Supplementary Materials for:
"Identification and Inference in Nonlinear Difference-In-Differences Models"
by Susan Athey & Guido Imbens

1 Introduction

In these supplementary materials we provide some details on the implementation of the methods developed in the paper. In addition we apply the different DID approaches using the data analyzed by Meyer, Viscusi, and Durbin (1995). These authors used DID methods to analyze the effects of an increase in disability benefits in the state of Kentucky, where the increase applied to high-earning but not low-earning workers. Next we do a small simulation study. Finally, we provide some additional proofs.

2 Details for Implementation of Estimators

In this section we discuss the implementation of the estimators in more detail. This is useful in particular in cases where the support conditions are not satisfied in the sample. Even if these support conditions (e.g., $Y_{10} \subset Y_{00}$) are satisfied in the population, it finite samples it may well be that there are values y such that $Y_{10,i} = y$ even though there are no observations with $Y_{00,j} = y$. This implies that some of the representations that are equivalent in the population may differ in the finite sample. Here we describe the precise implementation we use in the application and simulations, and in the software that is available on our website.

In all cases (continuous or discrete methods) let $\hat{\mathbb{Y}}_{gt}$ denote the full set of values observed in the subsample with $(G_i, T_i) = (g, t)$, and let $\hat{\mathbb{Y}} = \cup_{g,t} \hat{\mathbb{Y}}_{gt}$ be the union of these. In a finite sample these are all finite sets. For each (g, t) let $\underline{y}_{gt}$ and $\overline{y}_{gt}$ denote the minimum and maximum of the corresponding $\hat{\mathbb{Y}}_{gt}$, and similarly let $\underline{y}$ and $\overline{y}$ denote the minimum and maximum of $\hat{\mathbb{Y}}$. Let L be the cardinality of the set $\hat{\mathbb{Y}}$, and let $\lambda_1, \ldots, \lambda_L$ be the ordered points of support. The data can then be coded as four L-vectors π_{gt}, with $\pi_{gt,l} = \sum_{i=1}^{N} 1\{Y_{gt,i} = \lambda_l\}/N_{gt}$ the proportion of the sample at support point λ_l. As an estimator of the inverse distribution function we now use

$$\hat{F}^{-1}_{Y,gt}(q) = \min\{y \in \hat{\mathbb{Y}_{gt}} : \hat{F}_{Y,gt}(y) \geq q\},$$

so that $\hat{F}^{-1}_{Y,gt}(0) = \min(\hat{\mathbb{Y}}_{gt})$.

In addition, let $\hat{\mathbb{Y}}^* = \hat{\mathbb{Y}} \cup \{-\infty\}$, and let

$$\hat{F}^{(-1)}_{Y,gt}(q) = \max\{y \in \hat{\mathbb{Y}}^* : \hat{F}_{Y,gt}(y) \leq q\},$$

with $F_{Y,gt}(-\infty) = 0$.

2.1 The Continuous Model

For the continous CIC model we estimate the cumulative distribution function for Y_{11}^N as

$$\hat{F}_{Y_{11}^N}(y) = \begin{cases} 0 & \text{if } y < \underline{y}_{01} \\ \hat{F}_{Y,10}\left(\hat{F}_{Y,00}^{-1}\left(\hat{F}_{Y,01}(y)\right)\right) & \text{if } \underline{y}_{01} \leq y < \overline{y}_{01} \\ 1 & \text{if } \overline{y}_{01} \leq y, \end{cases}$$

using the representation from Theorem 3.1. This is a proper cumulative distribution function for a discrete distribution with points of support contained in $\hat{\mathbb{Y}}_{01}$. We use this distribution function to construct estimates of the mean of Y_{11}^N. Finally these are subtracted from the mean of Y_{11} to get an estimate of the average effect of the intervention.

To calculate the standard errors we do the following. First we estimate $f_{Y,01}(y)$ using kernel estimation with an Epanechnikov kernel, $k(a) = 1\{|a| < \sqrt{5}\} \cdot (1 - a^2/5) \cdot 3/(4 \cdot \sqrt{5})$, so that

$$\hat{f}_{Y,01}(y) = \frac{1}{h \cdot N_{01}} \sum_{i=1}^{N_{01}} k\left(\frac{Y_{01,i} - y}{h}\right).$$

The bandwith is choosen using Silverman's rule of thumb $h = 1.06 \cdot S_{Y_{01}} \cdot N_{01}^{-1/5}$, where $S_{Y_{01}} = \sqrt{\sum_{i=1}^{N_{01}} (Y_{01,i} - \bar{Y}_{01})^2}$ is the sample standard deviation of Y_{01}.

Next we estimate $P(y, z)$, $p(y)$, $Q(y, z)$, $q(y)$, $r(y)$, and $s(y)$ using (5.37)-(5.40):

$$\hat{P}(y, z) = \frac{1}{\hat{f}_{Y,01}(\hat{F}_{Y,01}^{-1}(\hat{F}_{Y,00}(z)))} \cdot \left(1\{y \leq z\} - \hat{F}_{Y,00}(z)\right), \qquad \hat{p}(y) = \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \hat{P}(y, Y_{10,i}),$$

$$\hat{Q}(y, z) = -\frac{1}{\hat{f}_{Y,01}(\hat{F}_{Y,01}^{-1}(\hat{F}_{Y,00}(z)))} \cdot \left(1\{\hat{F}_{Y,01}(y) \leq \hat{F}_{Y,00}(z)\} - \hat{F}_{Y,00}(z)\right), \qquad \hat{q}(y) = \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \hat{Q}(y, Y_{10,i}),$$

$$\hat{r}(y) = \hat{F}_{Y,01}^{-1}(\hat{F}_{Y,00}(y)) - \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \hat{F}_{Y,01}^{-1}(\hat{F}_{Y,00}(Y_{10,i})),$$

$$\hat{s}(y) = y - \frac{1}{N_{11}} \sum_{i=1}^{N_{11}} Y_{11,i},$$

Finally we estimate the asymptotic variance of $\sqrt{N}(\hat{\tau}^{\text{cic}} - \tau^{\text{cic}})$ as

$$\frac{1}{N_{00}} \sum_{i=1}^{N_{00}} \hat{p}(Y_{00,i})^2/\hat{\alpha}_{00} + \frac{1}{N_{01}} \sum_{i=1}^{N_{01}} \hat{q}(Y_{01,i})^2/\hat{\alpha}_{01} + \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \hat{r}(Y_{10,i})^2/\hat{\alpha}_{10} + \frac{1}{N_{11}} \sum_{i=1}^{N_{11}} \hat{s}(Y_{11,i})^2/\hat{\alpha}_{11},$$

where $\hat{\alpha}_{gt} = N_{gt}/N$.

For the bootstrap standard errors we bootstrap the sample conditional on N_{gt} for $g, t = 0, 1$. In the simulations and application we use $B = 1000$ bootstrap draws. Given the B bootstrap draws we calculate the difference between the 0.975 and 0.025 quantiles and divided that by 2×1.96 to get standard error estimates.

