# Switching to the New Norm: From Heuristics to Formal Tests using Integrable Empirical Processes

Tetsuya Kaji*
MIT

[Job Market Paper]

November 30, 2017

(Click here for the latest version)

**Abstract**

A frequent concern in empirical research is whether a handful of outlying observations have driven the key empirical findings. The current widespread practice in economics is to redo the statistical analysis adjusting for outliers and see if we obtain similar results, checking "robustness to outliers." However, such empirical practices have little theoretical justification, and researchers have had to rely on heuristic arguments. This paper constructs a formal statistical test of outlier robustness that accommodates many empirical settings. The key is to observe that statistics related to outlier robustness analysis are represented as $L$-statistics—integrals of empirical quantile functions with respect to sample selection measures—and to consider them in spaces equipped with appropriate norms. In particular, we characterize weak convergence of empirical distribution functions in the space of *bounded integrable* functions, establish the delta method for their inverses (empirical quantile functions) as maps from this space into the space of *integrable* functions, characterize weak convergence of random sample selection measures in the space of *bounded integrable* Lipschitz functions, and derive the delta method for $L$-statistics as maps from those spaces into a Euclidean space. As an empirical application, we revisit the outlier robustness analysis in Acemoglu et al. (2017) and demonstrate that our test can detect sensitivity of a parameter that was otherwise indiscernible had we relied on existing heuristics. Our theory of $L$-statistics is new and of independent interest; we propose other applications, including multiple testing problems and tests of higher-order Lorenz dominance.

*I thank Anna Mikusheva, Elena Manresa, Kengo Kato, Rachael Meager, Matthew Masten, Abhijit Banerjee, Daron Acemoglu, Isaiah Andrews, Hideatsu Tsukahara, Hidehiko Ichimura, Victor Chernozhukov, Jerry Hausman, Whitney Newey, Alberto Abadie, Michal Kolesár, Joshua Angrist, Andrey Malenko, Brendan K. Beare, and seminar participants at the MIT Econometrics Lunch, Kyoto University Econometrics Seminar, Keio University Econometrics Workshop, and the University of Tokyo Applied Statistics Workshop and Empirical Micro Research Seminar for helpful comments. I am also thankful to Daron Acemoglu and Pascual Restrepo for kindly sharing their data and code for reexamination of their paper. Email: `tkaji@mit.edu`.

# 1 Introduction

In empirical research in economics, a common concern is whether a handful of outlying observations may have driven crucial empirical findings. In estimating the effect of microcredit with experimental data, Augsburg et al. (2015) report that trimming 1% of the observations makes the effect of the loan program on business profits significant that is otherwise insignificant. The analysis of Acemoglu et al. (2001) on the effect of institutions on economic performance using differences in mortality rates prompted extensive discussion of whether outliers undermined the validity of mortality rates as instruments (Albouy, 2012; Acemoglu et al., 2012). Herndon et al. (2014) discuss whether the exclusion of some observations invalidates the findings in Reinhart and Rogoff (2010). Guthrie et al. (2012) find that a result in Chhaochharia and Grinstein (2009) is driven by outliers. De Long and Summers (1991) and de Long et al. (1992) find that machinery and equipment investment have a strong connection with economic growth, which is followed by discussion of whether outliers drove their findings (Auerbach et al., 1994; de Long and Summers, 1994).

It is thus considered an important characteristic of valid empirical findings that a small number of outliers do not affect the conclusion of analysis to a nonnegligible degree (Young, 2017). The common practice in empirical research is to carry out robustness checks by redoing the analyses on the sample that is adjusted for outliers (such as removal or winsorization) and comparing the results from the original ones relative to standard errors (e.g., Acemoglu et al., 2001, 2016, 2017; Agarwal et al., 2010; Alatas et al., 2016; Fabrizio et al., 2007). While such heuristic practices lack formal justification (as explained in Section 2.2), it is technically demanding to obtain the joint distribution of the estimates necessary to formalize the outlier robustness checks as statistical tests.

The main contribution of this paper is to develop a method to derive the joint distribution of full-sample and outlier-adjusted estimators for a wide range of sample selection procedures, including removal or winsorization at cutoffs that depend on the entire sample. With our results, we can test whether the parameter of interest changes its value significantly before and after such sample selection, enabling formal statistical investigation of outlier robustness checks. Many statistics related to outlier robustness analysis are represented as $L$-statistics—integrals of transformations of empirical quantile functions with respect to random sample selection measures. We develop a new empirical process theory tailored for these statistics, with an important innovation related to the choice of appropriate norms. Despite the long tradition of empirical process techniques in establishing asymptotic properties of $L$-statistics (Shorack and Wellner, 1986; Van der Vaart and Wellner, 1996; Koul, 2002), the literature has confined attention to empirical processes under the "uniform norm," which has imposed severe limitations to the range of applications; in particular, it did not cover some essential $L$-statistics that appear in outlier robustness analysis. In contrast, our theory employs appropriate norms and allows us to cover a very general form of $L$-statistics including them.

Our theoretical contribution consists of three key results: we consider empirical distribution functions in the space of bounded integrable functions and characterize weak convergence therein (Section 3.1); we consider empirical quantile functions in the space of integrable functions and establish the functional delta method for the map from distribution functions to quantile functions in these spaces (Section 3.2); we consider random sample selection measures in the space of bounded integrable Lipschitz functions and establish a functional delta method for $L$-statistics from these spaces (Section 3.3). Lastly, we obtain the formula for the joint asymptotic distribution of $L$-statistics and establish the validity of bootstrap for computing the asymptotic distribution.

The challenge in deriving asymptotic distributions of outlier-adjusted statistics is that the sample selection procedure often depends on the whole of the sample in ways that render classical multivariate central limit theorems inapplicable. As the first step of our analysis, we observe that many statistics related to outlier robustness analysis are given by $L$-statistics. An *$L$-statistic* is a quantity given by

$$\int_0^1 m(\mathbb{Q}_n(u))d\mathbb{K}_n(u),$$

where $m$ is a known continuously differentiable function, $\mathbb{Q}_n : (0,1) \to \mathbb{R}$ an empirical quantile function of some random variable $X_i$, and $\mathbb{K}_n : (0,1) \to \mathbb{R}$ a Lipschitz function that is possibly random.[1] Here, $\mathbb{K}_n$ represents the sample selection procedure (such as outlier removal or winsorization) that can be heavily dependent on the quantile function and other observations.

As a toy example, let us consider the problem of deriving the joint distribution of two sample means: the full sample mean and the $\alpha$-trimmed sample mean, the mean that drops $\alpha \in (0,1)$ portions of observations from both tails,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{and} \qquad \frac{1}{n-2\lfloor \alpha n\rfloor}\sum_{i=\lfloor \alpha n\rfloor+1}^{n-\lfloor \alpha n\rfloor} X_{(i)},$$

where $X_{(i)}$ denotes the $i$th smallest order statistic of $X_1, \dots, X_n$. Surprisingly, deriving the joint distribution of these statistics is a nontrivial problem, as the order statistics are highly dependent, and the trimmed mean cannot be represented as a simple sum of i.i.d. (or stationary) random variables, preventing the use of familiar central limit theorems. We tackle this problem by transforming them into the integral forms:

$$\int_0^1 \mathbb{Q}_n(u)du \qquad \text{and} \qquad \frac{n}{n-2\lfloor \alpha n\rfloor}\int_{\lfloor \alpha n\rfloor/n}^{1-\lfloor \alpha n\rfloor/n} \mathbb{Q}_n(u)du,$$

where $\mathbb{Q}_n$ is the empirical quantile function of $X_i$.[2] This formulation is susceptible to functional delta methods, once we know "weak convergence" of the empirical quantile process $\sqrt{n}(\mathbb{Q}_n - Q)$, where $Q : (0,1) \to \mathbb{R}$ is the population quantile function of $X_i$.

However, now we face a major difficulty; the empirical quantile process thus defined does not converge in the standard sense. If $X_i$ is supported on the whole of $\mathbb{R}$, the true quantile function $Q$ is unbounded on $(0,1)$. On the other hand, the empirical quantile function $\mathbb{Q}_n$ is by construction bounded on $(0,1)$ since for each $n$ there are only finitely many values $\mathbb{Q}_n$ can take (in particular, $X_1$ through $X_n$). It is clear, then, that the maximum distance between $\mathbb{Q}_n$ and $Q$ is infinity for every $n$, implying that $\mathbb{Q}_n$ does *not* converge to $Q$ uniformly. Corresponding to this point, Van der Vaart (1998, p. 317) notes that the functional delta method for $L$-statistics "is preferable in that it applies to more general statistics, but it...does not cover the simplest $L$-statistic: the sample mean." To

[1] The empirical quantile function is a generalized inverse of the empirical distribution function, in particular, $\mathbb{Q}_n(u) := \inf\{x \in \mathbb{R} : \mathbb{F}_n(x) \geq u\}$ where $\mathbb{F}_n(x) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_i \leq x\}$.

[2] Note that the trimmed mean can be further written as $\int_0^1 \mathbb{Q}_n dK_n$ for $K_n$ Lipschitz. See Section 2.

circumvent this issue, the literature on empirical processes has often confined itself to bounded or truncated quantile functions (Van der Vaart and Wellner, 1996, Chapter 3.9) or weighted empirical quantile processes that suitably down-weight the tails (Csörgő and Horváth, 1993). However, none of these methods work for our purpose, as we neither want to limit attention to bounded random variables nor can we expect that the random variables are weighted in such a nice way.

We solve this problem by considering the quantile functions in the $L_1$ space instead of in the traditional $L_\infty$ space. The important point is to realize that what we truly need is the convergence of *integrals* of quantile functions; uniform convergence of quantile functions, as often considered, is neither necessary nor sufficient for our purpose. In light of this, we first characterize weak convergence of empirical distribution processes in the space of bounded and integrable functions, and then establish the functional delta method for the map from such distribution functions $F$ to quantile functions $Q = F^{-1}$; this establishes weak convergence of empirical quantile processes in $L_1$. The key intuition in the proof is to observe that the $L_1$ norm is compatible with Fubini's theorem.

Given weak convergence of empirical quantile processes, we now proceed to weak convergence of the possibly random sample selection function $\mathbb{K}_n$ and the functional delta method for $(Q, K) \mapsto \int Q dK$. Note that

$$\int \mathbb{Q}_n d\mathbb{K}_n - \int Q dK = \int (\mathbb{Q}_n - Q) d\mathbb{K}_n + \int Q d(\mathbb{K}_n - K).$$

For the first term to converge whenever the sample average $\int \mathbb{Q}_n du$ does, we need $\mathbb{K}_n$ to be uniformly Lipschitz. For the second to be well-defined, we need that $\mathbb{K}_n$ converges to $K$ in $L_\infty$. Then by integration by parts, the second term can be approximately written as $-\int (\mathbb{K}_n - K) dQ$, meaning that $\mathbb{K}_n$ needs to converge to $K$ in $L_Q$, the space of functions integrable with respect to $Q$. This exercise reveals that the appropriate convergence of the sample selection function $\mathbb{K}_n$ can be established in, again, the space of bounded and integrable functions. Finally, we establish the functional delta method for the $L$-statistic, $(Q, K) \mapsto \int Q dK$ (more precisely, we allow transformations of quantile functions, $\int m(Q) dK$).

The utility of our functional delta method approach is not only the generality it brings but also that it implies the validity of the nonparametric bootstrap. This allows researchers to avoid making strong distributional assumptions to derive the asymptotic distributions of their estimators.

This paper characterizes the asymptotic distributions of many $L$-statistics in the form of Gaussian distributions. Note, however, that not all $L$-statistics converge weakly to Gaussian distributions; for example, the largest order statistic, appropriately scaled, often converges to some extreme value distribution (de Haan and Ferreira, 2006). In this sense, this paper can be seen as establishing the conditions under which a general form of $L$-statistics converges to a Gaussian distribution. The key to convergence toward Gaussian distributions is that the sample selection mechanisms become less and less dependent on $n$ as $n$ tends to infinity; in the outlier removal example, the threshold $\alpha$ does not approach 0 as $n$ tends to infinity. This assumption, however, may not be plausible in some applications. We note that our delta method results are potentially susceptible to generalizations to other nonstandard distributions; see Section 3.3.

Applying the theory developed thus far, we propose a test of outlier robustness that takes into account natural comovement of the two estimators. As an empirical application, we revisit the outlier robustness analysis discussed in Acemoglu et al. (2017) and carry out a formal statistical test as proposed in this paper. They estimate the effect of democracy on the GDP growth, and examine the sensitivity of their results to the removal of outliers on the residuals, in particular examining whether removing the extreme values in GDP growth would significantly change their findings. For all but one coefficient, the test is not rejected at 5% level, meaning that they exhibit robustness to such outlier removal. For the one rejected—persistence of the GDP growth—we show that the rejection would not have been "detectable" had we relied on the heuristic testing procedure commonly practiced in the literature.

The theory of $L$-statistics developed in this paper is itself new and of independent interest; it can be used to solve other econometric problems aside from outlier robustness analyses. Kaji and Kang (2017) define a class of risk measures subsuming Value-at-Risk and expected shortfall that can incorporate estimation errors into the risks being estimated. The asymptotic results in their paper use the theory developed in this paper. Kaji (2017) interprets quantile treatment effects as individual treatment effects that attain the lower bound of the total absolute treatment effect and proposes a variant of subgroup treatment effects to assess the heterogeneity of treatment effects. Again, the asymptotic properties follow from the results of the present paper. In the main text, we also discuss applications to multiple testing problems and tests of higher-order Lorenz dominance.

The rest of the paper is organized as follows. Section 2 defines the class of $L$-statistics considered in this paper and discusses how it subsumes many statistics widely used in economics. Section 2 also elaborates on the outlier robustness analysis and explains how outlier robustness can be tested using the asymptotic distributions of $L$-statistics. Section 3 describes the main theoretical contribution of the paper; it develops the asymptotic theory of $L$-statistics using integrable empirical processes and functional delta methods. The exposition is aimed to be minimal and intuitive, leaving most of the details to Appendices. Section 4 discusses an approach for testing outlier robustness and revisits the outlier robustness analysis of Acemoglu et al. (2017). Section 5 applies the asymptotic theory of $L$-statistics to other econometric problems: multiple testing problems, tests of higher-order Lorenz dominance, tail risk measures by Kaji and Kang (2017), and heterogeneous treatment effects by Kaji (2017). Section 6 reviews the related literature on empirical processes, $L$-statistics, and robust estimation. Finally, Section 7 concludes. All figures, tables, and proofs appear in the Appendices.

# 2 Setup and Motivation

## 2.1 $L$-statistics

This paper concerns statistics that are averages of functions of independent and identically distributed (i.i.d.) random variables, where each observation may be omitted or weighted differently from other observations. To fix this idea, let $X_i$ be an i.i.d. scalar random variable and $w_i$ a possibly random weight whose distribution is assumed to be bounded but is allowed to depend on all of $X_i$. Consider a statistic of the form

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} m(X_i)w_i,$$

where $m$ is some continuously differentiable function. For example, the sample average is such a statistic where $m$ is an identity and $w_i$ is identically one; the sample average from 1st to 99th percentiles (excluding the bottom and top 1% of observations) is also such a statistic for which $m$ is an identity and $w_i$ is the indicator of whether $X_i$ falls between the 1st and 99th percentiles of $X_1, \dots, X_n$. Thus, $w_i$ captures the idea that each $X_i$ may be weighted (or excluded) in an interdependent way.

Note that rearranging the summands does not affect the sum itself. In particular, let $X_{(i)}$ be the order statistic of $X_i$; $X_{(1)}$ represents the smallest observation, $X_{(2)}$ the second smallest, and so on. Denote by $w_{(i)}$ the weight corresponding to $X_{(i)}$ (so $w_i$ is sorted according to the order of $X_i$). Then, one may rewrite the average without loss of generality as

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} m(X_{(i)}) w_{(i)}.$$

This formulation is known as an *L-statistic*, where "$L$" stands for the fact that $\hat{\beta}$ is a *linear* combination of functions of order statistics $X_{(i)}$. Many statistics commonly used in economics are $L$-statistics, as will be shown in examples below.

We develop a method to derive the distribution of $\hat{\beta}$ using empirical quantile functions. Let $\mathbb{Q}_n(u)$, $u \in (0,1)$, be the empirical $u$-quantile of $X_i$, that is,

$$\mathbb{Q}_n(u) := \begin{cases} X_{(1)} & u \in \bigl(0, \frac{1}{n}\bigr], \\ X_{(i)} & u \in \bigl(\frac{i-1}{n}, \frac{i}{n}\bigr], \\ X_{(n)} & u \in \bigl(\frac{n-1}{n}, 1\bigr). \end{cases}$$

Using this, one can write

$$\hat{\beta} = \int_0^1 m(\mathbb{Q}_n(u)) d\mathbb{K}_n(u),$$

where $\mathbb{K}_n$ is the measure that assigns density $w_{(i)}$ to $u \in \bigl(\frac{i-1}{n}, \frac{i}{n}\bigr]$. Although the two representations are mathematically equivalent, the first representation as a sum of order statistics evokes the multivariate central limit theorems, while the second representation as an integral evokes the functional central limit theorems and functional delta methods. Correspondingly, there are two methods to derive the asymptotic distribution of $L$-statistics—the Hájek projection and the functional delta method—each of which covers nonoverlapping quantities (Van der Vaart, 1998, Chapter 22). For example, the Hájek projection covers the full sample average, while the functional delta method covers plug-ins of estimated quantile functions. However, nonoverlapping coverage can be problematic when we want the *joint* distribution of various $L$-statistics, as in the outlier robustness analysis. While we leave further comparison of the two methods to Section 6.2, this paper achieves substantial generalization of the second method that is enough to accommodate quite general forms of $L$-statistics useful for outlier robustness analyses and other problems.

