Triple Changes Estimator for Targeted Policies

Sina Akbari [1]   **Negar Kiyavash** [1]

Abstract

The renowned difference-in-differences (DiD) estimator relies on the assumption of 'parallel trends,' which may not hold in many practical applications. To address this issue, economists are increasingly considering the triple difference estimator as a more credible alternative. Both DiD and triple difference are limited to assessing average effects exclusively. An alternative avenue is offered by the changes-in-changes (CiC) estimator, which provides an estimate of the entire counterfactual distribution by relying on assumptions imposed on the distribution of potential outcomes. In this work, we extend the triple difference estimator to accommodate the CiC framework, presenting the 'triple changes estimator' and its identification assumptions, thereby expanding the scope of the CiC paradigm. Subsequently, we empirically evaluate the proposed framework and apply it to a study examining the impact of Medicaid expansion on children's preventive care.

1. Introduction

In the domains of econometrics and quantitative social sciences, the *difference-in-differences* (DiD) estimator and its extensions have emerged as indispensable tools for estimating causal effects in observational studies. With its roots traced back to the work of Snow (1855) and further popularized by seminal works such as Ashenfelter and Card (1984) and Card and Krueger (1993), DiD has provided a reliable method for evaluating the effect of policy interventions over time. In particular, DiD stands out in studies where inference based on controlling for confounders or using instrumental variables is deemed unsuitable, and where pre-treatment information is available. Researchers have actively pursued various extensions of the DiD estimator, aimed at enhancing its applicability and robustness (Athey

[1]EPFL, Switzerland. Correspondence to: Sina Akbari <sina.akbari@epfl.ch>.
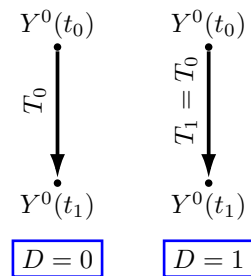
Figure 1. No drift assumption of CiC framework. Mappings T_0 and T_1 are assumed to be identical.

and Imbens, 2006; Sofer et al., 2016; Callaway et al., 2021; Roth and Sant'Anna, 2023).

The DiD estimator hinges on the assumption of parallel trends, which states that the (average) outcome of both the control and treatment groups share exactly the same evolution trend in the absence of the treatment. Mathematically speaking, this assumption translates to

$$\mathbb{E}[Y^0(t_1) - Y^0(t_0) \mid D = 1] = \mathbb{E}[Y^0(t_1) - Y^0(t_0) \mid D = 0], \quad (1)$$

where $Y^0(t)$ denotes the potential outcome associated with the absence of the treatment at time t, and D denotes the assigned treatment. In particular, $D = 1$ and $D = 0$ represent the treatment and control groups, respectively. The validity of this assumption can be challenged, and parallel trends might get violated due to unobserved time-varying factors or dynamic changes in the study context. This violation may result in biased estimates. As a response to potential deviations from parallel trends, researchers have turned to more flexible models, such as the *triple difference* framework (Gruber, 1994). The triple difference estimator can be formulated as the difference between two DiD estimators. Intuitively, the difference of two DiD estimators is unbiased, provided both estimators have the same bias. Indeed, sole purpose of of subtracting the second DiD estimator is to debias the first one. Despite the prevalent use of the triple difference estimator, especially over the last two decades (Raifman et al., 2018; Sakurai and Theodoro, 2020; Han et al., 2016; Chen et al., 2020; Tai-Seale et al., 2001), only recently a formal presentation of the framework and its identification assumptions was provided (Olden and Møen, 2022).

The triple difference framework guarantees only the identification of the *average treatment effect* (on the treated). Another issue that arises is that the triple difference estimator is biased if the outcomes in the pre- and post-treatment or among control and treatment groups are measured on a different scale. The *changes-in-changes* (CiC) framework (Athey and Imbens, 2006), on the other hand, is scale-invariant and yields the identification of the counterfactual outcome probability distribution. To do so, CiC framework requires imposing assumptions that differ from Eq. (1). It postulates that there exists a unique monotone mapping T_d which maps the probability measure over $Y^0(t_0)$ to that over $Y^0(t_1)$ within group $D = d$ for $d \in \{0, 1\}$ (see Figure 1.) Moreover, these two mappings are assumed to be identical. That is, $T_1(y) = T_0(y)$ for every y, or since the mappings are bijective,

$$T_1 \circ T_0^{-1} = \mathrm{Id}, \tag{2}$$

where Id represents the identity map. Eq. (2) states that there is *no drift* in the evolution trend across groups. Note that Eq. (2) was not explicitly stated in Athey and Imbens (2006) as an assumption but can be derived as a consequence of the following four assumptions:

Assumption 1.1 (Model assumption)**.** The potential outcomes $Y^0(t)$ can be modelled using a production function h of a latent variable U, which models the individual characteristics:

$$\forall t \in \{0, 1\}: \quad Y^0(t) = h(U; t).$$

In particular, 1.1 posits that h does not depend on the group assignment (D).

Assumption 1.2 (Strict monotonicity)**.** The function $h(\cdot; t)$ is strictly increasing in U for every $t \in \{t_0, t_1\}$.

Assumption 1.3 (Time invariance)**.** Within every subgroup, the distribution of the latent variable U does not change over time.

Assumption 1.4 (Latent support overlap)**.** The support of the latent variable U in the treatment group is a subset of its support in the control group.

In the case of a one-dimensional outcome, mappings T_d can be expressed as

$$T_d = F_{Y^0(t_1) \mid D=d}^{-1} \circ F_{Y^0(t_0) \mid D=d} \tag{3}$$

where $F_{Y^0(t) \mid D=d}(\cdot)$ represents the cumulative density function of $Y^0(t)$ in group $D = d$. Under regularity conditions, T_d is the unique monotone map that pushes forward the probability measure over $Y^0(t_0)$ to that over $Y^0(t_1)$ in group $D = d$ (Villani et al., 2009; Santambrogio, 2015). Combining Equations (2) and (3), it is straightforward to

identify the counterfactual distribution of $Y^0(t_1)$ in the treatment group, $F_{Y^0(t_1) \mid D=1}(y)$. Specifically,

$$\begin{aligned} F_{Y^0(t_1) \mid D=1}(y) = \\ F_{Y^0(t_0) \mid D=1} \circ F_{Y^0(t_0) \mid D=0}^{-1} \circ F_{Y^0(t_1) \mid D=0}(y), \end{aligned}$$

which matches Eq. (9) in the original work of Athey and Imbens (2006).

The no-drift assumption specified in Eq. (2) can be challenged in practice, especially in scenarios where the treatment or exposure is directed toward a specific subpopulation. This situation arises in studies on the impact of targeted interventions, such as a criminal justice initiative (D), on recidivism rates, where this intervention is exclusively administered to individuals with specific criminal histories or risk profiles. Naturally, it is expected that the time evolution of counterfactual recidivism rates will exhibit significant disparities between the control and treated groups. Similar challenges arise when the eligibility criteria is narrow within the context of social programs such as welfare or housing assistance, which target specific demographic groups.

Recognizing the strengths and weaknesses of both the triple difference and CiC frameworks, we propose a novel estimator that combines the best of both worlds. Our proposed 'triple changes' estimator aims to overcome the limitations of DiD and the triple difference estimator by leveraging the scale invariance and strong identification results of CiC, while introducing a more flexible mapping assumption that allows us to relax the constraint of (2). The triple changes framework is closely linked to data fusion and causal identification from multiple sources (Bareinboim and Pearl, 2016; Akbari et al., 2022; Abouei et al., 2024), as identification is achieved by combining datasets from separate sources, while this identification would not be achievable using one of the datasets alone. We briefly present the contributions of this work.

- We formally present and analyze the triple changes estimator as an extension to CiC, and provide the necessary assumptions for its point identification first in the scalar case. We then discuss how to generalize our results to high-dimensional outcomes by harnessing theory of optimal transport.

- We provide several partial identification results under relaxed versions of our proposed point identifiability assumptions. These results showcase the robustness of our identification results to slight violations of the assumptions. Further, we show the validity of analogous results for the classic CiC framework as a special case of our derivations.

- We introduce a finite-sample estimator for the average

treatment effect on the treated within our framework and analyze its asymptotic behaviour.

- We conduct an empirical evaluation of our estimator on both synthetic and real datasets.

This paper is organized as follows. Section 1.1 reviews the necessary background and the setup of the study. In Section 2, the identification of our estimand of interest is studied for a scalar outcome. Section 3 provides an estimator for the latter and studies its asymptotic properties. In Section 4, we extend our work to high-dimensional outcomes using the theory of optimal transport. Numerical evaluations are presented in Section 5.

1.1. Causal model

We consider a study where we have access to data from two sources, e.g., two states of the united states, or two cities, or any two separate populations. These two sources of data will be denoted by $S = s_0$ and $S = s_1$ throughout. We assume that a treatment (e.g., a health-care policy) is administered in one state, without loss of generality in $S = s_1$, and not in the other. In both states, the individuals are partitioned into two cohorts, namely, $D = d_1$ and $D = d_0$, signifying the individuals that are eligible and not eligible for receiving the treatment, respectively[1]. The outcome is measured in two time points, namely $t_0 < t_1$, where the eligible individuals in state s_1 receive the treatment at an infinitesimal amount of time after t_0[2]. We denote by $Y^{D=d_0}(t)$ and $Y^{D=d_1}(t)$ the potential outcome variables associated with the outcome at time t in the absence, and in the presence of treatment, respectively. To improve readability, we will often use the short-hands $Y^0(t)$ and $Y^1(t)$ for $Y^{D=d_0}(t)$ and $Y^{D=d_1}(t)$, respectively. We denote the observed outcome at time t by $Y(t)$. Throughout, we make the following standard consistency assumption (Rubin, 1980).

Assumption 1.5 (Consistency). At each time t, the realized outcome $Y(t)$ is determined as

$$Y(t) = \sum_d \mathbb{1}\{D = d\} \cdot Y^{D=d}(t),$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function.

The estimand of interest is the effect of treatment on the treated, i.e., the group corresponding to $S = s_1, D = d_1$. As such, we target learning the probability measure over the counterfactual outcome Y^0 within this subgroup:

$$F(Y^0 \mid S = s_1, \ D = d_1),$$

---

[1] In a classic controlled trial, these would correspond to the treatment and control groups, respectively.

[2] Note that we do not limit our setting to panel data. In particular, the individuals for which the outcome is measured may differ across time points.

where F denotes the cumulative density function. When clear from context, we use the shorthand $F_{Y^0|s_1,d_1}$ instead. Throughout, we assume that random variables are defined over a compact domain, and that densities are absolutely continuous with respect to the Lebesgue measure.

2. One-dimensional Estimator

We commence our analysis by considering cases where the outcome of interest, Y, is one-dimensional, i.e., a scalar.

In contrast to Athey and Imbens (2006), we posit that the outcome of an individual can be determined by a combination of the latent variable U (with a common support across groups), and the group to which the individual belongs. This adjustment relaxes the model assumption of the CiC framework, as outlined below.

A1' (Model assumption). *At each state $S = s$ and treatment group $D = d$, the potential outcomes $Y^0(t)$ can be determined through a production function $h_{s,d}(\,\cdot\,;t)$ of a latent variable U, which models the individual characteristics:*

$$\forall t \in \{t_0, t_1\}: \quad Y^0(t) = h_{s,d}(U;t).$$

A key distinction between our setup and the classic CiC framework lies in relaxing 1.1 to A1', allowing the production functions $h_{s,d}(\cdot)$ to be *group-specific*. In applications involving targeted treatment assignments, A1' emerges as a more sensible assumption. Assumptions 1.2 and 1.3 are adapted analogously:

A2' (Strict monotonicity). *The production functions $h_{s,d}(\,\cdot\,;t)$ are strictly increasing in U for every $t \in \{t_0, t_1\}$, and every s, d.*

A3' (Time invariance). *Within every subgroup, the distribution of the latent variable U does not change over time. That is, $\forall u, \forall s, d$,*

$$F_{U|S=s,D=d,T=t_1}(u) = F_{U|S=s,D=d,T=t_0}(u).$$

Additionally, we prefer to formulate the overlap assumption in relation to the potential outcomes rather than the latent variable U. This preference arises from the broader accessibility and interpretability of the support of the outcome, as opposed to that of latent characteristics.

A4' (Outcome support overlap). *The potential outcomes $Y^0(t_0)$ and $Y^0(t_1)$ are defined over domains $\mathbb{Y}_0$ and $\mathbb{Y}_1$, respectively, which are common across subgroups $S \in \{s_0, s_1\}, D \in \{d_0, d_1\}$. Moreover, $\mathbb{Y}_0 \subseteq \mathbb{Y}_1$.*

Assumptions A1' through A4' establish the existence of four distinct monotone maps that push forward the density of $Y^0(t_0)$ to that of $Y^0(t_1)$ in each subgroup. In particular, let $T_{s,d}$ denote the monotone map that pushes forward the density of $Y^0(t_0)$ in the group corresponding to $S = s, D = $
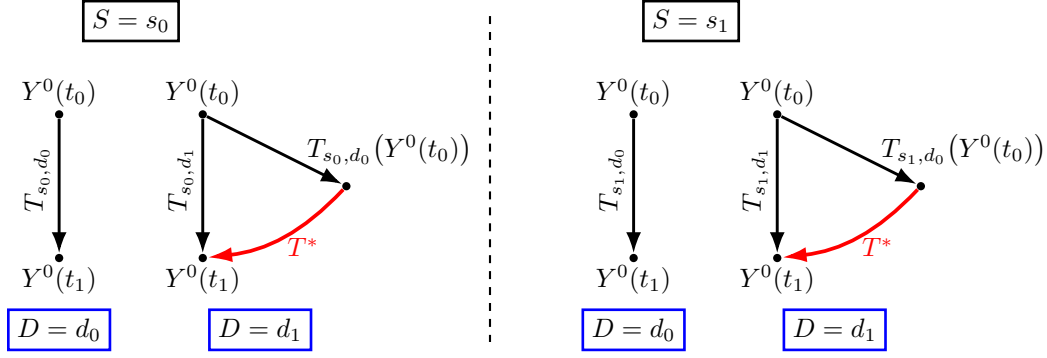
Figure 2. Illustration of the state-independent non-overlap effects assumption (2.1). The drift mapping T^* highlighted in red is identical among both states.

d to the density of $Y^0(t_1)$ in the same group (see Figure 2.) Specifically, $T_{s,d}$ can be expressed in terms of the $h(\cdot)$ functions as[3]

$$T_{s,d}(y) = h_{s,d}\big(h_{s,d}^{-1}(y;t_0);t_1\big), \qquad (4)$$

or equivalently, in terms of the cumulative density functions,

$$T_{s,d} = F_{Y^0(t_1)\,|\,s,d}^{-1} \circ F_{Y^0(t_0)\,|\,s,d}. \qquad (5)$$

It is noteworthy that under 1.5, $F_{Y^0(t)|s,t} = F_{Y(t)|s,t}$ for every subgroup except $\{s_1, d_1\}$. Consider the map $T_s^* = T_{s,d_1} \circ T_{s,d_0}^{-1}$. T_s^* can be interpreted as the non-linear drift between the maps from $Y^0(t_0)$ to $Y^0(t_1)$ among control and treatment groups in any state s (see Figure 2.) Identification in CiC is achieved by assuming that there is no drift, i.e., T_s^* is the identity map (see Eq. 2). We shall proceed by relaxing this assumption as follows.

Assumption 2.1 (State-independent drifts)**.** The drift between the mappings of potential outcomes $Y^0(t_0)$ to $Y^0(t_1)$ in the control and treatment groups is independent of the state. Formally,

$$T^* := T_{s_0,d_1} \circ T_{s_0,d_0}^{-1} \equiv T_{s_1,d_1} \circ T_{s_1,d_0}^{-1}.$$

Specifically, rather than assuming *no drift*, we have relaxed the assumption to *equal drift* across the two states.

Remark 2.2. As mentioned earlier, our focus is on policies that target specific sub-populations. For example, the Medicaid expansion discussed in Section 5 targets the low-income population. While the treated (low-income) and control (high-income) groups may have inherent differences, we expect these differences to be invariant across neighboring states in the US. Therefore, while it may not be plausible to impose the same mappings for the control and treatment groups (as in Eq. 2), it is more credible to impose the same correction map (T^* as introduced above) across states.

---

[3]For the purposes of this section, Eq. (4) can be considered as the definition of maps $T_{s,d}$.

2.1 can also be expressed in terms of the production functions, albeit at the cost of interpretability:

$$h_{s_0,d_1}\Big\{h_{s_0,d_1}^{-1}\big\{h_{s_0,d_0}\big(h_{s_0,d_0}^{-1}(\cdot;t_1);t_0\big);t_0\big\};t_1\Big\} =$$
$$h_{s_1,d_1}\Big\{h_{s_1,d_1}^{-1}\big\{h_{s_1,d_0}\big(h_{s_1,d_0}^{-1}(\cdot;t_1);t_0\big);t_0\big\};t_1\Big\}. \quad (6)$$

We are now ready to state our identification result. The complete set of proofs for our results can be found in Appendix D.

Theorem 2.3. *Under assumptions A1' - A4' and 1.5 - 2.1, the cumulative density function of the missing counterfactual $Y^0(t_1)$ in the group $S = s_1, D = d_1$ is identified as:*

$$F_{Y^0(t_1)|s_1,d_1}(y) =$$
$$F_{Y(t_0)|s_1,d_1} \circ T_{s_0,d_1}^{-1} \circ T_{s_0,d_0} \circ T_{s_1,d_0}^{-1}(y), \quad (7)$$

where $T_{s,d}$ is given by Eq. (5).

