Extreme Changes in Changes

Yuya Sasaki and Yulong Wang

Vanderbilt University, Nashville, TN; Syracuse University, Syracuse, NY

ABSTRACT

Policy analysts are often interested in treating the units with extreme outcomes, such as infants with extremely low birth weights. Existing changes-in-changes (CIC) estimators are tailored to middle quantiles and do not work well for such subpopulations. This article proposes a new CIC estimator to accurately estimate treatment effects at extreme quantiles. With its asymptotic normality, we also propose a method of statistical inference, which is simple to implement. Based on simulation studies, we propose to use our extreme CIC estimator for extreme quantiles, while the conventional CIC estimator should be used for intermediate quantiles. Applying the proposed method, we study the effects of income gains from the 1993 EITC reform on infant birth weights for those in the most critical conditions. This article is accompanied by a Stata command.

1. Introduction

The difference-in-differences (DID) approach is a widely employed empirical strategy for program evaluation in the presence of policy events in time. The common DID methods critically depend on parallel trend assumptions and focus on identifying (conditional) average effects. An alternative empirical strategy is the changes in changes (CIC) method proposed by Athey and Imbens (2006). At the cost of alternative distributional assumptions, the CIC gets around the common trend assumption and further can identify distributions of counterfactual outcomes as opposed to just their averages.

As are the cases with other quantile-based estimands, however, the existing CIC estimator only works for intermediate quantiles (i.e., \( q \in [\hat{q}, 1 - \hat{q}] \) for a certain \( \hat{q} > 0 \)) in theory. This limitation rules out causal inference for those individuals at the extreme top and extreme bottom quantiles. Yet, it is sometimes rather at extreme quantiles that treatment effects are more relevant to social policy analysis. For instance, policymakers often care about treating economically disadvantaged subpopulations like the poorest individuals characterized by the limit \( q \to 0 \). The treatment effects for these tail subpopulations could be substantially larger than that for the mid-sample subpopulations, and hence it is imperative for such policymakers to have methods with which they can accurately assess treatment effects at the subpopulations in the tail.

In this article, we propose an alternative CIC estimator that more accurately estimates the treatment effects at the tails, technically in the limits as \( q \to 0 \) and \( q \to 1 \). We also develop asymptotic normality for this estimator and propose an easy-to-construct confidence interval. Based on our simulation studies, we provide the following practical recommendation. For the intermediate quantiles \( q \in [\hat{q}, 1 - \hat{q}] \), use the existing estimator by Athey and Imbens (2006) along with its standard error. For the extreme quantiles \( q \not\in [\hat{q}, 1 - \hat{q}] \), on the other hand, use our proposed estimator along with its standard error. We suggest using the log-log plot to choose the switching point \( \hat{q} > 0 \) in Section 5.2, and demonstrate a combined use of both estimators in our empirical application.

With the proposed econometric method, we revisit the study by Hoynes, Miller, and Simon (2015) in which they use the 1993 event of EITC reform to evaluate the effects of income gains on infant birth weights. While they analyze average effects via the DID, we focus on the effects at the low quantiles to see if such income gains can improve infant birth weights, particularly for those at the most critical birth weight conditions. This empirical question is of interest because low infant birth weight is known to have long-lasting impacts on the health and economic well-being in adulthood (e.g., Currie 2011) as well as an immediate impact on infant mortality.

Literature. In contrast to the nowadays extensive body of literature on DID, the literature on CIC is relatively thin. Since its first proposal by Athey and Imbens (2006), the CIC framework has been extended to fuzzy treatment assignments (de Chaisemartin and D’Haultfœuille 2014), models with covariates (Melly and Santangelo 2015), continuous treatments (D’Haultfœuille, Hoderlein, and Sasaki 2023), and correction of attrition bias (Ghanem et al. 2022). To our best knowledge, however, no preceding paper investigates extreme quantiles in the context of CIC, despite the aforementioned policy relevance. On the other hand, there are a few papers that investigate treatment effects at extreme quantiles outside the context of CIC—see Chernozhukov (2005), Chernozhukov and Fernández-Val (2011), D’Haultfœuille, Maurel, and Zhang (2018), Zhang (2018), and Deuber et al. (2021) to list but a few. None of the existing

CONTACT Yulong Wang, jwong402@syr.edu Syracuse University, Syracuse, NY.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

© 2023 American Statistical Association
papers on extremal treatment effects consider DID or CIC frameworks.

**Organization.** Section 2 provides a review of CIC. Section 3 introduces the proposed method, and Section 4 derives asymptotic properties. Section 5 discusses some practical issues, and Section 6 extends the proposed method to allow for covariates. Section 7 shows simulation studies, and Section 8 presents the empirical application. Section 9 presents additional simulation results calibrated to the empirical dataset, and Section 10 concludes.

**Stata Command.** This article is accompanied by a Stata command, `ecic` (extreme changes in changes). The package can be installed from SSC archive with the following command line: `ssc install ecic`. After the installation, run `help ecic` for usage of the command.

### 2. The Changes in Changes

This section briefly reviews the CIC estimator following Athey and Imbens (2006). The goals here are to introduce the data-generating model and the treatment parameter of interest, as well as to fix notations to be used in the rest of this article.

Individual $i$ belongs to group $G^i \in \{0, 1\}$, where value of 0 (respectively, 1) indicates the control (respectively, treatment) group. Each individual is observed in one of the two time periods $T^i \in \{0, 1\}$. For each draw $i = 1, \ldots, n$ from the population, the group identity $G^i$ and time period $T^i$ are treated as random variables. Letting $Y^i$ denote a continuous outcome, econometricians observe a random sample of $(Y^i, G^i, T^i)$.

The underlying structure to generate $Y^i$ is as follows. Let $Y^i_0$ (respectively, $Y^i_1$) denote the potential outcome for individual $i$ under no treatment (respectively, under treatment). The potential outcome $Y^i_N$ under no treatment is generated by

$$Y^i_N = h(U, T),$$

where $U$ represents unobserved characteristics, $h(\cdot, t)$ is strictly increasing for each $t \in \{0, 1\}$, and $U \perp T|G$. Let $I^i \in \{0, 1\}$ indicate that individual $i$ receives a treatment. In the two-group-two-period setting, we have $I^i = G^i T^i$. The realized outcome $Y^i$ is generated by

$$Y^i = Y^i_N (1 - I^i) + Y^i_1 . I^i.$$

We now introduce the following short-hand notations:

$$Y^N_{gt} \sim Y_N|G = g, T = t$$

$$Y^1_{gt} \sim Y_I|G = g, T = t$$

$$Y_{gt} \sim Y|G = g, T = t.$$

For any distribution function $F$, we define its left-inverse $F^{-1}$ by $F^{-1}(q) = \inf\{y : F(y) \geq q\}$. In this setup and with these notations, Athey and Imbens (2006, Theorem 3.1) establish

$$F_{Y^N_{gt}}(y) = F_{Y_10} \circ F_{Y_01} \circ F_{Y_01}(y)$$

for all $y$, provided that $U|G = 1$ is a subset of the support of $U|G = 0$.