To wrap up, our objective is to derive the joint distribution of finitely many statistics of the form $\hat{\beta} = \int_0^1 m(\mathbb{Q}_n) d\mathbb{K}_n$; in particular, $\hat{\beta}_j = \int_0^1 m_j(\mathbb{Q}_{n,j}) d\mathbb{K}_{n,j}$, $j = 1, \dots, d$.

## 2.2 Outlier robustness analysis

This section clarifies the motivation of outlier robustness analysis and explains why $L$-statistics are useful for this purpose.

What is the problem in current heuristic practice? Let $\hat{\beta}_1$ be the key estimator on which our empirical findings are based. When we want to claim that $\hat{\beta}_1$ is "not the consequence of only a few outlying observations," we often compute another estimator $\hat{\beta}_2$ from the sample that excludes some outliers and then argue that their difference is not too large compared to the standard error estimated for $\hat{\beta}_1$. However, comparing the difference $\hat{\beta}_1 - \hat{\beta}_2$ to the marginal standard error of $\hat{\beta}_1$ does not make much sense from a statistical point of view. Naturally, $\hat{\beta}_1$ and $\hat{\beta}_2$ are based on almost identical sets of observations. Therefore, even if the contribution of outliers is fairly large, the difference of the two estimators can be, by construction, much smaller than the marginal standard error of $\hat{\beta}_1$. Moreover, it so happens that the asymptotic distribution of an efficient estimator is independent of its difference from another estimator; then, such empirical practices may not be susceptible to an interpretation as a meaningful statistical testing procedure of some hypothesis.

What does a researcher want to investigate by checking "robustness" to outliers? As an example, consider the problem of estimating the treatment effect of microcredit provision on households' business profits in rural villages in some country. Let $\beta_1$ be the true average treatment effect and suppose that its estimate $\hat{\beta}_1$ is significantly positive. We may then suggest policy implications such as "Since $\hat{\beta}_1$ is significantly positive, we recommend to expand availability of microcredit to all villages in this country." However, we are worried that such a finding may be mostly driven by some "outlying" observations. For example, we are concerned about the possibility that the treatment effects are largely positive for above-the-poverty-line households while they can be negative for poor or extremely poor households, aggregating to a modestly positive average treatment effect. If this is the case, despite the average effect being positive, we may not wish to implement microcredit as it may exacerbate economic inequalities. In another scenario, we may be concerned that some extreme data points are not representative of the true population; for example, some respondents with limited literacy may have mistakenly answered their incomes as unreasonably high (or low) figures, and that may be driving the treatment effect unreasonably positive and significant. If so, again, we may not wish to base our policy recommendations on such imprecise measures.

In this setting, let $X_i$ be (a part of) household $i$'s characteristics; $X_i$ can be a regressor or can be a dependent variable. We are worried about the robustness of our findings to outliers of $X_i$; let $\beta_2$ be the true average treatment effect on the population that excludes the outlying portion of $X_i$, e.g., $\mathbb{E}[Y_{i1} - Y_{i0} \mid X_i \leq c]$. Concerns about the first scenario can be formulated as "the average effect $\beta_1$ does not represent the average effect among 'typical' individuals, $\beta_2$." Then, the null hypothesis subject to be tested in the outlier robustness analysis can be formulated as

$$H_0 : |\beta_1 - \beta_2| \leq h$$

for some $h \geq 0$. In the second scenario, we are concerned that outliers may not be from the true data generating process of interest, and they may be affecting the estimate too much. However, if outliers affect the findings of the statistical analyses only to a negligible degree, then we may say that our findings are robust to such possibilities. Then, the null hypothesis we want to test is, again, $H_0 : |\beta_1 - \beta_2| \leq h$ for some $h \geq 0$.

The choice of $h$ in the null hypothesis is an important practical question, but we treat it as given in this paper. This $h$ should be based on how much error can be tolerated in applying the empirical findings, and hence should be determined on a case-by-case basis.[3] That being said, we list a few special cases later in this section where the choice of $h$ is necessarily determined by the characteristics of the model. To summarize, if we develop a way to test the above hypothesis, we can formalize many heuristic arguments carried out in empirical research.[4]

To relate $L$-statistics to our context, consider the regression equation

$$y_i = x_i\beta + \varepsilon_i, \qquad \mathbb{E}[x_i\varepsilon_i] = 0,$$

where we estimate $\beta$ by the ordinary least squares (OLS) regression. Let us first consider cases where we compare another OLS estimator that excludes outliers of $y_i$ as in Acemoglu et al. (2016) or Banerjee et al. (2014). Now we have two estimators:

$$\hat{\beta}_1 = \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i, \qquad \hat{\beta}_2 = \left(\frac{1}{n}\sum_{i=1}^n x_i^2 w_i\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i w_i, \tag{1}$$

where $w_i = \mathbb{1}\{y_{(\lfloor \tau n\rfloor+1)} \leq y_i \leq y_{(n-\lfloor \tau n\rfloor)}\}$. Denote by $\mathbb{F}_n$ and $\mathbb{Q}_n$ the empirical distribution and empirical quantile functions of $x_i y_i$. Then, they can also be written as

$$\hat{\beta}_1 = \int_0^1 \mathbb{Q}_n(u)d\mathbb{K}_{n,1}(u), \qquad \hat{\beta}_2 = \int_0^1 \mathbb{Q}_n(u)d\mathbb{K}_{n,2}(u),$$

where $\mathbb{K}_{n,1}$ and $\mathbb{K}_{n,2}$ are random measures that assign, respectively, density $(\frac{1}{n}\sum_{i=1}^n x_i^2)^{-1}$ to $(0,1)$ and density $(\frac{1}{n}\sum_{i=1}^n x_i^2 w_i)^{-1} w_i$ to $u \in (\mathbb{F}_n(x_i y_i) - 1/n, \mathbb{F}_n(x_i y_i)]$. Along the same line, we can represent the two-stage least squares (2SLS) estimators as $L$-statistics as well.

We might instead think that outlying observations have some information and want to winsorize $x_i$ as in Acemoglu et al. (2012). Here, winsorization of $x_i$ at quantile $\tau$ means replacing every $x_{(i)}$ for $i = 1, \dots, \lfloor \tau n\rfloor$ by $x_{(\lfloor \tau n\rfloor+1)}$, and every $x_{(i)}$ for $i = n - \lfloor \tau n\rfloor + 1, \dots, n$ by $x_{(n-\lfloor \tau n\rfloor)}$. Thus, winsorization replaces "outliers" with the closest value that is considered non-outlier. Then, we would have

$$\hat{\beta}_1 = \left(\frac{1}{n}\sum_{i=1}^n x_i^2\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i, \qquad \hat{\beta}_2 = \left(\frac{1}{n}\sum_{i=1}^n x_i^2 w_i^2\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i w_i, \tag{2}$$

where

$$w_i = \begin{cases} x_{(\lfloor \tau n\rfloor+1)}/x_i & x_i < x_{(\lfloor \tau n\rfloor+1)} < 0, \\ x_{(n-\lfloor \tau n\rfloor)}/x_i & x_i > x_{(n-\lfloor \tau n\rfloor)} > 0, \\ 1 & \text{otherwise.} \end{cases}$$

These can be written as

$$\hat{\beta}_1 = \int_0^1 \mathbb{Q}_n(u)d\mathbb{K}_{n,1}(u), \qquad \hat{\beta}_2 = \int_0^1 \mathbb{Q}_n(u)d\mathbb{K}_{n,2}(u),$$

[3] In the empirical application in Section 4, we use the severest null $h = 0$. If one cannot reject the hypothesis with $h = 0$, that can be considered a "strong" indicator of robustness.

[4] One may wish to "test" whether the outliers affect the significance of the estimates. However, significance depends by construction on data and hence is not solely determined by the population characteristics; therefore, bringing it up in the null hypothesis is difficult to justify from a statistical point of view.

where $\mathbb{K}_{n,1}$ and $\mathbb{K}_{n,2}$ are, again, random measures that assign density $(\frac{1}{n}\sum_{i=1}^n x_i^2)^{-1}$ and density $(\frac{1}{n}\sum_{i=1}^n x_i^2 w_i^2)^{-1} w_i$ to $u \in (\mathbb{F}_n(x_i y_i) - 1/n, \mathbb{F}_n(x_i y_i)]$. If we can derive the joint distribution of the involved $L$-statistics, we are able to formally test our hypothesis about the outlier robustness.

Now let us look at a few special cases of linear regression models in which outlier removal will not cause the coefficient to change. First, if we have $\mathbb{E}[\varepsilon_i \mid x_i] = 0$, then any sample selection conditional on $x_i$ does not change the value of $\beta$. Therefore, outlier removal based on $x_i$ is harmless, and we can use $h = 0$. Second, if the conditional distribution of $\varepsilon_i$ conditional on $x_i$ is symmetric around zero and we remove samples symmetrically by $\varepsilon_i$, it will not cause any bias on $\beta$ (in reality, we remove by $\hat{\varepsilon}_i$, which consistently estimates $\varepsilon$). Third, if $\varepsilon_i$ is independent of $x_i$, then the sample selection based on $\varepsilon_i$ does not introduce bias on $\beta$ except for the intercept. However, if we select samples based on $y_i$, the true value of $\beta$ will almost always change.[5]

### 2.3 Notes on the setup

The $L$-statistics introduced so far share an important feature that the random measure $\mathbb{K}_n$ is "well-behaved" (in the sense defined precisely in the next section). The intuition is that the selection or weighting mechanism does not depend on the sample size $n$, at least asymptotically. The results of this paper apply in such contexts. The next example does not possess this feature and thus falls outside the scope of the theory developed in this paper.[6]

**Example** (Extreme order statistics)**.** The minimum of the observations $X_1, \dots, X_n$ can be written as

$$X_{(1)} = \frac{1}{n}\sum_{i=1}^{(1/n)n} nX_{(i)} = \int_0^1 \mathbb{Q}_n(u) d\mathbb{K}_n(u),$$

where $\mathbb{Q}_n$ is the empirical quantile of $X_i$, and $\mathbb{K}_n$ assigns density $n$ on $(0, 1/n]$ and zero elsewhere. Then $\mathbb{K}_n$ "converges" to the measure that assigns mass 1 to $u = 0$ and zero elsewhere, which is not absolutely continuous with respect to the Lebesgue measure. $\square$

## 3 Overview of Main Results

This section describes the key ideas and theoretical contributions of this paper. The formal mathematical development and proofs are given in the Appendices.

We recall our setup from Section 2.1. Let $(X_{i,1}, \dots, X_{i,d})$ be an i.i.d. random vector and $(w_{i,1}, \dots, w_{i,d})$ vector of possibly random weights whose distribution is bounded but allowed to depend on all of $X_{i,j}$. Denote by $\mathbb{F}_{n,j}$ the (marginal) empirical distribution function of $X_{1,j}, \dots, X_{n,j}$. We want to know the joint distribution of

$$\hat{\beta}_j = \int_0^1 m_j(\mathbb{Q}_{n,j}(u)) d\mathbb{K}_{n,j}(u), \qquad j = 1, \dots, d,$$

where $\mathbb{Q}_{n,j} := \mathbb{F}_{n,j}^{-1}$ denotes the generalized inverse of $\mathbb{F}_{n,j}$, $m_j : \mathbb{R} \to \mathbb{R}$ are continuously differentiable functions, and $\mathbb{K}_{n,j}$ possibly random measures. We derive this using the

[5] Or, if one regards $\beta$ as a fixed structural parameter, then it can be put as "the plim of popular estimators does not coincide with the structural $\beta$ any more."

[6] This does not mean that extension to such cases is impossible. See the end of Section 3.

asymptotic distributions of $\mathbb{F}_{n,j}$ and $\mathbb{K}_{n,j}$ and applying the corresponding functional delta methods for the map $(F,K) \mapsto \int_0^1 m(F^{-1})dK$.

What would be a plausible derivative formula for the delta method? Let $Q$ and $K$ be the population counterparts of $\mathbb{Q}_n$ and $\mathbb{K}_n$ and suppress dependence on $j$. Informal calculation suggests that

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \sqrt{n}\biggl(\int_0^1 m(\mathbb{Q}_n)d\mathbb{K}_n - \int_0^1 m(Q)dK\biggr) \\
&= \int_0^1 \sqrt{n}[m(\mathbb{Q}_n) - m(Q)]d\mathbb{K}_n + \int_0^1 m(Q)d\Bigl(\sqrt{n}(\mathbb{K}_n - K)\Bigr) \\
&\approx \int_0^1 \sqrt{n}[m(\mathbb{Q}_n) - m(Q)]dK - \int_0^1 \sqrt{n}(\mathbb{K}_n - K)dm(Q) \\
&\approx \int_0^1 m'(Q)\sqrt{n}(\mathbb{Q}_n - Q)dK - \int_0^1 m'(Q)\sqrt{n}(\mathbb{K}_n - K)dQ, \qquad (3)
\end{aligned}$$

where the third "equality" follows from integration by parts and the fourth from a delta method. One of the main goals of this paper is to give the conditions under which this derivation can be justified. The purpose of this section is to provide an accessible introduction to the issues involved, while the rigorous treatment is left to the Appendices.

We proceed in three steps:

Step 1. Explore in what sense the empirical distribution function $\mathbb{F}_n$ must converge, and give sufficient conditions for such convergence. Along the way, we will also find the right notion of convergence for the empirical quantile function $\mathbb{Q}_n := \mathbb{F}_n^{-1}$.

Step 2. Under the stated conditions, show that functions of the empirical quantile function $\mathbb{Q}_n$ do indeed converge in the required sense, and characterize its asymptotic distribution. The key is the functional delta method for $F \mapsto m(Q) = m(F^{-1})$.

Step 3. Formulate a proper convergence notion for $\mathbb{K}_n$. Combining these results, show that our $L$-statistics converge to a normal random vector, and obtain its formula. The key is the functional delta method for $(Q,K) \mapsto \int_0^1 m(Q)dK$.

### 3.1 Step 1: Convergence of empirical processes

The empirical process literature (Shorack and Wellner, 1986; Van der Vaart and Wellner, 1996; Kosorok, 2008; Dudley, 2014) shows that the classical empirical process $\sqrt{n}(\mathbb{F}_n - F)$ converges to a Gaussian process in $L_\infty$. Due to the choice of this norm (the *uniform norm*), such results are referred to as *uniform central limit theorems*. To proceed with our agenda, however, such classical results turn out to be insufficient.

To understand the difficulty we face, consider the *empirical quantile process* in analogy with the empirical process for distribution functions,

$$\sqrt{n}(\mathbb{Q}_n(u) - Q(u)), \qquad u \in (0,1).$$

If the support of the underlying distribution $F$ is unbounded (which is necessary to accommodate many empirically relevant problems in economics), the true quantile function $Q$ is an unbounded function on $(0,1)$, while the empirical quantile function $\mathbb{Q}_n$ is bounded for every $n$ by construction. Therefore, it immediately follows that the empirical quantile process, with no restrictions on its range, never converges in the traditional uniform sense. This is why the previous literature has restricted its attention to convergence of quantile processes of truncated or bounded random variables (Van der Vaart, 1998; Van der Vaart and Wellner, 1996) or of weighted versions of quantile processes so they are effectively bounded (Csörgő and Horváth, 1993).

The first key idea of this paper is to switch to a new norm on the space of quantile functions. Recall that our eventual target is the statistics represented by the integral of the empirical quantile functions; we are not interested in any kind of inference that requires uniform convergence of $\mathbb{Q}_n$ (such as Kolmogorov-Smirnov type tests or uniform confidence bands around a quantile function). Then, the appropriate space for our quantile functions would naturally be the space of integrable functions, and the corresponding notion of convergence be $L_1$; the classical uniform norm $L_\infty$ appears neither appropriate nor desirable. Thus, we give up the uniform convergence and seek the conditions under which the empirical quantile process $\sqrt{n}(\mathbb{Q}_n - Q)$ converges weakly *in* $L_1$. In light of this, define the following space.

**Definition.** Let $\mathbb{B}$ be the Banach space of measurable functions $z$ from $(0,1)$ to $\mathbb{R}$ with the norm

$$\|z\|_{\mathbb{B}} := \int_0^1 |z(u)|du.$$

Not all probability distributions have a quantile function that is integrable. Precisely, a quantile function is integrable if and only if the corresponding probability distribution has a finite first moment (Lemma A.1). One sees therefore that even if the empirical distribution function $\mathbb{F}_n$ converges to the true distribution function $F$ in the uniform sense (which is indeed the case for every probability distribution regardless of how many moments it has), it might not be the case that the empirical quantile function $\mathbb{Q}_n$ converges to the true quantile function $Q$ in $L_1$. In other words, the inverse map $F \mapsto Q := F^{-1}$, when viewed as a map from $L_\infty$ to $L_1$, is not even continuous, let alone differentiable. This is why the classical uniform central limit theorems are not suitable for our purpose; we need to make use of a sufficiently strong norm on the space of distribution functions that ensures the existence of at least the first moment. Put together, the norm must be strong enough that the inverse map $F \mapsto Q$ be differentiable, but not so strong that it excludes many distributions of our potential interest.

The second key idea of this paper is to observe that the integrability of quantile functions is equivalent to the integrability of distribution functions by integration by parts. In particular, we require the distribution function $F$ to be "integrable" in the sense that its modification

$$\tilde{F}(x) = \begin{cases} F(x) & x < 0 \\ F(x) - 1 & x \geq 0 \end{cases}$$

is integrable. The adequacy of this norm is intuitively understood by observing that the quantile function is integrable if and only if the modification of the distribution function is integrable (Lemma A.1).

The precise definition of the norm is as follows.