Remark 2.4. To avoid unnecessarily heavy notation, we did not discuss the observed covariates. However, an identical analysis can be done after adjusting for the observed covariates, X. In particular, production functions may depend on the observed covariates, as long as their monotonicity in U is maintained for every x in the domain of X. A3', A4' and 2.1 need to be valid conditioned on X in this scenario, and Eq. (7) must hold when every term is conditioned on X.

Remark 2.5. We articulated our identifiability assumptions in accordance with the original work of Athey and Imbens (2006). An alternative way of presenting the assumptions would be to do it akin to the *quantile-quantile equiconfounding bias* assumption proposed by Ghassami et al. (2022). In our context, this would translate to directly assuming the existence of monotone maps $T_{s,d}$ based on Eq. (5) instead of drawing conclusions from A1'-A4' to establish it.

Remark 2.6. We formulated production functions to model the potential outcomes under no treatment (Y^0), whereas we left the other potential outcome, Y^1, unrestricted. Due

to symmetry, one could model Y^1 using production functions and leave Y^0 unrestricted. This scenario might arise for instance, in an study where 3 out of 4 cohorts receive treatment. One should exercise greater caution in such cases however, since in practice, the treatment may have effects that significantly alter the composition of the population under study. Under these circumstances, assuming monotone production functions for Y^1 could be a more drastic assumption.

2.1. Relaxing monotonicity

As mentioned earlier, A2' (and its counterpart, 1.2 in Athey and Imbens (2006)) is an untestable assumption, and might be drastic to impose in certain applications. The monotonicity of functions $h_{s,d}(\cdot; t)$ in U has two implications: (i) these functions are bijective, establishing a well-defined inverse for them; (ii) the property that $\mathbb{P}(h_{s,d}(U; t) \leq y) = \mathbb{P}(U \leq h_{s,d}^{-1}(y; t))$, which is repeatedly utilized in proving the point identification result in Theorem 2.3 (see Appendix D). While the bijectivity of $h_{s,d}(\cdot; t)$ appears to be essential for our framework to work, we can relax (ii). Let us first rephrase the monotonicity assumption in the equivalent form:

$$\langle u_0 - u_1, h_{s,d}(u_0; t) - h_{s,d}(u_1; t) \rangle \geq 0, \quad \forall u_0, u_1. \quad (8)$$

This assumption can be relaxed as follows.

Assumption 2.7 (ϵ-monotonicity). Functions $h_{s,d}^{-1}(\cdot; t)$ are well-defined. Additionally, for any u in the support of U,

$$\mathbb{P}\big(\langle U - u, h_{s,d}(U; t) - h_{s,d}(u; t) \rangle < 0 \mid s, d\big) \leq \frac{\epsilon}{2}. \quad (9)$$

For example, the monthly income of an individual in terms of her age after adjusting for the other covariates can be a ϵ-monotone function. In general, monthly income increases due to promotions and inflation. However, temporary unemployment and retirement can affect this trend. See Figure 3 for a visualization.

We expect that under this relaxation, point identification cannot be achieved. However, the following partial identification result holds.

Proposition 2.8. *Under assumptions A1', A3', A4' and 1.5 - 2.7, for any $y \in \mathbb{Y}_1$,*

$$F_{Y(t_0)|s_1,d_1} \circ \underline{\Phi}_{s_0,d_1}^{-1} \circ \underline{\Phi}_{s_0,d_0} \circ \underline{\Phi}_{s_1,d_0}^{-1}(y) - \epsilon$$
$$\leq F_{Y^0(t_1)|s_1,d_1}(y) \leq$$
$$F_{Y(t_0)|s_1,d_1} \circ \overline{\Phi}_{s_0,d_1}^{-1} \circ \overline{\Phi}_{s_0,d_0} \circ \overline{\Phi}_{s_1,d_0}^{-1}(y) + \epsilon,$$
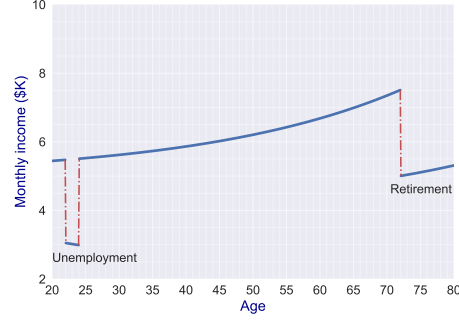

Figure 3. Gross monthly income versus age after adjusting for other covariates.

where

$$\underline{\Phi}_{s,d}(y) = F_{Y(t_1)|s,d}^{-1}\big(F_{Y(t_0)|s,d}(y) - \epsilon\big),$$
$$\underline{\Phi}_{s,d}^{-1}(y) = F_{Y(t_0)|s,d}^{-1}\big(F_{Y(t_1)|s,d}(y) - \epsilon\big),$$
$$\overline{\Phi}_{s,d}(y) = F_{Y(t_1)|s,d}^{-1}\big(F_{Y(t_0)|s,d}(y) + \epsilon\big),$$
$$\overline{\Phi}_{s,d}^{-1}(y) = F_{Y(t_0)|s,d}^{-1}\big(F_{Y(t_1)|s,d}(y) + \epsilon\big).$$

Remark 2.9. Note that strict monotonicity is a special case of 2.7, which corresponds to $\epsilon = 0$. Accordingly, Proposition 2.8 reduces to Theorem 2.3 with the choice of $\epsilon = 0$.

Remark 2.10. An analogous result applies to the original framework of CiC when 1.2 is relaxed to ϵ-monotonicity. See Appendix A.1 for details.

2.2. Relaxing time invariance

In certain applications, A3' may also be violated. In particular, in repeated cross-sections, it might be challenging to maintain the same distribution among the cases under study. Even if possible, this may both reduce the number of available samples and result in selection bias, making inference more complicated. As such, we consider a relaxation of A3' and derive a partial identification result in this setting. We assume that the distribution of the latent variable may change over time, but this change is bounded in Kolmogorov (aka KS) distance (Kolmogorov, 1933).

Assumption 2.11 (δ-invariance). Within every subgroup, the Kolmogorov distance of the distribution of the latent variable across $T = t_0$ and $T = t_1$ is bounded by δ. That is,

$$\sup_u \big| F_{U|S=s,D=d,T=t_1}(u) - F_{U|S=s,D=d,T=t_0}(u) \big| \leq \delta.$$

Note again that when $\delta = 0$, 2.11 reduces to A3'.

Proposition 2.12. *Under assumptions A1', A2', 2.11, A4',*

and 1.5-2.1, for any $y \in \mathbb{Y}_1$,

$$F_{Y(t_0)|s_1,d_1} \circ \underline{\Psi}^{-1}_{s_0,d_1} \circ \underline{\Psi}_{s_0,d_0} \circ \underline{\Psi}^{-1}_{s_1,d_0}(y) - \delta$$
$$\leq F_{Y^0(t_1)|s_1,d_1}(y) \leq$$
$$F_{Y(t_0)|s_1,d_1} \circ \overline{\Psi}^{-1}_{s_0,d_1} \circ \overline{\Psi}_{s_0,d_0} \circ \overline{\Psi}^{-1}_{s_1,d_0}(y) + \delta,$$

where

$$\underline{\Psi}_{s,d}(y) = F^{-1}_{Y(t_1)|s,d}\big(F_{Y(t_0)|s,d}(y) - \delta\big),$$
$$\underline{\Psi}^{-1}_{s,d}(y) = F^{-1}_{Y(t_0)|s,d}\big(F_{Y(t_1)|s,d}(y) - \delta\big),$$
$$\overline{\Psi}_{s,d}(y) = F^{-1}_{Y(t_1)|s,d}\big(F_{Y(t_0)|s,d}(y) + \delta\big),$$
$$\overline{\Psi}^{-1}_{s,d}(y) = F^{-1}_{Y(t_0)|s,d}\big(F_{Y(t_1)|s,d}(y) + \delta\big).$$

Remark 2.13. Even more generally, we can simultaneously relax A2' to 2.7 and A3' to 2.11. See Proposition A.3 in Appendix A for the partial identification result for this case. Moreover, Proposition A.6 provides an analogous result in the context of the original CiC framework.

2.3. Marginal contrasts and joint counterfactuals

Theorem 2.3 guarantees the identification of the probability density of the missing counterfactual, $Y^0(t_1)$ in the group corresponding to $S = s_1, D = d_1$. Having access to this density, we can compute any *marginal contrast estimand* (Franks et al., 2019). A marginal contrast estimand is an estimand that can be expressed as a functional of the marginal distribution of the counterfactual outcomes. This includes a vast majority of commonly used estimands, such as average treatment effects, conditional average treatment effects, quantile treatment effects, risk ratios, etc. However, in certain applications, more information is desired. Examples of estimands that are not marginal contrasts include the *distribution* of the treatment effect, individual-level treatment effects, or the quantiles of the treatment effect. As a concrete example, consider $Z := Y^1(t_1) - Y^0(t_1)$. The density of Z is not identifiable from merely the marginal densities of $Y^1(t_1)$ and $Y^0(t_0)$. In particular, the treatment may have non-zero effects on a fraction of the population even if the two potential outcomes have identical distributions. Estimands that are not marginal contrasts require stronger identifiability assumptions in general. In this section, we discuss a stronger version of A3' that can lead to the identification of such estimands.

A3" (Strong time invariance). *The latent variable U does not change over time, and there is no loss to follow-up.*

A3" is equivalent to a rank invariance assumption across time points within the same group (Doksum, 1974; Lehmann and D'Abrera, 2006). This is stronger than A3', in the sense that it completely rules out the possibility of any changes in the latent variable itself, or the cohort of the study. In contrast, A3' would allow for loss to follow-up, as long as

similar individuals were recruited for the study. A3' even accommodates a repeated cross-sections study. A3' also allows for the evolution of the latent variable, as long as its distribution among the study population remains the same. On the other hand, the stronger assumption A3" allows for the identification of a wider range of causal estimands. In particular, under A3", the joint density of the counterfactuals $\big(Y^0(t_1), Y^1(t_1)\big)$ is identified.

Proposition 2.14. *Under assumptions A1', A2', A3", A4', and 1.5 - 2.1, the joint density of $Y^0(t_1)$ and $Y^1(t_1)$ in the group corresponding to $S = s_1, D = d_1$ is identified as*

$$F_{Y^0(t_1),Y^1(t_1)|s_1,d_1}(y^0, y^1) =$$
$$F_{Y(t_0),Y(t_1)|s_1,d_1}\big(F^{-1}_{Y(t_0)|s_1,d_1} \circ F_{Y^0(t_1)|s_1,d_1}(y^0), y^1\big),$$

where $F_{Y^0(t_1)|s_1,d_1}(\cdot)$ is given by Equation (7).

3. Inference

In this section, we discuss the estimation aspect of our framework with a focus on the average effect of treatment on the treated. More formally, we consider the estimation of

$$\tau := \mathbb{E}[Y^1(t_1) - Y^0(t_1)|S = s_1, D = d_1].$$

Note that under the identifiability assumptions of Section 2, τ is identified as

$$\tau = \mathbb{E}[Y(t_1)|s_1, d_1] - \mathbb{E}[F^{-1}_{Y(t_1)|s_0,d_1} \circ$$
$$F_{Y(t_0)|s_0,d_1} \circ F^{-1}_{Y(t_0)|s_0,d_0} \circ F_{Y(t_1)|s_0,d_0} \circ$$
$$F^{-1}_{Y(t_1)|s_1,d_0} \circ F_{Y(t_0)|s_1,d_0}\big(Y(t_0)\big)|s_1, d_1]. \quad (10)$$

We make the following assumption on the data generating mechanism to render estimation feasible.

Assumption 3.1. Conditioned on $S = s, D = d, T = t$, variables $Y(t)$ are continuous random variables defined on a shared bounded domain $[\underline{y}, \overline{y}]$, with continuously differentiable density functions f_{sdt}, where f_{sdt} is bounded from above and away from 0, and $\partial f_{sdt}/\partial y$ is bounded. For all s, d, t, $p_{sdt} = \mathbb{P}(S = s, D = d, T = t) > 0$, and given s, d, t, the samples $Y_i(t)$ are independent draws from f_{sdt}.

Let N be the total number of observed outcome samples. Akin to Athey and Imbens (2006), we build a finite-sample estimator for τ based on empirical estimators of cumulative density functions. In particular, let $\{Y_{sd,i}(t)\}_{i=1}^{N_{sdt}}$ denote the independent samples of the outcome at time t in group $S = s, D = d$. We define

$$\hat{F}_{Y(t)|s,d}(y) := N^{-1}_{sdt} \sum_{i=1}^{N_{sdt}} \mathbb{1}\{Y_{sd,i}(t) \leq y\}, \quad \text{and,} \quad (11)$$

$$\hat{F}^{-1}_{Y(t)|s,d}(u) := \inf\{y \in \mathbb{Y}_t, \hat{F}_{Y(t)|s,d}(y) \geq u\}. \quad (12)$$

Finally, the estimator for τ is built as:

$$\hat{\tau} := N_{s_1 d_1 t_1}^{-1} \sum_{i=1}^{N_{s_1 d_1 t_1}} Y_{s_1 d_1, i}(t_1) -$$

$$N_{s_1 d_1 t_0}^{-1} \sum_{i=1}^{N_{s_1 d_1 t_0}} \hat{F}_{Y(t_1)|s_0,d_1}^{-1} \circ \hat{F}_{Y(t_0)|s_0,d_1} \circ$$

$$\hat{F}_{Y(t_0)|s_0,d_0}^{-1} \circ \hat{F}_{Y(t_1)|s_0,d_0} \circ \hat{F}_{Y(t_1)|s_1,d_0}^{-1} \circ$$

$$\hat{F}_{Y(t_0)|s_1,d_0}\big(Y_{s_1 d_1, i}(t_0)\big). \quad (13)$$

The following theorem establishes the consistency and asymptotic normality of $\hat{\tau}$.

Theorem 3.2. *Under 3.1, $\hat{\tau} - \tau = \mathcal{O}_p(N^{-\frac{1}{2}})$, and*

$$\sqrt{N}(\hat{\tau} - \tau) \xrightarrow{D} \mathcal{N}(0, \frac{V_0}{p_{s_1 d_1 t_1}} + \frac{V_1}{p_{s_0 d_1 t_1}} + \frac{V_2}{p_{s_0 d_1 t_0}} +$$

$$\frac{V_3}{p_{s_0 d_0 t_0}} + \frac{V_4}{p_{s_0 d_0 t_1}} + \frac{V_5}{p_{s_1 d_0 t_1}} + \frac{V_6}{p_{s_1 d_0 t_0}} + \frac{V_7}{p_{s_1 d_1 t_0}}), \quad (14)$$

where $\{V_i\}_{i=0}^7$ are given by Eq. (57).

Due to space limitations, the expression for the variance and the discussion on its estimation are postponed to Appendix B.

4. Optimal Transport Representation and High-dimensional Extension

So far, we focused on the special case of scalar outcome for ease of presentation. However, in many applications, the researcher may be interested in simultaneously analyzing the effect of a policy on several outcomes. For instance, in a study examining the impact of a job training program on both earnings, Y, and job satisfaction, Z, applying our method separately to each outcome would only identify the marginal distribution, and would simply overlook potential dependencies or interactions between earnings and job satisfaction. This is because the copula term is left unrestricted.

In this section, we discuss the generalization of our results to cover multi-dimensional outcomes. To this end, we first review the one-dimensional case from an optimal transport point of view. Let $\eta_{s,d}$ and $\mu_{s,d}$ denote the probability measures over $Y^0(t_0)$ and $Y^0(t_1)$ in the group corresponding to $S = s, D = d$, respectively. Brenier's theorem (Brenier, 1991) implies that there exists a unique mapping T such that $T_{\#}\eta_{s,d} = \mu_{s,d}$, i.e., T pushes forward $\eta_{s,d}$ to $\mu_{s,d}$, and T is the gradient of a convex function. Moreover, T is the optimal transport map with quadratic cost. More precisely, let Γ denote the space of joint distributions over $\big(Y^0(t_0), Y^0(t_1)\big)$ conditioned on $S = s, D = d$, that agree with the marginal densities $\eta_{s,d}$ and $\mu_{s,d}$. The optimization problem

$$\inf_{\gamma \in \Gamma} \int ||y_0 - y_1||^2 d\gamma(y_0, y_1) \quad (15)$$

has a unique solution γ^*, where $\big(Y^0(t_0), Y^0(t_1)\big) \sim \gamma^*$ if and only if $Y^0(t_0) \sim \eta_{s,d}$ and $Y^0(t_1) = T\big(Y^0(t_0)\big), \eta_{s,d} - a.s.$[4] It is straightforward to verify that a one-dimensional function is the gradient of a convex function if and only if it is monotone. Therefore, the monotonicity of $h_{s,d}(\cdot)$ and the monotonicity of $T_{s,d}(\cdot)$ (as a consequence of the latter) imply $T_{s,d} \equiv T$. In other words, the mapping $T_{s,d}$ is identified as the optimal transport map with quadratic cost that pushes forward $\eta_{s,d}$ to $\mu_{s,d}$.

Indeed, monotonicity can be slightly relaxed. The identifiability of the map $T_{s,d}$ in one dimension is guaranteed under *co-monotonicity* of $h_{s,d}(\cdot; t_0)$ and $h_{s,d}(\cdot; t_1)$:

$$\langle u_0 - u_1, h_{s,d}(u_0; t_0) - h_{s,d}(u_1; t_1) \rangle \geq 0, \ \forall u_0, u_1. \quad (16)$$

In order to achieve identifiability results in higher dimensions based on Brenier's theorem, we need to make sure that $T_{s,d}$ is the gradient of a convex function. It is known that a function is the gradient of a convex function if and only if it is *cyclically monotone* (Rockafellar, 1970). A function h is said to be cyclically monotone if for any sequence $x_0, \ldots, x_n$ in its domain,

$$\sum_{i=0}^n \langle h(x_i), x_i - x_{i+1} \rangle \geq 0, \quad (17)$$

where $x_{n+1} = x_0$. For $n = 1$, Eq. (17) reduces to Eq. (8). With this preliminary discussion in place, the identification result in higher dimensions can be stated as follows.