For each quantile $q \in (0, 1)$, the quantile effect of the treatment is thus identified by

$$\tau^CIC^\circ_q := F^{-1}_{Y^N_{11}}(q) - F^{-1}_{Y^N_{01}}(q) = F^{-1}_{Y_{11}}(q) - F^{-1}_{Y_{01}} \circ F_{Y_{00}} \circ F^{-1}_{Y_{10}}(q).$$

### 3. The Extreme Changes in Changes

The conventional estimator (Athey and Imbens 2006, p. 464) for $\tau^CIC^\circ_q$ performs well in middle quantiles, such as $q \in (0.05, 0.95)$, but may perform less desirably in extreme quantiles (e.g., $q \in (0.00, 0.05] \cup [0.95, 1.00)$), as are the case with common quantile estimators. Indeed, the asymptotic theory for the conventional estimator rules out extreme values of $q$. In this section, we present our proposed method of estimating $\tau^eCIC^\circ_q$ as $q \rightarrow q_n \rightarrow 1$ in the right tail. A symmetric argument applies to the limit on the other side of the distribution in the left tail as $q \rightarrow 0$. To stress the drifting sequence of limiting parameters we consider, we use the notation $\tau^eCIC^\circ_q$ for extreme CIC. In other words, 'e' in “eCIC” is used to remind readers that this parameter drifts with $q$ as the sample size increases.

Suppose that the distribution function $F_{Y^g_{gt}}$ of $Y^g_{gt}$ has regularly varying tails for each $\{g, t\} \in \{0, 1\}^2$. Specifically, we assume

$$1 - F_{Y^g_{gt}}(ty) \sim y^{-\alpha_{gt}} as t \rightarrow \infty$$

for all $y > 0$ and each $\{g, t\} \in \{0, 1\}^2$. Here, the parameter $\alpha_{gt} > 0$ is referred to as the Pareto exponent. Our extreme CIC estimation is built on estimating the Pareto exponent. We emphasize that this assumption is quite mild and most of the common families of parametric distributions as well as a large class of nonparametric distributions satisfy it. For example, the Student-t distribution with $v$ degrees of freedom satisfies this condition with $v$ being the Pareto exponent. See, for example, de Haan and Ferreira (2007, chap. 1) and Resnick (2007, chap. 2) for reviews on this condition.

Let $Y^{(1)}_{gt} \geq Y^{(2)}_{gt} \geq \cdots \geq Y^{(ng)}_{gt}$ denote the order statistics of the realized outcomes in the group $\{g, t\}$, where $n_{gt}$ denotes the subsample size in this group. Choose the largest $k_{gt} + 1$ of them, that is

$$Y^{(1)}_{gt} \geq Y^{(2)}_{gt} \geq \cdots \geq Y^{(k_{gt}+1)}_{gt}.$$

Then, $\alpha_{gt}$ can be estimated by the Hill estimator (Hill 1975)

$$\hat{\alpha}_{gt} = \left( \frac{1}{k_{gt}} \sum_{i=1}^{k_{gt}} \log \left( \frac{Y^{(i)}_{gt}}{Y^{(k_{gt}+1)}_{gt}} \right) \right)^{-1}.$$

As $q \rightarrow 1$, $F_{Y^g_{gt}}^{-1}(q)$ is estimated by

$$\hat{F}_{Y^g_{gt}}^{-1}(q) = Y^{(k_{gt}+1)}_{gt} \left( \frac{k_{gt}}{n_{gt} (1-q)} \right)^{1/\hat{\alpha}_{gt}}.$$

Moreover, the tail probability can be estimated by

$$1 - \hat{F}_{Y^g_{gt}}(y) = \frac{k_{gt}}{n_{gt}} \left( \frac{y}{Y^{(k_{gt}+1)}_{gt}} \right)^{-\hat{\alpha}_{gt}}$$

as $q \rightarrow 1$. See, for example, de Haan and Ferreira (2007, chap. 4).

By combining the identifying formula (2) with the component estimators (3)–(5), we obtain the following estimator for the extreme CIC, $\tau^eCIC^\circ_q$.

$$\hat{\tau}^eCIC^\circ_q = \hat{F}_{Y^g_{gt}}^{-1}(q) - \hat{F}_{Y_01} \circ \hat{F}_{Y_{00}} \circ \hat{F}_{Y_{10}}(q).$$
In this section, we derive limit distributional properties for the extreme CIC. Moreover, for some constants \( \rho > 0 \), \( c_{11} > 0 \), and \( c_{21} \neq 0 \),

\[
1 - F_{\text{gt}}(y) = c_{11} y^{-\alpha_{gt}} + c_{21} y^{-\alpha_{gt} - \theta_{\text{gt}} (1 + \alpha)} \quad \text{as} \quad y \to \infty.
\]

3. \( n_{11}/n_{\text{gt}} \to n_{11}/\lambda_{11} \in (0, \infty) \) and \( k_{11}/\lambda_{11} \to \lambda_{11}/\lambda_{11} \in (0, \infty) \) for each \( g, t \in \{0, 1\}^2 \).

4. \( k_{\text{gt}} \to \infty \) and \( k_{\text{gt}} = o \left( n_{\text{gt}}^{2\rho/(1+2\rho)} \right) \) for each \( g, t \in \{0, 1\}^2 \).

5. \( n_{\text{gt}}(1 - q) = o(k_{\text{gt}}) \) and \( \log[n_{\text{gt}}(1 - q)] = o(\sqrt{k_{\text{gt}}}) \) for each \( g, t \in \{0, 1\}^2 \).

6. \( F_{\text{gt}}^{-1}(\cdot) / F_{\text{gt}}^{-1} \circ F_{\text{gt}}^{-1}(\cdot) \rightarrow c \in (0, \infty) \).

We provide some discussions about these conditions. Following Athey and Imbens (2006), Condition 1 assumes random sampling within each time and treatment group, and independence across time periods and groups. Thus, it presumes repeated cross sections rather than panel data. Condition 2 imposes the regularly varying tail conditions on all four conditional distributions of the outcome. Note that, if a distribution has a regularly varying tail, then this distribution belongs to the domain of attraction of the extreme value distribution with a positive tail index. See, for example, de Haan and Ferreira (2007, chap. 1). Since we derive the convergence rate, the second-order Pareto tail approximation is inevitable. The second-order parameter \( \rho \) governs the distance between the true underlying distribution and the Pareto one. As remarked previously, this condition imposes a rather mild restriction and also satisfies the common support condition. Condition 3 requires that the sample sizes of all subsamples are asymptotically of the same order of magnitude.

Condition 4 specifies the order of the tail thresholds used in estimation. For simplicity of illustration, we select \( k_{\text{gt}} \) to be of a smaller order than \( n_{\text{gt}}^{2\rho/(1+2\rho)} \) so that the estimators incur negligible asymptotic biases relative to variances. This requirement is similar in spirit to under-smoothing bandwidths in kernel estimation or under-smoothing dimensions in sieve estimation. On the other hand, if we select \( k_{\text{gt}} \) to be of the same order of \( n_{\text{gt}}^{2\rho/(1+2\rho)} \), the asymptotic bias becomes non-negligible, whose expression is complicated. In particular, the asymptotic bias involves the second-order parameter \( \rho \) (e.g., de Haan and Ferreira 2007, chap. 3). Estimation of this parameter is challenging since it requires further restrictions on the underlying distribution (e.g., Cheng and Peng 2001; Hauesler and Segers 2007; Carpentier and Kim 2014), which are hard to interpret and hard to justify. Furthermore, such bias estimators entail slower rates of convergence. Given these limitations, we focus on the asymptotics based on undersmoothing for a better statistical inference.

Condition 5 imposes restrictions on the rate at which the quantile level \( q \) under investigation tends to the unit in the drifting sequence. In particular, \( q \) should tend to one sufficiently fast so that the quantile under investigation is extreme. Otherwise, the \( q \) quantile is not in the tail and can be better estimated by the standard CIC method. This condition is also common in the extreme quantile literature (e.g., de Haan and Ferreira 2007, chap. 4). Note that this condition allows for \( n_{\text{gt}}(1 - q) \to 0 \). When this happens, the other part of this condition implicitly imposes a lower bound of \( 1 - q \) and equivalently that the

\[
y_{\text{gt}}(1 - q) \to 0
\]
extrapolation cannot be pushed too far in the right tail (e.g., de Haan and Ferreira 2007, Remark 4.3.4). To see this, observe that the condition \(\log[n_{gt}(1 - q)] = o(\sqrt{k_{gt}})\) implies \(1 - q > n^{-1} \exp(-\varepsilon_0 \sqrt{k_{gt}})\) for each \(\varepsilon_0 > 0\).