**Definition.** Let $-\infty \le a < c < b \le \infty$ and $\mu$ be a positive Lebesgue-Stieltjes measure on $(a,b)$. Define the space $\mathbb{L}_\mu$ of $\mu$-measurable functions $z : (a,b) \to \mathbb{R}$ with limits $z(a) := \lim_{x\to a} z(x)$ and $z(b) := \lim_{x\to b} z(x)$, and the norm

$$\|z\|_{\mathbb{L}_\mu} := \|z\|_\infty \vee \|\tilde{z}\|_\mu := \left(\sup_{x\in(a,b)} |z(x)|\right) \vee \left(\int_a^b |\tilde{z}(x)| d\mu(x)\right)$$

where

$$\tilde{z}(x) := \begin{cases} z(x) - z(a) & x < c, \\ z(x) - z(b) & x \ge c. \end{cases}$$

**Definition.** Let $\mathbb{L}$ be the special case of $\mathbb{L}_\mu$ where $(a,b,c) = (-\infty,\infty,0)$ and $\mu$ be equal to the Lebesgue measure. The space of distribution functions is the subset $\mathbb{L}_\phi$ of $\mathbb{L}$ of functions $z$ that are monotone and cadlag with $z(-\infty) = 0$ and $z(+\infty) = 1$.

Note that we still require the distribution function to converge uniformly (the $L_\infty$ part of the norm); this ensures that the "inverse function" is well defined. Being the intersection of the familiar spaces $L_\infty$ and $L_1$, weak convergence in $\mathbb{L}$ implies convergence in both.

Henceforth we will focus on distributions $F$ that are members of $\mathbb{L}$ and prove weak convergence of empirical processes $\sqrt{n}(\mathbb{F}_n - F)$ *in* $\mathbb{L}$. Eventually, we want to show that this convergence of the empirical processes implies convergence of the empirical quantile processes $\sqrt{n}(\mathbb{Q}_n - Q)$ in $\mathbb{B}$. Since convergence in $\mathbb{L}$ is a stronger requirement than convergence in $L_\infty$, we cannot rely on classical results to show convergence in our norm; now we develop the conditions for our convergence. The next theorem gives the complete characterization of weak convergence in $\mathbb{L}_\mu$.

**Theorem 1** (Characterization of weak convergence in $\mathbb{L}_\mu$)**.** *The sequence of processes $X_n : \Omega \to \mathbb{L}_\mu$ converges weakly in $\mathbb{L}_\mu$ if and only if all of the following three conditions are met. (We denote $X_n(\omega)(t)$ by $X_n(t)$.)*

*(i) Every finite marginal $(X_n(t_1), \ldots, X_n(t_k))$ converges weakly in $\mathbb{R}^k$ for every $k$.*

*(ii) There exists a semimetric $\rho_1$ on $(a,b)$ such that $(a,b)$ is totally bounded in $\rho_1$ and $X_n$ is asymptotically uniformly $\rho_1$-equicontinuous in probability, that is, for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that*

$$\limsup_{n\to\infty} P\left(\sup_{\rho_1(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon\right) < \eta.$$

*(iii) There exists a semimetric $\rho_2$ on $(a,b)$ such that $(a,b)$ is totally bounded in $\rho_2$ and $X_n$ is asymptotically $(\rho_2,\mu)$-equiintegrable in probability, that is, for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that*

$$\limsup_{n\to\infty} P\left(\sup_{t\in\mathbb{R}} \int_{0<\rho_2(s,t)<\delta} |\tilde{X}_n(s)| d\mu(s) > \varepsilon\right) < \eta.$$

*Remark.* The classical empirical process literature shows that weak convergence in $L_\infty$ is connected to the Arzelá-Ascoli theorem (Van der Vaart and Wellner, 1996, Chapter 1.5). This is to say that for the sequence of stochastic processes to converge weakly uniformly, the elements of the sequence must be equally uniformly continuous. The upshot of the above theorem is that, in order for weak convergence in $L_1$ to take place additionally, the elements of the sequence must be "equally integrable" as well. This insight is reminiscent of the Dunford-Pettis theorem in functional analysis.

Despite its technical complexity, the conditions of the theorem are not necessarily difficult to check. It is known that the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ satisfies conditions (i) and (ii) (Van der Vaart and Wellner, 1996, Examples 2.1.3 and 2.5.4). The following proposition shows that if $F$ has slightly more than variance, then it also satisfies condition (iii).

**Proposition 2** (Convergence of empirical processes)**.** *Let $F$ be a probability distribution function on $\mathbb{R}$ with a $(2+\varepsilon)$th moment for some $\varepsilon > 0$.*[7] *Then the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ converges weakly in $\mathbb{L}$ to a Gaussian process with mean zero and covariance function* $\operatorname{Cov}(x, y) = F(x \wedge y) - F(x)F(y)$.

### 3.2 Step 2: Convergence of quantile processes

Now we proceed on to weak convergence of the empirical quantile process $\sqrt{n}(\mathbb{Q}_n - Q)$ in $\mathbb{B}$. This is established by showing that the inverse map $F \mapsto Q = F^{-1}$ is Hadamard differentiable as a map from $\mathbb{L}$ to $\mathbb{B}$. Weak convergence of the empirical quantile process then follows by the functional delta method.

The next theorem establishes Hadamard differentiability of the inverse map.

**Theorem 3** (Differentiability of the inverse map)**.** *Let $F \in \mathbb{L}_\phi$ be a distribution function that has at most finitely many jumps and is otherwise continuously differentiable with a strictly positive density. Then the inverse map $\phi : \mathbb{L}_\phi \to \mathbb{B}$, $\phi(F) = Q$, is Hadamard differentiable at $F$ tangentially to the set $\mathbb{L}_0$ of all continuous functions in $\mathbb{L}$. The derivative map, $\phi'_F : \mathbb{L}_0 \to \mathbb{B}$, is given by*

$$\phi'_F(z)(u) = -z(Q(u))Q'(u), \qquad u \in (0, 1).$$

Importantly, the derivative formula, we find, is the same as the one known in the literature for the uniform norm (Van der Vaart and Wellner, 1996, Section 3.9.4.2). Note that, although they are both about the "same" operator $\phi : F \mapsto Q$, the derivative formula need not be the same as we have changed the norm. The delta method states that the distribution of a function of a statistic is characterized by the derivative and the distribution of the statistic. Then, that the derivative formula stays unchanged reveals a relieving fact that we do not need to worry about the unboundedness of the quantile functions when it comes to integrating them; we may continue using the same old formula.

We summarize the main conclusion of this section.

**Proposition 4** (Convergence of quantile processes)**.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function. For a distribution function $F$ on $\mathbb{R}$ that has at most finitely many jumps and is otherwise continuously differentiable with strictly positive density*

[7] The classical central limit theorems only require finite variance. This marginal gap between the classical central limit theorems and the $L_1$ functional central limit theorem is mentioned in del Barrio et al. (1999). This is the "cost of generality" we pay in this paper. In some cases, however, it is possible to show that the second moment is sufficient, e.g., as in Shorack and Wellner (1986, Chapter 19) and **?**.

*such that $m(X)$ has a $(2+\varepsilon)$th moment for $X \sim F$ and some $\varepsilon > 0$, the process $\sqrt{n}(m(\mathbb{Q}_n) - m(Q))$ converges weakly in $\mathbb{B}$ to a Gaussian process with mean zero and covariance function* $\mathrm{Cov}(s,t) = m'(Q(s))Q'(s)m'(Q(t))Q'(t)(s \wedge t - st)$.

In addition to inversion, the proposition allows for transformation $m$. While we leave the formal treatment to the Appendices, we provide an intuitive discussion of this generalization in the remainder of this section.

Assume for simplicity that $m$ is increasing. Observe that by integration by parts,

$$\int m(x)dF = -\int \tilde{F}dm(x).$$

This indicates that existence of the expectation of the random variable $m(X)$ is equivalent to the distribution function $F$ of $X$ belonging to the space $\mathbb{L}_m$ of functions that are integrable with respect to $m$. Meanwhile, by the change of variables $u = F \circ m^{-1}(x)$,

$$\int m(x)dF = \int xdF \circ m^{-1} = \int (F \circ m^{-1})^{-1}du = \int m(Q)du.$$

Combine the results as follows. If $X$ is such that $m(X)$ has a finite first moment, then $F$ belongs to $\mathbb{L}_m$. This is equivalent to saying that $F \circ m^{-1}$ belongs to $\mathbb{L}$. Now we invoke Theorem 3 to find that its inverse $(F \circ m^{-1})^{-1}$ is in $\mathbb{B}$. Since $(F \circ m^{-1})^{-1} = m(Q)$, it follows that $m(Q)$ is in $\mathbb{B}$. Finally, if, in addition, $m(X)$ has (slightly more than) a variance, then its "empirical distribution function" $\mathbb{F}_n \circ m^{-1}$ converges weakly in $\mathbb{L}$ by Proposition 2 and hence the result follows by the delta method just established.

### 3.3 Step 3: Convergence of $L$-statistics

The last step is to show that the $L$-statistics of the form $\int m_j(\mathbb{Q}_{n,j})d\mathbb{K}_{n,j}$, $j = 1, \dots, d$, jointly converge weakly to a normal vector. Again, this is achieved by proving that the $L$-statistics, when seen as a map, are Hadamard differentiable. But for this, we need to take care of the randomness that arises from the measure $\mathbb{K}_n$.

By the informal exercise in (3), the appropriate notion of convergence for $\mathbb{K}_n$ is expected to involve integrability. It turns out that the norm developed in Section 3.1 does the right job. Here we recall the definition with specialization to the unit interval.

**Definition.** For a quantile function $Q : (0,1) \to \mathbb{R}$, denote by $\mathbb{L}_Q$ the Banach space of functions $\kappa : (0,1) \to \mathbb{R}$ with the norm

$$\|\kappa\|_{\mathbb{L}_Q} := \Big(\sup_{u \in (0,1)} |\kappa(u)|\Big) \vee \Big(\int_0^1 |\tilde{\kappa}(u)|dQ(u)\Big).$$

where $\tilde{\kappa}(u) := \kappa(u) - \kappa(0)\mathbb{1}\{u \le 1/2\} - \kappa(1)\mathbb{1}\{u > 1/2\}$. Define by $\mathbb{L}_{Q,M}$ the subset of $\mathbb{L}_Q$ of Lipschitz functions whose Lipschitz constants are uniformly bounded by $M$.

Now we are ready to show Hadamard differentiability of $L$-statistics. Fortunately, the derivative formula in the next theorem confirms our intuition in equation (3).

**Theorem 5** (Differentiability of $L$-statistics). *For each $M$, the maps $\lambda : \mathbb{B} \times \mathbb{L}_{Q,M} \to \mathbb{R}$ and $\tilde{\lambda} : \mathbb{B} \times \mathbb{L}_{Q,M} \to L_\infty(0,1)^2$,*

$$\lambda(Q,K) = \int_0^1 Q(u)dK(u) \qquad \textit{and} \qquad \tilde{\lambda}(Q,K)(s,t) = \int_s^t Q(u)dK(u),$$

*are Hadamard differentiable at $(Q, K) \in \mathbb{B}_\phi \times \mathbb{L}_{Q,M}$ uniformly over $\mathbb{L}_{Q,M}$ tangentially to the set $\mathbb{B} \times \mathbb{L}_{Q,0}$ where $\mathbb{L}_{Q,0}$ is the subset of $\mathbb{L}_Q$ of continuous functions $\kappa$ such that $Q(u)\kappa(u) \to 0$ as $u \to \{0,1\}$. The derivative is given by*

$$\lambda'_{Q,K}(z,\kappa) = \int_0^1 Q(u)d\kappa(u) + \int_0^1 z(u)dK(u),$$
$$\tilde{\lambda}'_{Q,K}(z,\kappa)(s,t) = \int_s^t Q(u)d\kappa(u) + \int_s^t z(u)dK(u),$$

*where $\int Qd\kappa$ is defined via integration by parts if $\kappa$ is of unbounded variation.*

Thus, for Hadamard differentiability of $L$-statistics, we require that the random "distribution function" $\mathbb{K}_n$ be uniformly Lipschitz, that is, when seen as a measure, $\mathbb{K}_n$ has a uniformly bounded density with respect to the Lebesgue measure.

Do selection measures such as outlier removal or winsorization satisfy this condition? If so, how can we verify it? Are there more primitive conditions that are easy to check? To answer these questions, consider the randomly weighted sum $\frac{1}{n}\sum_{i=1}^n X_i w_i$, or equivalently, $\frac{1}{n}\sum_{i=1}^n X_{(i)} w_{(i)}$ where $w_{(i)}$ is sorted according to the order of $X_i$. Using the empirical quantile function $\mathbb{Q}_n$ of $X_i$, we write this sum as an integral:

$$\int_0^1 \mathbb{Q}_n d\mathbb{K}_{n,0} \qquad \text{where} \qquad \mathbb{K}_{n,0}(u) := \frac{1}{n}\sum_{i=1}^{\lfloor nu \rfloor} w_{(i)} = \frac{1}{n}\sum_{i=1}^{n} w_{(i)} \times \begin{cases} 0 & u < \frac{i}{n}, \\ 1 & \frac{i}{n} \le u. \end{cases}$$

This function $\mathbb{K}_{n,0}$ is simple enough but the results developed in this paper require that this function be Lipschitz. We accomplish this by linearly interpolating $\mathbb{K}_{n,0}$, as $\mathbb{Q}_n$ is piecewise constant on $1/n$ intervals. In particular, we can replace the integral by

$$\int_0^1 \mathbb{Q}_n d\mathbb{K}_n \qquad \text{where} \qquad \mathbb{K}_n(u) := \frac{1}{n}\sum_{i=1}^{n} w_{(i)} \times \begin{cases} 0 & u < \frac{i-1}{n}, \\ n\big(u - \frac{i-1}{n}\big) & \frac{i-1}{n} \le u < \frac{i}{n}, \\ 1 & \frac{i}{n} \le u. \end{cases}$$

Since $\mathbb{F}_n(X_{(i)}) = i/n$, we can write $\mathbb{K}_n(u)$ as

$$\frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{1}\big\{0 \vee (nu - n\mathbb{F}_n(X_i) + 1) \wedge 1\big\}.$$

Therefore, as long as $w_i$ is bounded by some constant, this $\mathbb{K}_n$ is Lipschitz almost surely.

**Proposition 6** (Convergence of selection measures)**.** *Let $U_1, \dots, U_n$ be independent uniformly distributed random variables on $(0,1)$ and $w_{1,n}, \dots, w_{n,n}$ random variables bounded by some constant $M$ whose distribution can depend on $U_1, \dots, U_n$ and $n$. Define*

$$\mathbb{F}_n(u) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{U_i \le u\}, \qquad \mathbb{G}_n(u) := \frac{1}{n}\sum_{i=1}^{n} w_{i,n}\mathbb{1}\{U_i \le u\}.$$

*Let $I(u) := u$ and assume that $K(u) := \lim_{n\to\infty} \mathbb{E}[\mathbb{G}_n(u)]$ exists and is Lipschitz and differentiable. If $\sqrt{n}(\mathbb{G}_n - K)$ weakly converges jointly with $\sqrt{n}(\mathbb{F}_n - I)$ in $L_\infty$, then for the "selection" measure*

$$\mathbb{K}_n(u) := \frac{1}{n}\sum_{i=1}^{n} w_i \mathbb{1}\big\{0 \vee \big(nu - n\mathbb{F}_n(U_i) + 1\big) \wedge 1\big\},$$

*we have $\sqrt{n}(\mathbb{K}_n - K)$ converge weakly in $\mathbb{L}_Q$ for every quantile function $Q$ whose distribution has a $(2+c)$th moment for some $c > 0$.*

This means that most "well-behaved" sample selection measures converge in $\mathbb{L}_Q$; roughly speaking, if the empirical distribution of the selected sample $X_{1,n}, \dots, X_{m,n}$ converges in the traditional uniform sense together with that of the entire sample, then the selection measure $\mathbb{K}_n$ as defined in Section 2 converges in $\mathbb{L}_Q$. This can be verified as follows.

**Example 1** (Outlier robustness analysis)**.** Let $F$ be the true distribution of $x_i y_i$. In this example, $\mathbb{F}_n$ and $\mathbb{G}_n$ in Proposition 6 are

$$\mathbb{F}_n(u) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{F(x_i y_i) \le u\}, \qquad \mathbb{G}_n(u) := \frac{1}{n}\sum_{i=1}^n w_{i,n}\mathbb{1}\{F(x_i y_i) \le u\}.$$

Since both $y_{(\lfloor \tau n \rfloor + 1)}$ and $y_{(n - \lfloor \tau n \rfloor)}$ converge almost surely to $Q_y(\tau)$ and $Q_y(1-\tau)$, we have that for outlier removal at $\tau$- and $(1-\tau)$-quantiles (1), $w_{i,n}$ converges almost surely to $\mathbb{1}\{Q_y(\tau) \le y \le Q_y(1-\tau)\}$ and for winsorization at $\tau$- and $(1-\tau)$-quantiles (2), to

$$w_i = \begin{cases} Q_y(\tau)/y_i & y_i < Q_y(\tau) < 0, \\ Q_y(1-\tau)/y_i & y_i > Q_y(1-\tau) > 0, \\ 1 & \text{otherwise.} \end{cases}$$

Then, $\mathbb{F}_n$ and $\mathbb{G}_n$ jointly converge uniformly, respectively to an identity function and $\mathbb{E}[w_i \mid F(x_i y_i) = u]$. $\square$

Now we are ready to state the main result of this paper: the joint asymptotic distribution of general $L$-statistics.