A2'' (Co-cyclic monotonocity). *Functions $h_{s,d}(\cdot; t_0)$ and $h_{s,d}(\cdot; t_1)$ are co-cyclically monotone. That is, for any sequence $u_0, \ldots, u_n$ in their common domain,*

$$\sum_{i=0}^n \langle h_{s,d}(u_i; t_0), h_{s,d}(u_i; t_1) - h_{s,d}(u_{i+1}; t_1) \rangle \geq 0,$$

where $u_{n+1} = u_0$.

Theorem 4.1. *Under assumptions A1', A2'', A3', A4', and 1.5 - 2.1, the probability measure over the missing counterfactual, i.e., μ_{s_1, d_1}, is identified as*

$$\mu_{s_1, d_1} = (T^* \circ T_{s_1, d_0})_{\#} \eta_{s_1, d_1}, \quad (18)$$

where $T_{s,d}$ is the Brenier map that pushes forward $\eta_{s,d}$ to $\mu_{s,d}$ for every $s \in \{s_0, s_1\}, d \in \{d_0, d_1\}$, and T^* is the Brenier map that pushes forward $(T_{s_0, d_0 \#} \eta_{s_0, d_1})$ to μ_{s_0, d_1}.

Theorem 4.1 is in essence a generalization of Theorem 2.3. To see this, note that the Brenier map that pushes forward $\eta_{s,d}$ to $\mu_{s,d}$ in one dimension is precisely $F_{Y^0(t_1)|s,d}^{-1} \circ F_{Y^0(t_0)|s,d}$. However, the co-cyclic monotonicity assumption A2'' is not as easy to interpret as A2'. To address this challenge, we follow the proposition of Torous et al. (2021) based on the following result.

---

[4]Note that we have omitted the implicit conditioning on $S = s, D = d$ in our notation to improve readability.

(a) Linear production functions.
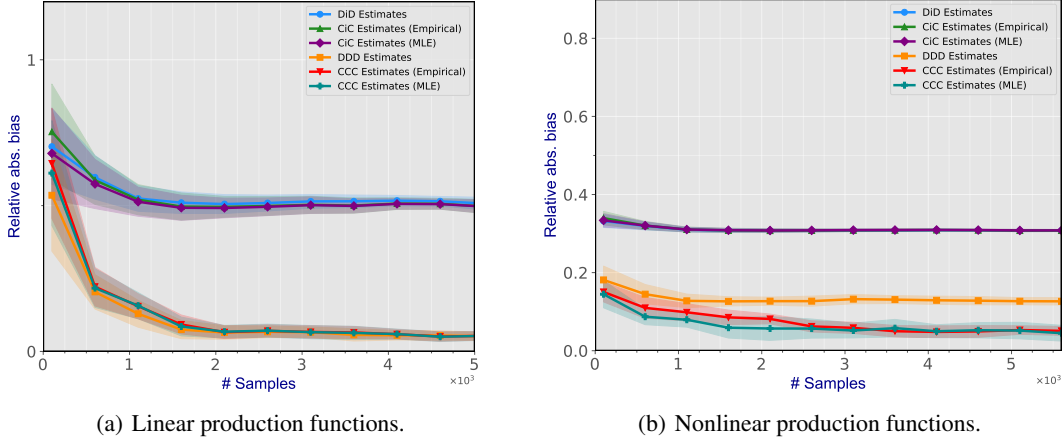

(b) Nonlinear production functions.

Figure 4. Relative bias of the estimators evaluated on (a) a linear, and (b) a non-linear model.

Proposition 4.2 (Saks and Yu, 2005). *Let K and F be a convex and a finite subset of $\mathbb{R}^d$ respectively. A function $T : K \to F$ is cyclically monotone if and only if it is monotone.*

Proposition 4.2 implies that in our setting, if the densities $\eta_{s,d}$ are supported on convex sets and $Y^0(t_1)$ is finite-valued in every subgroup, then A2" reduces to co-monotonicity (16). In other words, restricting $\eta_{s,d}$ and $\mu_{s,d}$ densities to be defined over convex and finite sets, respectively, Assumption A2" of Theorem 4.1 can be replaced by the easily interpretable assumption of Eq. (16).

5. Simulation Studies

Our empirical analysis is structured into two main sections[5]. In the first part, we compare the estimation error of the triple changes estimator and DiD, triple difference, and CiC estimators using synthetically generated datasets, where we know the ground truth: i.e., the treatment effect on the treated. To evaluate the performance of each estimator, we focus on the relative bias metric, defined as $\varepsilon = \left| 1 - \frac{\widehat{\tau}}{\tau} \right|$, where τ is the average treatment effect on the treated, given by (10), and $\widehat{\tau}$ represents the estimate provided by the respective estimator. In the second part, we apply our estimator to data from *National Survey on Children's Health* (NSCH)[6] to assess the effect of Medicaid expansion under Affordable Care Act (ACA) in the united states on preventive care for children.

---

[5]The code for reproducing the simulation results of this paper are accessible at https://github.com/SinaAkbarii/Triple-Changes.

[6]https://mchb.hrsa.gov/national-survey-childrens-health-questionnaires-datasets-supporting-documents

5.1. Synthetic data

We begin with a simple linear model where the latent variable U given s, d is sampled from a Gaussian distribution with mean $\nu_{s,d}$ and variance 1 (see Appendix E for a complete table of parameters $\nu_{s,d}$, as well as further details of our experiment setup,) and the production functions are defined as

$$ h_{s,d}(u; t) := 2u + \left(\frac{1+s}{4} + \frac{d-0.5}{2} \right)t. \quad (19) $$

The actual outcome at time t_1 in the treated group (s_1, d_1) is also sampled from a Gaussian distribution with mean 2.75 and variance 1.

CiC and triple changes estimators use estimates of the cumulative density functions and their inverses. We implemented two versions of each of these estimators, namely a nonparametric estimator using the empirical estimator of density functions (see Eq. 11 and 12,) and a model-based version using the maximum likelihood estimator given a class of distributions. The class of distributions in this section are specified correctly, i.e., the true model lies in the class. See appendix E for a setting where this class is misspecified.

In Figure 4, the triple difference and triple changes estimators are denoted by DDD and CCC, respectively. As depicted by Figure 4(a), DiD and CiC estimators exhibit persistent bias, not converging to zero even with increasing sample size. In contrast, the biases of triple difference and triple changes estimators approach zero as the sample size grows. Furthermore, since the triple difference estimator only relies on empirical averages, it shows slightly lower bias in small sample sizes.

To add non-linearities to the previous model, we modified the data generating mechanism, including two of the production functions. Specifically, for (s_0, d_1, t_1), and (s_1, d_1, t_1),

the production functions were modified to

$$h_{s,d}(u;t) = 0.1 \exp\Big(2u + \big(\frac{1+s}{4} + \frac{d-0.5}{2}\big)t\Big),$$

whereas the other six groups were generated according to Eq. (19). The outcomes for this model are illustrated in Figure 4(b). Notably, under the nonlinear model, the triple-difference estimator yields biased estimates, while the triple-changes estimators remain (asymptotically) unbiased. Furthermore, owing to the complexity of the model, empirical estimators of density functions exhibit more bias compared to their MLE-based counterparts, particularly in low-sample regimes.

5.2. Application to NSCH data

The Medicaid expansion under Affordable Care Act was designed to extend Medicaid coverage to more low-income citizens in the US. This expansion raised the income threshold for Medicaid eligibility to 138% of the federal poverty level (FPL). While Louisiana (s_1) adopted this expansion in 2016, the neighboring states of Mississippi and Texas (s_0) are yet to do so. We utilized publicly available anonymous NSCH data for the years 2016 (t_0) and 2017 (t_1) to assess the impact of Medicaid expansion on children's access to preventive healthcare, specifically analyzing the change in the frequency of doctor visits. Individuals with FPL $\leq 100\%$ were classified as eligible for the expansion (d_1), while those with FPL $\geq 140\%$ were considered not eligible (d_0). Data within the uncertainty margin between these thresholds was excluded.

Applying the triple difference and triple changes estimators on the frequency of doctor visits for preventive purposes with 1000 bootstraps resulted in means of 0.170 and 0.145, respectively, with 90% confidence intervals of $[-0.004, 0.344]$ and $[-0.01, 0.331]$, respectively. Our findings suggest that the adoption of Medicaid expansion in Louisiana increased the likelihood of children undergoing an annual preventive care visit, compared to those in non-expansion states. This aligns with the conclusion of Roy et al. (2020).

6. Concluding Remarks

We formally presented the triple changes estimator for assessing the treatment effect on the treated in observational studies. We derived a set of necessary assumptions for point identification and discussed partial identification results under relaxed versions of these assumptions. Future research avenues could include exploring partial identification under alternative assumptions, investigating extensions to time series settings, and conducting statistical analyses to enhance the robustness and applicability of the estimator.

Impact Statement

The primary contribution of this work is a novel estimator, the 'triple changes' estimator, which combines the strengths of both triple difference and changes-in-changes (CiC) methodologies. This work aims to enhance the methodological foundations of causal inference, thereby supporting research practices across various domains, including policy evaluation (e.g., assessing regional impacts of new legislation), public health (e.g., evaluating interventions across different communities), economics (e.g., analyzing effects of market reforms across sectors), and environmental studies (e.g., studying the impact of conservation efforts in different areas). We do not foresee any specific ethical concerns or negative societal implications arising directly from this methodological contribution.

References

Amir Mohammad Abouei, Ehsan Mokhtarian, and Negar Kiyavash. s-id: Causal effect identification in a subpopulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20302–20310, 2024.

Sina Akbari, Jalal Etesami, and Negar Kiyavash. Minimum cost intervention design for causal effect identification. In *International Conference on Machine Learning*, pages 258–289. PMLR, 2022.

Orley C Ashenfelter and David Card. Using the longitudinal structure of earnings to estimate the effect of training programs, 1984.

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

Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

Brantly Callaway, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. Difference-in-differences with a continuous treatment. *arXiv preprint arXiv:2107.02637*, 2021.

David Card and Alan B Krueger. Minimum wages and employment: A case study of the fast food industry in new jersey and pennsylvania, 1993.

Qihui Chen, Xueling Chu, Suzhen Wang, and Bo Zhang. A triple-difference approach to re-evaluating the impact of china's new cooperative medical scheme on incidences of

chronic diseases among older adults in rural communities. *Risk Management and Healthcare Policy*, pages 643–659, 2020.

Kjell Doksum. Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *The annals of statistics*, pages 267–277, 1974.

Alexander M Franks, Alexander D'Amour, and Avi Feller. Flexible sensitivity analysis for observational studies without observable implications. *Journal of the American Statistical Association*, 2019.

AmirEmad Ghassami, Alan Yang, David Richardson, Ilya Shpitser, and Eric Tchetgen Tchetgen. Combining experimental and observational data for identification and estimation of long-term causal effects. *arXiv preprint arXiv:2201.10743*, 2022.

Jonathan Gruber. The incidence of mandated maternity benefits. *The American economic review*, pages 622–641, 1994.

SeungHoon Han, Charles C Branas, and John M MacDonald. The effect of a sunday liquor-sales ban repeal on crime: A triple-difference analysis. *Alcoholism: Clinical and Experimental Research*, 40(5):1111–1121, 2016.

Andrey Kolmogorov. Sulla determinazione empirica di una legge didistribuzione. *Giorn Dell'inst Ital Degli Att*, 4: 89–91, 1933.

Erich Leo Lehmann and Howard JM D'Abrera. *Nonparametrics: statistical methods based on ranks*, volume 464. Springer New York, 2006.

Andreas Olden and Jarle Møen. The triple difference estimator. *The Econometrics Journal*, 25(3):531–553, 2022.

Julia Raifman, Ellen Moscoe, S Bryn Austin, Mark L Hatzenbuehler, and Sandro Galea. Association of state laws permitting denial of services to same-sex couples with mental distress in sexual minority adults: A difference-in-difference-in-differences analysis. *JAMA psychiatry*, 75(7):671–677, 2018.

Ralph Tyrell Rockafellar. Convex analysis, 1970.

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

Shreya Roy, Fernando A Wilson, Li-Wu Chen, Jungyoon Kim, and Fang Yu. The impact of medicaid expansion for adults under the affordable care act on preventive care for children: evidence from the southern united states. *Medical care*, 58(11):945–951, 2020.

Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593, 1980.

Michael Saks and Lan Yu. Weak monotonicity suffices for truthfulness on convex domains. In *Proceedings of the 6th ACM conference on Electronic commerce*, pages 286–293, 2005.

Sergio Naruhiko Sakurai and Maria Isabel Accoroni Theodoro. On the relationship between political alignment and government transfers: triple differences evidence from a developing country. *Empirical Economics*, 58:1107–1141, 2020.

Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

John Snow. *On the Mode of Communication of Cholera*. John Churchill, 1855.

Tamar Sofer, David B Richardson, Elena Colicino, Joel Schwartz, and Eric J Tchetgen Tchetgen. On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 31(3):348, 2016.

Ming Tai-Seale, Deborah Freund, and Anthony LoSasso. Racial disparities in service use among medicaid beneficiaries after mandatory enrollment in managed care: a difference-in-differences approach. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 38(1):49–59, 2001.

William Torous, Florian Gunsilius, and Philippe Rigollet. An optimal transport approach to causal inference. *arXiv preprint arXiv:2108.05858*, 2021.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Appendices

The appendices are organized as follows. We begin with providing a general partial identification result that was promised in the main text in Section A. We review the consistency and asymptotic normality of the estimator $\hat{\tau}$ given by Eq. (13) in Appendix B. Appendix C provides twelve lemmas that form the basis of the proof of Theorem 3.2. Appendix D includes the omitted proofs of the claims made in the main text. Some of these claims are corollaries of the general result provided in Appendix A. Finally, Appendix E provides further experimental results, as well as complementary details of the experimental setup used in the main text.

A. Supplementary Findings on Partial Identification of the Missing Potential Outcome

Propositions 2.8 and 2.12 given in the text are in effect corollaries of the result given below. We later provide an analogous result for the classic CiC framework. For completeness, we include the relaxed assumptions here.

Assumption A.1 (ϵ-monotonicity). Functions $h_{s,d}^{-1}(\cdot; t)$ are well-defined, and for any u in the support of U,

$$\mathbb{P}\big(\langle U - u, h_{s,d}(U; t) - h_{s,d}(u; t)\rangle < 0 \mid S = s, D = d, T = t\big) \leq \frac{\epsilon}{2}. \tag{20}$$

Remark A.2. A.1 is slightly different from 2.7 in that we have also conditioned on $T = t$ in Eq. (20). Clearly, under A3', the two are equivalent. However, we are relaxing A3' to 2.11 in this section, and it is imperative to highlight this difference.

Assumption 2.11 (δ-invariance). Within every subgroup, the Kolmogorov distance of the distribution of the latent variable across $T = t_0$ and $T = t_1$ is bounded by δ. That is,

$$\sup_u \big|F_{U|S=s,D=d,T=t_1}(u) - F_{U|S=s,D=d,T=t_0}(u)\big| \leq \delta.$$

Proposition A.3. *Under assumptions A1', A.1, 2.11, A4', and 1.5-2.1, for any $y \in \mathbb{Y}_1$,*

$$F_{Y(t_0)|s_1,d_1} \circ \underline{\Psi}_{s_0,d_1}^{-1} \circ \underline{\Psi}_{s_0,d_0} \circ \underline{\Psi}_{s_1,d_0}^{-1}(y) - \epsilon - \delta$$
$$\leq F_{Y^0(t_1)|s_1,d_1}(y) \leq$$
$$F_{Y(t_0)|s_1,d_1} \circ \overline{\Psi}_{s_0,d_1}^{-1} \circ \overline{\Psi}_{s_0,d_0} \circ \overline{\Psi}_{s_1,d_0}^{-1}(y) + \epsilon + \delta,$$

where

$$\underline{\Psi}_{s,d}(y) = F_{Y(t_1)|s,d}^{-1}\big(F_{Y(t_0)|s,d}(y) - \epsilon - \delta\big),$$
$$\underline{\Psi}_{s,d}^{-1}(y) = F_{Y(t_0)|s,d}^{-1}\big(F_{Y(t_1)|s,d}(y) - \epsilon - \delta\big),$$
$$\overline{\Psi}_{s,d}(y) = F_{Y(t_1)|s,d}^{-1}\big(F_{Y(t_0)|s,d}(y) + \epsilon + \delta\big),$$
$$\overline{\Psi}_{s,d}^{-1}(y) = F_{Y(t_0)|s,d}^{-1}\big(F_{Y(t_1)|s,d}(y) + \epsilon + \delta\big).$$

Proof. Suppose $t \in \{t_0, t_1\}$.