Condition 6 requires that the limit of the counterfactual outcome ratio is finite as \(q\) tends to the unit. For simplicity, we consider \(\varepsilon_0 \in (0, \infty)\). If \(\varepsilon_0\) is 0 or \(\infty\), however, the estimator \(\hat{F}_{k_{gt}}(\cdot)\) has a different convergence rate across \(g\) and \(t\), and consequently, we could ignore the estimation error for some pairs of \(g\) and \(t\).

The following theorem establishes the asymptotic normality for the extreme CIC estimator (6) under these conditions.

**Theorem 1.** If Conditions 1–6 are satisfied, then

\[
\frac{k_{11}^{1/2}}{\hat{F}_{k_{11}}(q)} \log d_{11} \left( \hat{r}_{q}^{CIC} - \hat{r}_{q}^{CIC} \right) \xrightarrow{d} \mathcal{N}(0, \Omega)
\]

holds, where \(d_{gt} = k_{gt}/(n_{gt}(1 - q))\) and

\[
\Omega = \alpha^{-2} \left[ 1 + \frac{1}{\varepsilon} \left( \frac{\lambda_{11/10}}{\eta_{11/10}} \right)^2 \left( \hat{\lambda}_{11/00} + \hat{\lambda}_{11/10} + \hat{\lambda}_{11/01} \right) \frac{\alpha_0^2}{\alpha_0^2 \alpha_0^2} \right].
\]

A proof is relegated to Appendix A.

In finite samples, the asymptotic variance can be estimated by substituting \(\hat{\alpha}_{gt}, \hat{\lambda}_{11/10}, \hat{\lambda}_{11/11/00}, \hat{\lambda}_{11/11/10}, \hat{\lambda}_{11/11/01}\), and \(\varepsilon_0\), respectively, in the formula of the asymptotic variance \(\Omega\) provided in the statement of Theorem 1. Under the same conditions, this estimator of \(\Omega\) is also consistent. The 95% confidence interval is then constructed as

\[
\hat{r}_{q}^{CIC} \pm 1.96k_{11}^{1/2} \log d_{11} \left[ \frac{\hat{F}_{k_{11}}(q)}{\sqrt{\hat{F}_{k_{11}}(q)}} \left( \hat{\lambda}_{11/00} + \hat{\lambda}_{11/10} + \hat{\lambda}_{11/01} \right) \frac{\alpha_0^2}{\alpha_0^2 \alpha_0^2} \right]^{1/2}
\]

(7)

In practice, it is recommended to replace \(d_{11}\) by \(d_{11} \uparrow d\) for some \(d > 1\) to ensure a positive value of the logarithm in (7). We set \(d = 10\) in the subsequent simulation studies and empirical application. Finally, we remark that \(\Omega\) simplifies to

\[
\Omega = \alpha^{-2} \left[ 1 + \frac{1}{\varepsilon} \left( \frac{\lambda_{11/10}}{\eta_{11/10}} \right)^2 \left( \hat{\lambda}_{11/00} + \hat{\lambda}_{11/10} + \hat{\lambda}_{11/01} \right) \right]
\]

in the special case where \(\alpha_{gt}\) is the same across \((g, t)\), say \(\alpha_{gt} = \alpha\) for all \((g, t)\), although we do not impose this restriction in the subsequent numerical analyses. This could happen if the treatment effect is a constant shift of the outcome. Given that \(\hat{\alpha}_{gt}\) is asymptotically independent across \(g\) and \(t\), we can perform the standard \(t\)-test for their equivalence.

### 5. Practical Issues

This section collects discussions on the remaining practical issues.

#### 5.1. Choice of \(k_{gt}\)

The number \(k_{gt}\) of order statistics is the key tuning parameter in our method. We propose to use the empirical choice rule proposed by Guillou and Hall (2001). We present the detailed procedure here for convenience of readers.

Since the identical algorithm applies to each pair of \(g\) and \(t\), we suppress these subscripts in this subsection for notational simplicity. Given a random sample \(\{Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)}\}\), we first sort them descendingly and denote the order statistics by \(Y^{(1)} \geq Y^{(2)} \geq \cdots \geq Y^{(n)}\). Define \(Z_i = i \log(Y^{(i)}/Y^{(i+1)})\) for \(i = 1, \ldots, n - 1\). For each \(k = 1, \ldots, n - 1\), construct

\[
T_k \equiv \left( \sum_{i=1}^k w_i^2 \right)^{-1/2} \hat{\xi}^{-1} U_k,
\]

where \(w_i = \text{sgn}(k - 2i + 1) |k - 2i + 1|, \hat{\xi} = 1/\hat{\alpha}_t\), and \(U_k \equiv \sum_{i=-[k/2]}^{[k/2]} Z_i\). When the Pareto tail approximation performs well, \(T_k\) should have its mean close to zero and variance close to one. Accordingly, we can minimize the following criteria based on a moving average of \(T_k^2\):

\[
C_k = \left( \frac{2[k/2] + 1}{2} \right)^{-1} \left( \sum_{j=-[k/2]}^{[k/2]} T_{k+j} \right)^{1/2}.
\]

The optimal value \(k^*\) of \(k\) is

\[
k^* = \min_{1 \leq k \leq n-1} \{ k : C_t > 1\text{ for all } t \geq k \}.
\]

#### 5.2. Extreme Quantiles

We now discuss how to define the domain \([q, 1]\) of \(q\) on which one may use the extreme CIC estimator \(\hat{r}_q^{CIC}\), as opposed to the conventional CIC estimator \(\hat{r}_q^{CIC}\). We suggest to make a scatter-plot of \(\{\log Y_{gt}^{(i)}\}_{i=1}^{n_{gt}}\) against \(\{\log \hat{Y}_{gt}^{(i)}\}_{i=1}^{n_{gt}}\), called the log-log plot. This plot is linear near small values of \(q\) if the tail is approximately Pareto, and our estimator is accurate where it appears linear. In this light, one can choose the boundary point \(q_t\) such that this log-log plot appears linear for \(i \in \{1, \ldots, \lfloor n_{gt}(1 - q) \rfloor \}\). We concretely illustrate this procedure in our empirical application in Section 8.

A researcher will face a tradeoff concerning the choice of \(q_t\) when the extreme CIC estimator \(\hat{r}_q^{CIC}\) is used for \(q \geq q_t\) and the conventional CIC estimator \(\hat{r}_q^{CIC}\) is used for \(q < q_t\). If a too small value (far from one) is chosen for the threshold \(q_t\) such that the log-log plot is nonlinear at \(\lfloor n_{gt}(1 - q) \rfloor\)th order statistic, then our extreme CIC estimator \(\hat{r}_q^{CIC}\) incurs a large bias for \(q \geq q_t\) close to \(q_t\), which may dominate the large variance of the conventional CIC estimator \(\hat{r}_q^{CIC}\) at such values of \(q\). With this tradeoff in mind, a researcher should avoid choosing a too small value of \(q_t\). On the other hand, if a too large value (close to one) is chosen for \(q_t\), then the conventional CIC estimator \(\hat{r}_q^{CIC}\) incurs a large variance for \(q < q_t\) close to \(q_t\), which may dominate the bias of our extreme CIC estimator \(\hat{r}_q^{CIC}\) at such values of \(q\). Thus, a researcher should avoid choosing a too large value of \(q_t\) too.
6. Extension: Covariates

Our proposed method can be easily extended to allow for covariates. Similarly to Athey and Imbens (2006, pp. 465–466), we first regress the outcome variable on the covariates and then apply the proposed extreme CIC estimator to the regression residuals. We formalize this procedure as follows.