2.2 Bounds for Discrete Model

For the discrete case we estimate the lower and upper bound on the cumulative distribution function using the representation in Theorem 4.1:

$$\hat{F}^{LB}_{Y^N_{11}}(y) = \begin{cases} 0 & \text{if } y < \underline{y}_{01} \\ \hat{F}_{Y,10}\left(\hat{F}^{(-1)}_{Y,00}\left(\hat{F}_{Y,01}(y)\right)\right) & \text{if } \underline{y}_{01} \le y < \overline{y}_{01} \\ 1 & \text{if } \overline{y}_{01} \le y, \end{cases}$$

$$\hat{F}^{UB}_{Y^N_{11}}(y) = \begin{cases} 0 & \text{if } y < \underline{y}_{01} \\ \hat{F}_{Y,10}\left(\hat{F}^{-1}_{Y,00}\left(\hat{F}_{Y,01}(y)\right)\right) & \text{if } \underline{y}_{01} \le y < \overline{y}_{01} \\ 1 & \text{if } \overline{y}_{01} \le y. \end{cases}$$

For the analytic standard errors we use the representation in(5.43)-(5.44), leading to the following estimators for the normalized variance of $\sqrt{N}(\hat{\tau}_{LB} - \tau_{LB})$ and $\sqrt{N}(\hat{\tau}_{UB} - \tau_{UB})$:

$$\frac{1}{N_{10}}\sum_{i=1}^{N_{10}}\left(\hat{\overline{k}}(Y_{10,i}) - \overline{\hat{\overline{k}}}(Y_{10})\right)^2 + \frac{1}{N_{11}}\sum_{i=1}^{N_{11}}\hat{s}(Y_{11,i})^2/\hat{\alpha}_{11},$$

and

$$\frac{1}{N_{10}}\sum_{i=1}^{N_{10}}\left(\hat{\underline{k}}(Y_{10,i}) - \overline{\hat{\underline{k}}}(Y_{10})\right)^2 + \frac{1}{N_{11}}\sum_{i=1}^{N_{11}}\hat{s}(Y_{11,i})^2/\hat{\alpha}_{11},$$

respectively, where

$$\overline{\hat{\overline{k}}}(Y_{10}) = \frac{1}{N_{10}}\sum_{i=1}^{N_{10}}\hat{\overline{k}}(Y_{10,i}) \qquad \text{and} \quad \overline{\hat{\underline{k}}}(Y_{10}) = \frac{1}{N_{10}}\sum_{i=1}^{N_{10}}\hat{\underline{k}}(Y_{10,i}).$$

The bootstrap variances are calculated similar to the description for the continuous CIC estimator.

2.3 The Discrete Model with Conditional Independence

For the discrete case with conditional independence we estimate the lower and upper bound on the cumulative distribution function using the representation in Theorem 4.2:

$$\hat{F}^{DCI}_{Y^N,11}(y) = \hat{F}^{LB}_{Y^N_{11}}(y)$$
$$+ \left(\hat{F}^{UB}_{Y^N_{11}}(y) - \hat{F}^{LB}_{Y^N_{11}}(y)\right) \cdot \frac{\hat{F}_{Y,01}(y) - \hat{F}_{Y,00}\left(\hat{F}^{(-1)}_{Y,00}\left(\hat{F}_{Y,01}(y)\right)\right)}{\hat{F}_{Y,00}\left(\hat{F}^{-1}_{Y,00}\left(\hat{F}_{Y,01}(y)\right)\right) - \hat{F}_{Y,00}\left(\hat{F}^{(-1)}_{Y,00}\left(\hat{F}_{Y,01}(y)\right)\right)},$$

if $\hat{F}_{Y,00}(\hat{F}^{-1}_{Y,00}(\hat{F}_{Y,01}(y))) - \hat{F}_{Y,00}(\hat{F}^{(-1)}_{Y,00}(\hat{F}_{Y,01}(y))) > 0$, otherwise $\hat{F}^{DCI}_{Y^N,11}(y) = \hat{F}^{LB}_{Y^N_{11}}(y)$.

In this case the estimator is a continuous function of the $\hat{\pi}_{gt}$, which we can write as $\hat{\tau}^{DCI} = g(\hat{\pi})$. However, although the function is continuous, it is not continuously differentiable everywhere. In such cases both the derivative from the left and the derivative from the right exist. To calculate standard errors we first estimate the variance/covariance matrix of the $4(L-1)$-vector $\hat{\pi} = (\hat{\pi}'_{00}, \hat{\pi}'_{01}, \hat{\pi}'_{10}, \hat{\pi}'_{11})'$ (dropping the last element of each vector π_{gt} because it is a linear combination of the others). We denote this covariance matrix by Σ_π, with estimated value $\hat{\Sigma}_\pi$. This covariance matrix is block diagonal with four non-zero $(L-1) \times (L-1)$ blocks on the diagonal, one corresponding to each $\hat{\pi}_{gt}$. The $(L-1) \times (L-1)$ block corresponding to $\hat{\pi}_{gt}$ has (j,k)th element equal to $\left(-\hat{\pi}_{gt,j} \cdot \hat{\pi}_{gt,k} + 1\{j=k\} \cdot \hat{\pi}_{gt,j}\right)/N_{gt}$. Second, in order to apply the delta method, we numerically calculate the derivative of the estimated average effect with respect to the non-zero elements of π. We stack these derivatives, combined with zeros for the derivatives with respect to the elements of π that are equal to zero into a $4(L-1)$-vector $\hat{\tau}_\pi$. Then we estimate the variance of $\hat{\pi}$ as

$$\hat{V}\left(\hat{\tau}^{DCI}\right) = \hat{\tau}'_\pi \hat{\Sigma}_\pi \hat{\tau}_\pi. \tag{A.1}$$

The bootstrap standard errors are calculated as before.

3 An Application to the Meyer-Viscusi-Durbin Injury-duration Data

In this section, we apply the different DID approaches using the data analyzed by Meyer, Viscusi, and Durbin (1995). These authors used DID methods to analyze the effects of an increase in disability benefits in the state of Kentucky, where the increase applied to high-earning but not low-earning workers. The outcome variable is the number of weeks a worker spent on disability. This variable is measured in whole weeks, with the exception of values of zero weeks which are recoded to 0.25 weeks. The distribution of injury durations is highly skewed, with a mean of 8.9, a median of 3, a minimum of 0.25 and a maximum of 182 weeks. In Table 1 we present summary statistics. Meyer, Viscusi and Durbin (1995) noted that their results were quite sensitive to the choice of specification; they found that the treatment led to a significant reduction in the length of spells when the outcome was the natural logarithm of the number of weeks, but not when the outcome is the number of weeks.

To interpret the assumptions required for the CIC model, first normalize $h(u,0) = u$. Then, we interpret u as the number of weeks an individual would desire to stay on disability if the individual faced the period 0 regulatory environment, taking into account the individual's wages, severity of injury, and opportunity cost of time. The distribution of $U|G=g$ should differ across the different earnings groups. The CIC model then requires two substantive assumptions. First, the distribution of U should stay the same over time within a group, which is plausible unless changes in disability programs lead to rapid adjustments in employment decisions. Second, the untreated "outcome function" $h(u,1)$ is monotone in u and is the same for both groups, ruling out, e.g., changes over time in the relationship between wages and the severity of injury in determining the desire for disability benefits.

In Tables 2 and 3 we present the results for the effect of the change in benefits on injury durations. The results for the effect on the treated are in Table 2, and the results for the effect on the control group are in Table 3. We present the results for five estimators: (*i*) the DID-level model, (*ii*) the DID-log model, (*iii*) the discrete CIC model with conditional independence, (*iv*) the discrete CIC model lower bound, and (*v*) the discrete CIC model upper bound. We present six statistics for each estimator and their standard errors based on 1,000 bootstrap replications. The first two statistics are (*i*) the average effect on weeks and (*ii*) the average effect on log weeks.[41] The next four are the difference in quantiles of the distribution of outcomes for the second period treatment group and the counterfactual distribution at following four quantiles: (*iii*) .25, (*iv*) 0.50, (*v*) 0.75, and (*vi*) 0.90.