$$\begin{aligned}
F_{Y^0(t)|s,d}(y) &= \mathbb{P}(h_{s,d}(U; t) \leq y \mid S = s, D = d, T = t) \\
&= \mathbb{P}(h_{s,d}(U; t) \leq y,\, U \leq h_{s,d}^{-1}(y; t) \mid S = s, D = d, T = t) \\
&+ \mathbb{P}(h_{s,d}(U; t) \leq y,\, U > h_{s,d}^{-1}(y; t) \mid S = s, D = d, T = t),
\end{aligned} \tag{21}$$

which implies the following inequalities:

$$\begin{aligned}
F_{Y^0(t)|s,d}(y) & \\
&\overset{(a)}{\leq} \mathbb{P}(U \leq h_{s,d}^{-1}(y; t) \mid S = s, D = d, T = t) \\
&+ \mathbb{P}(h_{s,d}(U; t) \leq y,\, U > h_{s,d}^{-1}(y; t) \mid S = s, D = d, T = t) \\
&\overset{(b)}{\leq} \mathbb{P}(U \leq h_{s,d}^{-1}(y; t) \mid S = s, D = d, T = t) + \frac{\epsilon}{2} \\
&= F_{U|s,d,t}(h_{s,d}^{-1}(y; t)) + \frac{\epsilon}{2},
\end{aligned} \tag{22}$$

where (a) is a basic probability manipulation and (b) follows from 2.7, and

$$
\begin{aligned}
F_{Y^0(t)|s,d}(y) \\
&\overset{(c)}{\geq} \mathbb{P}(h_{s,d}(U;t) \leq y,\, U \leq h_{s,d}^{-1}(y;t) \mid S = s, D = d, T = t) \\
&= \mathbb{P}(U \leq h_{s,d}^{-1}(y;t) \mid S = s, D = d, T = t) \\
&\quad - \mathbb{P}(h_{s,d}(U;t) > y,\, U \leq h_{s,d}^{-1}(y;t) \mid S = s, D = d, T = t) \\
&\overset{(d)}{\geq} \mathbb{P}(U \leq h_{s,d}^{-1}(y;t) \mid S = s, D = d, T = t) - \frac{\epsilon}{2} \\
&= F_{U|s,d,t}(h_{s,d}^{-1}(y;t)) - \frac{\epsilon}{2},
\end{aligned}
\tag{23}
$$

where (c) is a consequence of (21), and (d) follows from 2.7. To summarize, we got

$$
F_{U|s,d,t}(h_{s,d}^{-1}(y;t)) - \frac{\epsilon}{2} \leq F_{Y^0(t)|s,d}(y) \leq F_{U|s,d,t}(h_{s,d}^{-1}(y;t)) + \frac{\epsilon}{2},
\tag{24}
$$

and using 2.11,

$$
F_{U|s,d,\tilde{t}}(h_{s,d}^{-1}(y;t)) - \frac{\epsilon}{2} - \delta \leq F_{Y^0(t)|s,d}(y) \leq F_{U|s,d,\tilde{t}}(h_{s,d}^{-1}(y;t)) + \frac{\epsilon}{2} + \delta,
\tag{25}
$$

where $\tilde{t} \in \{t_0, t_1\}$. Eq. (25) can also be written as

$$
F_{Y^0(t)|s,d}(y) - \frac{\epsilon}{2} - \delta \leq F_{U|s,d,\tilde{t}}(h_{s,d}^{-1}(y;t)) \leq F_{Y^0(t)|s,d}(y) + \frac{\epsilon}{2} + \delta,
\tag{26}
$$

and consequently,

$$
F_{U|s,d,\tilde{t}}^{-1}\big(F_{Y^0(t)|s,d}(y) - \frac{\epsilon}{2} - \delta\big) \leq h_{s,d}^{-1}(y;t) \leq F_{U|s,d,\tilde{t}}^{-1}\big(F_{Y^0(t)|s,d}(y) + \frac{\epsilon}{2} + \delta\big).
\tag{27}
$$

Also, choosing $y = h_{s,d}(u;t)$ in Eq. (24), and applying $F_{Y^0(t)|s,d}^{-1}(\cdot)$ to its sides, and renaming t to $\tilde{t}$:

$$
F_{Y^0(\tilde{t})|s,d}^{-1}\big(F_{U|s,d,\tilde{t}}(u) - \frac{\epsilon}{2}\big) \leq h_{s,d}(u;\tilde{t}) \leq F_{Y^0(\tilde{t})|s,d}^{-1}\big(F_{U|s,d,\tilde{t}}(u) + \frac{\epsilon}{2}\big)
\tag{28}
$$

Finally, choose $u = h_{s,d}^{-1}(y;t)$ in Eq. (28), and utilize Eq. (27) to get

$$
F_{Y^0(\tilde{t})|s,d}^{-1}\big(F_{Y^0(t)|s,d}(y) - \epsilon - \delta\big) \leq h_{s,d}\big(h_{s,d}^{-1}(y;t);\tilde{t}\big) \leq F_{Y^0(\tilde{t})|s,d}^{-1}\big(F_{Y^0(t)|s,d}(y) + \epsilon + \delta\big).
\tag{29}
$$

Note that we utilized the fact that CDFs and their inverses are increasing.

In 2.1 (see Eq. 6), choose $y = h_{s_1,d_0}\big(h_{s_1,d_0}^{-1}(y';t_0);t_1\big)$, and take inverses which results in

$$
h_{s_1,d_1}\big(h_{s_1,d_1}^{-1}(y;t_1);t_0\big) = h_{s_0,d_1}\bigg(h_{s_0,d_1}^{-1}\Big(h_{s_0,d_0}\big\{h_{s_0,d_0}^{-1}\{h_{s_1,d_0}\big(h_{s_1,d_0}^{-1}(y;t_1);t_0\big);t_0\}; t_1\big\};t_1\Big);t_0\bigg),
\tag{30}
$$

Using the inequalities of (29),

$$
\begin{aligned}
&F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(F^{-1}_{Y^0(t_1)|s_0,d_0}\left\{F_{Y^0(t_0)|s_0,d_0}\left\{F^{-1}_{Y^0(t_0)|s_1,d_0}\left(F_{Y^0(t_1)|s_1,d_0}(y)-\epsilon-\delta\right)\right\}-\epsilon-\delta\right\}\right)-\epsilon-\delta\right)\\
&\leq F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(F^{-1}_{Y^0(t_1)|s_0,d_0}\left\{F_{Y^0(t_0)|s_0,d_0}\left\{h_{s_1,d_0}\left(h^{-1}_{s_1,d_0}(y;t_1);t_0\right)\right\}-\epsilon-\delta\right\}\right)-\epsilon-\delta\right)\\
&\leq F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(h_{s_0,d_0}\left\{h^{-1}_{s_0,d_0}\left\{h_{s_1,d_0}\left(h^{-1}_{s_1,d_0}(y;t_1);t_0\right);t_0\right\};t_1\right\}\right)-\epsilon-\delta\right)\\
&\leq h_{s_1,d_1}\left(h^{-1}_{s_1,d_1}(y;t_1);t_0\right)\\
&\leq F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(h_{s_0,d_0}\left\{h^{-1}_{s_0,d_0}\left\{h_{s_1,d_0}\left(h^{-1}_{s_1,d_0}(y;t_1);t_0\right);t_0\right\};t_1\right\}\right)+\epsilon+\delta\right)\\
&\leq F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(F^{-1}_{Y^0(t_1)|s_0,d_0}\left\{F_{Y^0(t_0)|s_0,d_0}\left\{h_{s_1,d_0}\left(h^{-1}_{s_1,d_0}(y;t_1);t_0\right)\right\}+\epsilon+\delta\right\}\right)+\epsilon+\delta\right)\\
&\leq F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(F^{-1}_{Y^0(t_1)|s_0,d_0}\left\{F_{Y^0(t_0)|s_0,d_0}\left\{F^{-1}_{Y^0(t_0)|s_1,d_0}\left(F_{Y^0(t_1)|s_1,d_0}(y)+\epsilon+\delta\right)\right\}+\epsilon+\delta\right\}\right)+\epsilon+\delta\right),
\end{aligned}
\tag{31}
$$

and finally applying (29) to the term $h_{s_1,d_1}\left(h^{-1}_{s_1,d_1}(y;t_1);t_0\right)$ itself, we get the two inequalities

$$
\begin{aligned}
&F^{-1}_{Y^0(t_0)|s_1,d_1}\left(F_{Y^0(t_1)|s_1,d_1}(y)-\epsilon-\delta\right)\leq\\
&F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(F^{-1}_{Y^0(t_1)|s_0,d_0}\left\{F_{Y^0(t_0)|s_0,d_0}\left\{F^{-1}_{Y^0(t_0)|s_1,d_0}\left(F_{Y^0(t_1)|s_1,d_0}(y)+\epsilon+\delta\right)\right\}+\epsilon+\delta\right\}\right)+\epsilon+\delta\right),
\end{aligned}
\tag{32}
$$

and

$$
\begin{aligned}
&F^{-1}_{Y^0(t_0)|s_0,d_1}\left(F_{Y^0(t_1)|s_0,d_1}\left(F^{-1}_{Y^0(t_1)|s_0,d_0}\left\{F_{Y^0(t_0)|s_0,d_0}\left\{F^{-1}_{Y^0(t_0)|s_1,d_0}\left(F_{Y^0(t_1)|s_1,d_0}(y)-\epsilon-\delta\right)\right\}-\epsilon-\delta\right\}\right)-\epsilon-\delta\right)\\
&\leq F^{-1}_{Y^0(t_0)|s_1,d_1}\left(F_{Y^0(t_1)|s_1,d_1}(y)+\epsilon+\delta\right),
\end{aligned}
\tag{33}
$$

which complete the proof by applying $F_{Y^0(t_0)|s_1,d_1}(\cdot)$ to their sides and applying 1.5.

$\square$

A.1. Partial identification for changes-in-changes

In this section, we present the partial identification result for the conventional framework of changes-in-changes under relaxed versions of monotonicity (1.2), and time invariance (1.3). Serving as a counterpart to Proposition A.3 within our triple changes framework, this result stands on its own merit, given the significance and wide applicability of changes-in-changes. We begin with stating the relaxed assumption.

Assumption A.4 (ϵ-monotonicity). Functions $h^{-1}(\cdot;t)$ are well-defined, and for any u in the support of U, and any $d \in \{d_0, d_1 1\}$,

$$
\mathbb{P}\Big(\langle U-u, h(U;t)-h(u;t)\rangle < 0 \mid D=d, T=t\Big) \leq \frac{\epsilon}{2}.
$$

Assumption A.5 (δ-invariance). Within every subgroup, the Kolmogorov distance of the distribution of the latent variable at $T=t_0$ and $T=t_1$ is bounded by δ. That is,

$$
\sup_u \big|F_{U|D=d,T=t_1}(u) - F_{U|D=d,T=t_0}(u)\big| \leq \delta.
$$

Proposition A.6. *Under assumptions 1.1, A.5, A.4, 1.4, and 1.5, for any y in the support of Y^0,*

$$
F_{Y(t_0)|d_1}\left(F^{-1}_{Y(t_0)|d_0}\left(F_{Y(t_1)|d_0}(y)-\epsilon-\delta\right)\right)-\epsilon-\delta\leq F_{Y^0(t_1)|d_1}(y)\leq F_{Y(t_0)|d_1}\left(F^{-1}_{Y(t_0)|d_0}\left(F_{Y(t_1)|d_0}(y)+\epsilon+\delta\right)\right)+\epsilon+\delta.
$$

Proof. Suppose $t \in \{t_0, t_1\}$.

$$
\begin{aligned}
F_{Y^0(t)|d}(y) &= \mathbb{P}(h(U;t) \le y \mid D = d, T = t) \\
&= \mathbb{P}(h(U;t) \le y,\, U \le h^{-1}(y;t) \mid D = d, T = t) \\
&\quad + \mathbb{P}(h(U;t) \le y,\, U > h^{-1}(y;t) \mid D = d, T = t),
\end{aligned}
\tag{34}
$$

which implies the following inequalities:

$$
\begin{aligned}
F_{Y^0(t)|d}(y) &\overset{(a)}{\le} \mathbb{P}(U \le h^{-1}(y;t) \mid D = d, T = t) + \mathbb{P}(h(U;t) \le y,\, U > h^{-1}(y;t) \mid D = d, T = t) \\
&\overset{(b)}{\le} \mathbb{P}(U \le h^{-1}(y;t) \mid D = d, T = t) + \frac{\epsilon}{2} \\
&= F_{U|d,t}(h^{-1}(y;t)) + \frac{\epsilon}{2},
\end{aligned}
\tag{35}
$$

where (a) is a basic probability manipulation and (b) follows from A.4, and

$$
\begin{aligned}
F_{Y^0(t)|d}(y) &\overset{(c)}{\ge} \mathbb{P}(h(U;t) \le y,\, U \le h^{-1}(y;t) \mid D = d, T = t) \\
&= \mathbb{P}(U \le h^{-1}(y;t) \mid D = d, T = t) - \mathbb{P}(h(U;t) > y,\, U \le h^{-1}(y;t) \mid D = d, T = t) \\
&\overset{(d)}{\ge} \mathbb{P}(U \le h^{-1}(y;t) \mid D = d, T = t) - \frac{\epsilon}{2} \\
&= F_{U|d,t}(h^{-1}(y;t)) - \frac{\epsilon}{2},
\end{aligned}
\tag{36}
$$

where (c) is a consequence of (34), and (d) follows from A.4. To summarize, we got

$$
F_{U|d,t}(h^{-1}(y;t)) - \frac{\epsilon}{2} \le F_{Y^0(t)|d}(y) \le F_{U|d,t}(h^{-1}(y;t)) + \frac{\epsilon}{2},
\tag{37}
$$

and using A.5,

$$
F_{U|d,\tilde{t}}(h^{-1}(y;t)) - \frac{\epsilon}{2} - \delta \le F_{Y^0(t)|d}(y) \le F_{U|d,\tilde{t}}(h^{-1}(y;t)) + \frac{\epsilon}{2} + \delta,
\tag{38}
$$

where $\tilde{t} \in \{t_0, t_1\}$. Eq. (38) can also be written as

$$
F_{Y^0(t)|d}(y) - \frac{\epsilon}{2} - \delta \le F_{U|d,\tilde{t}}(h^{-1}(y;t)) \le F_{Y^0(t)|d}(y) + \frac{\epsilon}{2} + \delta,
\tag{39}
$$

and consequently,

$$
F_{U|d,\tilde{t}}^{-1}\big(F_{Y^0(t)|d}(y) - \frac{\epsilon}{2} - \delta\big) \le h^{-1}(y;t) \le F_{U|d,\tilde{t}}^{-1}\big(F_{Y^0(t)|d}(y) + \frac{\epsilon}{2} + \delta\big).
\tag{40}
$$

Also, choosing $y = h(u;t)$ in Eq. (37), and applying $F_{Y^0(t)|d}^{-1}(\cdot)$ to its sides, and renaming t to $\tilde{t}$:

$$
F_{Y^0(\tilde{t})|d}^{-1}\big(F_{U|d,\tilde{t}}(u) - \frac{\epsilon}{2}\big) \le h(u;\tilde{t}) \le F_{Y^0(\tilde{t})|d}^{-1}\big(F_{U|d,\tilde{t}}(u) + \frac{\epsilon}{2}\big)
\tag{41}
$$

Finally, choose $u = h^{-1}(y;t)$ in (41) and use (40), choosing $\tilde{t} = t_0, t = t_1$ to get

$$
F_{Y^0(t_0)|d}^{-1}\big(F_{Y^0(t_1)|d}(y) - \epsilon - \delta\big) \le h\big(h^{-1}(y;t_1);t_0\big) \le F_{Y^0(t_0)|d}^{-1}\big(F_{Y^0(t_1)|d}(y) + \epsilon + \delta\big).
\tag{42}
$$

Note that we utilized the fact that CDFs and their inverses are increasing. Note that Eq. (42) holds both for $d = d_0, d_1$. Combining Equations (42) for $d = d_0$ and $d = d_1$ yields the following two inequalities, concluding the proof after applying $F_{Y^0(t_0)|d_1}$ to both sides.

$$
F_{Y^0(t_0)|d_0}^{-1}\big(F_{Y^0(t_1)|d_0}(y) - \epsilon - \delta\big) \le F_{Y^0(t_0)|d_1}^{-1}\big(F_{Y^0(t_1)|d_1}(y) + \epsilon + \delta\big),
$$

and,

$$
F_{Y^0(t_0)|d_1}^{-1}\big(F_{Y^0(t_1)|d_1}(y) - \epsilon - \delta\big) \le F_{Y^0(t_0)|d_0}^{-1}\big(F_{Y^0(t_1)|d_0}(y) + \epsilon + \delta\big).
$$

The result follows by 1.5. $\qquad\square$

B. Asymptotic Normality of $\hat{\tau}$

Due to space limitations, we discuss the details of our estimator and its asymptotic properties in this appendix. For simplicity of notation, we adopt F_{sdt} to denote $F_{Y(t)|s,d}$ throughout. We use the same convention for $F_{Y(t)|s,d}^{-1}$ and their corresponding estimators. In particular, the empirical estimators were defined as

$$\hat{F}_{sdt}(y) := N_{sdt}^{-1} \sum_{i=1}^{N_{sdt}} \mathbb{1}\{Y_{sd,i}(t) \leq y\}, \tag{43}$$

and

$$\hat{F}_{sdt}^{-1}(u) := \inf\{y \in \mathbb{Y}_t, \hat{F}_{sdt}(y) \geq u\}. \tag{44}$$

Our estimator for τ was also built as

$$\hat{\tau} := N_{s_1 d_1 t_1}^{-1} \sum_{i=1}^{N_{s_1 d_1 t_1}} Y_{s_1 d_1, i}(t_1) - N_{s_1 d_1 t_0}^{-1} \sum_{i=1}^{N_{s_1 d_1 t_0}} \hat{F}_{s_0 d_1 t_1}^{-1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0}(Y_{s_1 d_1, i}(t_0)). \tag{45}$$

For ease of notation, in what follows, we adopt the notation $\mathbb{E}_{sd}$ to denote expectation conditioned on $S = s, D = d$. We define the following functions:

$$\begin{aligned} g_2(\cdot) &:= F_{s_0 d_1 t_1}^{-1}(\cdot), \quad g_3(\cdot) := g_2 \circ F_{s_0 d_1 t_0}(\cdot), \quad g_4(\cdot) := g_3 \circ F_{s_0 d_0 t_0}^{-1}(\cdot), \\ g_5(\cdot) &:= g_4 \circ F_{s_0 d_0 t_1}(\cdot), \quad g_6(\cdot) := g_5 \circ F_{s_1 d_0 t_1}^{-1}(\cdot), \quad g_7(\cdot) := g_6 \circ F_{s_1 d_0 t_0}(\cdot), \end{aligned} \tag{46}$$

as well as

$$\begin{aligned} r_5(\cdot) &:= F_{s_1 d_0 t_0}(\cdot), \quad r_4(\cdot) := F_{s_1 d_0 t_1}^{-1} \circ r_5(\cdot), \quad r_3(\cdot) := F_{s_0 d_0 t_1} \circ r_4(\cdot), \\ r_2(\cdot) &:= F_{s_0 d_0 t_0}^{-1} \circ r_3(\cdot), \quad r_1(\cdot) := F_{s_0 d_1 t_0} \circ r_2(\cdot). \end{aligned} \tag{47}$$

