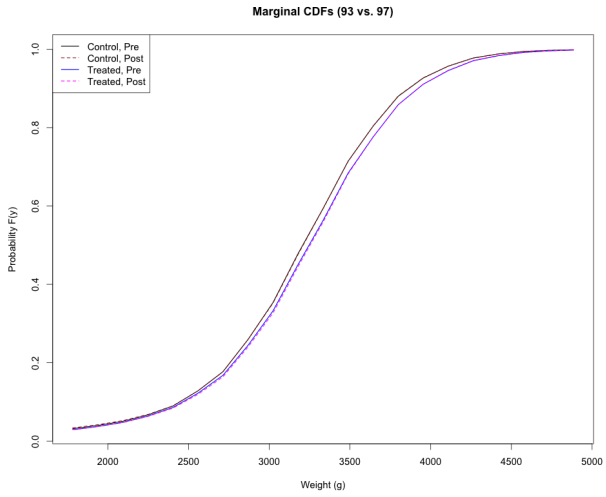
Figure 7: Marginal distribution and quantile functions of the pairwise samples.



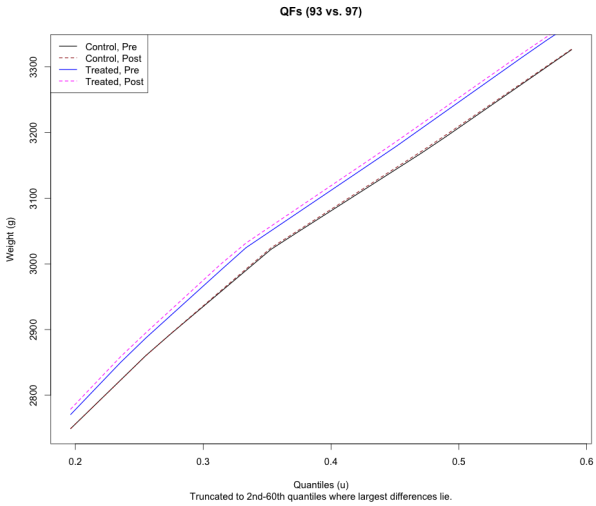
(a) Pre: 1993, Post: 1996



(b) Pre: 1993, Post: 1996



(c) Pre: 1993, Post: 1997



(d) Pre: 1993, Post: 1997

Figure 8: Marginal distribution and quantile functions of the pairwise samples.