**Proposition 7** (Convergence of $L$-statistics)**.** *Let $m_1, m_2 : \mathbb{R} \to \mathbb{R}$ be continuously differentiable functions and $F : \mathbb{R}^2 \to [0,1]$ be distribution function on $\mathbb{R}^2$ with marginal distributions $F_1, F_2$ that have at most finitely many jumps and are otherwise continuously differentiable with strictly positive marginal densities such that $m_1(X_1)$ and $m_2(X_2)$, $(X_1, X_2) \sim F$, have a $(2+\varepsilon)$th moment for some $\varepsilon > 0$. Along with independent and identically distributed random variables $X_{1,1}, \dots, X_{n,1}$ and $X_{1,2}, \dots, X_{n,2}$, let $w_{1,n,1}, \dots, w_{n,n,1}$ and $w_{1,n,2}, \dots, w_{n,n,2}$ be random variables bounded by some constant $M$ whose distribution can depend on $n$ and all of $X_{1,1}, \dots, X_{n,1}$ and $X_{1,2}, \dots, X_{n,2}$ such that the empirical distributions of $X_{i,1}$, $X_{i,2}$, $w_{i,n,1}X_{i,1}$, and $w_{i,n,2}X_{i,2}$ converge uniformly jointly to continuously differentiable distribution functions. Then, the normalized $L$-statistics*

$$\sqrt{n}\begin{pmatrix} \mathbb{E}_n[m_1(X_{i,1})w_{i,n,1}] - \mathbb{E}[m_1(X_{i,1})w_{i,n,1}] \\ \mathbb{E}_n[m_2(X_{i,2})w_{i,n,2}] - \mathbb{E}[m_2(X_{i,2})w_{i,n,2}] \end{pmatrix} \\ = \sqrt{n}\begin{pmatrix} \int_0^1 m_1(\mathbb{Q}_{n,1})d\mathbb{K}_{n,1} - \int_0^1 m_1(Q_1)dK_1 \\ \int_0^1 m_2(\mathbb{Q}_{n,2})d\mathbb{K}_{n,2} - \int_0^1 m_2(Q_2)dK_2 \end{pmatrix}$$

*where*

$$\mathbb{K}_{n,j}(u) := \frac{1}{n}\sum_{i=1}^n w_{i,n,j}\mathbb{1}\left\{0 \vee \left(u - \mathbb{F}_{n,j}(X_i) + \frac{1}{n}\right) \wedge \frac{1}{n}\right\},$$
$$K_j(u) := \lim_{n\to\infty} \mathbb{E}[w_{i,n,j} \mid F_j(X_{i,j}) \le u],$$

*converge weakly in $\mathbb{R}^2$ to a normal vector $(\xi_1, \xi_2)$ with mean zero and (co)variance*

$$\begin{aligned}\mathrm{Cov}(\xi_j, \xi_k) = \int_0^1 \int_0^1 & m_j'(Q_j(s))Q_j'(s)m_k'(Q_k(t))Q_k'(t)\times \\ & \Big([F_{jk}(s,t) - st] + [K_{jk}(s,t)F_{jk}(s,t) - stK_j(s)K_k(t)] \\ & - K_j(s)[F_{jk}(s,t) - st] - K_k(t)[F_{jk}(s,t) - st]\Big)dsdt,\end{aligned}$$

*where $F_{jk}(s,t) := \Pr(X_{i,j} \le Q_j(s), X_{i,k} \le Q_k(t))$ and $K_{jk}(s,t) := \lim_{n\to\infty} \mathbb{E}[w_{i,n,j}w_{i,n,k} \mid X_{i,j} \le Q_j(s), X_{i,k} \le Q_k(t)]$.*

In applications, one can compute the distribution either analytically, by parametric bootstrap, or by nonparametric bootstrap. Nonparametric bootstrap does not require distributional assumptions and can be quite convenient when one iteration of the estimation does not consume much time. The following is a procedure for the nonparametric bootstrap, stated for completeness.

**Proposition 8** (Validity of nonparametric bootstrap)**.** *In the assumptions stated in Proposition 7, assume further that $w_{i,n,j}$ represents sample selection based on a fixed number of empirical quantiles.*[8] *Then, the joint distribution of $(\hat\beta_1, \dots, \hat\beta_d)$ can be computed by nonparametric bootstrap. The algorithm is as follows. Here, $X_i$ denotes a vector $(X_{i,1}, \dots, X_{i,d})$.*

*i. Bootstrap $n$ (or fewer) random observations from $X_1, \dots, X_n$ with replacement.*

*ii. Compute the statistics $(\hat\beta_1^*, \dots, \hat\beta_d^*)$ for the bootstrapped sample.*

*iii. Repeat the above steps $S$ times.*

*iv. Use the empirical distribution of $(\hat\beta_1^*, \dots, \hat\beta_d^*)$ as the approximation to the theoretical asymptotic distribution of $(\hat\beta_1, \dots, \hat\beta_d)$.*

We have hitherto assumed that $X_i$ of interest is univariate. Multivariate cases, as in regressions, can be accommodated as follows.

**Example** (Multivariate regression)**.** Let $x_i = (1, x_{i1}, x_{i2})'$ and $\beta = (\beta_0, \beta_1, \beta_2)'$, and consider $y_i = x_i'\beta + \varepsilon_i$. The OLS estimator for $\beta$ is

$$\hat\beta = \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i y_i = \beta + \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^n x_i \varepsilon_i.$$

Therefore,

$$\sqrt{n}(\hat\beta_1 - \beta_1) = c_1 \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i + c_2 \frac{1}{\sqrt{n}}\sum_{i=1}^n x_{i,1}\varepsilon_i + c_3 \frac{1}{\sqrt{n}}\sum_{i=1}^n x_{i,2}\varepsilon_i + o_P(1)$$

for some constants $c_1$, $c_2$, and $c_3$. Thus, one can reduce the weak convergence of the vector $\sqrt{n}(\hat\beta - \beta)$ to the joint convergence of univariate empirical quantiles of $\varepsilon_i$, $x_{i,1}\varepsilon_i$, and $x_{i,2}\varepsilon_i$. □

[8] The assumption on convergence must be extended (from bivariate) to joint over all processes involved.

We note some possibilities for generalizing our results to other cases that are not considered in this paper. For instance, we do not explicitly consider cases where data are dependent (Dehling et al., 2002, 2014), where smoothed or estimated cdfs are substituted for empirical cdfs (Hall et al., 1999; Berg and Politis, 2009), or where a non-conventional convergence device such as extreme value theory is used for the stochastic processes (Einmahl, 1992; Rootzén, 2009; Drees and Rootzén, 2010, 2016). The part that requires additional work is the proof that $\mathbb{F}_n$ and $\mathbb{K}_n$ in each case converge weakly in our space $\mathbb{L}_\mu$. Fortunately, since we have completely characterized weak convergence in $\mathbb{L}_\mu$ in Theorem 1 without relying particularly on the central limit theorem structure (see Section 6.1 and Appendix A.4), half of such work is already taken care of. Once convergence of $\mathbb{F}_n$ and $\mathbb{K}_n$ is established, weak convergence of their transformations follows immediately by the Hadamard differentiability of the maps proved herein.

This concludes the overview of the key ideas and main results of the paper. Interested readers may consult the Appendices for general statements and proofs.

# 4 Application to Outlier Robustness Analysis

## 4.1 Test of Robustness to Outliers

We apply the results developed in Section 3 to the problem described in Section 2.2 and construct a statistical test of outlier robustness analysis. We briefly recall our setup from Section 2.2. Let $\beta_1$ be the parameter of interest and $\hat{\beta}_1$ its estimator. Denote by $\hat{\beta}_2$ the estimator that is computed with outlier-adjusted sample, i.e., the sample that excludes or winsorizes outliers. Since outlier removal or winsorization can change the true parameter in the population, we let $\beta_2$ be the true parameter from the outlier-adjusted population. The null hypothesis we want to test is given by

$$H_0 : \|\beta_1 - \beta_2\| \leq h$$

for a fixed $h \geq 0$.

We assume that $h$ is a scalar while $\beta$ can be a vector, and we take the norm $\|\cdot\|$ to be the Mahalanobis distance between $\beta_1$ and $\beta_2$, that is, $[(\hat{\beta}_1 - \hat{\beta}_2)'\Sigma^{-1}(\hat{\beta}_1 - \hat{\beta}_2)]^{1/2}$ where $\Sigma$ is either an identity, the covariance matrix of $\hat{\beta}_1 - \hat{\beta}_2$, or some other positive definite symmetric matrix. The natural test statistic to use is $\|\hat{\beta}_1 - \hat{\beta}_2\|$.

Let $\alpha \in (0,1)$ be the size of the test. According to the main result, the variance $\Sigma$ of the difference $\hat{\beta}_1 - \hat{\beta}_2$ can be estimated either by the analytic formula or by the bootstrap. Note that if $h > 0$, the null hypothesis is composite; hence the definition of critical values includes taking supremum over the set of point null hypotheses. In particular, the critical value $c_\alpha$ in a general case must satisfy

$$\sup_{\|v\|\leq 1} \Pr\bigl(\|hv + \xi\|^2 > c_\alpha\bigr) \leq \alpha,$$

where $\xi \sim N(0, \Sigma)$. If $\beta$ is a scalar, it reduces to finding $c_\alpha$ such that

$$\Pr\bigl((h + \xi)^2 > c_\alpha\bigr) = \alpha$$

for $\xi \sim N(0, \sigma^2)$ where $\sigma^2$ is the variance of $\hat{\beta}_1 - \hat{\beta}_2$.

### 4.2 Empirical Application to the Effect of Democracy on Growth

Now we apply this test to reinvestigate the outlier robustness analysis in Acemoglu et al. (2017). The aim of their paper is to answer the long-standing question of whether democracy affects economic growth in a negative or positive way. To address difficulties arising from the effect of DGP dynamics and endogenous selection into democracy, Acemoglu et al. (2017) conduct three analyses that guard against different possibilities and find very similar results: after 25 years from permanent democratization, GDP per capita is about 20% higher than it would be otherwise. The three analyses in Acemoglu et al. (2017) consist of fixed effects regression on a dynamic panel that models GDP dynamics, treatment effects analysis that does not impose parametric assumptions on the GDP process, and IV fixed effects regression on the same dynamic panel instrumenting a wave of democratization. Acemoglu et al. (2017) then check robustness of their results to outliers for the two panel regressions. In this section, we estimate the joint distribution of the baseline and outlier-removed estimates in Acemoglu et al. (2017) and conduct a test of outlier robustness as developed above.

The first regression equation is given by:

$$\log GDP_{i,t} = \beta_0 Democracy_{i,t} + \sum_{s=1}^{4} \beta_s \log GDP_{i,t-s} + \alpha_i + \delta_t + \varepsilon_{i,t},$$

where $i$ represents a country, $t$ a year, and $Democracy_{i,t}$ the indicator of democracy at country $i$ in year $t$. Here, Acemoglu et al. (2017) assume sequential exogeneity, which means the error term is mean independent with all contemporary and past variables, namely democracy, the GDP, and fixed effects:

$$\mathbb{E}[\varepsilon_{i,t} \mid \log GDP_{i,t-s}, Democracy_{i,t-u}, \alpha_i, \delta_t : s = 1, \ldots, t, u = 0, \ldots, t] = 0 \text{ for all } i \text{ and } t.$$

The data consist of 6,336 observations. The original paper examines two more specifications, but we omit them as the results of reexamination are similar.

In the third analysis, Acemoglu et al. (2017) use the regional wave of democratization as an instrument. The first-stage equation is now

$$Democracy_{i,t} = \sum_{s=1}^{4} \pi_s\, WaveOfDemocracy_{i,t-s} + \sum_{s=1}^{4} \phi_s \log GDP_{i,t-s} + \theta_i + \eta_t + v_{i,t},$$

where $WaveOfDemocracy_{i,t}$ is the instrument that is constructed by indicators of democracy of nearby countries that share similar political history as country $i$. The assumption needed for this IV model is the exclusion restriction:

$$\mathbb{E}[\varepsilon_{i,t} \mid \log GDP_{i,t-s}, WaveOfDemocracy_{i,t-s}, \alpha_i, \delta_t : s = 1, \ldots, t] = 0 \text{ for all } i \text{ and } t.$$

Since the panel data is unbalanced, each country has a varied number of observations. Let $t_i$ be the year of a country $i$'s first appearance in the sample and $T_i$ be the number of observations country $i$ has. Then, $i$'s array of time observations consists of $(i, t_i), (i, t_i + 1), \ldots, (i, t_i + T_i - 1)$.

Aside from the regression coefficients, Acemoglu et al. (2017) report three more parameters. The first is the *long-run effect of democracy* defined as $\beta_5 := \beta_0/(1 - \beta_1 - \beta_2 - \beta_3 - \beta_4)$, which represents the impact on $\log GDP_{i,\infty}$ of the transition from non-democracy $D_{i,t-1} = 0$ to permanent democracy $D_{i,t+s} = 1$ for every $s \geq 0$. The second parameter is the *effect of transition to democracy after 25 years* given by $\beta_6 := e_{25}$, where $e_j = \beta_0 + \beta_1 e_{j-1} + \beta_2 e_{j-2} + \beta_3 e_{j-3} + \beta_4 e_{j-4}$ and $e_0 = e_{-1} = e_{-2} = e_{-3} = 0$, which represents the impact on log $GDP_{i,25}$ of the same transition. The third parameter is *persistence of the GDP process* defined to be $\beta_7 := \beta_1 + \beta_2 + \beta_3 + \beta_4$, which represents how persistently a unit change in log *GDP* would remain.

To check robustness of their results to outliers, Acemoglu et al. (2017) carry out the same regression but exclude some observations that have large residuals. For notational convenience, let

$$x_{i,t} := \begin{bmatrix} Democracy_{i,t} \\ \log GDP_{i,t-1} \\ \vdots \\ \log GDP_{i,t-4} \\ \mathbb{1}_{i=1} \\ \vdots \\ \mathbb{1}_{i=N} \\ \mathbb{1}_{t=0} \\ \vdots \\ \mathbb{1}_{t=T} \end{bmatrix}, \quad \beta := \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_4 \\ \alpha_1 \\ \vdots \\ \alpha_N \\ \delta_1 \\ \vdots \\ \delta_T \end{bmatrix}, \quad z_{i,t} := \begin{bmatrix} WaveOfDemocracy_{i,t-1} \\ \vdots \\ WaveOfDemocracy_{i,t-4} \\ \log GDP_{i,t-1} \\ \vdots \\ \log GDP_{i,t-4} \\ \mathbb{1}_{i=1} \\ \vdots \\ \mathbb{1}_{i=N} \\ \mathbb{1}_{t=0} \\ \vdots \\ \mathbb{1}_{t=T} \end{bmatrix}, \quad \pi := \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_4 \\ \phi_1 \\ \vdots \\ \phi_4 \\ \theta_1 \\ \vdots \\ \theta_N \\ \eta_1 \\ \vdots \\ \eta_T \end{bmatrix}.$$

Outliers are defined in their paper by $|\hat{\varepsilon}_{i,t}| \geq 1.96\,\hat{\sigma}_\varepsilon$, where $\hat{\sigma}_\varepsilon$ is the estimate of the homoskedastic standard error of $\varepsilon$,[9]

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1}(y_{i,t} - x_{i,t}'\hat{\beta})^2,$$

and, for the IV model, also by $|\hat{v}_{i,t}| \geq 1.96\,\hat{\sigma}_v$, where

$$\hat{\sigma}_v^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1}(x_{i1,t} - z_{i,t}'\hat{\pi})^2.$$

This means that they are concerned whether tail observations in the GDP might have disproportionate effects on the estimates. Defining outliers based on $\hat{\varepsilon}$ but not on $y$, even if they are interested in the effects of outliers of the GDP, is a reasonable choice since, under some assumptions, sample selection based on $\hat{\varepsilon}$ does not affect the true parameters while selection based on the dependent variable log *GDP* would almost certainly bias the true parameters.

Let $\mathbb{F}_{n,xy}$ be the vector of marginal empirical distribution functions of $\frac{1}{T_i}\sum_t x_{i,t}y_{i,t}$ and $\mathbb{Q}_{n,xy}$ the vector of marginal empirical quantile functions of $\frac{1}{T_i}\sum_t x_{i,t}y_{i,t}$. Note that, with $w_{i,t,n} = \mathbb{1}\{|\hat{\varepsilon}_{i,t}| \geq 1.96\hat{\sigma}_\varepsilon\}$, the full-sample and outlier-removed OLS estimators are

[9] The purpose of computing the homoskedastic standard error $\hat{\sigma}_\varepsilon$ is normalization. Acemoglu et al. (2017) do allow for heteroskedasticity and use heteroskedasticity-robust estimators for inference.

written as

$$\begin{aligned}
\hat{\beta}^1_{\text{OLS}} &= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}x_{i,t}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}y_{i,t} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}x_{i,t}'\right)^{-1}\int_0^1 \mathbb{Q}_{n,xy}(u)du, \\
\hat{\beta}^2_{\text{OLS}} &= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}x_{i,t}'w_{i,t,n}\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}y_{i,t}w_{i,t,n} \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}x_{i,t}'w_{i,t,n}\right)^{-1}\int_0^1 \mathbb{Q}_{n,xy}(u)d\mathbb{K}_{n,xy}(u),
\end{aligned}$$

where $\mathbb{K}_{n,xy}$ is the vector of random selection measures whose $j$th element assigns density $\sum_t x_{i,t,j}y_{i,t}w_{i,t,n}/\sum_t x_{i,t,j}y_{i,t}$ to $u \in (\mathbb{F}_{n,xy,j}(\frac{1}{T_i}\sum_t x_{i,t,j}y_{i,t})-1/n, \mathbb{F}_{n,xy,j}(\frac{1}{T_i}\sum_t x_{i,t,j}y_{i,t})]$. Assume that the cdfs of $\frac{1}{T_i}\sum_t x_{i,t}y_{i,t}$ are smooth with $(2+c)$th moments for some $c > 0$ and $\hat{\sigma}_\varepsilon$ has a well-defined limit. Since each density $\sum_t x_{i,t,j}y_{i,t}w_{i,t,n}/\sum_t x_{i,t,j}y_{i,t}$ of $\mathbb{K}_{n,xy}$ Then, our results indicate that the joint distribution of two vectors

$$\int_0^1 \mathbb{Q}_{n,xy}(u)du \qquad \text{and} \qquad \int_0^1 \mathbb{Q}_{n,xy}(u)d\mathbb{K}_{n,xy}(u)$$

converges and can be estimated by nonparametric bootstrap. Since $\hat{\beta}^1_{\text{OLS}}$ and $\hat{\beta}^2_{\text{OLS}}$ converge to fixed combinations of elements of these vectors, their joint distribution can also be estimated by nonparametric bootstrap, as we will do.