Accordingly, define $Q_0(y) = y - \mathbb{E}_{s_1 d_1}[Y(t_1)]$, and

$$\begin{aligned} q_1(y, z) &:= \frac{\mathbb{1}\{F_{s_0 d_1 t_1}(y) \leq r_1(z)\} - r_1(z)}{f_{s_0 d_1 t_1} \circ F_{s_0 d_1 t_1}^{-1} \circ r_1(z)}, \quad Q_1(y) := \mathbb{E}_{s_1 d_1}[q_1(y, Y(t_0))] \\ q_2(y, z) &:= (\mathbb{1}\{y \leq r_2(z)\} - F_{s_0 d_1 t_0} \circ r_2(z)) \cdot (g_2' \circ F_{s_0 d_1 t_0} \circ r_2(z)), \quad Q_2(y) := \mathbb{E}_{s_1 d_1}[q_2(y, Y(t_0))] \\ q_3(y, z) &:= \frac{\mathbb{1}\{F_{s_0 d_0 t_0}(y) \leq r_3(z)\} - r_3(z)}{f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ r_3(z)} \cdot (g_3' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ r_3(z)), \quad Q_3(y) := \mathbb{E}_{s_1 d_1}[q_3(y, Y(t_0))] \\ q_4(y, z) &:= (\mathbb{1}\{y \leq r_4(z)\} - F_{s_0 d_0 t_1} \circ r_4(z)) \cdot (g_4' \circ F_{s_0 d_0 t_1} \circ r_4(z)), \quad Q_4(y) := \mathbb{E}_{s_1 d_1}[q_4(y, Y(t_0))] \\ q_5(y, z) &:= \frac{\mathbb{1}\{F_{s_1 d_0 t_1}(y) \leq r_5(z)\} - r_5(z)}{f_{s_1 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ r_5(z)} \cdot (g_5' \circ f_{s_1 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ r_5(z)) \quad Q_5(y) := \mathbb{E}_{s_1 d_1}[q_5(y, Y(t_0))], \\ q_6(y, z) &:= (\mathbb{1}\{y \leq z\} - F_{s_1 d_0 t_0}(z)) \cdot (g_6' \circ F_{s_1 d_0 t_0}(z)), \quad Q_6(y) := \mathbb{E}_{s_1 d_1}[q_6(y, Y(t_0))], \\ Q_7(y) &:= g_7(y) - \mathbb{E}_{s_1 d_1}[g_7(Y(t_0))]. \end{aligned} \tag{48}$$

Also define

$$\hat{\mu}_0 := \frac{1}{N_{s_1 d_1 t_1}} \sum_{i=1}^{N_{s_1 d_1 t_1}} Q_0(Y_{s_1 d_1, i}(t_1)), \tag{49}$$

and,

$$\hat{\mu}_1 := \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_1 t_1}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_1 t_1}} q_1(Y_{s_0 d_1 t, j}(t_1), Y_{s_1 d_1, i}(t_0)), \qquad \hat{\mu}_1 := \frac{1}{N_{s_0 d_1 t_1}} \sum_{j=1}^{N_{s_0 d_1 t_1}} Q_1(Y_{s_0 d_1 t, j}(t_1)), \tag{50}$$

$$\hat{\mu}_2 = \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_1 t_0}} q_2(Y_{s_0 d_1 t, j}(t_0), Y_{s_1 d_1, i}(t_0)), \qquad \hat{\mu}_2 = \frac{1}{N_{s_0 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_1 t_0}} Q_2(Y_{s_0 d_1 t, j}(t_0)), \quad (51)$$

$$\hat{\mu}_3 = \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_0 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} q_3(Y_{s_0 d_0 t, j}(t_0), Y_{s_1 d_1, i}(t_0)), \qquad \hat{\mu}_3 = \frac{1}{N_{s_0 d_0 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} Q_3(Y_{s_0 d_0 t, j}(t_0)), \quad (52)$$

$$\hat{\mu}_4 = \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_0 t_1}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_1}} q_4(Y_{s_0 d_0 t, j}(t_1), Y_{s_1 d_1, i}(t_0)), \qquad \hat{\mu}_4 = \frac{1}{N_{s_0 d_0 t_1}} \sum_{j=1}^{N_{s_0 d_0 t_1}} Q_4(Y_{s_0 d_0 t, j}(t_1)), \quad (53)$$

$$\hat{\mu}_5 = \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_1 d_0 t_1}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_1 d_0 t_1}} q_5(Y_{s_0 d_1 t, j}(t_1), Y_{s_1 d_1, i}(t_0)), \qquad \hat{\mu}_5 = \frac{1}{N_{s_1 d_0 t_1}} \sum_{j=1}^{N_{s_1 d_0 t_1}} Q_5(Y_{s_0 d_1 t, j}(t_1)), \quad (54)$$

$$\hat{\mu}_6 = \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_1 d_0 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_1 d_0 t_0}} q_6(Y_{s_0 d_1 t, j}(t_0), Y_{s_1 d_1, i}(t_0)), \qquad \hat{\mu}_6 = \frac{1}{N_{s_1 d_0 t_0}} \sum_{j=1}^{N_{s_1 d_0 t_0}} Q_6(Y_{s_0 d_1 t, j}(t_0)), \quad (55)$$

and,

$$\hat{\mu}_7 := \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} Q_7\big(Y_{s_1 d_1, i}(t_0)\big). \tag{56}$$

Finally, define the variance terms $\{V_0, \dots, V_7\}$ as

$$\begin{aligned}
V_0 &:= \mathbb{E}_{s_1 d_1}[Q_0\big(Y(t_1)\big)^2], \\
V_1 &:= \mathbb{E}_{s_0 d_1}[Q_1\big(Y(t_0)\big)^2], \\
V_2 &:= \mathbb{E}_{s_0 d_1}[Q_2\big(Y(t_1)\big)^2], \\
V_3 &:= \mathbb{E}_{s_0 d_0}[Q_3\big(Y(t_1)\big)^2], \\
V_4 &:= \mathbb{E}_{s_0 d_0}[Q_4\big(Y(t_0)\big)^2], \\
V_5 &:= \mathbb{E}_{s_0 d_1}[Q_5\big(Y(t_0)\big)^2], \\
V_6 &:= \mathbb{E}_{s_0 d_1}[Q_6\big(Y(t_1)\big)^2], \\
V_7 &:= \mathbb{E}_{s_1 d_1}[Q_7\big(Y(t_0)\big)^2].
\end{aligned} \tag{57}$$

For the sake of completeness, we review the consistency and asymptotic normality result from the main text for $\hat{\tau}$:

Theorem 3.2. *Under 3.1, $\hat{\tau} - \tau = \mathcal{O}_p(N^{-\frac{1}{2}})$, and*

$$\sqrt{N}(\hat{\tau} - \tau) \overset{D}{\to} \mathcal{N}(0, \frac{V_0}{p_{s_1 d_1 t_1}} + \frac{V_1}{p_{s_0 d_1 t_1}} + \frac{V_2}{p_{s_0 d_1 t_0}} +$$

$$\frac{V_3}{p_{s_0 d_0 t_0}} + \frac{V_4}{p_{s_0 d_0 t_1}} + \frac{V_5}{p_{s_1 d_0 t_1}} + \frac{V_6}{p_{s_1 d_0 t_0}} + \frac{V_7}{p_{s_1 d_1 t_0}}), \quad (14)$$

where $\{V_i\}_{i=0}^{7}$ are given by Eq. (57).

In order to estimate the asymptotic variance, one can substitute expectations with sample averages. This involves utilizing empirical distributions and their inverses for cumulative distribution functions, as well as employing any uniformly consistent non-parametric estimator for the density functions.

Twelve lemmas are in order within the next section which will be used to prove Theorem 3.2 in Section D.

C. Preliminary Lemmas

Throughout this subsection, we suppress the subscripts sdt and replace N_{sdt} by N whenever it does not affect the statements. Note that since $N_{sdt}/N \to p_{sdt} > 0$ (see 3.1), if a term is $O_p(N_{sdt}^{-\alpha})$ $(o_p(N_{sdt}^{-\alpha}))$, it is also $O_p(N^{-\alpha})$ $(o_p(N^{-\alpha}))$. For the following four lemmas, let X be a random variable with twice continuously differentiable cdf $F(\cdot)$, where F has bounded first and second derivatives. Suppose the pdf of X is bounded away from 0. Let $\hat{F}$ and $\hat{F}^{-1}$ be empirical estimators of the CDF of X and its inverse, analogous to Eq. (43) and Eq. (44).

Lemma C.1 (Athey and Imbens, 2006). *For any $\delta < 1/2$,*

$$\sup_x N^\delta \cdot \left| \hat{F}(x) - F(x) \right| \xrightarrow{P} 0, \qquad \sup_{u \in [0,1]} N^\delta \cdot \left| \hat{F}^{-1}(u) - F^{-1}(u) \right| \xrightarrow{P} 0.$$

The following lemma is a slight generalization of the latter.

Lemma C.2. *Let $g(\cdot)$ be a continuously differentiable function with a bounded derivative. For any $\delta < \frac{1}{2}$,*

$$\sup_x N^\delta \cdot \left| g \circ \hat{F}(x) - g \circ F(x) \right| \xrightarrow{P} 0, \qquad \sup_{u \in [0,1]} N^\delta \cdot \left| g \circ \hat{F}^{-1}(u) - g \circ F^{-1}(u) \right| \xrightarrow{P} 0.$$

Proof.

$$\sup_x N^\delta \cdot \left| g \circ \hat{F}(x) - g \circ F(x) \right| \leq \sup_{x,\tilde{x}} N^\delta \cdot \left| g'(\tilde{x}) \right| \cdot \left| \hat{F}(x) - F(x) \right| \leq \sup_{x,\tilde{x}} \left| g'(\tilde{x}) \right| \cdot \sup_x N^\delta \cdot \left| \hat{F}(x) - F(x) \right|,$$

and the result follows from Lemma C.1. The proof for the second part is identical. $\square$

Lemma C.3. *Let $\{X_i\}_{i=1}^k$ be random variables with twice differentiable cdfs $\{F_i(\cdot)\}_{i=1}^k$, where F_is have bounded first and second derivatives. Suppose also that all X_is have pdfs that are bounded away from 0. Then for any $\delta < \frac{1}{2}$,*

$$\sup_x N^\delta \cdot \left| \hat{g}_1 \circ \cdots \circ \hat{g}_k(x) - g_1 \circ \cdots \circ g_k(x) \right| \xrightarrow{P} 0,$$

where $g_i \in \{F_i, F_i^{-1}\}$, and $\hat{g}_i$ is the empirical estimator of g_i, analogous to Equations (43) and (44).

Proof. Using triangle inequality,

$$\sup_x N^\delta \cdot \left| \hat{g}_1 \circ \cdots \circ \hat{g}_k(x) - g_1 \circ \cdots \circ g_k(x) \right| \leq$$

$$\sup_x N^\delta \cdot \left| \hat{g}_1 \circ \cdots \circ \hat{g}_k(x) - g_1 \circ \hat{g}_2 \circ \cdots \circ \hat{g}_k(x) \right|$$

$$+ \sup_x N^\delta \cdot \left| g_1 \circ \hat{g}_2 \circ \cdots \circ \hat{g}_k(x) - g_1 \circ g_2(x) \circ \hat{g}_3 \circ \cdots \circ \hat{g}_k(x) \right|$$

$$+ \ldots$$

$$+ \sup_x N^\delta \cdot \left| g_1 \circ \cdots \circ g_{k-1} \circ \hat{g}_k(x) - g_1 \circ \cdots \circ g_k(x) \right|,$$

and every term on the right hand side is $o_p(1)$ due to Lemma C.2. $\square$

Lemma C.4. *Let $g(\cdot)$ be a twice continuously differentiable function, with bounded first and second derivatives $g'(\cdot)$ and $g''(\cdot)$, respectively. Then for all $0 < \eta < 5/7$,*

$$\sup_{u \in [0,1]} N^\eta \cdot \left| g \circ \hat{F}^{-1}(u) - g \circ F^{-1}(u) + \frac{g' \circ f \circ F^{-1}(u)}{f \circ F^{-1}(u)} \left(\hat{F} \circ F^{-1}(u) - u \right) \right| \xrightarrow{P} 0. \tag{58}$$

17

Proof. Using triangle inequality,

$$\sup_{u\in[0,1]} N^{\eta} \cdot \left| g \circ \hat{F}^{-1}(u) - g \circ F^{-1}(u) + \frac{g' \circ f \circ F^{-1}(u)}{f \circ F^{-1}(u)} \big(\hat{F} \circ F^{-1}(u) - u \big) \right|$$

$$\leq \sup_{u\in[0,1]} N^{\eta} \cdot \left| g \circ \hat{F}^{-1}(u) - g \circ F^{-1} \circ \hat{F} \circ \hat{F}^{-1}(u) + \frac{g' \circ f \circ \hat{F}^{-1}(u)}{f \circ \hat{F}^{-1}(u)} \big(\hat{F} \circ \hat{F}^{-1}(u) - F \circ \hat{F}^{-1}(u) \big) \right|$$

$$+ \sup_{u\in[0,1]} N^{\eta} \cdot \left| \frac{g' \circ f \circ F^{-1}(u)}{f \circ F^{-1}(u)} \big(\hat{F} \circ F^{-1}(u) - u \big) - \frac{g' \circ f \circ \hat{F}^{-1}(u)}{f \circ \hat{F}^{-1}(u)} \big(\hat{F} \circ \hat{F}^{-1}(u) - F \circ \hat{F}^{-1}(u) \big) \right|$$

$$+ \sup_{u\in[0,1]} N^{\eta} \cdot \left| g \circ F^{-1} \circ \hat{F} \circ \hat{F}^{-1}(u) - g \circ F^{-1}(u) \right|, \quad (59)$$

where we shall show each of the three terms on the right hand side are $o_p(1)$. The first term is bounded as

$$\sup_{x} N^{\eta} \cdot \left| g(x) - g \circ F^{-1} \circ \hat{F}(x) + \frac{g' \circ f(x)}{f(x)} \big(\hat{F}(x) - F(x) \big) \right|$$

$$= \sup_{x} N^{\eta} \cdot \left| g \circ F^{-1} \big(F(x) \big) - g \circ F^{-1} \big(\hat{F}(x) \big) - \frac{g' \circ f(x)}{f \circ F^{-1} \big(F(x) \big)} \big(F(x) - \hat{F}(x) \big) \right|$$

$$\leq \sup_{\tilde{x},x} N^{\eta} \cdot \left| \Big(- \frac{1}{f(\tilde{x})^3} \frac{\partial f}{\partial x}(\tilde{x}) g' \circ F^{-1}(\tilde{x}) + \frac{1}{f(\tilde{x})^2} g'' \circ F^{-1}(\tilde{x}) \Big) \big(F(x) - \hat{F}(x) \big)^2 \right|, \quad (60)$$

where we used the expansion of the function $g \circ F^{-1}(\cdot)$ around the point $F(x)$. The latter is $o_p(1)$ for every $\eta < 1$ due to Lemma C.1 and smoothness assumptions over g and f. Next, the second term of Eq. (59) can be bounded using triangle inequality by

$$\sup_{u\in[0,1]} N^{\eta} \cdot \left| \frac{g' \circ f \circ F^{-1}(u)}{f \circ F^{-1}(u)} \big(\hat{F} \circ F^{-1}(u) - u \big) - \frac{g' \circ f \circ \hat{F}^{-1}(u)}{f \circ \hat{F}^{-1}(u)} \big(\hat{F} \circ \hat{F}^{-1}(u) - F \circ \hat{F}^{-1}(u) \big) \right|$$

$$\leq \sup_{u\in[0,1]} N^{\eta} \cdot \left| \frac{g' \circ f \circ F^{-1}(u)}{f \circ F^{-1}(u)} \big(\hat{F} \circ F^{-1}(u) - u \big) - \frac{g' \circ f \circ \hat{F}^{-1}(u)}{f \circ \hat{F}^{-1}(u)} \big(\hat{F} \circ F^{-1}(u) - u \big) \right|$$

$$+ \sup_{u\in[0,1]} N^{\eta} \cdot \left| \frac{g' \circ f \circ \hat{F}^{-1}(u)}{f \circ \hat{F}^{-1}(u)} \big(\hat{F} \circ F^{-1}(u) - u \big) - \frac{g' \circ f \circ \hat{F}^{-1}(u)}{f \circ \hat{F}^{-1}(u)} \big(\hat{F} \circ \hat{F}^{-1}(u) - F \circ \hat{F}^{-1}(u) \big) \right|$$

$$\leq \sup_{u\in[0,1]} N^{\eta/2} \cdot \left| \hat{F} \circ F^{-1}(u) - u \right| \cdot \sup_{u\in[0,1]} N^{\eta/2} \cdot \left| \frac{g' \circ f \circ F^{-1}(u)}{f \circ F^{-1}(u)} - \frac{g' \circ f \circ \hat{F}^{-1}(u)}{f \circ \hat{F}^{-1}(u)} \right|$$

$$+ \sup_{x} \left| \frac{g' \circ f(x)}{f(x)} \right| \cdot \sup_{u\in[0,1]} N^{\eta} \left| \hat{F} \circ F^{-1}(u) - u - \big(\hat{F} \circ \hat{F}^{-1}(u) - F \circ \hat{F}^{-1}(u) \big) \right|, \quad (61)$$

where the first term is $o_p(1)$ due to Lemma C.1, Lemma A.3 of Athey and Imbens (2006), and the smoothness assumptions over f and g, and the second term is $o_p(1)$ as shown in the proof of Lemma A.6 of Athey and Imbens (2006). Finally, consider the third term of Eq. (59):

$$\sup_{u\in[0,1]} N^{\eta} \cdot \left| g \circ F^{-1} \circ \hat{F} \circ \hat{F}^{-1}(u) - g \circ F^{-1}(u) \right|. \quad (62)$$