Consider the linear model

\[ W_{gt}^i = \left( X_{gt}^i \right)^{\hat{\beta}_g} + Y_{gt}^i, \]

where \( W_{gt}^i \) denotes the outcome variable for the \( i \)th individual in group \( g \) and time \( t \), and \( X_{gt}^i \) denotes the covariate vector. The coefficient \( \hat{\beta}_g \) can be different across \( g \) and \( t \), and hence the above regression can be conducted separately for each \( g \) and \( t \). For notational simplicity, we continue using \( \hat{\beta}_g \) as effective observations and construct the proposed extreme CIC estimator based on them. Specifically, we order the residuals as

\[ \hat{Y}_{gt}^{(1)} \geq \hat{Y}_{gt}^{(2)} \geq \ldots \geq \hat{Y}_{gt}^{(g-1)} \]

and replace \( \{ Y_{gt}^{(i)} \}_{i=1}^{k_g+1} \) with \( \{ \hat{Y}_{gt}^{(i)} \}_{i=1}^{k_g+1} \) in (3)–(6).

In additional to Conditions 1–6, we require the following additional condition.

**Condition**

7. \[ \max_{1 \leq i \leq n_g} \sqrt{k_g} \frac{|Y_{gt}^i - Y_{gt}^j|}{1 + |Y_{gt}^i|} = o_p(1) \] for all \( g \) and \( t \).

Condition 7 is proposed recently by Girard, Stupfler, and Usseglio-Carleve (2021), who study an estimator of tail factors according to the linear model. Generated first are the group and time period indicators. Similarly to Athey and Imbens (2006, pp. 465–466), we first regress the outcome variable on the covariates and then apply the proposed extreme CIC estimator to the regression residuals. We formulae this procedure as follows.

Consider the linear model

\[ Y_{gt}^i = \beta_{gt} + \epsilon_{gt} \]

where \( Y_{gt}^i \) denotes the outcome variable for the \( i \)th individual in group \( g \) and time \( t \), and \( \beta_{gt} \) denotes the coefficient. A potential outcome \( Y_{gt}^i \) can be different across \( g \) and \( t \), and hence the above regression can be conducted separately for each \( g \) and \( t \). For notational simplicity, we continue using \( \beta_{gt} \) as effective observations and construct the proposed extreme CIC estimator based on them. Specifically, we order the residuals as

\[ \hat{\epsilon}_{gt}^{(1)} \geq \hat{\epsilon}_{gt}^{(2)} \geq \ldots \geq \hat{\epsilon}_{gt}^{(g-1)} \]

and replace \( \{ \epsilon_{gt}^{(i)} \}_{i=1}^{k_g+1} \) with \( \{ \hat{\epsilon}_{gt}^{(i)} \}_{i=1}^{k_g+1} \) in (3)–(6).

The following corollary summarizes the result.

**Corollary 1.** Consider the linear regression model (9). If Conditions 1–7 are satisfied, then the estimator \( \hat{\tau}_q^{\text{CIC}} \) based on \( \{ \hat{\epsilon}_{gt}^{(i)} \}_{i=1}^{k_g+1} \) has the same asymptotic distribution as in Theorem 1.

A proof is relegated to Appendix A.

7. Simulations

We use the following data generating design based on our baseline model. Generated first are the group and time period indicators according to

\[ G^i \sim \text{Bernoulli}(\pi_G) \quad \text{and} \quad T^i \sim \text{Bernoulli}(\pi_T). \]

To allow for the endogenous dependence between the group \( G^i \) and the unobservables \( U^i \), we in turn generate \( U^i \) conditionally on \( G^i \) as follows.

\[ U^i \sim \begin{cases} \text{Beta}(\pi_A, \pi_B) & \text{if } G^i = 0 \\ \text{Uniform}(0, 1) & \text{if } G^i = 1. \end{cases} \]

Here, we use the uniform distribution under \( G^i = 1 \) for ease of analytic tractability of both the Pareto exponent and the quantile treatment effects and for the purpose of accurate evaluations of simulation results with analytically known true parameter values. We also remark that the conditional independence assumption \( U^i \perp T^i | G^i \) of Athey and Imbens (2006) is satisfied in this design by construction. In this two-group-two-period setting, the treatment indicator is in turn defined by \( F = GT \).

The potential outcomes are generated through the model

\[ Y^i = \begin{cases} h^N(U^i, T^i) = F^{-1}_y(U^i) + T^i & \text{and} \quad (10) \\ h^2(U^i, T^i) = F^{-1}_q(U^i) + U^i + 1, & \quad (11) \end{cases} \]

where \( F^{-1}_y \) denotes the quantile function of the Student-t distribution with \( \alpha \) degrees of freedom. Now, the observed outcomes are in turn generated by

\[ Y^i = Y_{gt}^i (1 - I^i) + Y^i \cdot I^i. \]

There are three notable features in this data generating process. First, \( F_{Y_N} \) and \( F_{Y_q} \) in (10)–(11) have Pareto exponents of \( \alpha \). Second, the second term \( U \) on the right-hand side of (11), but not of (10), causes heterogeneous treatment effects characterized as follows

\[ \tau_{q}^{\text{CIC}} = F_{Y_{q}'}^{-1}(q) - F_{Y_{q}'}^{-1}(q) = q. \]

Finally, we remark that the monotonicity assumption of Athey and Imbens (2006) for the identification is satisfied in this model.

We evaluate the finite sample performance of our proposed extreme CIC estimator \( \hat{\tau}_q^{\text{CIC}} \) given in (6) along with its standard error estimator (7). The order statistics \( k_g \) are chosen based on Guillasol and Hall (2001) for each subsample \( g, \delta \) as described in Section 4.1. We also present simulation results for the conventional estimator \( \hat{\tau}_q^{\text{CIC}} \) of Athey and Imbens (2006, p. 464) with its standard error estimator (Athey and Imbens 2006, pp. 464–465). For the standard error estimation for \( \hat{\tau}_q^{\text{CIC}} \), we use Epanechnikov kernel and Silverman’s rule of thumb for bandwidth selection. We focus on the top quintile \( q \in [0.80, 1.00] \).

Figure 1 shows Monte Carlo averages and inter-quartile ranges of the estimates under the design with \( (\pi_G, \pi_T, \pi_A, \pi_B, \alpha) = (0.1, 0.5, 1.0, 2.0, 10) \). The dashed curves on the left column of the figure indicate the average estimates based on the conventional estimator \( \hat{\tau}_q^{\text{CIC}} \). The dotted curves on the right column of the figure indicate the average estimates based on our proposed estimator \( \hat{\tau}_q^{\text{CIC}} \). In each panel, the shaded regions indicate the inter-quartile ranges of the estimates by the respective methods. The results are shown for the top quintile \( q \in [0.80, 1.00] \) and for sample sizes \( N \in [2500, 5000] \). The solid curves indicate the true treatment effects. Observe that the conventional estimator \( \hat{\tau}_q^{\text{CIC}} \) tends to give biased estimates as \( q \to 1 \). On the other hand, our proposed estimator \( \hat{\tau}_q^{\text{CIC}} \) yields...
Figure 1. Monte Carlo averages and inter-quartile ranges (shaded) of the estimates based on the conventional estimator $\hat{\tau}^{CIC}_q$ (dashed curves on the left column) and our proposed estimator $\hat{\tau}^{eCIC}_q$ (dotted curves on the right column) for the top quintile $q \in [0.80, 1.00)$ under the design with $(\pi_G, \pi_T, \pi_A, \pi_B, \alpha) = (0.1, 0.5, 1.0, 2.0, 10)$. The true treatment effects are indicated by the solid curves.