Similarly, let $\mathbb{F}_{n,zy}$ and $\mathbb{Q}_{n,zy}$ be the vectors of marginal empirical distribution functions and marginal empirical quantile functions of $\frac{1}{T_i}\sum_t z_{i,t}y_{i,t}$. The full-sample and outlier-removed IV estimators are written as

$$\begin{aligned}
\hat{\beta}^1_{\text{IV}} &= \left(\overline{xz'}(\overline{zz'})^{-1}\overline{zx'}\right)^{-1}\overline{xz'}(\overline{zz'})^{-1}\int_0^1 \mathbb{Q}_{n,zy}(u)du, \\
\hat{\beta}^2_{\text{IV}} &= \left(\overline{xz'\tilde{w}}(\overline{zz'\tilde{w}})^{-1}\overline{zx'\tilde{w}}\right)^{-1}\overline{xz'\tilde{w}}(\overline{zz'\tilde{w}})^{-1}\int_0^1 \mathbb{Q}_{n,zy}(u)d\mathbb{K}_{n,zy}(u),
\end{aligned}$$

where $\tilde{w}_{i,n} = \mathbb{1}\{|\hat{\varepsilon}_i| \geq 1.96\hat{\sigma}_\varepsilon \text{ and } |\hat{v}_i| \geq 1.96\hat{\sigma}_v\}$,

$$\begin{aligned}
(\overline{zx'})' = \overline{xz'} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}z_{i,t}', & \overline{zz'} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} z_{i,t}z_{i,t}', \\
(\overline{zx'\tilde{w}})' = \overline{xz'\tilde{w}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} x_{i,t}z_{i,t}'\tilde{w}_{i,t,n}, & \overline{zz'\tilde{w}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{t=t_i}^{t_i+T_i-1} z_{i,t}z_{i,t}'\tilde{w}_{i,t,n},
\end{aligned}$$

and $\mathbb{K}_{n,zy}$ is the vector of random selection measures whose $j$th element assigns density $\sum_t z_{i,t,j}y_{i,t}\tilde{w}_{i,t,n}/\sum_t z_{i,t,j}y_{i,t}$ to $u \in (\mathbb{F}_{n,zy,j}(\frac{1}{T_i}\sum_t z_{i,t,j}y_{i,t}) - 1/n, \mathbb{F}_{n,zy,j}(\frac{1}{T_i}\sum_t z_{i,t,j}y_{i,t})]$. Again, if $\frac{1}{T_i}\sum_t z_{i,t}y_{i,t}$ has smooth cdfs with $(2+c)$th moments and $\hat{\sigma}_v$ has a well-defined limit, our results imply that the joint distribution of $\hat{\beta}^1_{\text{IV}}$ and $\hat{\beta}^2_{\text{IV}}$ can be derived by nonparametric bootstrap.

In a simple case where $\varepsilon$ (and $v$) is independent of the covariates, outlier removal according to this criterion will not change the true values of the coefficients at least asymptotically; therefore, it is sensible to set the allowed bias $h$ to the most conservative choice, zero. Letting $\beta_j^1$ and $\beta_j^2$ be the full-sample and outlier-removed true coefficients respectively, we are testing the null hypothesis in which $\beta_j^1$ is postulated to be identical to $\beta_j^2$, i.e., $H_0 : \beta_j^1 = \beta_j^2$.

The $L$-statistics formula used in our paper is visualized in Figures 1a and 1b. In particular, Figures 1a and 1b show selected elements of empirical quantile functions $\mathbb{Q}_{n,xy}$ and sample selection functions $\mathbb{K}_{n,xy}$ used in OLS estimators. The blue line in Figure 1a is the empirical quantile function of the time average of $Democracy_{i,t} \times \log GDP_{i,t}$, which is the first element of $\mathbb{Q}_{n,xy}$; the solid orange line is the sample selection function for outlier removals, which is the first element of $\mathbb{K}_{n,xy}$; the dashed orange line is the identity function (the sample selection function for the baseline estimator). Similarly, Figure 1b shows the empirical quantile and sample selection functions for the time average of $\log GDP_{i,t-1} \times \log GDP_{i,t}$. Figures 1c and 1d depict selected elements of empirical quantile functions $\mathbb{Q}_{n,zy}$ and sample selection functions $\mathbb{K}_{n,zy}$ used in IV estimators. Now, the first element of $\mathbb{Q}_{n,zy}$ is the empirical quantile function of the time average of $WaveOfDemocracy_{i,t} \times \log GDP_{i,t}$, which we represent with the blue line in Figure 1c. The sample selection function for this time average is the solid orange line, which is less steep than that in Figure 1a; this is due to the additional removal of observations for large first-stage errors. Figure 1b shows the empirical quantile and sample selection functions for the time average of products of log GDP and its lag for IV estimators.

Outlier selection criteria are visualized in Figures 2a to 2d; they indicate that there is no "crazy" observations that can drastically change the analysis but instead error distributions are as smoothly distributed as normal distributions. Figure 2a gives the histogram of estimated errors $\hat{\varepsilon}_{i,t}$ of the OLS regression. The dotted line indicates the threshold of outliers, $1.96\hat{\sigma}_\varepsilon$ and $-1.96\hat{\sigma}_\varepsilon$. The blue observations are included in the outlier-adjusted sample and the red are excluded. Figure 2b gives the two-dimensional histogram of estimated errors $(\hat{v}_{i,t}, \hat{\varepsilon}_{i,t})$; the blue observations in the rectangle are included in the outlier-removed sample while the red outside the rectangle are not. Figure 2c and Figure 2d show the marginal distributions of $\hat{v}_{i,t}$ and $\hat{\varepsilon}_{i,t}$; some observations in the blue bars are excluded because of the other error falling outside the cutoff.

We carry out nonparametric bootstrap by randomly drawing countries $i$. All fixed effects are replaced by their corresponding dummy variables. Each draw of country $i$ adds a $T_i$ number observations to the bootstrap sample; equivalently, we treat each sum over time, in particular, $\frac{1}{T_i}\sum_t x_{i,t}y_{i,t}$, $\frac{1}{T_i}\sum_t y_{i,t}y_{i,t-s}$, and $\frac{1}{T_i}\sum_t z_{i,t}y_{i,t}$, as an observation in the bootstrap in order to exploit the i.i.d. structure needed for our theory. Here, the bootstrap consists of 10,000 iterations. In each iteration for OLS regression, we draw 175 random countries with replacement; for IV regression, 174 random countries with replacement.

Our reexamination of Acemoglu et al. (2017) mostly reconfirms robustness to outliers of the results found in Acemoglu et al. (2017) with the most stringent choice of a hypothesis ($h = 0$). However, there is one coefficient, persistence of the GDP process, for which the hypothesis of outlier robustness is rejected. Table 1 lists the estimates and $p$-values for the hypotheses that outliers have no effect on the parameters. Column 1 shows the baseline OLS estimates of key parameters that use the full sample. The figures in Column 2 are the outlier-removed OLS estimates that remove observations with $|\hat{\varepsilon}_{i,t}| \geq 1.96\hat{\sigma}_\varepsilon$.

Column 3 provides the baseline IV estimates, and column 4 the outlier-removed IV estimates, removing observations with $|\hat{\varepsilon}_{i,t}| \geq 1.96\hat{\sigma}_\varepsilon$ or $|\hat{v}_{i,t}| \geq 1.96\hat{\sigma}_v$. Columns 5 to 8 illustrate the utility of our results in formal tests of outlier robustness analysis. Column 5 gives the $p$-values of the hypotheses that the two parameters estimated by columns 1 and 2 are identical, $H_0 : \beta_j^1 = \beta_j^2$, using the standard error of the difference of two estimators estimated by bootstrap. Column 6 gives the "$p$-values" of the same hypotheses, but uses the standard error of the marginal distribution of the baseline OLS estimates. These results can be considered as $p$-values of the "heuristic arguments" explained in the introduction. Column 7 shows the $p$-values of the same hypotheses calculated with IV estimates, using the standard error of the difference. Column 8 lists the "$p$-values" using the marginal standard error of the baseline IV estimates. We see that the identity of persistence of the GDP process is rejected in formal tests while accepted in heuristic tests at the 5% level. We note that the magnitudes of persistence are very close in both regressions (0.96 and 0.97), so if we allow bias $h$ of, say, 0.01, the hypothesis will not be rejected. The point of this paper is that, even when we end up accepting the robustness hypothesis, such results should be rooted in correct statistical reasoning.

Positive correlation of baseline and outlier-adjusted estimators can be visualized by our bootstrap results. Figures 3a and 3b illustrate the joint distributions of baseline and outlier-removed OLS estimators, $(\hat{\beta}_0^1, \hat{\beta}_0^2)$ and $(\hat{\beta}_7^1, \hat{\beta}_7^2)$. Figures 3c and 3d show the joint distributions of baseline and outlier-removed IV estimators, $(\hat{\beta}_0^1, \hat{\beta}_0^2)$ and $(\hat{\beta}_7^1, \hat{\beta}_7^2)$. For the contour plots, we use the kernel density estimators for ease of visualization (instead of scatter-plotting the bootstrap points). We see that the estimators are positively correlated, which is anticipated by the fact that they are based on similar sets of samples. The figures illustrate why we need the joint distributions to statistically test our null hypotheses. Graphically, the tests examine if each red star in the figures is close enough to the 45 degree line shown as the black dotted line.

To see whether there were countries that were consistently labeled as outliers, we present the histograms of numbers of removal in Figures 4a and 4b. If any observation with index $i$ is removed in an iteration, we increment the "number of removal" for country $i$. There is the largest spike at 6,000–6,200, with the second largest one at 0 in each figure. This means that more than half the countries experience about 6,000 removals throughout the bootstrap, while a little fewer than half do not undergo any removal. In other words, there is no small portion of countries that is consistently marked as an outlier, while there is a slight tendency to remove a certain subset of countries. We interpret this as follows: there are likely to be no "outliers" in the sample in the sense that they potentially come from a different data-generating process, while there are observations that happen to be relatively more extreme than the rest, which is a natural consequence of random observations.

# 5 Applications to Other Econometric Problems

Our results on $L$-statistics are new and of independent interest. As $L$-statistics appear in many places in economics, the results can be applied to other problems aside from the outlier robustness analysis. We discuss two applications and briefly describe two more applications from Kaji and Kang (2017) and Kaji (2017).

## 5.1 Multiple testing with dependence

Economists often contend with tens or hundreds of statistical inference problems in a single research project (Banerjee et al., 2015b; Casey et al., 2012; Anderson, 2008). As a consequence, economists devote increasing attention to the simultaneous inference problem. The *simultaneous inference problem* refers to the issue that statistical discoveries often arise purely by chance when many hypotheses are tested simultaneously and individually. For instance, if one tests a hundred hypotheses at size 5% each, then even when all of the null hypotheses are true, we expect that about five of them come out rejected (if the hypotheses were jointly independent).

If we value even a single statistical discovery out of a large number of hypotheses, then procedures that control the probability of obtaining even one false positive, the *familywise error rate (FWER)*, turn out to be too conservative for practical use in many contexts. Therefore, statisticians have proposed alternative forms of error control (Lehmann and Romano, 2005; Romano et al., 2010). Among them, the *false discovery rate (FDR)* is an increasingly popular concept (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Yekutieli, 2007; Romano et al., 2008).

To illustrate the utility of our results in this setting, suppose we have many hypotheses to test, and some of the test statistics are based on different subgroups. Consider, for example, the effects of productivity shocks on rice yields among subgroups classified by quartiles of land ownership (Demont, 2013); relationship between the wage and crop yield instrumented by the indicator of rainfall being above and below certain percentiles (Jayachandran, 2006); or the effect of access to microcredit on business revenue with and without individuals who are above the 99th percentile in business revenue (Augsburg et al., 2015). Assuming that these statistics are asymptotically linear, we are interested in $d$ statistics $\hat{\beta}_1, \dots, \hat{\beta}_d$ of the form

$$\hat{\beta}_j = \frac{1}{n} \sum_{i=1}^{n} m_j(X_i) w_{i,j} + o_P\left(\frac{1}{\sqrt{n}}\right) = \int_0^1 m_j(\mathbb{Q}_n) d\mathbb{K}_n + o_P\left(\frac{1}{\sqrt{n}}\right),$$

where $w_{i,j}$ is an indicator of subgroup $j$, $\mathbb{Q}_n$ the empirical quantile function of $X_i$, and $\mathbb{K}_n$ the random measure that assigns density $w_{i,j}$ to $(\mathbb{F}_n(x_i) - 1/n, \mathbb{F}_n(x_i)]$ for the empirical distribution function $\mathbb{F}_n$ of $X_i$.[10]

Many early applications of multiple testing procedures in economics overlooked the issue of dependence among such test statistics and relied on procedures that assumed independence. But if we can estimate the joint distribution of these statistics, then we can safely rely on the multiple testing procedures that exploit the knowledge of the dependence structure, such as Yekutieli (2007) and Romano et al. (2008).

## 5.2 Testing higher degree Lorenz dominance

For an income or wealth variable $X$ with quantile function $Q$, the *Lorenz curve* is the function

$$L_Q(\tau) := \frac{1}{\mathbb{E}[X]} \int_0^\tau Q(u) du.$$

It is customary to interpret $L_Q(\tau)$ as the fraction of total income or wealth held by the lowest $\tau$-fraction (Lorenz, 1905; Csörgő, 1983). The value of the Lorenz curve at a

[10]The *(asymptotic) influence function* of $X_i$ corresponds to the function $m_j(\cdot) + \beta_j$. Note that each function $m_j$ may differ as sample selection may change the influence function.

specific point $\tau$ is called the *Lorenz share at* $\tau$ (Bhattacharya, 2005). The *Gini coefficient* is then defined as

$$G_Q := 1 - 2\int_0^1 L_Q(\tau)d\tau = \frac{1}{\mathbb{E}[X]}\int_0^1 (2\tau - 1)Q(\tau)d\tau.$$

The Gini coefficient is one of the most popular inequality indices used in economics since its introduction by Gini (1912). Both the Lorenz curve and the Gini coefficient can be estimated by replacing $\mathbb{E}[X]$ and $Q$ with their sample analogues $\mathbb{E}_n[X]$ and $\mathbb{Q}_n$. Such estimators are $L$-statistics.

The Lorenz curve is a continuous visualization of inequality over the income distribution. As such, comparing Lorenz curves across within-country, cross-country, or counterfactual income distributions has become a way to "uniformly" assess differences or changes in economic inequalities (Bishop et al., 1991, 1993; Morelli et al., 2015; Fellman, 2002). This led an important inequality comparison concept, *Lorenz dominance* (Dasgupta et al., 1973; Lambert, 2001). Namely, a Lorenz curve $L_1$ is said to *Lorenz dominate* another Lorenz curve $L_2$ if $L_1(\tau) \geq L_2(\tau)$ for every $u \in (0, 1)$. If this is the case, the society with income distribution $L_1$ is considered to be "more equal" than that with $L_2$.

While conceptually simple and appealing, these concepts are criticized for being too restrictive; Lorenz curves often cross in data. Thus, in order to obtain a finer (partial) ordering of distributions, generalized versions of Lorenz dominance are proposed (Aaberge, 2009).[11] The *kth degree downward Lorenz curve* puts more emphasis on income transfers to the poor and less emphasis on transfers to the rich; it is defined for some $k \geq 2$ by

$$L_Q^k(\tau) := \int_\tau^1 L_Q^{k-1}(u)du = \frac{1}{(k-1)!\mathbb{E}[X]}\int_\tau^1 (u - \tau)^{k-1}Q(u)du,$$

where $L_Q^1 := 1 - L_Q$. The higher the value of $k$, the larger the emphasis put on the poor. We say that a Lorenz curve $L_1$ *kth degree downward Lorenz dominates* another $L_2$ if the corresponding $k$th degree downward Lorenz curve $L_1^k$ dominates $L_2^k$; intuitively, the society with income distribution $L_1$ is more equal than that with $L_2$ when additional emphasis is put on the poorer population. Likewise, the *kth degree upward Lorenz curve* is defined by

$$\tilde{L}_Q^k(\tau) := \int_0^\tau \tilde{L}_Q^{k-1}(u)du = \frac{1}{(k-1)!\mathbb{E}[X]}\int_0^\tau (\tau - u)^{k-1}Q(u)du,$$

where $\tilde{L}_Q^1 := L_Q$, and puts more emphasis on the rich. The Lorenz curve $L_1$ *kth degree upward Lorenz dominates* another $L_2$ if $\tilde{L}_1^k \geq \tilde{L}_2^k$ uniformly. Again, the natural sample analogue estimators of higher degree Lorenz curves are $L$-statistics.