Because of the extreme skewness of the distribution of outcomes, we focus on the results for the mean of the logarithm of the number of weeks and on quantiles. First, for the DID-log model the prediction of the effect of the treatment on the treated is $E[\ln(Y_{11}^I)] - E[\ln(Y_{11}^N)] = 0.191$. Note that the CIC-discrete and DID-log estimates are comparable, including their precision.[42] For the 25th percentile, the DID-level model yields an estimate of -0.77, while the CIC-discrete model yields a lower bound of 0. Despite the fact that the DID-level parameter is not precisely estimated, the large and negative point estimate highlights the fact that the choice of model can matter for predictions.

Second, let us compare the effects for the treated and the controls. Under the DID-level model the effect on levels is restricted to be the same for the treated and the controls. For the DID-log model the effect on the average logarithm is restricted to be the same. For both groups it is equal to 0.191. The CIC models allow for an unrestricted difference between the effect of the treatment on the treated and control groups; using the CIC-discrete model with the conditional independence assumption, the difference is 0.183-0.211=-0.0273 with a standard error of .0114, so that the difference (while small) is significant at the 95% level.

Finally, consider the bounds on the CIC-discrete estimates. Imbens and Manski (2004)

---

[41] There is one exception. The average on log weeks is not reported for the standard DID model in levels. The reason is that in order for the estimator to be well defined one would have to restrict the effect on levels to be greater than -0.25. If it is less than or equal to 0.25, the predicted outcome for individuals in the first period treatment group who currently have a duration of 0.25 would be negative and we could not take logarithms. Recall that for the DID-level model we estimate the counterfactual distribution as $Y_{11}^N \sim \hat{k}^{DID}(Y_{10})$, where $\hat{k}^{DID}(y) = y + \overline{Y}_{01} - \overline{Y}_{00}$. Rather than use an ad hoc modification we do not report estimates for the effect on log durations under the DID model in levels.

[42] Recall that all standard errors are computed using bootstrapping, so they are comparable; however, note that the asymptotic distributions of the quantile estimates from discrete distributions are not normal, and bootstrapping is not necessarily formally justified.

show how to construct confidence intervals for the average treatment effect in cases where only bounds are identified. Their approach leads to a 95% confidence interval for the average treatment effect on the treated group of $[-0.06, 0.83]$.[43,44] Observe that the upper bound of the estimate for the treatment effect is positive and significantly different from zero, with an estimate of 0.584 and a standard error of 0.15. This is the estimate that would be obtained if we ignored the fact that the outcome is discrete and estimated the average treatment effect directly from (3.9). On the other hand, if we continued to ignore discreteness but instead estimated the average treatment effect using (5.36) (an approach that is equivalent if the data are continuous), we would obtain the lower bound, which is not significantly different from zero. Thus, dealing directly with discreteness of the data can be important, even when the outcome takes on a substantial number of values.

4 A Small Simulation Study

In this section we investigate the finite sample properties of the various estimators and methods for inference by simulation. Our conclusions from these simulation experiments are: (*i*) with continuous data and when the relevant assumptions are satisfied, the asymptotic distributions can approximate the finite sample distributions well, and (*ii*) with discrete data, even if the assumption of no ties in the distribution is formally satisfied, analytic standard errors can be misleading and bootstrap standard errors are more likely to lead to confidence intervals with good coverage rates, (*iii*) with discrete data the continuous estimator as given in (gnored the fact that the outcome is discrete and estimated the average treatment effect directly from (3.9) can be severely downward biased. More simulations would be useful for further understanding the finite sample properties of these procedures.

We create three sets of artificial data. In the first (the "continuous" data) the outcome is continuous on the interval $[0, 1]$. The four densities are: (*i*) $f_{Y,00}(y) = 1.75 - 1.50 \times y$, (*ii*) $f_{Y,01}(y) = 0.75 + 0.50 \times y$, (*iii*) $f_{Y,10}(y) = 0.80 + 0.40 \times y$, and (*iv*) $f_{Y,11}(y) = 0.50 + 1.00 \times y$. The four subsample sizes N_{00}, N_{01}, N_{10}, and N_{11} are all equal to 100 (so the total sample size is 400). For the second set of artificial data (the "discrete" data) we round the continuous outcomes up to the next 0.1, leading to a range of values $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. For the third set of artificial data (the "binary" data) we round up the continuous outcomes to

[43] This is essentially calculated as the lower bound minus 1.645 times its standard error and the upper bound plus 1.645 times its standard error – note the use of 1.645 rather than 1.96 for a 95% confidence interval. See Imbens and Manski (2004) for more details and Chernozhukov, Hong and Tamer(2004) for alternative approaches for confidence intervals in the presence of partially identified parameters.

[44] We could potentially narrow the bounds substantially by incorporating covariates, following the approach suggested in Section 4.3 although the corresponding independence assumption might be difficult to justify for many observed covariates about workers. For this reason, we leave this exercise for future work.

the next multiple of 0.5. In each case we simulate 10,000 data sets, apply the four estimators (continous CIC discrete CIC with conditional independence, discrete CIC lower bound and discrete CIC upper bound) for the average effect on the treated, and calculate analytic and bootstrap standard errors (with the bootstrap standard errors again based on 1,000 bootstrap replications). The analytic standard errors are based on Theorem 5.4 for the discrete models and on Theorem 5.1 for the continuous models.

In Table 4 we report results for six summary statistics: (*i*) the average bias ($\hat{\tau}^{CIC} - \tau^{CIC}$, $\hat{\tau}^{DCI} - \tau^{DCI}$, $\hat{\tau}^{LB} - \tau^{LB}$, or $\hat{\tau}^{UB} - \tau^{UB}$ as appropriate), (*ii*) the median bias, (*iii*) root-mean-squared-error, (*iv*) median-absolute-error, (*v*) 95% coverage rates based on analytic standard errors, and (*vi*) 95% confidence intervals based on bootstrap standard errors.[45] First consider the results for the continuous data. The four point estimates are all very similar, irrespective of whether we use the continuous model, the discrete model with conditional independence or the lower or upper bound. All four have very little bias. The coverage rate for the continuous models are close the nominal values, both based on analytic and based on bootstrap standard errors. For the other three models the bootstrap standard errors are still very accurate, but the analytic standard errors lead to confidence intervals with considerable undercoverage. This undercoverage is to be expected, since the analytic standard errors for the discrete estimator rely on having many observations for each realization of the outcome, an assumption that does not hold when the data are continuous.

Next, consider the case with discrete data. The continuous estimator is severely biased in this case, as expected given that (5.36) yields the lower bound of the average treatment effect. Although the conditional independence assumption is not formally satisfied, it is close enough to being satisfied for the discrete CIC estimator with conditional independence to perform very well. This is true both in terms of bias and in terms of the coverage rate of confidence intervals based on analytic standard errors. The lower and upper bound estimators also do well in terms of bias, but not as well in terms of coverage rates for analytic confidence intervals. Although the assumption concerning no ties in the distribution functions (Assumption 5.2) is formally satisfied, the sample sizes are sufficiently small that the data cannot rule out such ties, as demonstrated by the lack of coverage for the confidence intervals based on the analytic standard errors for the bounds estimators. The bootstrap confidence intervals perform much better,

---

[45] For the continuous data all four estimators estimate the same object, the true average treatment effect corresponding to the continuous data generating process used. This is equal to -0.1093. For the discrete and binary data we compare the continuous and discrete estimator under conditional independence to this true value as well. In these two cases (discrete and binary data) we compare the bounds to their population values. For the discrete data the bounds are -0.1593 and -0.0705. For the binary data the bounds are -0.1875 and 0.0375. The coverage rates reported in the tables refer the the rate at which the corresponding intervals cover these values.

Finally, consider the binary data. Here neither the continuous CIC model nor the discrete CIC model with conditional independence performs very well. Assumption 5.2 is now satisfied both formally and in terms of the data being able to reject ties, and the estimators for the bounds perform well in terms of bias and in terms of coverage rates for confidence intervals based on analytic standard errors.