Figure 2 shows Monte Carlo frequencies that the true treatment effects are covered by the 95% confidence intervals. The dashed curves indicate the results based on the conventional estimator $\hat{\tau}^{CIC}_q$ and the dotted curves indicate the results based on our proposed estimator $\hat{\tau}^{eCIC}_q$. Again, the results are shown for the top quintile $q \in [0.80, 1.00)$ and for sample sizes $N \in \{2500, 5000\}$. Observe that the coverage frequency based on the conventional method deviates away from the nominal probability of 0.95 as $q \to 1$. In contrast, the coverage frequency based on our proposed method is close to the nominal probability of 0.95 in the extreme quantiles, and outperforms that by the conventional method for most points $q \in [0.80, 1.00)$ in the top quintile. We remark again that we ran many additional simulations with varying design parameter values $(\pi_G, \pi_T, \pi_A, \pi_B, \alpha)$, and the results indicate similar patterns across sets of simulations.

In light of these simulation results, we provide the following practical recommendation. Use the conventional estimator $\hat{\tau}^{CIC}_q$ of Athey and Imbens (2006, p. 464) along with its standard

significantly less biased estimates even in the limit $q \to 1$. We ran many additional simulations with varying design parameter values $(\pi_G, \pi_T, \pi_A, \pi_B, \alpha)$, and the results indicate similar patterns across sets of simulations.
8. EITC and Extremely Low Birth Weights

There is a long history in health economics research to study causes and prevention of low infant birth weight. It is an important topic from policy viewpoint because low infant birth weight has been identified to have long-lasting impacts on the health and economic well-being of individuals in adulthood (e.g., Currie 2011) as well as they are well known to have immediate impact on infant mortality. Some economic and behavioral factors affecting infant birth weight include maternal smoking (e.g., Almond, Chay, and Lee 2005; Currie, Neidell, and Schmieder 2009), maternal stress (e.g., Camacho 2008; Evans and Garthwaite 2014; Aizer, Stroud, and Buka 2016), and economic resources (e.g., Hoynes, Miller, and Simon 2015), among others.

With studies of average effects as in most of the existing empirical studies, it still remains unknown if these causal factors would have positive impacts on the most vulnerable subpopulation, namely those infants born with extremely low birth weights. There are a few papers (Chernozhukov and Fernández-Val 2011; Zhang 2018; Sasaki and Wang 2022) that study extreme quantiles of infant birth weights, but causal interpretations of their estimation results require to assume exogeneity of the explanatory variable of interest conditional on other observed covariates. In empirical settings admitting a changes-in-changes design, on the other hand, we can handle flexible endogeneity in the treatment choice and study treatment effects for the most vulnerable subpopulation at the extremely low quantiles using the method proposed in this article.

Hoynes, Miller, and Simon (2015) use the difference-in-differences (DID) design based on EITC reform (Omnibus Reconciliation Act of 1993, OBRA93) to evaluate the effects of income gains through the EITC on infant health outcomes. They find significant average effects of income shocks on the incidence of low birth weight and the average infant birth weight. In this article, we aim to complement the work of Hoynes, Miller, and Simon (2015) by analyzing the heterogeneous effects of the income gains through the EITC on infant birth weight at extremely low quantiles, as opposed to those on average.

Following the prior work by Hoynes, Miller, and Simon (2015), we use the U.S. Vital Statistics Natality Data, 1989–1999. We also adopt their DID design for our extreme CIC analysis by following their two key assumptions. First, the effects of the EITC on infant birth weights run through the cash available to the family which arrives through tax refunds and the cash is spent over the subsequent 12 months. Second, we focus on the effects during the sensitive development stage in the three months prior to birth. Consequently, following the cash-in-hand assignment rule of Hoynes, Miller, and Simon (2015, Table 1), we include births in May 1994 or after in the “Post” group ($T = 1$) associated with the policy event of OBRA93. The eligibility criteria for the EITC includes the requirement that a taxpayer has a qualifying child. In this light, we include all the second- or higher-order live births as the treatment group ($G = 1$). The sample sizes are $n_{00} = 2372001$, $n_{01} = 1287185$, $n_{10} = 2652321$, and $n_{11} = 1325598$.

Hoynes, Miller, and Simon (2015) define subpopulations by year, state, parity, education, race, and mother’s age. Then, they
treat such a subpopulation as a unit of observation, and use the average birth weight within a subpopulation as the outcome value for the unit. However, this procedure will not allow us to analyze individual heterogeneity with the quantile treatment effect because aggregation eliminates individual heterogeneity. Hence, we use each birth as a unit of observation unlike Hoynes, Miller, and Simon (2015). Otherwise, we follow their empirical approach as follows. First, we use year and state fixed effects. Since Hoynes, Miller, and Simon (2015) use parity, education, race, and mother’s age to define their subgroups of aggregation, we instead use this list of variables as covariates in our analysis. To accommodate these covariates, the extended method introduced in Section 6 is employed. Second, we focus on single women with high school education or less as in Hoynes, Miller, and Simon (2015).

To determine the switching point $\tilde{q}$ between our extreme CIC estimator and the conventional CIC estimator, we draw the log-log plots for $-\hat{Y}_{00}$, $-\hat{Y}_{01}$, $-\hat{Y}_{10}$, and $-\hat{Y}_{11}$ in Figure 3. Observe in each figure that the plot is reasonably linear up to around the 2.5th or 5th percentile, and thereby starts to curve downward. In light of the discussion in Section 5.2 and noting that our current focus is on the left tail, we choose the switching point $\tilde{q}$ such that the log-log plot is linear for the ranks $i \in \{1, \ldots, [n_{1g}\tilde{q}]\}$. We therefore define the 2.5th percentile as our switching point.

Figure 4 illustrates estimates and confidence intervals for $\hat{\tau}_{q \text{CIC}}$. The estimates by our proposed method for the extreme quantiles $q \in (0.00, 0.025]$ are indicated by dotted curves, and the estimates by Athey and Imbens (2006) for the intermediate quantiles $q \in (0.025, 0.200]$ are indicated by the dashed curves. The gray shades indicate pointwise 95% confidence intervals.

Observe that the point estimates are unambiguously positive for all the quantiles $q \in (0.00, 0.200]$. Furthermore, these income effects are statistically significant at each quantile $q \in (0.00, 0.200]$. Therefore, we can conclude that income gains will causally improve the infant birth weights at low quantiles.

While Hoynes, Miller, and Simon (2015) discover positive effects of the EITC income gains on average, we further find positive effects at the low quantiles in particular. This progress in empirical research is important as causal effects for extremely low infant birth weights are more relevant to policy analysis. Low infant birth weight is known to have long-lasting impacts on the health and economic well being in adulthood (e.g., Currie 2011) as well as they are known to have immediate impact on infant mortality. Our findings focusing on the low quantiles imply that income support during pregnancy may help mitigate these adverse health and economic outcomes. We want to stress that, for us to reach this important empirical conclusion, both the conventional estimator $\hat{\tau}_{q \text{CIC}}$ by Athey and Imbens (2006) and our proposed estimator $\hat{\tau}_{q \text{CIC}}$ along with their standard errors are indispensable.