Comparison of Lorenz curves in applied work has relied mostly on visual inspection. To formally test dominance calls for uniform inference on the Lorenz curves; in particular, we require the joint distribution of every Lorenz share indexed by the income quantile $\tau \in (0, 1)$. The nonparametric uniform test of the standard (first degree) Lorenz dominance

[11] These extensions differ from the *generalized Lorenz curve* in Shorrocks (1983).

is recently established in the literature (Barrett et al., 2014).[12] The present paper allows the extension of the uniform test to arbitrary degree Lorenz dominance.

The testing procedure is as follows. Since $(u-\tau)^{k-1} \leq 1$ for $\tau \leq u \leq 1$, we have

$$\int_\tau^1 \big|(u-\tau)^{k-1}Q\big|du \leq \int_0^1 |Q|du.$$

Therefore, the uniform convergence of the estimated $k$th degree downward Lorenz curve (uniformly over $\tau$) follows by the $L_1$ convergence of the empirical quantile $\mathbb{Q}_n$. Suppose one has estimates of two $k$th degree downward Lorenz curves $\hat{L}_1^k$ and $\hat{L}_2^k$,

$$\hat{L}_j^k(\tau) := \frac{1}{(k-1)!\mathbb{E}_n[X_j]}\int_\tau^1 (u-\tau)^{k-1}\mathbb{Q}_{n,j}(u)du.$$

The assumptions posited here are that both income distributions $X_j$, $j=1,2$, have a $(2+\varepsilon)$th moment for some $\varepsilon > 0$, and that both empirical quantile processes $\sqrt{n}(\mathbb{Q}_{n,j} - Q_j)$, $j=1,2$, converge weakly in a joint manner. As in Barrett et al. (2014), we consider the null hypothesis that $L_1$ dominates $L_2$:

$$H_0 : L_1^k(\tau) \geq L_2^k(\tau), \qquad \forall \tau \in (0,1).$$

Rewrite this hypothesis as

$$H_0 : \sup_{\tau\in(0,1)} (L_2^k - L_1^k) \leq 0.$$

Thus, the test statistic is $\sup_\tau(\hat{L}_2^k - \hat{L}_1^k)$, and the critical value $c_\alpha$ must satisfy

$$\Pr\bigg(\sup_{\tau\in(0,1)} (\hat{L}_2^k - \hat{L}_1^k) > c_\alpha\bigg) \leq \alpha$$

under the null hypothesis. Although involvement of a supremum makes analytic calculation of the asymptotic distribution difficult, we may rely on the nonparametric bootstrap to obtain the critical value.

Let $(X_{1,1}, X_{1,2}), \dots, (X_{n,1}, X_{n,2})$ be the random variables representing households' incomes in two different economic states (be it within-country, cross-country, or counterfactual). In particular, $X_{i,1}$ is the income of a household in one economy, and $X_{i,2}$ is that in another. We denote them as a pair $X_i = (X_{i,1}, X_{i,2})$ since they can be dependent (as in within-country or counterfactual comparison), but they can also be treated separately if they are clearly independent (as in cross-country comparison); they can even have different sample sizes. Let $X_1^*, \dots, X_n^*$ denote the bootstrapped sample. Compute the bootstrap Lorenz curves by

$$\hat{L}_j^{k*}(\tau) := \frac{1}{(k-1)!\mathbb{E}_n[X_{i,j}^*]}\int_\tau^1 (u-\tau)^{k-1}\mathbb{Q}_{n,j}^*(u)du.$$

Then, compute the quantity $\sup_\tau(\hat{L}_2^{k*} - \hat{L}_1^{k*})$. We can use the bootstrap $(1-\alpha)$-quantile of this quantity as the critical value for $\sup_\tau(\hat{L}_2^k - \hat{L}_1^k)$ to test for $k$th order Lorenz dominance.

[12] Bishop et al. (1988) and Arora and Jain (2006) present tests of (generalized) Lorenz dominance on finitely many points.

## 5.3 Controlling tail risk with estimation errors

In the context of risk measurement in finance, Kaji and Kang (2017) develop a method to incorporate the estimation error into the risk to be estimated. For example, the *expected shortfall* is becoming popular in financial trading and banking regulation, which is defined as the expected return in the worst event of probability $\alpha$, typically 5%. Letting $X$ be the return of a portfolio, the expected shortfall $ES_\alpha$ is defined by

$$\sup_{E\in\mathcal{F}} \Big\{ \Pr(E) : \mathbb{E}[-X \mid E] \geq ES_\alpha \Big\} \leq \alpha,$$

where $\mathcal{F}$ is the set of events. Algebra reveals that

$$ES_\alpha = -\int_0^\alpha Q(u)du,$$

where $Q$ is the population quantile function of $X$. The true expected shortfall cannot be observed, so in practice an estimated quantity is used. Suppose for simplicity that observations of returns are i.i.d.; then a natural estimator is (the negative of) the sample mean of observations of $X$ below the $\alpha$-quantile, which is an $L$-statistic. The estimated expected shortfall, however, does not satisfy the above equation because of the estimation error. Instead, consider the $(1-\alpha)$-confidence set of the estimator and let $\overline{ES_\alpha}$ be its upper bound, that is,

$$\Pr\big(\overline{ES_\alpha} \geq ES_\alpha\big) \geq 1-\alpha.$$

Then, by the Bonferroni inequality,

$$\sup_E \Pr\big(E \wedge \mathbb{E}[-X \mid E] \geq \overline{ES_\alpha}\big) \leq \sup_E \Big\{ \Pr(E) : \mathbb{E}[-X \mid E] \geq ES_\alpha \Big\} + \alpha \leq 2\alpha.$$

This enables us to control the risk (the probability of "bad" events) by an observable quantity $\overline{ES_\alpha}$. Kaji and Kang (2017) generalize this idea and define a class of risk measures called the *tail risk measure* to which we can apply this bound. Many tail risk measures admit representations as $L$-statistics, and thus are susceptible to the use of our theory.

## 5.4 Assessing outcome-dependent heterogeneity in treatment effects

Kaji (2017) proposes a method to assess outcome-dependent heterogeneity in treatment effects. Let $Y_0$ be the outcome of an individual in the control group and $Y_1$ be that of an individual in the treatment group. Letting $X$ represent the characteristics to control for, the conditional average treatment effect $\mathbb{E}[Y_1 - Y_0 \mid X]$ is frequently used in empirical research, especially in ones with randomized controlled trials. Meanwhile, the average treatment effects conditional on outcome variables such as $\mathbb{E}[Y_1 - Y_0 \mid Y_0 \in A]$ cannot be estimated since the joint distribution of $Y_1$ and $Y_0$ is not identified (Heckman et al., 1997). However, it often happens that treatment effects conditional outcomes are of interest. Taking the microcredit example from Section 2.2, if the treatment effect for households with originally high business profits ($Y_0$ is large) is positive and that for households with low profits ($Y_0$ is small) is negative, then even if the average treatment effect is positive, one may not wish to implement the microcredit. The common practice to assess such heterogeneity is the quantile treatment effect (Banerjee et al., 2015a; Augsburg et al., 2015; Tarozzi et al., 2015). However, quantile treatment effects cannot in general be interpreted as individual treatment effects. Kaji (2017) interprets quantile treatment effects as a distribution of individual treatment effects that attains the minimal sum of absolute individual treatment effects, and proposes the integral function of quantile treatment effects as an alternative measure of outcome-dependent subgroup treatment effects that has better asymptotic properties. These asymptotic results are derived using results of the present paper.

# 6 Literature Review

We review three strands of the literature related to this paper: (1) empirical and quantile processes in statistics/econometrics, (2) $L$-statistics in statistics and risk measures in finance, and (3) outlier detection and robust estimation in statistics.

## 6.1 Empirical and quantile processes

Central limit theorems in Banach spaces such as $L_p$ are a classical topic of which Ledoux and Talagrand (1991) provide an excellent exposition. It is known that tightness and the limiting properties of the Banach-valued random variables are closely tied to the structure and geometry of the Banach spaces. Among Banach spaces, however, the uniform space $L_\infty$ attracts independent attention, not only because of its own statistical importance, but for its mathematical complication epitomized by the fact that pre-Gaussianity alone does not immediately imply central limit theorems. Such difficulty called for direct characterization of asymptotic tightness and developed into rich literature—in an effort to show tightness—including tail bounds, entropy theory, and the Vapnik-Červonenkis theory (Van der Vaart and Wellner, 1996; Dudley, 2014). This paper contributes to the literature by directly characterizing asymptotic tightness in $L_1$ in combination with $L_\infty$, enabling the establishment of weak convergence of processes that are not necessarily sample averages of i.i.d. Banach-valued random variables (thereby preventing the use of central limit theorems). In the particular context of this paper, notwithstanding the i.i.d. assumption, we needed such characterization to show weak convergence of $\mathbb{K}_n$.

The study of quantile processes is as old as that of empirical processes (Csörgő, 1983), but to the best of our knowledge the study is limited to quantile processes of *bounded* random variables and *weighted (standardized)* quantile processes (Shorack and Wellner, 1986; Csörgő and Horváth, 1988, 1990, 1993; Csörgő et al., 1986, 1993; Koul, 2002). This paper is the first to show the weak convergence of raw quantile processes of unbounded random variables directly in $L_1$. It is also novel that the functional delta method is proved for the inverse map with norms replaced by more appropriate ones. In this respect, this work is related to Beutner and Zähle (2010); they consider the weighted sup norm on the space of distribution functions and establish the functional delta method for risk functionals. Their paper and ours share the similar idea that the use of a new norm gives a new functional delta method, but their work is closer in spirit to the literature on weighted suprema of empirical processes. Although less clearly related, we note that a non-uniform norm for empirical processes has been occasionally considered in probability theory as well; e.g., Dudley (1997).

Some readers may wish to associate the results with the quantile regression literature popularized in economics (Koenker, 2005). This literature, initiated by Koenker and Bassett (1978), reinvigorated the old work by Laplace (1812), and yielded many important results, including Koenker and Xiao (2002), Chernozhukov (2005), Chernozhukov and Hansen (2005), Angrist et al. (2006), Firpo et al. (2009), Chernozhukov et al. (2010), and Belloni and Chernozhukov (2011). There is an obvious relationship between the conditional mean and conditional quantiles when some simplifying assumptions hold. Let $y_i = x_i'\beta + \varepsilon_i$ and $\mathbb{E}[\varepsilon_i \mid x_i] = 0$ for every $x_i$. Then, by the change of variables,

$$x_i'\beta = \mathbb{E}[y_i \mid x_i] = \int_0^1 Q_{y_i}(u \mid x_i)du = \int_0^1 x_i'\beta(u)du.$$

So, if $x_i$ is sufficiently rich, it must be that $\beta = \int_0^1 \beta(u)du$. Although this type of relationship is known and used in survival analysis (Cox et al., 2013) (and a remotely related one used in economics; Chernozhukov et al., 2013), it does not offer much to our purpose, as there is no guarantee that the relationship continues to hold in their sample analogues. It is, however, an important direction of future research to examine if weak convergence in $L_1$ takes place for quantile regression estimators on the whole of $(0, 1)$.[13]

Also, the paper does not explicitly consider weakly dependent samples, although as noted at the end of Section 3.3 it would be straightforward to extend the results to subsume such cases, e.g., by incorporating results of Dehling et al. (2002, 2014). Other extensions potentially useful for measuring financial risk are the application of extreme value theory for the tail empirical processes (Einmahl, 1992; Rootzén, 2009; Drees and Rootzén, 2010, 2016) or the use of smoothed or other explicitly estimated empirical distributions (Hall et al., 1999; Berg and Politis, 2009).

### 6.2 $L$-statistics

$L$-statistics are an old topic in statistics, especially in the study of location estimation (Van der Vaart, 1998, Chapter 22). There are two major ways to prove the asymptotic normality of $L$-statistics: the Hájek projection and the functional delta method. The difficulty of showing asymptotic normality lies in that the summands are intercorrelated with each other in a complicated way. The Hájek projection projects the summands onto the space of independent observations, thereby pulling the situation back to ones of classical central limit theorems. This requires, however, an effort to find the projection and to show that the residual of the projection goes away. This can be a hard task when the statistic of interest involves complicated estimation procedures or comes from nontrivial structural models.

The functional delta method, on the other hand, directly deals with the complicated intercorrelation in the raw form of an empirical quantile function. Therefore, it is more general than the Hájek projection, yields simple representation of the asymptotic distribution, and proves the validity of bootstrap at much less or no cost. The cons of this method are that the empirical processes literature usually requires uniform convergence, which unavoidably entails boundedness of the processes. All this has led to standardization methods using bounded quantile processes, as seen in the Chibisov-O'Reilly theorem. This paper, in combination with giving $L_1$ convergence of quantile processes, tackles this thorny issue by extending the functional delta method to $L_1$ processes and $L$-statistics.

[13]One notable difference from our setup is that they involve optimization over a class of functions to obtain a process.

### 6.3 Outlier detection and robust estimation

In this paper, we refer to observations lying in either tail as "outliers." The classical subfield of statistics, *outlier detection*, defines outliers instead as observations arising from a different unknown data-generating process and thereby subject to elimination (Hawkins, 1980; Aggarwal, 2013). Starting from a null hypothesis about the true data-generating process, the literature develops a way to detect observations that fall outside of the behavior permissible under the null. Distributional assumptions in the null hypothesis may vary from a complete description of the data generating process to only the tail behaviors to the proximity or temporal models. Outlier detection in a regression framework is also studied by Chatterjee and Hadi (1988) and Gujarati and Porter (2009).

Despite the concern for outliers and their removal, the paper has not much to share with this literature. This paper does not make assumptions on distributions but on existence of moments, and the null we aim to reject is about specific parameters rather than outliers themselves. Nevertheless, if one dares to draw a connection, one can say that the paper provides a new way to formulate the null hypothesis in outlier detection. If one has a particular parameter in mind that should not be largely affected by any few observations, then by conducting the outlier robustness test for that parameter one can detect the outliers when the test is rejected.

The final literature we review is on robust estimation in statistics (Hampel et al., 1986; Koul, 2002; Maronna et al., 2006; Huber and Ronchetti, 2009). These works concern estimators that are robust against deviations from an ideal model, especially when highly influential (erroneous) outliers are introduced. Similarly to this paper, they mostly consider estimators whose deviation from the true parameters is represented by the sum of influence functions of observations. To estimate asymptotic distributions of robust estimators, they often rely on empirical process theory (Koul, 2002, Chapter 4); in this sense, this paper can be considered an extension of their asymptotic theory. Their motivation goes further in finding the best influence function under some criteria and construct the best estimator for the purpose of robust estimation. Although such robust estimation has been considered in the economics discipline (Krasker, 1980), it is rather a minor subject—possibly due to the resulting estimators' non-straightforward interpretability.

## 7 Conclusion

Motivated by a need for formal tests of outlier robustness, this paper develops substantially generalized asymptotic theory of $L$-statistics. In particular, observing that essential for convergence of $L$-statistics is not the uniform convergence of empirical quantile processes but $L_1$ convergence, we establish the theory of $L$-statistics through the development of the theory of integrable empirical processes. The highlights of this theoretical development are the new norms introduced to the spaces of functions.

First, we consider distribution functions in the space of bounded integrable functions. Distribution functions need to converge uniformly in order for their inverses to be well-defined, and they also need to converge in $L_1$ in order for their inverses (quantile functions) to be integrable. We characterize weak convergence in this space by asymptotic uniform equicontinuity and equiintegrability in probability. Uniform equicontinuity is needed for uniform convergence, and equiintegrability for $L_1$ convergence. Using this, we show that empirical processes converge in this norm if the underlying distribution $F$ has a $(2+\varepsilon)$th moment for some small $\varepsilon > 0$.

Second, we consider quantile functions in the space of integrable functions, and derive weak convergence using the functional delta method. The key to the proof is the compatibility of the $L_1$ norm with Fubini's theorem. This is in contrast to classical results such as the Chibisov-O'Reilly theorem that use the $L_2$ norm.

Then, we consider sample selection functions in the set of Lipschitz functions in the space of bounded integrable functions. We need the Lipschitz property to make $\int \mathbb{Q}_n d\mathbb{K}_n$ converge whenever $\int \mathbb{Q}_n du$ does, boundedness to ensure that the Lebesgue-Stieltjes integral with respect to $K$ is well-defined, and integrability to ensure convergence of the integral itself. We derive weak convergence of sample selection functions by another application of our earlier results.

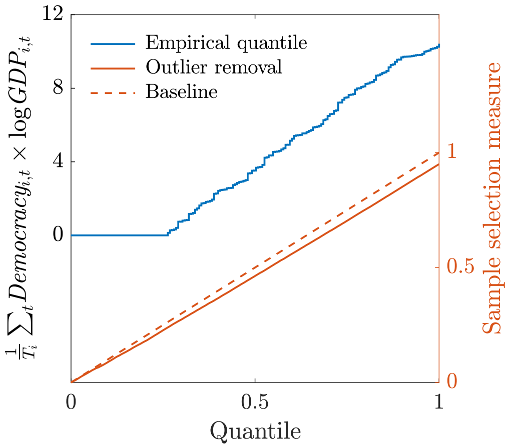
Finally, we derive weak convergence of $L$-statistics using the functional delta method on the map from quantile functions and sample selection functions to $L$-statistics. This can be seen as a generalization of the results on Wilcoxon statistics to subsume unbounded functions. As a byproduct of our functional delta method approach, we derive validity of nonparametric bootstrap.

Using our results, we construct a formal test of outlier robustness analysis. We apply our test to Acemoglu et al. (2017) and contrast heuristic arguments to formal tests. For one of the parameters, we "discover" sensitivity to outliers that could not have been discovered by heuristics.