5 Conclusion

The application presented in the paper show that the approach used to estimate the effects of a policy change can lead to results that differ from one another, in magnitude, significance, and even in sign. Thus, the restrictive assumptions required for standard DID methods can have significant policy implications. Even when one applies the more general classes of models proposed in this paper, however, it will be important to justify such assumptions carefully.

The simulations demonstrate that it can be important to take account of the discrete nature of the data, even when there are a substantial number of values that the outcomes can take on. They also show that the methods for dealing with discrete data can be effective in obtaining credible inferences.

6 Additional Proofs

Proof of Lemma A.1: First consider (i). Define $Y(u) = \{y \in Y | g(y) \geq u\}$. Since $g(y)$ is continuous from the right, $Y(u)$ is a closed set. Hence $g^{-1}(u) = \inf\{y \in Y(u)\} \in Y(u)$ and thus $g(g^{-1}(u)) \geq u$.

Second, consider (ii). By the definition in the proof for (i), $Y(g(y)) = \{y' \in Y | g(y') \geq g(y)\}$. By this definition $y \in Y(g(y))$. Hence $g^{-1}(g(y)) = \inf\{y \in Y(g(y))\} \leq y$.

Next, consider (iii). By (ii), $g^{-1}(g(y)) \leq y$. Because $g(\cdot)$ is nondecreasing it follows that $g(g^{-1}(g(y))) \leq g(y)$. Also, $g(g^{-1}(u)) \geq u$ so $g(g^{-1}(g(y))) \geq g(y)$. Hence it must be that $g(g^{-1}(g(y))) = g(y)$.

Next, consider (iv). First, $g^{-1}(g(y)) \leq y$ for all y, and thus $g^{-1}(g(g^{-1}(u)))) \leq g(g^{-1}(u))$. Second, $g(g^{-1}(u)) \geq u$, and thus $g^{-1}(g(g^{-1}(u))) \geq g^{-1}(u)$. Hence it follows that $g^{-1}(g(g^{-1}(u))) = g^{-1}(u)$.

Finally, consider (v). First we show that $u \leq g(y)$ implies $g^{-1}(u) \leq y$:

$$
\begin{aligned}
u \leq g(y) &\Rightarrow g^{-1}(u) \leq g^{-1}(g(y)) \\
&\Rightarrow g(g^{-1}(u)) \leq g(g^{-1}(g(y))) = g(y).
\end{aligned}
$$

Since $u \leq g(y)$ it follows that $y \in Y(u)$ and thus $g^{-1}(u) \leq y$.

Next we show that $g^{-1}(u) \leq y$ implies $u \leq g(y)$:

$$
g^{-1}(u) \leq y \Rightarrow g(g^{-1}(u)) \leq g(y).
$$

By (i) $u \leq g(g^{-1}(u))$ so that

$$u \leq g(g^{-1}(u)) \leq g(y),$$

which finishes the proof. □

Proof of Lemma A.5: The $\omega(a_N)$ in Lemma A.4 is $o_p(a_N^{-1/2})$. Hence if we take $a_N = N^{-\delta}$, it follows that for uniform U,

$$\sup_{0\leq u,u+v\leq 1,0\leq v\leq N^{-\delta}} N^{1/2} \cdot \left| \hat{F}_U(u+v) - \hat{F}_U(u) - (F_U(u+v) - F_U(u)) \right| = o_p(N^{-\delta/2}).$$

Now let $Y_i = F_Y^{-1}(U_i)$, and $\hat{F}_Y(y) = \sum_i 1\{Y_i \leq y\}/N$. Then $\hat{F}_Y(y) = \hat{F}_U(F_Y(y))$ and $\hat{F}_Y(y + x) = \hat{F}_U(F_Y(y+x))$. Hence

$$\sup_{y,x+y\in\mathbb{Y},0\leq x\leq N^{-\delta}} N^{1/2} \cdot \left| \hat{F}_Y(y+x) - \hat{F}_Y(y) - (F_Y(y+x) - F_Y(y)) \right|$$

$$= \sup_{y,x+y\in\mathbb{Y},0\leq x\leq N^{-\delta}} N^{1/2} \cdot \left| \hat{F}_U(F_Y(y+x)) - \hat{F}_U(F_Y(y)) - (F_U(F_Y(y+x)) - F_U(F_Y(y))) \right|$$

Transforming (y, x) to (u, v) where $u = F_Y(y)$ and $v = F_Y(y+x) - F_Y(y)$ so that $0 \leq x \leq N^{-\delta}$ implies $0 \leq v \leq N^{-\delta}\overline{f}_Y$, this can be bounded from above by

$$\sup_{u,u+v\in[0,1],0\leq v\leq N^{-\delta}\overline{f}_Y} N^{1/2} \cdot \left| \hat{F}_U(u+v) - \hat{F}_U(u) - (F_U(u+v) - F_U(u)) \right| = o_p(N^{-\delta/2}).$$

Thus,

$$\sup_{y,y+x\in\mathbb{Y},x\leq N^{-\delta}} N^{\eta} \left| \hat{F}_Y(y+x) - \hat{F}_Y(y) - (F_Y(y+x) - F_Y(y)) \right| = o_p(N^{\eta-1/2-\delta/2} = o_p(1),$$

because $\delta > 2\eta - 1$. □

Proof of Lemma A.6: By the TI,

$$\sup_q N^{\eta} \cdot \left| \hat{F}_Y^{-1}(q) - F_Y^{-1}(q) + \frac{1}{f_Y(F_Y^{-1}(q))} \left(\hat{F}_Y(F_Y^{-1}(q)) - q \right) \right| \tag{A.2}$$

$$\leq \sup_q N^{\eta} \cdot \left| \hat{F}_Y^{-1}(q) - F_Y^{-1}(\hat{F}_Y(\hat{F}_Y^{-1}(q))) + \frac{1}{f_Y(\hat{F}_Y^{-1}(q))} (\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q))) \right| \tag{A.3}$$

$$+ \sup_q N^{\eta} \cdot \left| \frac{1}{f_Y(F_Y^{-1}(q))} (\hat{F}_Y(F_Y^{-1}(q)) - q) - \frac{1}{f_Y(\hat{F}_Y^{-1}(q))} (\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q))) \right| \tag{A.4}$$

$$+ \sup_q N^{\eta} \cdot \left| F_Y^{-1}(\hat{F}_Y(\hat{F}_Y^{-1}(q))) - F_Y^{-1}(q) \right| \tag{A.5}$$

First, consider (A.3). Since $\hat{F}_Y^{-1}(q) \in Y$,

$$\sup_q N^\eta \cdot \left| \hat{F}_Y^{-1}(q) - F_Y^{-1}(\hat{F}_Y(\hat{F}_Y^{-1}(q))) + \frac{1}{f_Y(\hat{F}_Y^{-1}(q))}(\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q)))\right|$$

$$\leq \sup_y N^\eta \cdot \left| y - F_Y^{-1}(\hat{F}_Y(y)) + \frac{1}{f_Y(y)}(\hat{F}_Y(y) - F_Y(y))\right|$$

Expanding $F_Y^{-1}(\hat{F}_Y(y))$ around $F_Y(y)$ we have, for some $\tilde{y}$ in the support of Y,

$$F_Y^{-1}(\hat{F}_Y(y)) = y + \frac{1}{f_Y(F_Y^{-1}(F_Y(y)))}(\hat{F}_Y(y) - F_Y(y)) - \frac{1}{2f_Y(\tilde{y})^3}\frac{\partial f_Y}{\partial y}(\tilde{y})(\hat{F}_Y(y) - F_Y(y))^2.$$