Since $|\hat{F} \circ \hat{F}^{-1}(u) - u| < 1/N$ for all u (Athey and Imbens, 2006), and due to smoothness assumptions over $g \circ F^{-1}$, this term converges uniformly to zero. $\quad\square$

Lemma C.5. *Let $\{\hat{\mu}_k\}_{k=1}^6$ and $\{\hat{\mu}_k\}_{k=1}^6$ be given by Equations (50) through (55). Under 3.1, for any $k \in \{1,\ldots,6\}$,*

$$\sqrt{N}(\hat{\mu}_k - \hat{\mu}_k) \xrightarrow{P} 0.$$

Proof. We give the proof for $k = 1$. The proof is identical for the rest. $\hat{\mu}_1$ is a twp-sample V-statistic, and standard V-statistic theory implies

$$\hat{\mu}_1 = \frac{1}{N_{s_0 d_1 t_1}} \sum_{i=1}^{N_{s_0 d_1 t_1}} \mathbb{E}_{s_1 d_1}\left[q_1\left(Y_{s_0 d_1, i}(t_1), Y(t_0)\right)\right] + \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \mathbb{E}_{s_0 d_1}\left[q_1\left(Y(t_0), Y_{s_1 d_1, i}(t_0)\right)\right] + o_p(N^{-1/2})$$

$$= \frac{1}{N_{s_0 d_1 t_1}} \sum_{i=1}^{N_{s_0 d_1 t_0}} Q_1\left(Y_{s_0 d_1, i}(t_1)\right) + 0 + o_p(N^{-1/2}) = \hat{\mu}_1 + o_p(N^{-1/2}),$$

which completes the proof. $\qquad\square$

Lemma C.6. *Define $\hat{\mu}_3$ as given by Eq. (52). Under 3.1,*

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0}\left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0}\left(Y_{s_1 d_1, i}(t_0)\right) + \hat{\mu}_3 \right| \xrightarrow{P} 0. \quad (63)$$

Proof. For ease of notation, define $\hat{q}_i := \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0}\left(Y_{s_1 d_1, i}(t_0)\right)$, and $q_i := F_{s_0 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ F_{s_1 d_0 t_0}\left(Y_{s_1 d_1, i}(t_0)\right)$. Also, define $g(\cdot) := F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0}(\cdot)$. Using triangle inequality, Eq. (63) can be bounded as

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ \hat{F}_{s_0 d_0 t_0}^{-1}(\hat{q}_i) - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i) + \hat{\mu}_3 \right| \le$$

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ \hat{F}_{s_0 d_0 t_0}^{-1}(\hat{q}_i) - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i) \right.$$

$$\left. + \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_0 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} \frac{\mathbb{1}\{F_{s_0 d_0 t_0}\left(Y_{s_0 d_0, j}(t_0)\right) \le \hat{q}_i\} - \hat{q}_i}{f_{s_0 d_0 t_0}\left(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)\right)} \left(g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)\right) \right|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_0 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} \frac{\mathbb{1}\{F_{s_0 d_0 t_0}\left(Y_{s_0 d_0, j}(t_0)\right) \le \hat{q}_i\} - \hat{q}_i}{f_{s_0 d_0 t_0}\left(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)\right)} \left(g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)\right) - \hat{\mu}_3 \right| \quad (64)$$

The first term on the right hand side of (64) can be bounded by

$$\sqrt{N} \cdot \frac{1}{N_{s_1 d_1 t_0}} \cdot \sum_{i=1}^{N_{s_1 d_1 t_0}} \left| g \circ \hat{F}_{s_0 d_0 t_0}^{-1}(\hat{q}_i) - g \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i) \right.$$

$$\left. + \frac{1}{N_{s_0 d_0 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} \frac{\mathbb{1}\{F_{s_0 d_0 t_0}\left(Y_{s_0 d_0, j}(t_0)\right) \le \hat{q}_i\} - \hat{q}_i}{f_{s_0 d_0 t_0}\left(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)\right)} \left(g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)\right) \right| \le$$

$$\sqrt{N} \cdot \sup_u \left| g \circ \hat{F}_{s_0 d_0 t_0}^{-1}(u) - g \circ F_{s_0 d_0 t_0}^{-1}(u) \right.$$

$$\left. + \frac{1}{N_{s_0 d_0 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} \frac{\mathbb{1}\{F_{s_0 d_0 t_0}\left(Y_{s_0 d_0, j}(t_0)\right) \le u\} - u}{f_{s_0 d_0 t_0}\left(F_{s_0 d_0 t_0}^{-1}(u)\right)} \left(g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(u)\right) \right|$$

$$= \sqrt{N} \cdot \sup_u \left| g \circ \hat{F}_{s_0 d_0 t_0}^{-1}(u) - g \circ F_{s_0 d_0 t_0}^{-1}(u) + \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(u) - u}{f_{s_0 d_0 t_0}\left(F_{s_0 d_0 t_0}^{-1}(u)\right)} \left(g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(u)\right) \right| \quad (65)$$

Lemma C.4 implies that the latter is $o_p(1)$.

The second term on the right hand side of (64) can also be rephrased as follows.

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_0 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} \frac{\mathbb{1}\{F_{s_0 d_0 t_0}(Y_{s_0 d_0,j}(t_0)) \leq \hat{q}_i\} - \hat{q}_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)) - \hat{\mu}_3 \right|$$

$$= \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_0 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} \frac{\mathbb{1}\{F_{s_0 d_0 t_0}(Y_{s_0 d_0,j}(t_0)) \leq \hat{q}_i\} - \hat{q}_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_0 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_0 t_0}} \frac{\mathbb{1}\{F_{s_0 d_0 t_0}(Y_{s_0 d_0,j}(t_0)) \leq q_i\} - q_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(q_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i)) \right|$$

$$= \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i) - \hat{q}_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i) - q_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(q_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i)) \right|, \quad (66)$$

which can be bounded further by

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i) - \hat{q}_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i) - q_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)) \right|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i) - q_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i)) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i) - q_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(q_i))} (g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i)) \right| \quad (67)$$

The first term of Eq. (67) is bounded by

$$\sqrt{N} \cdot \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \left| \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i) - \hat{q}_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} - \frac{\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(q_i) - q_i}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(\hat{q}_i))} \right| \cdot \left| g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(\hat{q}_i) \right| \leq$$

$$\sqrt{N} \cdot \sup_u \left| \frac{g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(u)}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(u))} \right| \cdot \sup_y \left| \hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_1 t_0}^{-1} \circ \hat{F}_{s_0 d_1 t_1}(y) - \hat{F}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_1 t_0}^{-1} \circ \hat{F}_{s_0 d_1 t_1}(y) \right.$$

$$\left. - \left(\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_0} \circ F_{s_0 d_1 t_0}^{-1} \circ F_{s_0 d_1 t_1}(y) - F_{s_0 d_0 t_0} \circ F_{s_0 d_1 t_0}^{-1} \circ F_{s_0 d_1 t_1}(y) \right) \right| =$$

$$\sqrt{N} \cdot \sup_u \left| \frac{g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(u)}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(u))} \right| \cdot \sup_y \left| \hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_1 t_0}^{-1} \circ \hat{F}_{s_0 d_1 t_1}(y) - \right.$$

$$\hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_0} \circ F_{s_0 d_1 t_0}^{-1} \circ F_{s_0 d_1 t_1}(y)$$

$$\left. - \left(F_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_1 t_0}^{-1} \circ \hat{F}_{s_0 d_1 t_1}(y) - F_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_0} \circ F_{s_0 d_1 t_0}^{-1} \circ F_{s_0 d_1 t_1}(y) \right) \right| =$$

$$\sqrt{N} \cdot \sup_u \left| \frac{g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(u)}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(u))} \right| \cdot \sup_y \left| \hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{m}(y) - \hat{F}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ m(y) \right.$$

$$\left. - \left(F_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{m}(y) - F_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ m(y) \right) \right|, \quad (68)$$

where $m(\cdot) = F_{s_0 d_0 t_1} \circ F_{s_1 d_1 t_1}^{-1} \circ F_{s_1 d_0 t_0}(\cdot)$, and $\hat{m}(\cdot) = \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_1 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0}(\cdot)$. The term $\sup_u \left| \frac{g' \circ f_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_0}^{-1}(u)}{f_{s_0 d_0 t_0}(F_{s_0 d_0 t_0}^{-1}(u))} \right|$ is bounded due to 3.1. Choose $\delta = 1/3$. Due to Lemma C.3, $\hat{m}(y) - m(y) = o_p(N^{-\delta})$ and $F_{s_0 d_0 t_0}^{-1}$ has the required

smoothness properties due to 3.1, $F_{s_0d_0t_0}^{-1} \circ \hat{m}(y) - F_{s_0d_0t_0}^{-1} \circ m(y) = o_p(N^{-\delta})$. Therefore, Lemma A.5 of Athey and Imbens (2006) with the choice of $\eta = 1/2$ implies that the right hand side of Eq. (68) is $o_p(1)$.

Finally, the second term of Eq. (67) is upper-bounded by

$$
\sqrt{N} \cdot \frac{1}{N_{s_1d_1t_0}} \sum_{i=1}^{N_{s_1d_1t_0}} \left| \hat{F}_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1}(q_i) - q_i \right| \cdot \left| \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1}(\hat{q}_i)}{f_{s_0d_0t_0}\left(F_{s_0d_0t_0}^{-1}(\hat{q}_i)\right)} - \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1}(q_i)}{f_{s_0d_0t_0}\left(F_{s_0d_0t_0}^{-1}(q_i)\right)} \right|
$$

$$
\leq \left(N^{1/4} \sup_y \left| \hat{F}_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1}\left(m(y)\right) - m(y) \right| \right)
$$

$$
\times \left(N^{1/4} \sup_y \left| \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ \hat{m}(y)}{f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ \hat{m}(y)} - \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ m(y)}{f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ m(y)} \right| \right)
$$

$$
= \left(N^{1/4} \sup_y \left| \hat{F}_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1}\left(m(y)\right) - F_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1}\left(m(y)\right) \right| \right)
$$

$$
\times \left(N^{1/4} \sup_y \left| \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ \hat{m}(y)}{f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ \hat{m}(y)} - \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ m(y)}{f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ m(y)} \right| \right)
$$

$$
= \left(N^{1/4} \sup_y \left| \hat{F}_{s_0d_0t_0}(y) - F_{s_0d_0t_0}(y) \right| \right)
$$

$$
\times \left(N^{1/4} \sup_y \left| \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ \hat{m}(y)}{f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ \hat{m}(y)} - \frac{g' \circ f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ m(y)}{f_{s_0d_0t_0} \circ F_{s_0d_0t_0}^{-1} \circ m(y)} \right| \right), \quad (69)
$$

where the first term is $o_p(1)$ due to Lemma C.1, and the second term is $o_p(1)$ due to Lemma C.3 and Assumption 3.1. $\square$

Lemma C.7. *Define $\hat{\mu}_1$ as given by Eq.* (50). *Under 3.1,*

$$
\sqrt{N} \cdot \left| \frac{1}{N_{s_1d_1t_0}} \sum_{i=1}^{N_{s_1d_1t_0}} \hat{F}_{s_0d_1t_1}^{-1} \circ \hat{F}_{s_0d_1t_0} \circ \hat{F}_{s_0d_0t_0}^{-1} \circ \hat{F}_{s_0d_0t_1} \circ \hat{F}_{s_1d_0t_1}^{-1} \circ \hat{F}_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right) \right.
$$

$$
\left. - \frac{1}{N_{s_1d_1t_0}} \sum_{i=1}^{N_{s_1d_1t_0}} F_{s_0d_1t_1}^{-1} \circ \hat{F}_{s_0d_1t_0} \circ \hat{F}_{s_0d_0t_0}^{-1} \circ \hat{F}_{s_0d_0t_1} \circ \hat{F}_{s_1d_0t_1}^{-1} \circ \hat{F}_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right) + \hat{\mu}_1 \right| \xrightarrow{P} 0. \quad (70)
$$

Proof. The proof is analogous to that of Lemma C.6. To adapt the proof, we just need to replace the following definitions: Define $\hat{q}_i \coloneqq \hat{F}_{s_0d_1t_0}^{-1} \circ \hat{F}_{s_0d_0t_0}^{-1} \circ \hat{F}_{s_0d_0t_1} \circ \hat{F}_{s_1d_0t_1}^{-1} \circ \hat{F}_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right)$, $q_i \coloneqq F_{s_0d_1t_0}^{-1} \circ F_{s_0d_0t_0}^{-1} \circ F_{s_0d_0t_1} \circ F_{s_1d_0t_1}^{-1} \circ F_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right)$, and $g(\cdot)$ is the identity function. Finally, the subscripts $s_0d_0t_0$ are replaced by $s_0d_1t_1$. The rest of the proof is identical. $\square$

Lemma C.8. *Define $\hat{\mu}_5$ as given by Eq.* (54). *Under 3.1,*

$$
\sqrt{N} \cdot \left| \frac{1}{N_{s_1d_1t_0}} \sum_{i=1}^{N_{s_1d_1t_0}} F_{s_0d_1t_1}^{-1} \circ F_{s_0d_1t_0} \circ F_{s_0d_0t_0}^{-1} \circ F_{s_0d_0t_1} \circ \hat{F}_{s_1d_0t_1}^{-1} \circ \hat{F}_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right) \right.
$$

$$
\left. - \frac{1}{N_{s_1d_1t_0}} \sum_{i=1}^{N_{s_1d_1t_0}} F_{s_0d_1t_1}^{-1} \circ F_{s_0d_1t_0} \circ F_{s_0d_0t_0}^{-1} \circ F_{s_0d_0t_1} \circ F_{s_1d_0t_1}^{-1} \circ \hat{F}_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right) + \hat{\mu}_5 \right| \xrightarrow{P} 0. \quad (71)
$$

Proof. The proof is analogous to that of Lemma C.6. To adapt the proof, we just need to replace the following definitions: Define $\hat{q}_i \coloneqq \hat{F}_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right)$, $q_i \coloneqq F_{s_1d_0t_0}\left(Y_{s_1d_1,i}(t_0)\right)$, and $g(\cdot) \coloneqq F_{s_0d_1t_1}^{-1} \circ F_{s_0d_1t_0} \circ F_{s_0d_0t_0}^{-1} \circ F_{s_0d_0t_1}(\cdot)$. Finally, the subscripts $s_0d_0t_0$ are replaced by $s_1d_0t_1$. The rest of the proof is identical. $\square$

Lemma C.9. *Define $\hat{\mu}_2$ as given by Eq. (51). Under 3.1,*

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right)\right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) - \hat{\mu}_2 \right| \xrightarrow{P} 0. \quad (72)$$

Proof. For ease of notation, define $\hat{q}_i := \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right)$, and $q_i := F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ F_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right)$. Also define $g(\cdot) := F_{s_0 d_1 t_1}^{-1}(\cdot)$. Using triangle inequality and plugging in $\hat{\mu}_2$, Eq. (72) can be bounded as

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ \hat{F}_{s_0 d_1 t_0}(\hat{q}_i) - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ F_{s_0 d_1 t_0}(\hat{q}_i) - \hat{\mu}_2 \right|$$

$$\leq \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \left(g \circ \hat{F}_{s_0 d_1 t_0}(\hat{q}_i) - g \circ F_{s_0 d_1 t_0}(\hat{q}_i)\right) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_1 t_0}} \left(\mathbb{1}\{Y_{s_0 d_1, j}(t_0) \leq \hat{q}_i\} - F_{s_0 d_1 t_0}(\hat{q}_i)\right) \cdot g' \circ F_{s_0 d_1 t_0}(\hat{q}_i) \right|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_1 t_0}} \left(\mathbb{1}\{Y_{s_0 d_1, j}(t_0) \leq \hat{q}_i\} - F_{s_0 d_1 t_0}(\hat{q}_i)\right) \cdot g' \circ F_{s_0 d_1 t_0}(\hat{q}_i) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \frac{1}{N_{s_0 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_1 t_0}} \left(\mathbb{1}\{Y_{s_0 d_1, j}(t_0) \leq q_i\} - F_{s_0 d_1 t_0}(q_i)\right) \cdot g' \circ F_{s_0 d_1 t_0}(q_i) \right| \quad (73)$$

The first term on the right hand side of Eq. (73) can be bounded by

$$\sqrt{N} \cdot \sup_y \left| \left(g \circ \hat{F}_{s_0 d_1 t_0}(y) - g \circ F_{s_0 d_1 t_0}(y)\right) - \frac{1}{N_{s_0 d_1 t_0}} \sum_{j=1}^{N_{s_0 d_1 t_0}} \left(\mathbb{1}\{Y_{s_0 d_1, j}(t_0) \leq y\} - F_{s_0 d_1 t_0}(y)\right) \cdot g' \circ F_{s_0 d_1 t_0}(y) \right|$$

$$= \sqrt{N} \cdot \sup_y \left| g \circ \hat{F}_{s_0 d_1 t_0}(y) - g \circ F_{s_0 d_1 t_0}(y) - \left(\hat{F}_{s_0 d_1 t_0}(y) - F_{s_0 d_1 t_0}(y)\right) \cdot g' \circ F_{s_0 d_1 t_0}(y) \right|$$

$$\overset{(a)}{\leq} \sqrt{N} \cdot \sup_u \left| g''(u) \right| \cdot \sup_y \left| \hat{F}_{s_0 d_1 t_0}(y) - F_{s_0 d_1 t_0}(y) \right|^2$$

$$= \sqrt{N} \cdot \sup_u \left| g''(u) \right| \cdot \sup_y \left| \hat{F}_{s_0 d_1 t_0}(y) - F_{s_0 d_1 t_0}(y) \right|^2, \quad (74)$$

where $g''(u) = -\frac{1}{f_{s_0 d_1 t_1}(y)^3} \frac{\partial f_{s_0 d_1 t_1}}{\partial y}(u)$. Note that in (a) we used the expansion of $g \equiv F_{s_0 d_1 t_1}^{-1}$ around $F_{s_0 d_1 t_0}(y)$. The right hand side of Eq. (74) is $o_p(1)$ due to Assumption 3.1, and Lemma C.1.