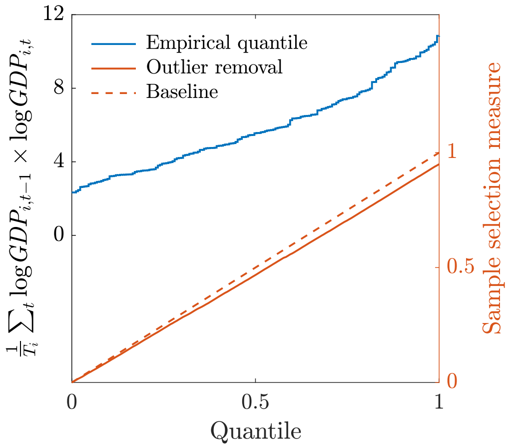
Our theory of $L$-statistics is itself new and of independent interest. As applications other than outlier robustness analysis, we explained multiple testing problems, tests of higher-order Lorenz dominance, estimation of tail risk measures by Kaji and Kang (2017), and estimation of bounds on outcome-dependent treatment effects by Kaji (2017).

# Appendices

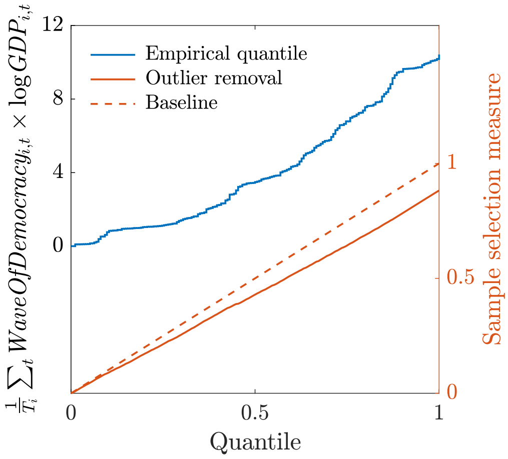
## A.1 Figures and Tables



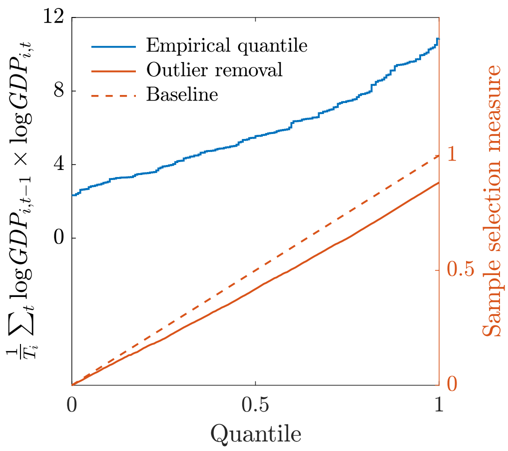

(a) Empirical distribution and sample selection functions for OLS estimators. $\frac{1}{T_i}\sum_t Democracy_{i,t} \times \log GDP_{i,t}$ [in 100 units].




(b) Empirical distribution and sample selection functions for OLS estimators. $\frac{1}{T_i}\sum_t \log GDP_{i,t-1} \times \log GDP_{i,t}$ [in 100,000 units].

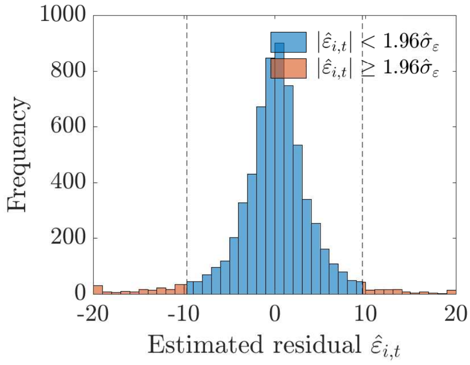



(c) Empirical distribution and sample selection functions for IV estimators. $\frac{1}{T_i}\sum_t WaveOfDemocracy_{i,t} \times \log GDP_{i,t}$ [in 100 units].

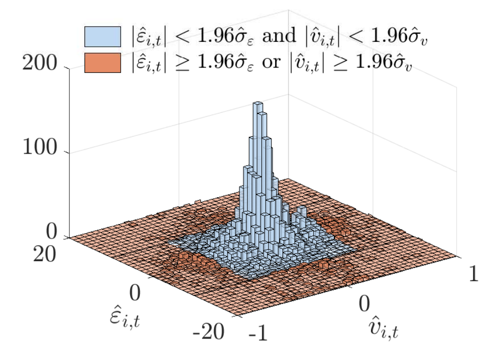



(d) Empirical distribution and sample selection functions for IV estimators. $\frac{1}{T_i}\sum_t \log GDP_{i,t-1} \times \log GDP_{i,t}$ [in 100,000 units].

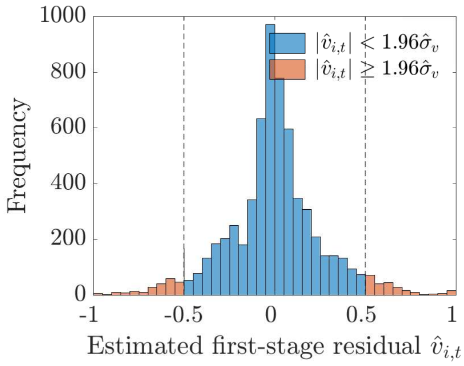
Figure 1: Empirical distribution and sample selection functions for OLS and IV estimators.



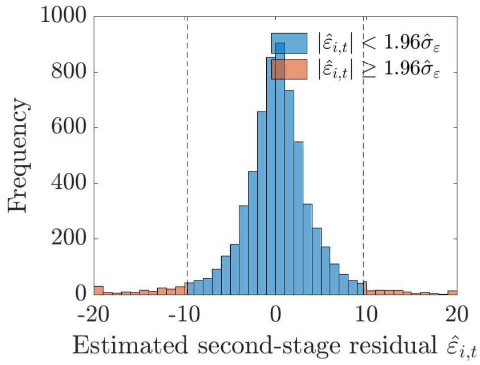

(a) Distribution of estimated residuals $\hat{\varepsilon}_{i,t}$ for OLS regression.




(b) Joint distribution of estimated residuals $(\hat{v}_{i,t}, \hat{\varepsilon}_{i,t})$ for IV regression.




(c) Distribution of estimated first-stage residuals $\hat{v}_{i,t}$ for IV regression.



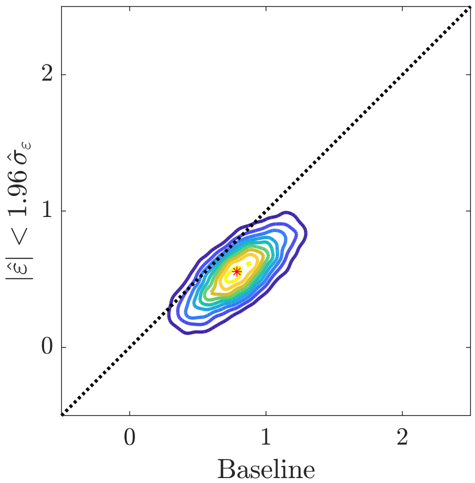

(d) Distribution of estimated second-stage residuals $\hat{\varepsilon}_{i,t}$ for IV regression.

Figure 2: Distributions of residuals of OLS and IV regressions. Clusters at the boundaries indicate how many observations fall outside of the range.

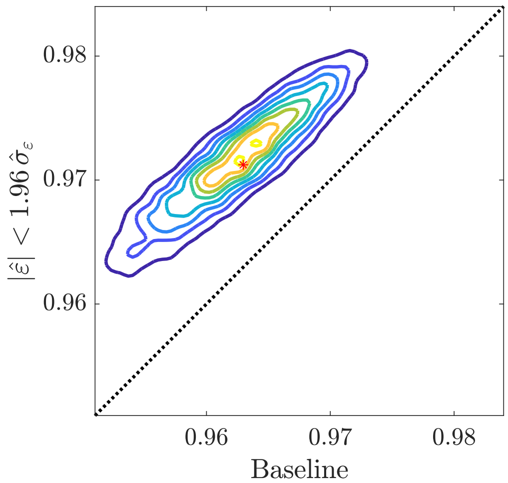
Table 1: Estimates and comparison of formal and heuristic $p$-values for the outlier robustness test in Acemoglu et al. (2017).

| | | Estimate | | | | $p$-value for $H_0 : \beta_j^1 = \beta_j^2$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | OLS | | IV | | OLS | | IV | |
| | Notation | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Democracy | $\beta_0$ | 0.79 | 0.56 | 1.15 | 0.66 | 0.15 | 0.32 | 0.20 | 0.41 |
| | | (0.23) | (0.20) | (0.59) | (0.44) | | | | |
| log GDP first lag | $\beta_1$ | 1.24 | 1.23 | 1.24 | 1.23 | 0.60 | 0.74 | 0.70 | 0.82 |
| | | (0.04) | (0.02) | (0.04) | (0.03) | | | | |
| log GDP second lag | $\beta_2$ | −0.21 | −0.20 | −0.21 | −0.20 | 0.75 | 0.85 | 0.84 | 0.91 |
| | | (0.05) | (0.03) | (0.05) | (0.04) | | | | |
| log GDP third lag | $\beta_3$ | −0.03 | −0.03 | −0.03 | −0.03 | 0.97 | 0.98 | 0.89 | 0.93 |
| | | (0.03) | (0.02) | (0.03) | (0.03) | | | | |
| log GDP fourth lag | $\beta_4$ | −0.04 | −0.03 | −0.04 | −0.03 | 0.26 | 0.45 | 0.28 | 0.46 |
| | | (0.02) | (0.01) | (0.02) | (0.02) | | | | |
| Long-run effect of democracy | $\beta_5$ | 21.24 | 19.32 | 31.52 | 22.63 | 0.72 | 0.79 | 0.46 | 0.63 |
| | | (7.32) | (8.54) | (18.49) | (18.14) | | | | |
| Effect of democracy after 25 years | $\beta_6$ | 16.90 | 13.00 | 24.87 | 15.47 | 0.29 | 0.46 | 0.27 | 0.49 |
| | | (5.32) | (5.02) | (13.53) | (10.82) | | | | |
| Persistence of GDP process | $\beta_7$ | 0.96 | 0.97 | 0.96 | 0.97 | 0.0002 | 0.12 | 0.004 | 0.20 |
| | | (0.01) | (0.005) | (0.01) | (0.005) | | | | |
| Number of observations | | 6,336 | 6,044 | 6,309 | 5,579 | | | | |
| Number of countries | $n$ | 175 | 175 | 174 | 174 | | | | |
| Minimum number of years | $\min_i T_i$ | 6 | 5 | 6 | 5 | | | | |
| Maximum number of years | $\max_i T_i$ | 47 | 47 | 47 | 47 | | | | |
| Average number of years | | 36.2 | 34.5 | 36.3 | 32.1 | | | | |
| Median number of years | | 42 | 40 | 42 | 34.5 | | | | |
| Number of bootstrap iterations | | | | | | 10,000 | 10,000 | 10,000 | 10,000 |

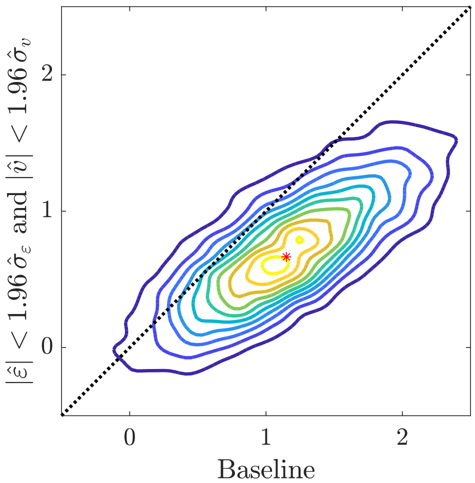
* (1) and (3) Baseline estimates; (2) Estimates on sample $|\hat{\varepsilon}_{i,t}| < 1.96\,\hat{\sigma}_\varepsilon$; (4) Estimates on sample $|\hat{\varepsilon}_{i,t}| < 1.96\,\hat{\sigma}_\varepsilon$ and $|\hat{v}_{i,t}| < 1.96\,\hat{\sigma}_v$; (5) and (7) give the $p$-values of the formal tests that use the standard errors of $\hat{\beta}_j^1 - \hat{\beta}_j^2$; (6) and (8) gives the "$p$-values" of the heuristic tests that use the marginal standard errors of $\hat{\beta}_j^1$. Some numbers in Columns (1) and (2) are slightly different from Acemoglu et al. (2017) since we use our own bootstrap results to compute standard errors.



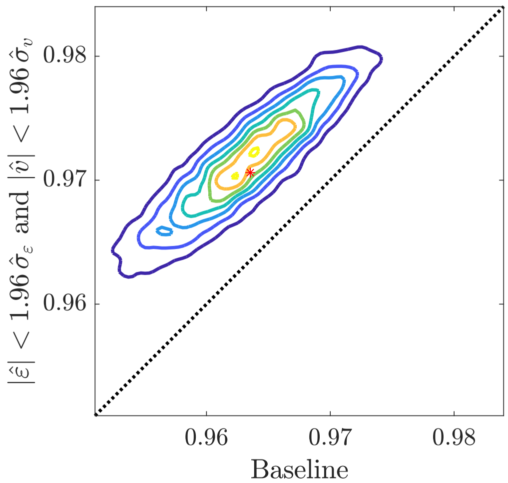

(a) Distribution of full-sample and outlier-removed OLS estimators for the effect of democracy $\beta_0$. $p = 0.15$.




(b) Distribution of full-sample and outlier-removed OLS estimators for persistence of GDP $\beta_7$. $p = 0.0002$.

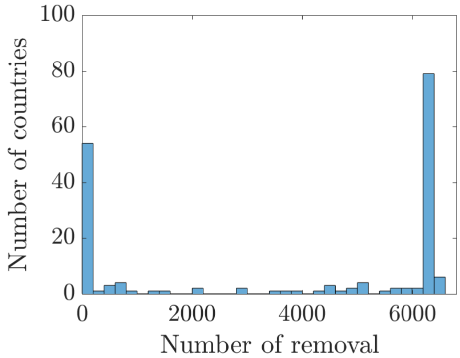



(c) Distribution of full-sample and outlier-removed IV estimators for the effect of democracy $\beta_0$. $p = 0.20$.

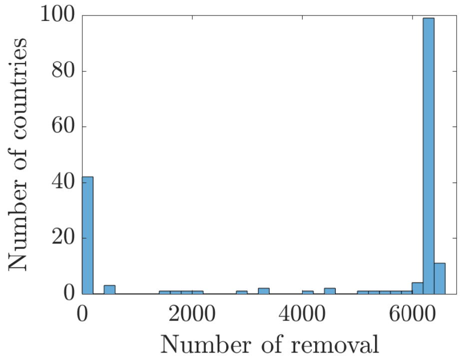



(d) Distribution of full-sample and outlier-removed IV estimators for persistence of GDP $\beta_7$. $p = 0.004$.

Figure 3: Joint distributions of full-sample and outlier-removed OLS and IV estimators for Acemoglu et al. (2017). Outliers are defined by $|\hat{\varepsilon}_{i,t}| \geq 1.96\,\hat{\sigma}_\varepsilon$ or $|\hat{v}_{i,t}| \geq 1.96\,\hat{\sigma}_v$. The black dotted lines indicate the 45-degree line. Nonparametric bootstrap is repeated for 10,000 times, randomly sampling across $i$. The contours drawn are of kernel density estimators.




(a) Histogram of numbers of removal that countries had in 10,000 iterations of OLS bootstrap.




(b) Histogram of numbers of removal that countries had in 10,000 iterations of IV bootstrap.

Figure 4: Histogram of numbers of removal that countries had in 10,000 iterations of bootstrap.

## A.2 Proofs of Results Stated in the Main Text

*Proof of Theorem 1.* If $(a, b)$ is totally bounded in $\rho_1$ and $\rho_2$, then so it is in $\rho := \rho_1 \vee \rho_2$. Then the theorem follows in combination of Theorems A.5 and A.6. ■

*Proof of Proposition 2.* This proposition is proved as Proposition A.2. ■

*Proof of Theorem 3.* This theorem is proved as Lemma A.7 and its extension in Theorem A.9. ■

*Proof of Proposition 4.* This proposition is proved as Proposition A.10. ■

*Proof of Theorem 5.* This theorem is proved as Theorem A.11. ■

*Proof of Proposition 6.* This proposition is proved as Proposition A.13. ■

*Proof of Proposition 7.* This proposition is proved as Proposition A.15. ■

*Proof of Proposition 8.* This proposition is proved as Proposition A.20. ■

## A.3 Mathematical Preliminaries

The *inverse function* $f^{-1} : \mathbb{R} \to \mathbb{R}$ of a function $f : \mathbb{R} \to \mathbb{R}$ is defined by the left-continuous generalized inverse, i.e.,

$$f^{-1}(y) := \inf\{x \in \mathbb{R} : f(x) \geq y\}.$$

This inverse is often denoted by $f^{\leftarrow}$ or $f^{-}$ in the literature (Dudley, 1997; Embrechts and Hofert, 2013). While we keep the notation $f^{-1}$ for this, when we refer to the right-continuous generalized inverse, we use the notation $f^{\rightarrow}$, that is,

$$f^{\rightarrow}(y) := \sup\{x \in \mathbb{R} : f(x) \leq y\}.$$

For properties of generalized inverses, see Feng et al. (2012) and Embrechts and Hofert (2013).

Let $-\infty \leq a < b \leq +\infty$. The Lebesgue-Stieltjes measure $\mu$ on $(a, b)$ associated with an increasing function $m : (a, b) \to \mathbb{R}$ assigns to an open interval $(c, d)$ the measure $m(d-) - m(c+)$, where $m(\cdot-)$ is the left limit and $m(\cdot+)$ the right limit. Conversely, a function $m : (a, b) \to \mathbb{R}$ associated with the Lebesgue-Stieltjes measure $\mu$ on $(a, b)$ is any function such that $\mu((c, d]) = m(d) - m(c)$ for every $c < d$.[14] Because of this relationship, we often denote both the function and the measure by the same letter.