9. Simulations Based on Empirical Data

Section 7 presents simulation studies based on data generated from an artificial design. In this section, we present additional simulation studies with resamples from the empirical data which we use in Section 8.

Let $\{\hat{Y}_{gi}^{\text{ng}}\}_{i=1}^{n_{gi}}$ denote the residualized sample we obtain in Section 8 for each $g$ and $t$. For each $g \in \{0, 1\}$, we draw a one-percent sub-sample of size $[0.01 \cdot n_{gi}]$ from $\{\hat{Y}_{00}^{\text{ng}}\}_{i=1}^{n_{gi}} \cup \{\hat{Y}_{10}^{\text{ng}}\}_{i=1}^{n_{gi}}$ with replacement, and define this subsample as a simulated sample of $Y_{gi}$. Similarly, from each $g \in \{0, 1\}$, we draw a one-percent sub-sample of size $[0.01 \cdot n_{gi}]$ from $\{\hat{Y}_{01}^{\text{ng}}\}_{i=1}^{n_{gi}} \cup \{\hat{Y}_{11}^{\text{ng}}\}_{i=1}^{n_{gi}}$ with replacement, and define this subsample as a simulated sample of $Y_{gi}$. Since we pool the source samples between the control and the treatment groups for each $t$, the true quantile treatment effect $\tau_{q \text{CIC}}$ is zero for all $q$ by construction. Recall from Section 8 that the original sample sizes are $n_{00} = 2372001$, $n_{01} = 1287185$, $n_{10} = 2652321$, and $n_{11} = 1325598$. Hence, simulation sample sizes are $[0.01 \cdot n_{00}] = 23720$, $[0.01 \cdot n_{01}] = 12871$, $[0.01 \cdot n_{10}] = 26523$, and $[0.01 \cdot n_{11}] = 13255$. Under this empirical Monte Carlo design, we run the same set of estimation and inference as in Section 7, except that we focus on the bottom quintile $q \in (0.00, 0.20]$ as opposed to the top quintile $q \in (0.80, 1.00)$.

The top row of Figure 5 shows Monte Carlo averages and inter-quartile ranges of the estimates, analogously to Figure 1 in Section 7. The dashed curve on the left panel indicates the average estimates based on the conventional estimator $\hat{\tau}_{q \text{CIC}}$. The dotted curve on the right panel indicates the average estimates based on our proposed estimator $\hat{\tau}_{q \text{CIC}}$. In each panel, the shaded region indicates the inter-quartile ranges of the estimates by the respective methods. The solid curves indicate the true treatment effects. Since the true treatment effects are homogeneously zero for all $q$ under the current data generating design, there is little bias in the both estimators. Therefore, the inter-quartile ranges are nicely symmetric for the both estimators. This feature of the results contrasts with that in Section 7, where nontrivial biases exist for the conventional estimator $\hat{\tau}_{q \text{CIC}}$ at the extreme quantiles.

The bottom row of Figure 5 shows Monte Carlo frequencies that the true treatment effects are covered by the 95% confidence intervals, analogously to Figure 2 in Section 7. The dashed curve indicates the results based on the conventional estimator $\hat{\tau}_{q \text{CIC}}$ and the dotted curve indicates the results based on our proposed estimator $\hat{\tau}_{q \text{CIC}}$. The results are shown at the extreme quantiles $q \in (0.00, 0.20]$ in the bottom quintile. Although the conventional estimator does not suffer from bias under the current design, statistical inference based on it still suffers from size distortions. Our proposed extreme CIC estimator $\hat{\tau}_{q \text{CIC}}$ yields substantially less size distortions than the conventional estimator $\hat{\tau}_{q \text{CIC}}$ for the entire quintile $q \in (0.00, 0.20]$.  

10. Summary and Discussions 

In this article, we propose a new CIC estimator to accurately estimate the treatment effects at extreme/tail quantiles. We also derive its asymptotic normality result for statistical inference. Our proposal of these new methods is motivated by the fact that policy analysts are often interested in treating subpopulations near tails of the distributions of outcome variables (e.g., extremely poor individuals and infants with extremely low birth weights) while existing CIC estimators are tailored to middle quantiles.

Simulation studies demonstrate that the new extreme CIC estimator along with its standard error estimator performs better than the conventional method in the tails. Based on our
observations of these results, we propose to use our proposed CIC estimator for extreme quantiles, while the conventional CIC estimation should be used for intermediate quantiles.

Applying the proposed method to U.S. Vital Statistics Natality Data, we study the effects of income gains from the 1993 EITC reform on infant birth weights for those in the most critical conditions. We find significant positive effects of the income gains on infant birth weights for the subpopulation at the low quantiles of birth weight.

Finally, we remind the readers that this article is accompanied by a Stata command, ecic (extreme changes in changes). The package can be installed from SSC archive with the following command line: ssc install ecic. After the installation, run help ecic for usage of the command.
Figure 4. Estimates and 95% confidence intervals for $\tau_{CICq}$ of infant birth weight for $q \in (0.000, 0.200)$. The sample consists of infants born between 1989 and 1999 from unmarried black mothers who have completed 12 years of education. The results for the extreme quantiles $q \in (0.000, 0.025)$ are based on the proposed method. The results for the middle quantiles $q \in (0.025, 0.200)$ are based on Athey and Imbens (2006).

Figure 5. Top: Monte Carlo averages and inter-quartile ranges (shaded) of the estimates based on the conventional estimator $\hat{\tau}_{CICq}$ (dashed curves on the left column) and our proposed estimator $\hat{\tau}_{eCICq}$ (dotted curves on the right column) for the bottom quintile $q \in (0.00, 0.20)$. The true treatment effects are indicated by the solid curves. Bottom: Monte Carlo frequencies of coverage of the true treatment effects by the 95% confidence intervals for the bottom quintile $q \in (0.00, 0.20)$. The dashed and dotted curves indicate the results based on the conventional estimator $\hat{\tau}_{CICq}$ and our proposed estimator $\hat{\tau}_{eCICq}$, respectively.
Appendices

Appendix A: Proof of Theorem 1

Proof. For succinctness, we use the short-hand notation $F_{gt}(\cdot)$ for $F_{Y_{gt}}(\cdot)$, and accordingly use the short-hand notation $F_{gt}^{-1}(\cdot)$ for $F_{Y_{gt}}^{-1}(\cdot)$. Under Conditions 1, 2, and 4, we have

$$\sqrt{k_{gt}} \left( \hat{a}_{gt} - a_{gt} \right) \equiv \Gamma_{gt} \xrightarrow{d} \mathcal{N} \left( 0, \sigma_{gt}^2 \right) \quad (12)$$

for all $(g, t) \in \{0, 1\}^2$ – see Hill (1975). Moreover, under Conditions 1, 2, 4, and 5, we have

$$\sqrt{\log \frac{d_{gt}}{\alpha_{gt}}} \left( \frac{F_{gt}^{-1}(q)}{F_{gt}^{-1}(t)} - 1 \right) \equiv \Delta_{gt} \xrightarrow{d} \mathcal{N} \left( 0, \sigma_{gt}^2 \right). \quad (13)$$

for all $(g, t) \in \{0, 1\}^2$ by Theorem 4.3.8 in de Haan and Ferreira (2007), where $d_{gt} \equiv k_{gt} / (n_{gt} (1 - q))$. Given the independence of $\{Y_{gt}\}$ across $(g, t)$ under Condition 1, $\{\hat{F}_{gt}^{-1}(q), \hat{a}_{gt}\}$ are also independent across $(g, t)$. Thus, it suffices to derive the limit of the second item in (6), that is,