By Lemma A.2 we have that for all $\delta < 1/2$, $N^\delta \cdot \sup_y |\hat{F}_Y(y) - F_Y(y)| \xrightarrow{p} 0$. Hence for $\eta < 1$ we have $N^\eta \cdot \sup_y |\hat{F}_Y(y) - F_Y(y)|^2 \xrightarrow{p} 0$. This in combination with that fact that both the derivative of density is bounded and the density is bounded away from zero, we have

$$\sup_y N^\eta \cdot \left| F_Y^{-1}(\hat{F}_Y(y)) - y - \frac{1}{f_Y(y)}(\hat{F}_Y(y) - F_Y(y))\right| \leq \sup_{y,\tilde{y}} N^\eta \left| \frac{1}{f_Y(\tilde{y})^3}\frac{\partial f_Y}{\partial y}(\tilde{y})(\hat{F}_Y(y) - F_y(y))^2\right| \xrightarrow{p} 0,$$

which proves that (A.3) converges to zero in probability.
Second, consider (A.4). By the TI,

$$\sup_q N^\eta \cdot \left| \frac{1}{f_Y(F_Y^{-1}(q))}(\hat{F}_Y(F_Y^{-1}(q)) - q) - \frac{1}{f_Y(\hat{F}_Y^{-1}(q))}(\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q)))\right|$$

$$\leq \sup_q N^\eta \cdot \left| \frac{1}{f_Y(F_Y^{-1}(q))}(\hat{F}_Y(F_Y^{-1}(q)) - q) - \frac{1}{f_Y(\hat{F}_Y^{-1}(q))}(\hat{F}_Y(F_Y^{-1}(q)) - q)\right|$$

$$+ \sup_q N^\eta \cdot \left| \frac{1}{f_Y(\hat{F}_Y^{-1}(q))}(\hat{F}_Y(F_Y^{-1}(q)) - q) - \frac{1}{f_Y(\hat{F}_Y^{-1}(q))}(\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q)))\right|$$

$$\leq \sup_q N^{\eta/2} \cdot \left| \frac{1}{f_Y(F_Y^{-1}(q))} - \frac{1}{f_Y(\hat{F}_Y^{-1}(q))}\right| \cdot \sup_q N^{\eta/2} \cdot \left| (\hat{F}_Y(F_Y^{-1}(q)) - q)\right| \tag{A.6}$$

$$+ \frac{1}{\underline{f}} \sup_q N^\eta \cdot \left| (\hat{F}_Y(F_Y^{-1}(q)) - q) - (\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q)))\right|. \tag{A.7}$$

Since $\sup_y N^{\eta/2}|\hat{F}_Y^{-1}(q) - F_Y^{-1}(q)|$ converges to zero by Lemma A.3, and since $f_Y(y)$ is continuously differentiable and bounded away from zero, it follows that $\sup_y N^{\eta/2}|1/f_Y(\hat{F}_Y^{-1}(q)) - 1/f_Y(F_Y^{-1}(q))|$ converges to zero. Also, $\sup_{q\in[0,1]} N^{\eta/2}|\hat{F}_Y(F_Y^{-1}(q)) - q| = \sup_{y\in\mathbb{Y}} N^{\eta/2}|\hat{F}_Y(y) - F_Y(y)|$, and by Lemma A.2 this converges to zero. Hence (A.6) converges to zero.
Next, consider (A.7). By the TI

$$\sup_q N^\eta \cdot \left| (\hat{F}_Y(F_Y^{-1}(q)) - q) - (\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q)))\right|$$

$$\leq \sup_q N^\eta \cdot \left| \hat{F}_Y(F_Y^{-1}(q)) - \hat{F}_Y(F_Y^{-1}(\hat{F}_Y(\hat{F}_Y^{-1}(q))))\right| \tag{A.8}$$

$$+\sup_q N^\eta \cdot \left|\hat{F}_Y(\hat{F}_Y^{-1}(q)) - q\right| \tag{A.9}$$

$$+\sup_q N^\eta \cdot \left|(\hat{F}_Y(F_Y^{-1}(\hat{F}_Y(\hat{F}_Y^{-1}(q)))) - \hat{F}_Y(\hat{F}_Y^{-1}(q))) - (\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q)))\right|. \tag{A.10}$$

For (A.8):

$$\sup_q N^\eta \cdot \left|\hat{F}_Y(F_Y^{-1}(q)) - \hat{F}_Y(F_Y^{-1}(\hat{F}_Y(\hat{F}_Y^{-1}(q))))\right| \leq \sup_q N^\eta \cdot \left|\hat{F}_Y(F_Y^{-1}(q)) - \hat{F}_Y(F_Y^{-1}(q + 1/N))\right|$$

$$\leq \sup_q N^\eta \cdot \left|\hat{F}_Y(F_Y^{-1}(q)) - \hat{F}_Y(F_Y^{-1}(q) + 1/(\underline{f}N))\right|$$

$$\leq \sup_q N^\eta \cdot \left|\hat{F}_Y(F_Y^{-1}(q)) - \hat{F}_Y(F_Y^{-1}(q) + 1/(\underline{f}N)) - \left(F_Y(F_Y^{-1}(q)) - F_Y(F_Y^{-1}(q) + 1/(\underline{f}N))\right)\right|$$

$$+\sup_q N^\eta \cdot \left|F_Y(F_Y^{-1}(q)) - F_Y(F_Y^{-1}(q) + 1/(\underline{f}N))\right|$$

$$\leq \sup_y N^\eta \cdot \left|\hat{F}_Y(y) - \hat{F}_Y(y + 1/(\underline{f}N)) - \left(F_Y(y) - F_Y(y + 1/(\underline{f}N))\right)\right| \tag{A.11}$$

$$+\sup_q N^\eta \cdot \left|F_Y(y) - F_Y(y + 1/(\underline{f}N))\right| \tag{A.12}$$

(A.11) converges to zero by Lemma A.5. (A.12) converges to zero because $\left|F_Y(y) - F_Y(y + 1/(\underline{f}N))\right| \leq \bar{f}/(\underline{f}N)$. Hence (A.8) converges to zero.
The second term, (A.9), converges to zero because of (A.2).
For (A.10), note that

$$\sup_q N^\eta \cdot \left|(\hat{F}_Y(F_Y^{-1}(\hat{F}_Y(\hat{F}_Y^{-1}(q)))) - \hat{F}_Y(\hat{F}_Y^{-1}(q))) - (\hat{F}_Y(\hat{F}_Y^{-1}(q)) - F_Y(\hat{F}_Y^{-1}(q)))\right|$$

$$\leq \sup_y N^\eta \cdot \left|\hat{F}_Y(F_Y^{-1}(\hat{F}_Y(y))) - \hat{F}_Y(y) - \left(\hat{F}_Y(y) - F_Y(y)\right)\right|. \tag{A.13}$$

$$= \sup_y N^\eta \cdot \left|\hat{F}_Y(y + F_Y^{-1}(\hat{F}_Y(y)) - y) - \hat{F}_Y(y) - \left(F_Y(y + F_Y^{-1}(\hat{F}_Y(y)) - y) - F_Y(y)\right)\right|,$$