Now consider the second term of Eq. (73), which is equal to

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \left(\hat{F}_{s_0 d_1 t_0}(\hat{q}_i) - F_{s_0 d_1 t_0}(\hat{q}_i)\right) \cdot g' \circ F_{s_0 d_1 t_0}(\hat{q}_i) \right.$$

$$\left. - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \left(\hat{F}_{s_0 d_1 t_0}(q_i) - F_{s_0 d_1 t_0}(q_i)\right) \cdot g' \circ F_{s_0 d_1 t_0}(q_i) \right|. \quad (75)$$

22

The latter can be bounded using triangle inequality by:

$$\sqrt{N}\Big|\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}\big(\hat{F}_{s_0d_1t_0}(\hat{q}_i)-F_{s_0d_1t_0}(\hat{q}_i)\big)\cdot g'\circ F_{s_0d_1t_0}(\hat{q}_i)$$

$$-\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}\big(\hat{F}_{s_0d_1t_0}(q_i)-F_{s_0d_1t_0}(q_i)\big)\cdot g'\circ F_{s_0d_1t_0}(\hat{q}_i)\Big|$$

$$+\sqrt{N}\Big|\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}\big(\hat{F}_{s_0d_1t_0}(q_i)-F_{s_0d_1t_0}(q_i)\big)\cdot g'\circ F_{s_0d_1t_0}(\hat{q}_i)-$$

$$\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}\big(\hat{F}_{s_0d_1t_0}(q_i)-F_{s_0d_1t_0}(q_i)\big)\cdot g'\circ F_{s_0d_1t_0}(q_i)\Big|. \quad (76)$$

The first term on the right hand side of Eq. (76) is bounded by

$$\sqrt{N}\cdot\sup_{y}\big|g'\circ F_{s_0d_1t_0}(y)\big|\cdot\sup_{y}\big|\hat{F}_{s_0d_1t_0}\circ\hat{m}(y)-F_{s_0d_1t_0}\circ m(y)-\big(\hat{F}_{s_0d_1t_0}\circ m(y)-F_{s_0d_1t_0}\circ m(y)\big)\big|, \quad (77)$$

where $m(y)=F_{s_0d_0t_0}^{-1}\circ F_{s_0d_0t_1}\circ F_{s_1d_0t_1}^{-1}\circ F_{s_1d_0t_0}(y)$, and $\hat{m}(y)=\hat{F}_{s_0d_0t_0}^{-1}\circ\hat{F}_{s_0d_0t_1}\circ\hat{F}_{s_1d_0t_1}^{-1}\circ\hat{F}_{s_1d_0t_0}(y)$. The latter is $o_p(1)$ due to Assumption 3.1 and Lemma A.5 of Athey and Imbens (2006) with the choice of $\eta=1/2, \delta=1/3$. Note that $\hat{m}(y)-m(y)=o_p(N^{-\delta})$ due to Lemma C.3 for any $\delta<1/2$, and the conditions of Lemma A.5 of (Athey and Imbens, 2006) are satisfied.

Finally, the second term of Eq. (76) is bounded by

$$\sqrt{N}\cdot\sup_{y}\big|\hat{F}_{s_0d_1t_0}(y)-F_{s_0d_1t_0}(y)\big|\cdot\max_{i}\big|g'\circ F_{s_0d_1t_0}(\hat{q}_i)-g'\circ F_{s_0d_1t_0}(q_i)\big|$$

$$\leq\Big(N^{1/4}\sup_{y}\big|\hat{F}_{s_0d_1t_0}(y)-F_{s_0d_1t_0}(y)\big|\Big)\cdot\Big(N^{1/4}\sup_{y}\big|g'\circ F_{s_0d_1t_0}\circ\hat{m}(y)-g'\circ F_{s_0d_1t_0}\circ m(y)\big|\Big), \quad (78)$$

where the first term is $o_p(1)$ due to Lemma C.1, and second term is $o_p(1)$ due to Assumption 3.1 and the fact that $\hat{m}(y)-m(y)=o_p(N^{-1/4})$ as above. $\square$

Lemma C.10. *Define $\hat{\mu}_4$ as given by Eq. (53). Under 3.1,*

$$\sqrt{N}\cdot\Big|\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}F_{s_0d_1t_1}^{-1}\circ F_{s_0d_1t_0}\circ F_{s_0d_0t_0}^{-1}\circ\hat{F}_{s_0d_0t_1}\circ\hat{F}_{s_0d_1t_1}^{-1}\circ\hat{F}_{s_0d_1t_0}\big(Y_{s_1d_1,i}(t_0)\big)$$

$$-\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}F_{s_0d_1t_1}^{-1}\circ F_{s_0d_1t_0}\circ F_{s_0d_0t_0}^{-1}\circ F_{s_0d_0t_1}\circ\hat{F}_{s_0d_1t_1}^{-1}\circ\hat{F}_{s_0d_1t_0}\big(Y_{s_1d_1,i}(t_0)\big)-\hat{\mu}_4\Big|\xrightarrow{P}0. \quad (79)$$

Proof. The proof is analogous to that of Lemma C.9. To adapt the proof, we just need to replace the following definitions: Define $\hat{q}_i:=\hat{F}_{s_1d_0t_1}^{-1}\circ\hat{F}_{s_1d_0t_0}\big(Y_{s_1d_1,i}(t_0)\big)$, $q_i:=F_{s_1d_0t_1}^{-1}\circ F_{s_1d_0t_0}\big(Y_{s_1d_1,i}(t_0)\big)$, and $g(\cdot):=F_{s_0d_1t_1}^{-1}\circ F_{s_0d_1t_0}\circ F_{s_0d_0t_0}^{-1}(\cdot)$. The rest of the proof is identical. $\square$

Lemma C.11. *Define $\hat{\mu}_6$ as given by Eq. (55). Under 3.1,*

$$\sqrt{N}\cdot\Big|\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}F_{s_0d_1t_1}^{-1}\circ F_{s_0d_1t_0}\circ F_{s_0d_0t_0}^{-1}\circ F_{s_0d_0t_1}\circ F_{s_1d_0t_1}^{-1}\circ\hat{F}_{s_1d_0t_0}\big(Y_{s_1d_1,i}(t_0)\big)$$

$$-\frac{1}{N_{s_1d_1t_0}}\sum_{i=1}^{N_{s_1d_1t_0}}F_{s_0d_1t_1}^{-1}\circ F_{s_0d_1t_0}\circ F_{s_0d_0t_0}^{-1}\circ F_{s_0d_0t_1}\circ F_{s_1d_0t_1}^{-1}\circ F_{s_1d_0t_0}\big(Y_{s_1d_1,i}(t_0)\big)-\hat{\mu}_6\Big|\xrightarrow{P}0. \quad (80)$$

Proof. For ease of notation, define $y_i := Y_{s_1 d_1, i}(t_0)$, and $g(\cdot) := F^{-1}_{s_0 d_1 t_1} \circ F_{s_0 d_1 t_0} \circ F^{-1}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_1} \circ F^{-1}_{s_1 d_0 t_1}(\cdot)$. Using triangle inequality and plugging in $\hat{\mu}_6$, Eq. (80) can be bounded as

$$
\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ \hat{F}_{s_1 d_0 t_0}(y_i) - \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} g \circ F_{s_1 d_0 t_0}(y_i) - \hat{\mu}_6 \right|
$$

$$
= \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \left(g \circ \hat{F}_{s_1 d_0 t_0}(y_i) - g \circ F_{s_1 d_0 t_0}(y_i) \right. \right.
$$

$$
\left. \left. - \frac{1}{N_{s_1 d_0 t_0}} \sum_{j=1}^{N_{s_1 d_0 t_0}} \left(\mathbb{1}\{Y_{s_1 d_0, j}(t_0) \le y_i\} - F_{s_1 d_0 t_0}(y_i) \right) \cdot g' \circ F_{s_1 d_0 t_0}(y_i) \right) \right|
$$

$$
\le \sqrt{N} \cdot \sup_y \left| g \circ \hat{F}_{s_1 d_0 t_0}(y) - g \circ F_{s_1 d_0 t_0}(y) - \frac{1}{N_{s_1 d_0 t_0}} \sum_{j=1}^{N_{s_1 d_0 t_0}} \left(\mathbb{1}\{Y_{s_1 d_0, j}(t_0) \le y\} - F_{s_1 d_0 t_0}(y) \right) \cdot g' \circ F_{s_1 d_0 t_0}(y) \right|
$$

$$
= \sqrt{N} \cdot \sup_y \left| g \circ \hat{F}_{s_1 d_0 t_0}(y) - g \circ F_{s_1 d_0 t_0}(y) - \left(\hat{F}_{s_1 d_0 t_0}(y) - F_{s_1 d_0 t_0}(y) \right) \cdot g' \circ F_{s_1 d_0 t_0}(y) \right|
$$

$$
\overset{(a)}{\le} \sqrt{N} \cdot \sup_u \left| g''(u) \right| \cdot \sup_y \left| \hat{F}_{s_1 d_0 t_0}(y) - F_{s_1 d_0 t_0}(y) \right|^2
$$

$$
= \sqrt{N} \cdot \sup_u \left| g''(u) \right| \cdot \sup_y \left| \hat{F}_{s_1 d_0 t_0}(y) - F_{s_1 d_0 t_0}(y) \right|^2, \quad (81)
$$

where in (a) we used the expansion of g around $F_{s_1 d_0 t_0}(y)$. The right hand side of Eq. (81) is $o_p(1)$ due to Assumption 3.1, and Lemma C.1. $\qquad \square$

Lemma C.12. *Let* $\{\hat{\mu}_i\}_{i=1}^7$ *be given by Equations* (50) *through* (56). *Under* 3.1,

$$
\sqrt{N} \cdot \left(\frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \hat{F}^{-1}_{s_0 d_1 t_1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}^{-1}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}^{-1}_{s_1 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0) \right) \right.
$$

$$
- \mathbb{E}_{s_1 d_1} \left[F^{-1}_{s_0 d_1 t_1} \circ F_{s_0 d_1 t_0} \circ F^{-1}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_1} \circ F^{-1}_{s_1 d_0 t_1} \circ F_{s_1 d_0 t_0} \left(Y(t_0) \right) \right]
$$

$$
\left. + \hat{\mu}_1 - \hat{\mu}_2 + \hat{\mu}_3 - \hat{\mu}_4 + \hat{\mu}_5 - \hat{\mu}_6 - \hat{\mu}_7 \right) \overset{P}{\to} 0. \quad (82)
$$

Proof. Using triangle inequality, the absolute value of the left hand side of Eq. (82) is bounded by

$$
\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \hat{F}^{-1}_{s_0 d_1 t_1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}^{-1}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}^{-1}_{s_1 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0) \right) \right.
$$

$$
- \mathbb{E}_{s_1 d_1} \left[F^{-1}_{s_0 d_1 t_1} \circ F_{s_0 d_1 t_0} \circ F^{-1}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_1} \circ F^{-1}_{s_1 d_0 t_1} \circ F_{s_1 d_0 t_0} \left(Y(t_0) \right) \right]
$$

$$
\left. + \hat{\mu}_1 - \hat{\mu}_2 + \hat{\mu}_3 - \hat{\mu}_4 + \hat{\mu}_5 - \hat{\mu}_6 - \hat{\mu}_7 \right| + \sqrt{N} \cdot \sum_{i=1}^{6} |\hat{\mu}_i - \hat{\mu}_i|. \quad (83)
$$

The term $\sqrt{N} \cdot \sum_{i=1}^{6} |\hat{\mu}_i - \hat{\mu}_i|$ is $o_p(1)$ due to Lemma C.5. The first term of Eq. (83) is bounded by $\qquad \square$

$$\sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \hat{F}_{s_0 d_1 t_1}^{-1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$- \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) + \hat{\mu}_1 \bigg|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$- \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) - \hat{\mu}_2 \bigg|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ \hat{F}_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$- \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) + \hat{\mu}_3 \bigg|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$- \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) - \hat{\mu}_4 \bigg|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$- \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) + \hat{\mu}_5 \bigg|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ \hat{F}_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$- \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ F_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) - \hat{\mu}_6 \bigg|$$

$$+ \sqrt{N} \cdot \left| \frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ F_{s_1 d_0 t_0} \left(Y_{s_1 d_1, i}(t_0)\right) \right.$$

$$\left. - \mathbb{E}_{s_1 d_1} \left[F_{s_0 d_1 t_1}^{-1} \circ F_{s_0 d_1 t_0} \circ F_{s_0 d_0 t_0}^{-1} \circ F_{s_0 d_0 t_1} \circ F_{s_1 d_0 t_1}^{-1} \circ F_{s_1 d_0 t_0} \left(Y(t_0)\right) \right] - \hat{\mu}_7 \right|, \quad (84)$$

where the first through sixth terms are $o_p(1)$ due to Lemmas C.7, C.9, C.6, C.10, C.8 and C.11, respectively, and the seventh term is 0 by definition of $\hat{\mu}_7$ (see Eq. 56.)

D. Omitted Proofs

Theorem 2.3. *Under assumptions A1' - A4' and 1.5 - 2.1, the cumulative density function of the missing counterfactual* $Y^0(t_1)$ *in the group* $S = s_1, D = d_1$ *is identified as:*

$$F_{Y^0(t_1)|s_1, d_1}(y) = F_{Y(t_0)|s_1, d_1} \circ T_{s_0, d_1}^{-1} \circ T_{s_0, d_0} \circ T_{s_1, d_0}^{-1}(y), \tag{7}$$

where $T_{s,d}$ is given by Eq. (5).

Proof.

$$
\begin{aligned}
F_{Y^0(t)|s,d}(y) &= \mathbb{P}(h_{s,d}(U;t) \leq y \mid S = s, D = d) \\
&= \mathbb{P}(U \leq h_{s,d}^{-1}(y;t) \mid S = s, D = d) \\
&= F_{U|s,d}(h_{s,d}^{-1}(y;t)),
\end{aligned}
\tag{85}
$$

and consequently,

$$
h_{s,d}^{-1}(y;t) = F_{U|s,d}^{-1}\big(F_{Y^0(t)|s,d}(y)\big).
\tag{86}
$$

Choose $y = h_{s,d}(u;t)$ in Eq. (85), and apply $F_{Y^0(t)|s,d}^{-1}(\cdot)$ to both sides:

$$
h_{s,d}(u;t) = F_{Y^0(t)|s,d}^{-1}\big(F_{U|s,d}(u)\big)
\tag{87}
$$

Finally, combine Eq. (86) and (87) to get

$$
h_{s,d}\big(h_{s,d}^{-1}(y;t);\tilde{t}\big) = F_{Y^0(\tilde{t})|s,d}^{-1}\big(F_{Y^0(t)|s,d}(y)\big).
\tag{88}
$$

In 2.1 (see Eq. 6), choose $y = h_{s_1,d_0}\big(h_{s_1,d_0}^{-1}(y';t_0);t_1\big)$, and take inverses which results in

$$
h_{s_1,d_1}\big(h_{s_1,d_1}^{-1}(y;t_1);t_0\big) = h_{s_0,d_1}\left(h_{s_0,d_1}^{-1}\left(h_{s,d}\Big\{h_{s,d}^{-1}\big\{h_{s_1,d_0}\big(h_{s_1,d_0}^{-1}(y;t_1);t_0\big);t_0\big\};t_1\Big\};t_1\right);t_0\right),
\tag{89}
$$

or equivalently, in terms of cumulative density functions,

$$
\begin{aligned}
F_{Y^0(t_0)|s_1,d_1}^{-1}&\big(F_{Y^0(t_1)|s_1,d_1}(y)\big) = \\
&F_{Y^0(t_0)|s_0,d_1}^{-1}\left(F_{Y^0(t_1)|s_0,d_1}\left(F_{Y^0(t_1)|s,d}^{-1}\big\{F_{Y^0(t_0)|s,d}\big\{F_{Y^0(t_0)|s_1,d_0}^{-1}\big(F_{Y^0(t_1)|s_1,d_0}(y)\big)\big\}\big\}\right)\right),
\end{aligned}
\tag{90}
$$

which results in the identification of the density of the missing counterfactual:

$$
\begin{aligned}
F_{Y^0(t_1)|s_1,d_1}(y) = \\
F_{Y^0(t_0)|s_1,d_1} \circ F_{Y^0(t_0)|s_0,d_1}^{-1} \circ F_{Y^0(t_1)|s_0,d_1} \circ F_{Y^0(t_1)|s_0,d_0}^{-1} \circ F_{Y^0(t_0)|s_0,d_0} \circ F_{Y^0(t_0)|s_1,d_0}^{-1} \circ F_{Y^0(t_1)|s_1,d_0}(y).
\end{aligned}
$$