$$\hat{A}_n = \hat{F}_{01}^{-1} (\hat{F}_{00}^{-1}(q)) = \frac{Y^{(k_{01}+1)}}{k_{10}} \left( \frac{1 - \alpha_{00/\alpha_{01}}}{\alpha_{00/\alpha_{01}}} \right) \left( \frac{n_{01} \alpha_{01}}{n_{10} \alpha_{00}} \right)^{1/\alpha_{00}} \left( \frac{n_{10}}{n_{01}} \right) \times \left( \frac{1}{\alpha_{01}} \right)^{1/\alpha_{00}} \left( \frac{n_{01}}{n_{10}} \right) \left( q \right)^{1/\alpha_{00/\alpha_{01}}},$$

and we are going to linearize $\hat{A}_n / A_n - 1$ around zero. First, note that we have

$$\sqrt{k_{gt}} \left( \frac{Y^{(k_{01}+1)}}{F_{gt}^{-1}(1 - k_{gt}/n_{gt})} - 1 \right) \equiv \Delta_{gt} \xrightarrow{d} \mathcal{N} \left( 0, \sigma_{gt}^2 \right) \quad (14)$$

from Theorem 2.4.8 in de Haan and Ferreira (2007) and our Condition 4. Second, we decompose $\hat{A}_n / A_n$ as

$$\log \left( \hat{A}_n / A_n \right) = \log \left( \frac{Y^{(k_{01}+1)}}{F_{01}^{-1}(1 - k_{01}/n_{01})} \right)$$

$$+ \left( \frac{\hat{a}_{00}}{\alpha_{01}} \log \left( \frac{Y^{(k_{01}+1)}}{F_{01}^{-1}(1 - k_{01}/n_{01})} \right) \right)$$

$$- \left( \frac{1}{\alpha_{01}} \right) \left( \frac{k_{01}}{n_{01}} \right) \log \left( \frac{1}{\alpha_{00}} \right)$$

$$+ \left( \frac{\hat{a}_{10}}{\alpha_{01}} \right) \left( \frac{1 - \alpha_{10/\alpha_{01}}}{\alpha_{10/\alpha_{01}}} \right) \left( \frac{1}{n_{01}} \right) \log \left( \frac{k_{10}}{n_{10}} \right) \left( q \right)^{1/\alpha_{00/\alpha_{01}}},$$

$$= I_{1n} + I_{2n} + I_{3n} + I_{4n}.$$

For the first term, $I_{1n}$, we have

$$I_{1n} = \log \left( \frac{Y^{(k_{01}+1)}}{F_{01}^{-1}(1 - k_{01}/n_{01})} \right) = \log \left( 1 + k_{01}^{-1/2} \Delta_{01} \right),$$

and

$$\Delta_{01} = k_{01}^{-1/2} \Delta_0 + \alpha \left( k_{01}^{-1/2} \right)$$

by (14). For the second term, $I_{2n}$, we decompose it as

$$I_{2n} = \left( \frac{\hat{a}_{00}}{\alpha_{01}} \right) \log \left( \frac{F_{10}^{-1}(1 - k_{01}/n_{10})}{F_{00}^{-1}(1 - k_{00}/n_{00})} \right)$$

$$+ \left( \frac{a_{00}}{\alpha_{01}} \right) \log \left( \frac{F_{10}^{-1}(1 - k_{10}/n_{00})}{F_{00}^{-1}(1 - k_{00}/n_{00})} \right)$$

$$= \left( \frac{a_{00}}{\alpha_{01}} \right) \log \left( \frac{F_{10}^{-1}(1 - k_{10}/n_{00})}{F_{00}^{-1}(1 - k_{00}/n_{00})} \right)$$

by (12) and (14). For term $I_{3n}$, we rewrite it as

$$I_{3n} = \left( \frac{1}{\alpha_{00}} \right) \log \left( \frac{k_{00}}{n_{00}} \right) \left( q \right)^{1/\alpha_{00/\alpha_{01}}},$$

by (12). For term $I_{4n}$, we rewrite it as

$$I_{4n} = \left( \frac{a_{00}}{\alpha_{01}} \right) \log \left( \frac{k_{00}}{n_{00}} \right) \left( q \right)^{1/\alpha_{00/\alpha_{01}}},$$

by (12). Conditions 4 and 5 imply that $d_{gt} = k_{gt} / (n_{gt} (1 - q)) \to \infty$ for all $g, t \in \{0, 1\}^2$. Moreover, Condition 2 implies that $F_{gt}^{-1}(1 - k_{gt}/n_{gt}) = O((k_{gt}/n_{gt})^{-1/\alpha_{gt}})$ for all $(g, t)$. Then, using L'Hopital's rule, Condition 3, and $d_{gt} \to \infty$, we obtain

$$\frac{1}{\log(d_{11})} \log \left( \frac{F_{10}^{-1}(1 - k_{10}/n_{10})}{F_{00}^{-1}(1 - k_{00}/n_{00})} \right)$$

$$= O \left( \frac{1}{\alpha_{00}} \log \left( k_{10}/n_{10} \right) \left( q \right)^{1/\alpha_{00/\alpha_{01}}} \right)$$

$$+ \frac{1}{\log(d_{11})} \log \left( \frac{k_{10}}{n_{10}} \right) \left( q \right)^{1/\alpha_{00/\alpha_{01}}},$$

$$= o(1).$$
Now, using the above derivations, we obtain

\[
\frac{\sqrt{K_{11}}}{\log(d_{11})}I_{1n} = \mathcal{O}_p \left( \frac{1}{\log(d_{11})} \right) = \mathcal{O}_p(1),
\]

\[
\frac{\sqrt{K_{11}}}{\log(d_{11})}I_{2n} = \mathcal{O}_p \left( \frac{1}{\log(d_{11})} \log \left( \frac{F_{11}^{-1}(1 - \frac{k_{11}}{n_{11}})}{F_{01}^{-1}(1 - \frac{k_{11}}{n_{11}})} \right) \right) = \mathcal{O}_p(1),
\]

\[
\frac{\sqrt{K_{11}}}{\log(d_{11})}I_{3n} = \mathcal{O}_p \left( \frac{1}{\log(d_{11})} \frac{k_{01}n_{00}}{n_{11}k_{00}} \right) = \mathcal{O}_p(1),
\]

\[
\frac{\sqrt{K_{11}}}{\log(d_{11})}I_{4n} = \log \left( \frac{d_{10}}{d_{11}} \right) \left[ \frac{(k_{11}k_{00})^{1/2}}{\alpha_{10}\sigma_{01}} \frac{\Gamma_{00}}{\alpha_{10}\sigma_{01}} - \frac{(k_{11}/k_{10})^{1/2}}{\alpha_{01}\sigma_{01}} \frac{\Gamma_{10}}{\alpha_{10}\sigma_{01}} \right] + \mathcal{O}_p(1).
\]

Now, combining \(I_{1n}, I_{2n}, I_{3n}, \) and \(I_{4n} \) together, and using the fact that \(\exp(x) = 1 + x + O(x^2) \) as \( x \to 0 \), we obtain

\[
\frac{\sqrt{K_{11}}}{\log(d_{11})} \left( \hat{\lambda}_n - \lambda \right) = \log \left( \frac{d_{10}}{d_{11}} \right) \left[ \frac{(k_{11}k_{00})^{1/2}}{\alpha_{10}\sigma_{01}} \frac{\Gamma_{00}}{\alpha_{10}\sigma_{01}} - \frac{(k_{11}/k_{10})^{1/2}}{\alpha_{01}\sigma_{01}} \frac{\Gamma_{10}}{\alpha_{10}\sigma_{01}} \right] + \mathcal{O}_p(1).
\]

by independence among \(\Gamma_{00}, \Gamma_{10}, \) and \(\Gamma_{01}\), Condition 3, and (12).