Using the inequality $Pr(A) \leq Pr(A|B) + Pr(not\ B)$, we can write

$$\Pr\left(\sup_y N^\eta \cdot \left|\hat{F}_Y(y + F_Y^{-1}(\hat{F}_Y(y)) - y) - \hat{F}_Y(y) - \left(F_Y(y + F_Y^{-1}(\hat{F}_Y(y)) - y) - F_Y(y)\right)\right| \geq \varepsilon\right)$$

$$\leq \Pr\left(\sup_y N^\eta \cdot \left|\hat{F}_Y(y + F_Y^{-1}(\hat{F}_Y(y)) - y) - \hat{F}_Y(y) - \left(F_Y(y + F_Y^{-1}(\hat{F}_Y(y)) - y) - F_Y(y)\right)\right| \geq \varepsilon\right. \tag{A.14}$$

$$\left|\sup_y N^\delta |\hat{F}_Y(y) - F_Y(y)| \leq 1/\underline{f}\right)$$

$$+\Pr\left(\sup_y N^\delta |\hat{F}_Y(y) - F_Y(y)| \leq 1/\underline{f}\right). \tag{A.15}$$

Since $N^\delta|\hat{F}_Y(y) - F_Y(y)| \leq 1/\underline{f}$ implies that $|F_Y^{-1}(\hat{F}_Y(y)) - y| \leq N^{-\delta}$, (A.14) converges to zero by Lemma A.5 if we choose $\delta = (2/3)\eta$. Since $\delta = (2/3)\eta$ and $\eta < 5/7$ implies that $\delta < 1/2$, Lemma A.2 implies that (A.15) converges to zero. Thus (A.10) converges to zero. Combined with the convergence of (A.8) and (A.9) this implies that (A.7) converges to zero. This in turn combined with the convergence of (A.6) implies that (A.4) converges to zero.

Third, consider (A.5). Because $|\hat{F}_Y(\hat{F}_Y^{-1}(q)) - q| < 1/N$ for all q by (A.2), this term converges to zero uniformly in q. Hence all three terms (A.3)-(A.5) converge to zero, and therefore (A.2) converges to zero. □

Proof of Lemma A.10: By the TI

$$\sup_{y_1\in\mathbb{Y}_1, y_2\in\mathbb{Y}_2}\left|\hat{h}_1(y_1)\hat{h}_2(y_2) - h_1(y_1)h_2(y_2)\right|$$

$$= \sup_{y_1\in\mathbb{Y}_1, y_2\in\mathbb{Y}_2}\left|h_2(y_2)(\hat{h}_1(y_1) - h_1(y_1)) + (h_2(y_2) - h_2(y_1))(\hat{h}_1(y_1) - h_1(y_1)) + h_1(y_1)(\hat{h}_2(y_2) - h_2(y_2))\right|$$

$$\leq \left|h_2(y_2)(\hat{h}_1(y_1) - h_1(y_1))\right| + \left|(h_2(y_2) - h_2(y_1))(\hat{h}_1(y_1) - h_1(y_1))\right| + \left|h_1(y_1)(\hat{h}_2(y_2) - h_2(y_2))\right|$$

$$\leq \overline{h}_2 \sup_{y_1\in\mathbb{Y}_1}\left|\hat{h}_1(y_1) - h_1(y_1))\right| + \sup_{y_2\in\mathbb{Y}_2}|h_2(y_2) - h_2(y_1)| \sup_{y_1\in\mathbb{Y}_1}\left|\hat{h}_1(y_1) - h_1(y_1)\right| + \overline{h}_1 \sup_{y_2\in\mathbb{Y}_2}\sup_{y_2\in\mathbb{Y}_2}\left|\hat{h}_2(y_2) - h_2(y_2)\right|.$$

All terms are $o_p(1)$. □

Proof of Lemma A.11: By the TI

$$\sup_{y\in\mathbb{Y}}\left|\hat{h}_2(\hat{h}_1(y)) - h_2(h_1(y))\right|$$

$$\leq \sup_{y\in\mathbb{Y}}\left|\hat{h}_2(\hat{h}_1(y)) - h_2(\hat{h}_1(y))\right| + \sup_{y\in\mathbb{Y}}\left|h_2(\hat{h}_1(y)) - h_2(h_1(y))\right|. \tag{A.16}$$

The first term in (A.16) is bounded from above by $\sup_{y\in\mathbb{Y}}|\hat{h}_2(y) - h_2(y)|$ which is $o_p(1)$. Using an MVT the second term in (A.16) is for some $\lambda \in [0,1]$ equal to

$$\sup_{y\in\mathbb{Y}}\left|\frac{\partial h_2}{\partial y}(h_1(y) + \lambda(\hat{h}_1(y) - h_1(y)))\left(\hat{h}_1(y) - h_1(y)\right)\right| \leq \overline{h}'_2 \sup_{y\in\mathbb{Y}}\left|\hat{h}_1(y) - h_1(y)\right| = o_p(1).$$

Hence (A.16) is $o_p(1)$. □

Proof of Theorem 5.3: We will prove that

$$\hat{\tau}_q^{\text{cic}} = \tau_q^{\text{cic}} + \frac{1}{N_{00}}\sum_{i=1}^{N_{00}} p_q(Y_{00,i}) + \frac{1}{N_{01}}\sum_{i=1}^{N_{01}} q_q(Y_{01,i}) + \frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i}) + \frac{1}{N_{11}}\sum_{i=1}^{N_{11}} s_q(Y_{11,i}) + o_p(N^{-1/2}),$$

and thus has an asymptotically linear representation. Then the result follows directly from the fact that $\frac{1}{N_{00}}\sum_{i=1}^{N_{00}} p_q(Y_{00,i})$, $\frac{1}{N_{01}}\sum_{i=1}^{N_{01}} q_q(Y_{01,i})$, $\frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i})$, and $\frac{1}{N_{11}}\sum_{i=1}^{N_{11}} s_q(Y_{11,i})$ all have expectation zero, variances equal to V_q^p, V_q^q, V_q^r, and V_q^s respectively and zero covariances. To prove this assertion is sufficient to show that

$$\hat{F}^{-1}_{Y,01}(\hat{F}_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) = F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))) \tag{A.17}$$

$$+\frac{1}{N_{00}}\sum_{i=1}^{N_{00}} p_q(Y_{00,i}) + \frac{1}{N_{01}}\sum_{i=1}^{N_{01}} q_q(Y_{01,i}) + \frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i}) + o_p(N^{-1/2}).$$

By the TI

$$\left|\hat{F}^{-1}_{Y,01}(\hat{F}_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) - F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))) - \frac{1}{N_{00}}\sum_{i=1}^{N_{00}} p_q(Y_{00,i}) - \frac{1}{N_{01}}\sum_{i=1}^{N_{01}} q_q(Y_{01,i}) - \frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i})\right|$$

$$\leq \left|\hat{F}^{-1}_{Y,01}(\hat{F}_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) - F^{-1}_{Y,01}(\hat{F}_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) - \frac{1}{N_{01}}\sum_{i=1}^{N_{01}} q_q(Y_{01,i})\right| \tag{A.18}$$

$$+\left|F^{-1}_{Y,01}(\hat{F}_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) - F^{-1}_{Y,01}(F_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) - \frac{1}{N_{00}}\sum_{i=1}^{N_{00}} p_q(Y_{00,i})\right| \tag{A.19}$$

$$+\left|F^{-1}_{Y,01}(F_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) - F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))) - \frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i})\right|. \tag{A.20}$$