The rest follows from 1.5. $\qquad\square$

Proposition 2.8. *Under assumptions A1', A3', A4' and 1.5 - 2.7, for any $y \in \mathbb{Y}_1$,*

$$
\begin{aligned}
F_{Y(t_0)|s_1,d_1} \circ \underline{\Phi}_{s_0,d_1}^{-1} \circ \underline{\Phi}_{s_0,d_0} \circ \underline{\Phi}_{s_1,d_0}^{-1}(y) - \epsilon \\
\leq F_{Y^0(t_1)|s_1,d_1}(y) \leq \\
F_{Y(t_0)|s_1,d_1} \circ \overline{\Phi}_{s_0,d_1}^{-1} \circ \overline{\Phi}_{s_0,d_0} \circ \overline{\Phi}_{s_1,d_0}^{-1}(y) + \epsilon,
\end{aligned}
$$

where

$$
\begin{aligned}
\underline{\Phi}_{s,d}(y) &= F_{Y(t_1)|s,d}^{-1}\big(F_{Y(t_0)|s,d}(y) - \epsilon\big), \\
\underline{\Phi}_{s,d}^{-1}(y) &= F_{Y(t_0)|s,d}^{-1}\big(F_{Y(t_1)|s,d}(y) - \epsilon\big), \\
\overline{\Phi}_{s,d}(y) &= F_{Y(t_1)|s,d}^{-1}\big(F_{Y(t_0)|s,d}(y) + \epsilon\big), \\
\overline{\Phi}_{s,d}^{-1}(y) &= F_{Y(t_0)|s,d}^{-1}\big(F_{Y(t_1)|s,d}(y) + \epsilon\big).
\end{aligned}
$$

Proof. Under A3', 2.7 is equivalent to A.1. Further, A3' implies that 2.11 holds for $\delta = 0$. The result follows from Proposition A.3 with $\delta = 0$. $\qquad\square$

Proposition 2.12. *Under assumptions A1', A2', 2.11, A4', and 1.5-2.1, for any $y \in \mathbb{Y}_1$,*

$$F_{Y(t_0)|s_1,d_1} \circ \underline{\Psi}^{-1}_{s_0,d_1} \circ \underline{\Psi}_{s_0,d_0} \circ \underline{\Psi}^{-1}_{s_1,d_0}(y) - \delta$$
$$\leq F_{Y^0(t_1)|s_1,d_1}(y) \leq$$
$$F_{Y(t_0)|s_1,d_1} \circ \overline{\Psi}^{-1}_{s_0,d_1} \circ \overline{\Psi}_{s_0,d_0} \circ \overline{\Psi}^{-1}_{s_1,d_0}(y) + \delta,$$

where

$$\underline{\Psi}_{s,d}(y) = F^{-1}_{Y(t_1)|s,d}\big(F_{Y(t_0)|s,d}(y) - \delta\big),$$
$$\underline{\Psi}^{-1}_{s,d}(y) = F^{-1}_{Y(t_0)|s,d}\big(F_{Y(t_1)|s,d}(y) - \delta\big),$$
$$\overline{\Psi}_{s,d}(y) = F^{-1}_{Y(t_1)|s,d}\big(F_{Y(t_0)|s,d}(y) + \delta\big),$$
$$\overline{\Psi}^{-1}_{s,d}(y) = F^{-1}_{Y(t_0)|s,d}\big(F_{Y(t_1)|s,d}(y) + \delta\big).$$

Proof. A2' implies that A.1 holds for $\epsilon = 0$. The result follows from Proposition A.3 with $\epsilon = 0$. □

Proposition 2.14. *Under assumptions A1', A2', A3'', A4', and 1.5 - 2.1, the joint density of $Y^0(t_1)$ and $Y^1(t_1)$ in the group corresponding to $S = s_1, D = d_1$ is identified as*

$$F_{Y^0(t_1),Y^1(t_1)|s_1,d_1}(y^0, y^1) =$$
$$F_{Y(t_0),Y(t_1)|s_1,d_1}\big(F^{-1}_{Y(t_0)|s_1,d_1} \circ F_{Y^0(t_1)|s_1,d_1}(y^0), y^1\big),$$

where $F_{Y^0(t_1)|s_1,d_1}(\cdot)$ is given by Equation (7).

Proof. Under the listed assumptions, latent variable U is a function of $Y^0(t_0)$:

$$Y^0(t_0) = h_{s_1,d_1}(U; t_0) \Rightarrow U = h^{-1}_{s_1,d_1}(Y^0(t_0); t_0),$$

and consequently, $Y^0(t_1)$ can be expressed as a function of $Y^0(t_0)$:

$$Y^0(t_1) = h_{s_1,d_1}(U; t_1) = h_{s_1,d_1}\big(h^{-1}_{s_1,d_1}(Y^0(t_0); t_0); t_1\big). \tag{91}$$

Analogous to the proof of Theorem 2.3 (see Equations 86 and 87), the following hold for every s, d, t:

$$h^{-1}_{s_1,d_1}(y; t) = F^{-1}_{U|s_1,d_1}\big(F_{Y^0(t)|s_1,d_1}(y)\big), \quad \text{and} \quad h_{s_1,d_1}(u; t) = F^{-1}_{Y^0(t)|s_1,d_1}\big(F_{U|s_1,d_1}(u)\big),$$

and combining the two, therefore, for any s, d:

$$h_{s_1,d_1}\big(h^{-1}_{s_1,d_1}(y; t_0); t_1\big) = F^{-1}_{Y^0(t_1)|s_1,d_1} \circ F_{Y^0(t_0)|s_1,d_1}(y). \tag{92}$$

From Equations (91) and (92),

$$Y^0(t_1) = F^{-1}_{Y^0(t_1)|s_1,d_1} \circ F_{Y^0(t_0)|s_1,d_1}\big(Y^0(t_0)\big). \tag{93}$$

Note that CDFs and their inverses are strictly increasing, and so is $Y^0(t_1)$ in terms of $Y^0(t_0)$. As a result,

$$F_{Y^0(t_1),Y^1(t_1)|s_1,d_1}(y^0, y^1)$$
$$= \mathbb{P}(Y^0(t_1) \leq y^0, Y^1(t_1) \leq y^1 \mid S = s_1, D = d_1)$$
$$= \mathbb{P}\big(F^{-1}_{Y^0(t_1)|s_1,d_1} \circ F_{Y^0(t_0)|s_1,d_1}\big(Y^0(t_0)\big) \leq y^0, Y^1(t_1) \leq y^1 \mid S = s_1, D = d_1\big)$$
$$= \mathbb{P}\big(Y^0(t_0) \leq F^{-1}_{Y^0(t_0)|s_1,d_1} \circ F_{Y^0(t_1)|s_1,d_1}(y^0), Y^1(t_1) \leq y^1 \mid S = s_1, D = d_1\big)$$
$$= F_{Y^0(t_0),Y^1(t_1)|s_1,d_1}\big(F^{-1}_{Y^0(t_0)|s_1,d_1} \circ F_{Y^0(t_1)|s_1,d_1}(y^0), y^1\big)$$
$$= F_{Y(t_0),Y(t_1)|s_1,d_1}\big(F^{-1}_{Y(t_0)|s_1,d_1} \circ F_{Y^0(t_1)|s_1,d_1}(y^0), y^1\big).$$

Since the assumptions of Theorem 2.3 are satisfied, $F_{Y^0(t_1)|s_1,d_1}(\cdot)$ is identified as in Equation (7) of Theorem 2.3. □

Theorem 3.2. *Under 3.1, $\hat{\tau} - \tau = \mathcal{O}_p(N^{-\frac{1}{2}})$, and*

$$\sqrt{N}(\hat{\tau} - \tau) \xrightarrow{D} \mathcal{N}(0, \frac{V_0}{p_{s_1 d_1 t_1}} + \frac{V_1}{p_{s_0 d_1 t_1}} + \frac{V_2}{p_{s_0 d_1 t_0}} +$$

$$\frac{V_3}{p_{s_0 d_0 t_0}} + \frac{V_4}{p_{s_0 d_0 t_1}} + \frac{V_5}{p_{s_1 d_0 t_1}} + \frac{V_6}{p_{s_1 d_0 t_0}} + \frac{V_7}{p_{s_1 d_1 t_0}}), \quad (14)$$

where $\{V_i\}_{i=0}^{7}$ are given by Eq. (57).

Proof. Consistency. The estimators $\hat{F}$ converge uniformly to their corresponding F, and the estimators $\hat{F}^{-1}$ converge uniformly to their corresponding F^{-1} (see Lemma C.1.) As such, $\hat{F}^{-1}_{s_0 d_1 t_1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}^{-1}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}^{-1}_{s_1 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_0}(y)$ converges to $F^{-1}_{s_0 d_1 t_1} \circ F_{s_0 d_1 t_0} \circ F^{-1}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_1} \circ F^{-1}_{s_1 d_0 t_1} \circ F_{s_1 d_0 t_0}(y)$ uniformly in y. It follows from the law of large numbers and 3.1 that $\frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \hat{F}^{-1}_{s_0 d_1 t_1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}^{-1}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}^{-1}_{s_1 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_0}(Y_{s_1 d_1, i}(t_0))$ converges to $\mathbb{E}_{s_1 d_1}[F^{-1}_{s_0 d_1 t_1} \circ F_{s_0 d_1 t_0} \circ F^{-1}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_1} \circ F^{-1}_{s_1 d_0 t_1} \circ F_{s_1 d_0 t_0}(Y(t_0))]$. Further, the law of large numbers also implies the convergence of $\frac{1}{N_{s_1 d_1 t_1}} \sum_{i=1}^{N_{s_1 d_1 t_1}} Y_{s_1 d_1, i}(t_1)$ to $\mathbb{E}_{s_1 d_1}[Y(t_1)]$, which completes the proof of consistency.

Asymptotic normality. Note that each of $\{\hat{\mu}_i\}_{i=0}^{7}$ is a sample averages of i.i.d. random variables. Under 3.1, the central limit theorem applies and

$$\sqrt{N}(\hat{\mu}_0 + \hat{\mu}_1 - \hat{\mu}_2 + \hat{\mu}_3 - \hat{\mu}_4 + \hat{\mu}_5 - \hat{\mu}_6 - \hat{\mu}_7) \xrightarrow{D} \mathcal{N}(0, V_0/p_{s_1 d_1 t_1} + V_1/p_{s_0 d_1 t_1} + V_2/p_{s_0 d_1 t_0}$$
$$+ V_3/p_{s_0 d_0 t_0} + V_4/p_{s_0 d_0 t_1} + V_5/p_{s_1 d_0 t_1} + V_6/p_{s_1 d_0 t_0} + V_7/p_{s_1 d_1 t_0}).$$
$$(94)$$

Also, due to Lemma C.12 and definition of $\hat{\mu}_0$ (Eq. 49),

$$\sqrt{N} \cdot |(\hat{\tau} - \tau) - (\hat{\mu}_0 + \hat{\mu}_1 - \hat{\mu}_2 + \hat{\mu}_3 - \hat{\mu}_4 + \hat{\mu}_5 - \hat{\mu}_6 - \hat{\mu}_7)|$$

$$= \sqrt{N} \cdot \left(\frac{1}{N_{s_1 d_1 t_0}} \sum_{i=1}^{N_{s_1 d_1 t_0}} \hat{F}^{-1}_{s_0 d_1 t_1} \circ \hat{F}_{s_0 d_1 t_0} \circ \hat{F}^{-1}_{s_0 d_0 t_0} \circ \hat{F}_{s_0 d_0 t_1} \circ \hat{F}^{-1}_{s_1 d_0 t_1} \circ \hat{F}_{s_1 d_0 t_0}(Y_{s_1 d_1, i}(t_0)) \right.$$

$$- \mathbb{E}_{s_1 d_1}[F^{-1}_{s_0 d_1 t_1} \circ F_{s_0 d_1 t_0} \circ F^{-1}_{s_0 d_0 t_0} \circ F_{s_0 d_0 t_1} \circ F^{-1}_{s_1 d_0 t_1} \circ F_{s_1 d_0 t_0}(Y(t_0))]$$

$$\left. + \hat{\mu}_1 - \hat{\mu}_2 + \hat{\mu}_3 - \hat{\mu}_4 + \hat{\mu}_5 - \hat{\mu}_6 - \hat{\mu}_7 \right) = o_p(1), \quad (95)$$

which concludes the proof. $\square$

Theorem 4.1. *Under assumptions A1', A2", A3', A4', and 1.5 - 2.1, the probability measure over the missing counterfactual, i.e., μ_{s_1, d_1}, is identified as*

$$\mu_{s_1, d_1} = (T^* \circ T_{s_1, d_0})_\# \eta_{s_1, d_1}, \quad (18)$$

where $T_{s,d}$ is the Brenier map that pushes forward $\eta_{s,d}$ to $\mu_{s,d}$ for every $s \in \{s_0, s_1\}, d \in \{d_0, d_1\}$, and T^* is the Brenier map that pushes forward $(T_{s_0, d_0 \#} \eta_{s_0, d_1})$ to μ_{s_0, d_1}.

Proof. T^* is the Brenier map that pushes forward $(T_{s_0, d_0 \#} \eta_{s_0, d_1})$ to μ_{s_0, d_1}. By 2.1, T^* is also the Brenier map that pushes forward $(T_{s_1, d_0 \#} \eta_{s_1, d_1})$ to μ_{s_1, d_1}. The result follows by uniqueness of these maps under the given assumptions. $\square$

E. Further on Empirical Evaluations

We begin with providing the complete details of the data generating mechanism we have used for our illustrations on synthetic data in the main text. We later provide an additional result which is postponed to this appendix due to space limitations.
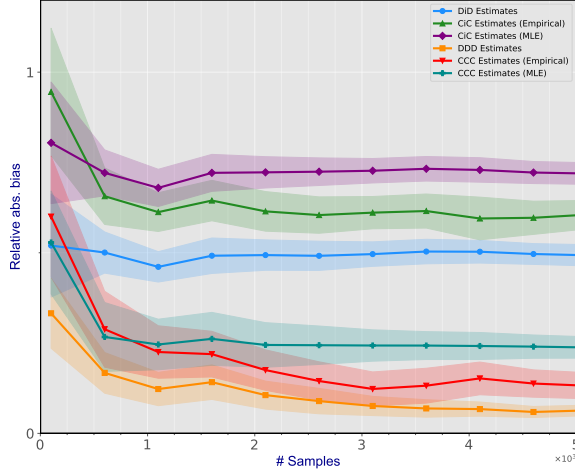
Figure 5. Relative bias of the estimators evaluated on a linear model. The models for maximum likelihood estimation are misspecified.

E.1. Data generating mechanism

To produce Figure 4(a), we sampled the latent variable U given s, d from a Gaussian distribution with mean $\nu_{s,d}$ and variance 1, where $\nu_{s,d}$s were chosen as follows:

$$\nu_{s_0,d_0} = 0,$$
$$\nu_{s_0,d_1} = 0.25,$$
$$\nu_{s_1,d_0} = -0.25,$$
$$\nu_{s_1,d_1} = 0.5.$$

We also generated the $Y^0(t)$ counterfactuals as

$$Y^0(t) = h_{s,d}(u; t),$$

where $h_{s,d}$ was defined as a linear function:

$$h_{s,d}(u; t) := 2u + \left(\frac{1+s}{4} + \frac{d - 0.5}{2}\right)t.$$

Finally, the actual outcome $Y(t_1)$ in the group (s_1, d_1) was sampled from a Gaussian distribution with mean 2.75 and variance 1, and the true ATT was calculated as $\tau = \mathbb{E}_{s_1 d_1}[Y(t_1)] - \mathbb{E}_{s_1 d_1}[Y^0(t_1)]$.

For producing Figure 4(b), the model above was slightly modified as follows. First, we sampled U from a Gaussian distribution with the following means and variances given each group:

$$\nu_{s_0,d_0} = 0, \quad \sigma_{s_0,d_0} = 1,$$
$$\nu_{s_0,d_1} = 0.25, \quad \sigma_{s_0,d_1} = 1,$$
$$\nu_{s_1,d_0} = -0.25, \quad \sigma_{s_1,d_0} = 1,$$
$$\nu_{s_1,d_1} = -0.5, \quad \sigma_{s_1,d_1} = 1.25.$$

Second, to add non-linearities to the model, two of the production functions were modified as:

$$h_{s_0,d_1}(u; t_1) = 0.1 \exp\left(2u + \frac{1}{2}\right), \quad \text{and,}$$
$$h_{s_0,d_1}(u; t_1) = 0.1 \exp\left(2u + \frac{3}{4}\right).$$

Note that these are the same functions as the ones given in the text, after evaluating at $d = 1, t = 1$, and their corresponding $s \in \{0, 1\}$. Moreover, the corresponding maximum likelihood estimators were accordingly adapted to the choice of log-linear distribution instead of Gaussian. Third, the actual outcome was sampled from a Gaussian distribution with mean 10 and variance 1.

E.2. A complementary plot

Throughout the main text, the models were correctly specified when using maximum likelihood estimators. For the sake of completeness, herein, we provide the estimation results under a setting similar to that of Figure 4(a), with the distinction that we sample the latent variables U from an exponential distribution. Specifically, U given s, d is sampled from an exponential distribution with parameters specified as follows.

$$\lambda_{s_0,d_0} = 1,$$
$$\lambda_{s_0,d_1} = 2,$$
$$\lambda_{s_1,d_0} = 3,$$
$$\lambda_{s_1,d_1} = 1.$$

The outcome in the treated groups is also sampled from an exponential distribution with parameter $\frac{4}{15}$. However, we use the Gaussian distribution for maximum likelihood estimation, causing a model mismatch. As a result, it is expected that the CiC and CCC estimators that are based on MLE are biased due to model misspecification. Figure 5 coincides with our expectations. Note that even the CiC estimates with MLE exhibit larger biases compared to their empirical estimator counterpart. As such, caution is advised when using model-based estimators. Note that since the production functions are linear, DDD is unbiased in large samples.