Finally, using (13) with \( q = t = 1 \) and the condition that \( F_{11}^{-1}(q)/\Lambda_n \to \zeta \), we obtain

\[
\frac{k_{11}^{1/2}}{\log d_{11}} \left( \frac{z_{q}^{CIC} - q_{q}^{CIC}}{F_{11}^{-1}(q)} \right) = \frac{k_{11}^{1/2}}{\log d_{11}} \left( \frac{\hat{F}_{11}^{-1}(q) - F_{11}^{-1}(q)}{F_{11}^{-1}(q)} \right) = \frac{\hat{\Lambda}_n - \Lambda_n}{\hat{F}_{11}^{-1}(q)} - \frac{\hat{\Lambda}_n}{\hat{F}_{11}^{-1}(q)} + \mathcal{O}_p \left( \frac{\hat{\Lambda}_n}{\Lambda_n} \right) \frac{\sigma_{10}^2}{\sigma_{01}^2} \frac{\alpha_{01}^2}{\alpha_{10}^2} \frac{\sigma_{01}^2}{\alpha_{10}^2} \frac{\sigma_{01}^2}{\alpha_{10}^2}.
\]

where

\[
\Omega = \alpha_{11}^{-2} + \left( \frac{1}{2} \right) \left( \frac{\lambda_{11}/\alpha_{11}}{\alpha_{11}/\alpha_{11}} \right)^2 \left( \lambda_{11}/\alpha_{01} + \lambda_{11}/\alpha_{11} + \lambda_{11}/\alpha_{01} \right) \frac{\sigma_{01}^2}{\alpha_{10}^2} \frac{\sigma_{01}^2}{\alpha_{10}^2}.
\]

This completes the proof.

**Appendix B: Proof of Corollary 1**

**Proof.** The proof follows once we establish (12)–(14). Our Condition 7 is the same as Girard, Stupler, and Üsséglio-Carleve (2021, eq. (2)). Our Condition 2 is sufficient for their second-order Pareto tail condition \( C_2 (\gamma, \rho, A) \). Then (12) and (14) directly follow from their Corollary 2.1. Using the same proof of Theorem 4.3.8 in de Haan and Ferreira (2007), (13) further follows from (12), (14), and our Condition 2.

**Supplementary Materials**

The supplementary materials include files for our simulations and empirical application.

**Acknowledgments**

We thank the editor (Ivan Canay), associate editor, two anonymous referees, Alfonso Flores-Lagunes, Hilary Hoyne, Doug Miller, and David Simon for useful advice about our empirical application. We benefited from discussions with Jon Roth. All remaining errors are ours. A stata command, ecic (extreme changes in changes), associated with this article can be installed from SSC archive with the following command line: `ssc install ecic`.

**Disclosure Statement**

The authors report there are no competing interests to declare.

**References**

Aizer, A., Stroud, L., and Buka, S. (2016), "Maternal Stress and Child Outcomes: Evidence from Siblings," Journal of Human Resources, 51, 523–555. [818]

Almond, D., Chay, K. Y., and Lee, D. S. (2005), "The Costs of Low Birth Weight," Quarterly Journal of Economics, 120, 1031–1083. [818]

Athey, S., and Imbens, G. W. (2006), "Identification and Inference in Nonlinear Difference-in-Differences Models," Econometrica, 74, 431–497. [812,813,814,815,816,817,818,819,821]

Camacho, A. (2008), "Stress and Birth Weight: Evidence from Terrorist Attacks," American Economic Review, 98, 511–15. [818]

Carpentier, A., and Kim, A. K. H. (2014), "Adaptive and Minimax Optimal Estimation of the Tail Coefficient," Statistical Sinica, 25, 1133–1144. [814]

Cheng, S., and Peng, L. (2011), "Confidence Intervals for the Tail Index," Bernoulli, 7, 751–760. [814]

Chernozhukov, V. (2005), "Extremal Quantile Regression," Annals of Statistics, 33, 806–839. [812]

Chernozhukov, V., and Fernández-Val, I. (2011), "Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks," Review of Economic Studies, 78, 559–589. [812,818]

Currie, J. (2011), "Inequality at Birth: Some Causes and Consequences," American Economic Review, 101, 1–22. [812,818,819]

Currie, J., Neidell, M., and Schmieder, J. F. (2009), "Air Pollution and Infant Health: Lessons from New Jersey," Journal of Health Economics, 28, 688–703. [818]

de Chaisemartin, C., and D’Haultfouillle, X. (2014), "Puffy Changes-in-Changes," CREST. [812]

de Haan, L., and Ferreira, A. (2007), Extreme Value Theory: An Introduction, New York: Springer. [812,813,814,815,822,823]

Deuber, D., Li, J., Engelke, S., and Maathuis, M. H. (2021), "Estimation and Inference of Extremal Quantile Treatment Effects for Heavy-Tailed Distributions," arXiv preprint arXiv:2110.06627. [812]

D’Haultfouillle, X., Hoderlein, S., and Sasaki, Y. (2023), "Nonparametric Difference-in-Differences in Repeated Cross-Sections with Continuous Treatments," Journal of Econometrics, 234, 664–690. [812]

Deuber, D., Li, J., Engelke, S., and Maathuis, M. H. (2021), "Estimation and Inference of Extremal Quantile Treatment Effects for Heavy-Tailed Distributions," arXiv preprint arXiv:2110.06627. [812]

D’Haultfouillle, X., Maurel, A., and Zhang, Y. (2018), "Extremal Quantile Regressions for Selection Models and the Black–White Wage Gap," Journal of Econometrics, 203, 129–142. [812]

Evans, W. N., and Garthwaite, C. L. (2014), "Giving Mom a Break: The Impact of Higher EITC Payments on Maternal Health," American Economic Journal: Economic Policy, 6, 258–90. [818]
Ghanem, D., Hirshleifer, S., Kedagni, D., and Ortiz-Becerra, K. (2022), “Correcting Attrition Bias Using Changes-in-Changes,” arXiv preprint arXiv:2203.12740. [812]

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2021), “Extreme Conditional Expectile Estimation in Heavy-Tailed Heteroscedastic Regression Models,” Annals of Statistics, 49, 3358–3382. [816,823]

Guillou, A., and Hall, P. (2001), “A Diagnostic for Selecting the Threshold in Extreme Value Analysis,” Journal of the Royal Statistical Society, Series B, 63, 293–305. [815,816]

Haeusler, E., and Segers, J. (2007), “Assessing Confidence Intervals for the Tail Index by Edgeworth Expansions for the Hill Estimator,” Bernoulli, 13, 175–194. [814]

Hill, B. M. (1975), “A Simple General Approach to Inference about the Tail of a Distribution,” Annals of Statistics, 3, 1163–1174. [813,822]

Hoyes, H., Miller, D., Simon, D. (2015), “Income, the Earned Income Tax Credit, and Infant Health,” American Economic Journal: Economic Policy, 7, 172–211. [812,818,819]

Melly, B., and Santangelo, G. (2015), The Changes-in-Changes Model with Covariates, 1, Bern: Universität Bern. [812]

Resnick, S. I. (2007), Heavy-Tail Phenomena: Probabilistic and Statistical Modeling, New York: Springer. [813]

Sasaki, Y., and Wang, Y. (2022), “Fixed-k Inference for Conditional Extremal Quantiles,” Journal of Business & Economic Statistics, 40, 829–837. [818]

Zhang, Y. (2018), “Extremal Quantile Treatment Effects,” Annals of Statistics, 46, 3707–3740. [812,818]