We will prove that (A.18), (A.19) and (A.20) are all $o_p(1)$ which will finish the proof. Finally, consider (A.20). Using a MVT we can write

$$F^{-1}_{Y,01}(F_{Y,00}(\hat{F}_{Y,10}(q))) = F^{-1}_{Y,01}(F_{Y,00}(F_{Y,10}(q)))$$

$$-\frac{f_{Y,00}(F^{-1}_{Y,10}(q) + \lambda(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q)))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q) + \lambda(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q)))}(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q)))$$

for some $\lambda \in [0,1]$. Therefore

$$\left|F^{-1}_{Y,01}(F_{Y,00}(\hat{F}^{-1}_{Y,10}(q))) - F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))) - \frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i})\right|.$$

$$= \left|-\frac{f_{Y,00}(F^{-1}_{Y,10}(q) + \lambda(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q))))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q) + \lambda(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q)))}(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q))) - \frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i})\right|.$$

$$\leq \left|\begin{array}{c}\left(\frac{f_{Y,00}(F^{-1}_{Y,10}(q)))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))))} - \frac{f_{Y,00}(F^{-1}_{Y,10}(q)+\lambda(\hat{F}^{-1}_{Y,10}(q)-F^{-1}_{Y,10}(q)))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q)+\lambda(\hat{F}^{-1}_{Y,10}(q)-F^{-1}_{Y,10}(q)))}\right) \\ \cdot(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q)))\end{array}\right|. \tag{A.21}$$

$$+\left|-\frac{f_{Y,00}(F^{-1}_{Y,10}(q))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q))))}(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q))) - \frac{1}{N_{10}}\sum_{i=1}^{N_{10}} r_q(Y_{10,i})\right|. \tag{A.22}$$

The second term, (A.22) is equal to:

$$\left| - \frac{f_{Y,00}(F^{-1}_{Y,10}(q))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q)))}(\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q))) \right.$$

$$\left. + \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \frac{f_{Y,00}(F^{-1}_{Y,10}(q))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q)))f_{Y,10}(F^{-1}_{Y,10}(q))} (1\{F_{10}(Y_{10,i}) \leq q\} - q) \right|$$

$$\leq \sup_q \left| \frac{f_{Y,00}(F^{-1}_{Y,10}(q))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q)))} \right| \sup_q \left| \begin{array}{c} (\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q))) - \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \frac{1}{f_{Y,10}(F^{-1}_{Y,10}(q))} \\ \cdot (1\{F_{10}(Y_{10,i}) \leq q\} - q) \end{array} \right|$$

$$= \sup_q \left| \frac{f_{Y,00}(F^{-1}_{Y,10}(q))}{f_{Y,01}(F^{-1}_{Y,01}(F_{Y,00}(F^{-1}_{Y,10}(q)))} \right| \sup_q \left| (\hat{F}^{-1}_{Y,10}(q) - F^{-1}_{Y,10}(q))) - \frac{1}{f_{Y,10}(F^{-1}_{Y,10}(q))} \left(\hat{F}_{Y,10}(F^{-1}_{Y,10}(q)) - q \right) \right|,$$

which is $o_p(N^{-1/2})$ by Lemma A.6. □

Proof of Theorem 5.5: Define $\hat{\mu}^p$, $\hat{\mu}^q$, $\hat{\mu}^r$, and $\hat{\mu}^s$ as before. Then by the same argument as in Lemma A.8 we have $\sqrt{N}(\hat{\tau}^{\text{cic}} - \hat{\mu}^p - \hat{\mu}^q - \hat{\mu}^r - \hat{\mu}^s) = o_p(1)$. It is also still true that $\sqrt{N}_0 \cdot \hat{\mu}^p \xrightarrow{d} N(0, V^p)$, $\sqrt{N}_0 \cdot \hat{\mu}^q \xrightarrow{d} N(0, V^q)$, $\sqrt{N}_1 \cdot (\hat{\mu}^r) \xrightarrow{d} N(0, V^r)$, and $\sqrt{N}_1 \cdot \hat{\mu}^s \xrightarrow{d} N(0, V^s)$. Hence $\sqrt{N} \cdot (\hat{\tau}^{\text{cic}} - \tau^{\text{cic}})$ is asymptotically normal with mean zero, and all that is left is to sort out the part of the variance coming from the correlations between the four terms $\sqrt{N}_0 \cdot \hat{\mu}^p$, $\sqrt{N}_0 \cdot \hat{\mu}^q$, $\sqrt{N}_1 \cdot \hat{\mu}^r$ and $\sqrt{N}_1 \cdot \hat{\mu}^s$, which are no longer all uncorrelated.

Even in the panel case it still holds that $E[p(Y_{00,j}, Y_{10,i})|Y_{10,i}] = 0$, implying that $\sqrt{N}_0 \cdot \hat{\mu}^p$ is uncorrelated with $\sqrt{N}_{10} \cdot \hat{\mu}_{10}$ and $\sqrt{N}_{11} \cdot \hat{\mu}^s$. Similarly, $E[q(Y_{01,j}, Y_{10,i})|Y_{10,i}] = 0$, implying that $\sqrt{N}_{01} \cdot \hat{\mu}^q$ is uncorrelated with $\sqrt{N}_{10} \cdot \hat{\mu}_{10}$ and $\sqrt{N}_{11} \cdot \hat{\mu}^s$. The only two remaining correlations are the ones between $\sqrt{N}_{10} \cdot \hat{\mu}^r$ and $\sqrt{N}_{11} \cdot \hat{\mu}^s$, and between $\sqrt{N}_{00} \cdot \hat{\mu}^p$ and $\sqrt{N}_{01} \cdot \hat{\mu}^q$. The first correlation is

$$\mathbb{E}\left[\sqrt{N}_1 \cdot \hat{\mu}^r \cdot \sqrt{N}_1 \cdot \hat{\mu}^s\right] = \mathbb{E}\left[\frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} \cdot r(Y_{10,i}) \cdot s(Y_{11,j})\right]$$

$$= \mathbb{E}\left[\frac{1}{N_1} \sum_{i=1}^{N_1} r(Y_{10,i}) \cdot (Y_{11,i} - \mathbb{E}[Y_{11}])\right] = \mathbb{E}[r(Y_{10}) \cdot s(Y_{11})] = C^{rs}.$$

The second correlation is

$$\mathbb{E}\left[\sqrt{N}_{00} \cdot \hat{\mu}^p \cdot \sqrt{N}_{01} \cdot \hat{\mu}^q\right] = \mathbb{E}\left[\frac{1}{N_0 N_1^2} \sum_{i=1}^{N_0} \sum_{j=1}^{N_0} \sum_{k=1}^{N_1} \sum_{l=1}^{N_1} p(Y_{00,i}, Y_{10,k}) \cdot q(Y_{01,j}, Y_{10,l})\right].$$

Terms with $i \neq j$ are equal to zero since $E[p(Y_{00,j}, Y_{10,i})|Y_{10,i}] = 0$. Thus the correlation reduces to

$$\mathbb{E}\left[\frac{1}{N_0 N_1^2} \cdot \sum_{i=1}^{N_0} \sum_{k=1}^{N_1} \sum_{l=1}^{N_1} p(Y_{00,i}, Y_{10,k}) \cdot q(Y_{01,i}, Y_{10,l})\right]$$

$$= \mathbb{E}\left[\frac{1}{N_0 N_1^2} \cdot \sum_{i=1}^{N_0}\sum_{k=1}^{N_1}\sum_{l=1,l\neq k}^{N_1} p(Y_{00,i},Y_{10,k}) \cdot q(Y_{01,i},Y_{10,l})\right]$$

$$+\mathbb{E}\left[\frac{1}{N_0 N_1^2} \cdot \sum_{i=1}^{N_0}\sum_{k=1}^{N_1} p(Y_{00,i},Y_{10,k}) \cdot q(Y_{01,i},Y_{10,k})\right]$$

$$= \frac{N_1 - 1}{N_1} \cdot \mathbb{E}\left[\mathbb{E}[p(Y_{00},Y_{10})|Y_{00}] \cdot \mathbb{E}[q(Y_{01},Y_{10})|Y_{01}]\right] + o_p(1) = C^{pq} + o_p(1).$$

□

Proof of Theorem 5.6: Convergence of $\hat{C}^{rs}$ to C^{rs} follows directly from a law of large numbers since $\hat{C}^{rs}$ is a simple sample average. Next, consider $\hat{C}^{pq}$. By the TI we have

$$\left|\hat{C}^{pq} - C^{pq}\right| \leq \left|\frac{1}{N_{00}}\sum_{i=1}^{N_{00}}\left\{\left[\frac{1}{N_{10}}\sum_{j=1}^{N_{10}}\hat{p}(Y_{00,i},Y_{10,j})\right]\cdot\left[\frac{1}{N_{10}}\sum_{j=1}^{N_{10}}\hat{q}(Y_{01,i},Y_{10,j})\right]\right\}\right. \tag{A.23}$$

$$\left. -\frac{1}{N_{00}}\sum_{i=1}^{N_{00}}\left\{\left[\frac{1}{N_{10}}\sum_{j=1}^{N_{10}}p(Y_{00,i},Y_{10,j})\right]\cdot\left[\frac{1}{N_{10}}\sum_{j=1}^{N_{10}}q(Y_{01,i},Y_{10,j})\right]\right\}\right|$$

$$+\left|\frac{1}{N_{00}}\sum_{i=1}^{N_{00}}\left\{\left[\frac{1}{N_{10}}\sum_{j=1}^{N_{10}}p(Y_{00,i},Y_{10,j})\right]\cdot\left[\frac{1}{N_{10}}\sum_{j=1}^{N_{10}}q(Y_{01,i},Y_{10,j})\right]\right\}\right. \tag{A.24}$$

$$\left. -\frac{1}{N_{00}}\sum_{i=1}^{N_{00}}\left\{\mathbb{E}\left[p(Y_{00,i},Y_{10})|Y_{00,i}\right]\cdot\mathbb{E}\left[q(Y_{01,i},Y_{10})|Y_{01,i}\right]\right\}\right|$$

$$+\left|\frac{1}{N_{00}}\sum_{i=1}^{N_{00}}\left\{\mathbb{E}\left[p(Y_{00,i},Y_{10})|Y_{00,i}\right]\cdot\mathbb{E}\left[q(Y_{01,i},Y_{10})|Y_{01,i}\right]\right\} - C_0\right|. \tag{A.25}$$

Expression (A.23) is $o_p(1)$ by the fact that $\sup_{y,z}|\hat{p}(y,z) - p(y,z)|$ and $\sup_{y,z}|\hat{q}(y,z) - q(y,z)|$ converge to zero. Expression (A.24) converges to zero by boundedness of $p(y,z)$ and $q(y,z)$. As a sample average expression (A.25) converges to zero by a law of large numbers. Thus, $\hat{C}^{pq} - C^{pq} = o_p(1)$. □

Proof of Theorem 5.7: The proof goes along the same line as the proof for Theorem 5.4. The estimation error in the control sample does not affect the asymptotic distribution, and the distribution of the bounds are to first order equal to the distribution of $(1/N_1)\sum_i Y_{11,i} - (1/N_1)\sum_i \underline{k}(Y_{10,i})$ and $(1/N_1)\sum_i Y_{11,i} - (1/N_1)\sum_i \overline{k}(Y_{10,i})$ respectively. Both components are still normally distributed in large samples, and hence their difference is normally distributed. The only difference is that the two components, $(1/N_1)\sum_i Y_{11,i}$ and either $(1/N_1)\sum_i \hat{\underline{k}}(Y_{10,i})$ or $(1/N_1)\sum_i \hat{\overline{k}}(Y_{10,i})$ are now correlated. □

Table 1: Summary Statistics

	mean weeks	(s.d.)	mean logs	(s.d.)	25th perc.	50th perc.	75th perc.	90th perc.
Control Group, First Period (N=1703)	6.27	(12.43)	1.13	(1.22)	1.00	3.00	7.00	12.00
Control Group, Second Period (N=1527)	7.04	(16.12)	1.13	(1.27)	1.00	3.00	7.00	14.00
Treatment Group, First Period (N=1233)	11.18	(28.99)	1.38	(1.30)	2.00	4.00	8.00	17.00
Treatment Group, Second Period (N=1161)	12.89	(28.25)	1.58	(1.30)	2.00	5.00	10.00	23.00

Table 2: Estimate of Effect of Treatment on the Treated Group

	mean weeks	(s.e.)	mean logs	(s.e.)	25th perc.	(s.e.)	50th perc.	(s.e.)	75th perc.	(s.e.)	90th perc.	(s.e.)
DID-level	0.95	(1.20)	–	–	-0.77	(0.58)	0.23	(0.63)	1.23	(0.81)	5.23	(2.36)
DID-logs	1.63	(1.26)	0.19	(0.07)	-0.02	(0.31)	0.97	(0.39)	1.94	(0.72)	5.87	(2.42)
CIC disc ci	0.39	(1.49)	0.18	(0.07)	0.00	(0.26)	1.00	(0.51)	2.00	(0.77)	5.00	(2.55)
CIC disc lower	0.07	(1.54)	0.14	(0.12)	0.00	(0.51)	1.00	(0.51)	1.00	(1.02)	4.00	(2.81)
CIC disc upper	1.08	(1.58)	0.58	(0.15)	1.00	(0.51)	2.00	(0.51)	2.00	(0.77)	5.00	(2.55)

Table 3: Estimate of Effect of Treatment on the Control Group

	mean weeks	(s.e.)	mean logs	(s.e.)	25th perc.	(s.e.)	50th perc.	(s.e.)	75th perc.	(s.e.)	90th perc.	(s.e.)
DID-level	0.95	(1.30)	–	–	1.72	(1.24)	1.72	(1.35)	1.72	(1.29)	-0.28	(1.60)
DID-logs	0.61	(0.74)	0.19	(0.07)	0.22	(0.56)	0.66	(0.69)	1.53	(0.63)	0.63	(1.49)
CIC disc ci	0.92	(0.88)	0.21	(0.07)	1.00	(0.26)	1.00	(0.51)	2.00	(0.77)	1.00	(2.04)
CIC disc lower	0.31	(0.85)	0.05	(0.06)	0.00	(0.51)	0.00	(0.51)	1.00	(0.77)	0.00	(2.04)
CIC disc upper	1.56	(0.89)	0.46	(0.07)	1.00	(0.51)	1.00	(0.77)	3.00	(0.77)	2.00	(2.30)

Table 4: Simulations for Average Treatment Effect

Continuous Data	bias		rmse	mae	coverage rate 95% ci	
	mean	median			analytic ci	bootstrap ci
continuous model	0.002	0.001	0.052	0.035	0.957	0.946
discrete model with cond indep	0.003	0.001	0.052	0.035	0.852	0.946
discrete model lower bound	0.003	0.001	0.052	0.035	0.852	0.946
discrete model upper bound	0.002	0.001	0.052	0.035	0.853	0.946
Discrete Data (10 values)	bias		rmse	mae	coverage rate 95% ci	
	mean	median			analytic ci	bootstrap ci
continuous model	0.048	0.048	0.072	0.003	0.864	0.856
discrete model with cond indep	0.007	0.006	0.051	0.001	0.944	0.943
discrete model lower bound	0.009	0.008	0.050	0.001	0.845	0.944
discrete model upper bound	0.009	0.009	0.054	0.001	0.853	0.944
Binary Data	bias		rmse	mae	coverage rate 95% ci	
	mean	median			analytic ci	bootstrap ci
continuous model	0.147	0.148	0.151	0.148	0.005	0.046
discrete model with cond indep	0.066	0.066	0.075	0.066	0.527	0.531
discrete model lower bound	0.000	0.001	0.024	0.017	0.948	0.950
discrete model upper bound	0.001	0.001	0.035	0.024	0.947	0.951