The following lemma is used throughout Appendices.

**Lemma A.1.** *Let $F$ be a probability distribution on $\mathbb{R}$, $\tilde{F}(x) := F(x) - \mathbb{1}\{x \geq 0\}$, and $Q := F^{-1}$ the quantile function. For $p > 0$ we have (i) ⇔ (ii) ⇔ (iii) ⇒ (iv) ⇔ (v), where*

[14] Note that the function constructed from a Lebesgue-Stieltjes measure is cadlag, while the Lebesgue-Stieltjes measure can be given to any increasing but not necessarily cadlag function. This asymmetry introduces minor adjustments to the change of variables for Lebesgue-Stieltjes integrals. See Falkner and Teschl (2012).

*(i) $F$ has a pth moment;*

*(ii) $Q$ is in $L_p(0,1)$;*

*(iii) $|x|^{p-1}\tilde{F}$ is integrable;*

*(iv) $|x|^p\tilde{F}$ converges to $0$ as $x \to \pm\infty$;*

*(v) $u^{1/p}(1-u)^{1/p}Q$ converges to $0$ as $u \to \{0,1\}$.*

*Proof.* We prove the following directions in order: (i) $\Rightarrow$ (iv), (iii) $\Rightarrow$ (iv), (i) $\Leftrightarrow$ (iii), (i) $\Leftrightarrow$ (ii), and (iv) $\Leftrightarrow$ (v). The second claim seems unnecessary, but will be used in proving the third claim.

(i) $\Rightarrow$ (iv). For $M > 0$, note that

$$\int_{\mathbb{R}} |x|^p dF \geq \int_{[-M,M]} |x|^p dF + M^p|\tilde{F}(-M)| + M^p|\tilde{F}(M)|.$$

Since the left-hand side (LHS) is finite, one may take $M$ large enough that

$$\int_{\mathbb{R}} |x|^p dF - \int_{[-M,M]} |x|^p dF$$

is smaller than an arbitrarily small positive number, which then bounds the two non-negative terms. Hence $|x|^p\tilde{F}(x) \to 0$ as $x \to \pm\infty$.

(iii) $\Rightarrow$ (iv). Suppose that $|x|^{p-1}|\tilde{F}|$ is integrable but $|x|^pF$ does not vanish as $x \to -\infty$, that is, there exist a constant $c > 0$ and a sequence $0 > x_1 > x_2 > \cdots \to -\infty$ such that $|x_i|^pF(x_i) \geq c$. Since $F \to 0$, one may take a subsequence such that

$$|x_i|^pF(x_{i+1}) \leq 2^{-i}.$$

By monotonicity of $F$,

$$\begin{aligned}
p\int_{-\infty}^{0} |x|^{p-1}F(x)dx &\geq F(x_1)\int_{x_1}^{0} p|x|^{p-1}dx + F(x_2)\int_{x_2}^{x_1} p|x|^{p-1}dx + \cdots \\
&= |x_1|^pF(x_1) + \big(|x_2|^p - |x_1|^p\big)F(x_2) + \big(|x_3|^p - |x_2|^p\big)F(x_3) + \cdots \\
&\geq c + \sum_{i=1}^{\infty}(c - 2^{-i}) = \infty,
\end{aligned}$$

which is a contradiction. Hence $|x|^pF$ must vanish. Deduce similarly that $|x|^p(1-F) \to 0$ as $x \to +\infty$.

(i) $\Leftrightarrow$ (iii). Note that $dF = d\tilde{F}$ for $x \neq 0$. Integration by parts yields

$$\int_{\mathbb{R}} |x|^p dF = \int_{\mathbb{R}} |x|^p d\tilde{F} = \Big[|x|^p\tilde{F}\Big]_{-\infty}^{\infty} + p\int_{-\infty}^{\infty} |x|^{p-1}|\tilde{F}|dx.$$

If the LHS is finite (i), then the first term in the right-hand side (RHS) is 0 (iv) and hence the second term is finite (iii). Conversely, if the second term is finite (iii), then the first term is 0 (iv) and hence the LHS is finite (i).

(i) $\Leftrightarrow$ (ii). By the change of variables,

$$\int_{\mathbb{R}} |x|^p dF = \int_{0}^{1} |Q|^p du.$$

Hence the LHS is finite if and only if the RHS is.

(iv) $\Leftrightarrow$ (v). Let $u = F(x)$. Then, $\lim_{x\to-\infty} |x|^p \tilde{F} = \lim_{u\to 0}(u^{1/p}Q)^p = 0$. Convergence of the other tail can be shown analogously. ■

*Remark.* One-sided implication in the lemma is strict. One can construct $\tilde{F}$ such that (iv) holds but (iii) does not. Let $\tilde{F}$ satisfy $|x|^{p-1}\tilde{F} \approx 1/p_{\lceil |x| \rceil}$ where $p_n$ denotes the $n$th prime number. Then, it is not integrable since the sum of the reciprocals of the primes diverges, but $|x|^p \tilde{F} \to 0$ since the primes only increase at a logarithmic speed.

*Remark.* Similar properties in norms play an important role in characterizing the asymptotic behaviors in general Banach spaces (Ledoux and Talagrand, 1991).

## A.4 Tightness of Bounded Integrable Processes

First, we consider the stochastic processes that are integrable with respect to a general measure. The main objective of this section is to develop the conditions for the sequence of integrable processes to converge weakly in the corresponding $L_1$ space. Most exposition of this section parallels the flow of arguments of Van der Vaart and Wellner (1996, Chapter 1.5).

**Definition.** Let $(T, \mathcal{T}, \mu)$ be a measure space where $T$ is an arbitrary set, $\mathcal{T}$ a $\sigma$-field on $T$, and $\mu$ a $\sigma$-finite signed measure on $\mathcal{T}$. Let $\mathbb{L}_\mu$ be the Banach space of bounded and $\mu$-integrable functions $z : T \to \mathbb{R}$, that is,

$$\|z\|_{\mathbb{L}_\mu} := \|z\|_T \vee \|z\|_\mu := \left(\sup_{t\in T} |z(t)|\right) \vee \left(\int_T |z||d\mu|\right) < \infty,$$

where $|d\mu|$ represents integration with respect to the total variation measure of $\mu$.[15]

In the main text, special cases of this are used for the distribution functions and the sample selection measures. General construction allows us to accommodate many other cases, including the following.

**Example** ($p$th moment)**.** Let $\mathbb{F}_n$ be the empirical distribution of a real-valued random variable. By integration by parts, the sample $p$th moment is given by

$$\int x^p d\mathbb{F}_n = -\int \tilde{\mathbb{F}}_n dx^p \qquad \text{where} \qquad \tilde{\mathbb{F}}_n(x) = \begin{cases} \mathbb{F}_n(x) & x < 0, \\ \mathbb{F}_n(x) - 1 & x \geq 0. \end{cases}$$

Then it is natural to consider $\tilde{\mathbb{F}}_n$ as a stochastic process integrable with respect to the $\sigma$-finite signed measure $\mu((a, b]) := b^p - a^p$ on $\mathbb{R}$. □

For processes represented as the sum of i.i.d. random variables, such as the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ itself, one can easily prove weak convergence in this space by the combination of classical central limit theorems (CLT) (Van der Vaart and Wellner, 1996; Dudley, 2014; Ledoux and Talagrand, 1991), as shown in the next proposition. However, for other types of processes that are not an average of i.i.d. variables, notably the random

[15]Integration with respect to the total variation measure is often denoted with $d|\mu|$. However, because we sometimes mix Lebesgue-Stieltjes integrals, we denote the total variation integration by $|d\mu|$ so it not be confused with the Lebesgue-Stieltjes integration with respect to the "function" $|\mu|$.

measure process $\sqrt{n}(\mathbb{K}_n - K)$ in this paper, we are unable to resort to CLT-type results. Therefore a more general, direct way of showing weak convergence will be developed subsequently.

**Proposition A.2.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a function of locally bounded variation and $\mu$ the Lebesgue-Stieltjes measure associated with $m$. For a probability distribution $F$ on $\mathbb{R}$ such that $m(X)$ has a $(2+c)$th moment for $X \sim F$ and some $c > 0$, the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ converges weakly in $\mathbb{L}_\mu$ to a Gaussian process with mean zero and covariance function* $\mathrm{Cov}(x, y) = F(x \wedge y) - F(x)F(y)$.

*Proof.* The marginal convergence is trivial. According to Van der Vaart and Wellner (1996, Example 2.5.4), the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ converges weakly in $L_\infty$. In light of Van der Vaart and Wellner (1996, Proposition 2.1.11), it suffices to show that for $X_i \sim F$ and $Z_i(x) := \mathbb{1}\{X_i \leq x\} - F(x)$,

$$\Pr(\|Z_i\|_\mu > t) = o(t^{-2}) \qquad \text{as} \qquad t \to \infty$$

and

$$\int_{\mathbb{R}} \big(\mathbb{E}[Z_i(x)^2]\big)^{1/2} |d\mu| = \int_{\mathbb{R}} \big(F(x)[1 - F(x)]\big)^{1/2} |d\mu| < \infty.$$

Since a function of locally bounded variation can be written as the difference of two increasing functions, $m$ can be assumed without loss of generality increasing, that is, $\mu$ be a positive measure. Observing $Z_i(x) = (\mathbb{1}\{X_i \leq x\} - \mathbb{1}\{0 \leq x\}) - (F(x) - \mathbb{1}\{0 \leq x\})$, find[16]

$$\|Z_i\|_\mu \leq |m(X_i+) - m(0+)| + \int_{\mathbb{R}} |\tilde{F} \circ m^{-1}| dx.$$

Therefore, the first condition is satisfied if $\tilde{F} \circ m^{-1}(t) = o(t^{-2})$, which is the case if $m(X_i)$ has a variance by Lemma A.1. Secondly, if $m(X_i)$ has a $(2+c)$th moment, then by Lemma A.1 again, $|x|^{1+c} \tilde{F} \circ m^{-1}$ is integrable and $|x|^{2+c} \tilde{F} \circ m^{-1} \to 0$ as $x \to \pm\infty$. Therefore, $\tilde{F} \circ m^{-1}(x) = o(1/|x|^{2+c})$ and hence $[(F \circ m^{-1})(1 - F \circ m^{-1})]^{1/2}$ is integrable, which means that $[F(1-F)]^{1/2}$ is integrable with respect to $\mu$. Thus, the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ converges weakly in $L_1(\mu)$ as well, as desired. ∎

As in the classical empirical process literature, we first characterize weak convergence in $\mathbb{L}_\mu$ by asymptotic tightness plus weak convergence of marginals. For this purpose, we consider a sequence of random elements taking values in $\mathbb{L}_\mu$, that is, $X_n : \Omega \to \mathbb{L}_\mu$. Following Van der Vaart and Wellner (1996), there are some generalizations we allow in our setup. We consider the generalized version of a sequence, a *net* $X_\alpha$ indexed by an arbitrary directed set, rather than a sequence $X_n$ indexed by natural numbers. Note that a sequence is a special case of a net. Moreover, we allow the sample space $\Omega$ to be different for each element in a net, that is, we consider $X_\alpha : \Omega_\alpha \to \mathbb{L}_\mu$. Finally, we allow that each element in the net is not necessarily measurable, so, when we seek rigor, we shall call $X_\alpha$ a net of "arbitrary maps from $\Omega_\alpha$ to $\mathbb{L}_\mu$" in lieu of "random elements taking values in $\mathbb{L}_\mu$." There is also a caveat on the notation: when we write $X(t)$ for a map $X : \Omega \to \mathbb{L}_\mu$, $t$ is understood to be an element of $T$ and we regard $X(t)$ as a map from $\Omega$ to $\mathbb{R}$ indexed by $T$; when we explicitly use $\omega \in \Omega$ in the discussion, we write $X(t, \omega)$.

The following lemmas and theorem characterize weak convergence in our space.

[16]See Falkner and Teschl (2012) for the change of variables formula.

**Lemma A.3.** *Let $X_\alpha : \Omega_\alpha \to \mathbb{L}_\mu$ be asymptotically tight. Then, it is asymptotically measurable if and only if $X_\alpha(t)$ is asymptotically measurable for every $t \in T$.*

**Lemma A.4.** *Let $X$ and $Y$ be tight Borel measurable maps into $\mathbb{L}_\mu$. Then, $X$ and $Y$ are equal in Borel law if and only if all corresponding marginals of $X$ and $Y$ are equal in law.*

**Theorem A.5.** *Let $X_\alpha : \Omega_\alpha \to \mathbb{L}_\mu$ be arbitrary. Then, $X_\alpha$ converges weakly to a tight limit if and only if $X_\alpha$ is asymptotically tight and the marginals $(X_\alpha(t_1), \dots, X_\alpha(t_k))$ converge weakly to a limit for every finite subset $t_1, \dots, t_k$ of $T$. If $X_\alpha$ is asymptotically tight and its marginals converge weakly to the marginals $(X(t_1), \dots, X(t_k))$ of a stochastic process $X$, then there is a version of $X$ with sample paths in $\mathbb{L}_\mu$ and $X_\alpha \rightsquigarrow X$.*

*Proofs.* Since our norm is stronger than the uniform norm, Lemmas A.3 and A.4 follow as corollaries to Van der Vaart and Wellner (1996, Lemmas 1.5.2 and 1.5.3). Now we prove the theorem.

Necessity is immediate. We prove sufficiency. If $X_\alpha$ is asymptotically tight and its marginals converge weakly, then $X_\alpha$ is asymptotically measurable by Lemma A.3. By Prohorov's theorem (Van der Vaart and Wellner, 1996, Theorem 1.3.9), $X_\alpha$ is relatively compact. Take any subnet in $X_\alpha$ that is convergent. Its limit point is unique by Lemma A.4 and the assumption that every marginal converges weakly. Thus, $X_\alpha$ converges weakly. The last statement is another consequence of Prohorov's theorem. ■

Although we consider a different norm, our space contains the same elements as the classical literature (e.g., empirical processes). Therefore, weak convergence of marginals can easily be established by the classical results such as the multivariate central limit theorems. Hence the question that remains is how to establish asymptotic tightness.

The space of interest $\mathbb{L}_\mu$ is the intersection of the uniform space (with respect to the norm $\|\cdot\|_T$) and the $L_1$ space (with respect to the norm $\|\cdot\|_\mu$). As such, asymptotic tightness in $\mathbb{L}_\mu$ is equivalent to joint satisfaction of asymptotic tightness in each space. Again, tightness in the uniform space can be established by classical results. Following Van der Vaart and Wellner (1996), we characterize tightness in $L_1$ in two ways: through the finite approximation and by the Dunford-Pettis theorem. The second characterization connects asymptotic tightness in $L_1$ to asymptotic equiintegrability of the sample paths. In light of this, define the following.

**Definition.** For a $\mu$-measurable semimetric $\rho$ on $T$,[17] a function $f : T \to \mathbb{R}$ is *uniformly $\rho$-continuous and $(\rho, \mu)$-integrable* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $t \in T$

$$\Bigl(\sup_{\rho(s,t)<\delta} |f(s) - f(t)|\Bigr) \vee \Bigl(\int_{0<\rho(s,t)<\delta} |f(s)||d\mu(s)|\Bigr) < \varepsilon.$$

**Definition.** For a $\mu$-measurable semimetric $\rho$ on $T$, the net $X_\alpha : \Omega_\alpha \to \mathbb{L}_\mu$ is *asymptotically uniformly $\rho$-equicontinuous and $(\rho, \mu)$-equiintegrable in probability* if for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that

$$\limsup_\alpha P^*\biggl(\sup_{t\in T}\Bigl[\Bigl(\sup_{\rho(s,t)<\delta} |X_\alpha(s) - X_\alpha(t)|\Bigr) \vee \Bigl(\int_{0<\rho(s,t)<\delta} |X_\alpha(s)||d\mu(s)|\Bigr)\Bigr] > \varepsilon\biggr) < \eta.$$

[17]We define a semimetric to be *$\mu$-measurable* if every open set thereby induced is measurable with respect to $\mu$.

*Remark.* The nomenclature *equiintegrable* is based on the fact that the standard definition of $\mu$-equiintegrability (or "uniform" $\mu$-integrability) in functional analysis roughly coincides with $(\mu, \mu)$-equiintegrability defined herein, albeit $\mu$ is not a metric. Here we prefer the prefix *equi-* over the (arguably more popular) qualifier *uniformly* for the meaning "equally among the class of functions" in order to maintain coherence with *uniformly equicontinuous.*

**Theorem A.6.** *The following are equivalent.*

*(i) A net $X_\alpha : \Omega_\alpha \to \mathbb{L}_\mu$ is asymptotically tight.*

*(ii) $X_\alpha(t)$ is asymptotically tight in $\mathbb{R}$ for every $t \in T$, $\|X_\alpha\|_\mu$ is asymptotically tight in $\mathbb{R}$, and for every $\varepsilon, \eta > 0$ there exists a finite $\mu$-measurable partition $T = \bigcup_{i=1}^k T_i$ such that*

$$\limsup_\alpha P^*\left(\left[\sup_{1\le i\le k}\sup_{s,t\in T_i}|X_\alpha(s) - X_\alpha(t)|\right] \vee \sum_{i=1}^k \inf_{x\in\mathbb{R}}\int_{T_i}|X_\alpha - x||d\mu| > \varepsilon\right) < \eta. \quad (4)$$

*(iii) $X_\alpha(t)$ is asymptotically tight in $\mathbb{R}$ for every $t \in T$ and there exists a $\mu$-measurable semimetric $\rho$ on $T$ such that $(T, \rho)$ is totally bounded and $X_\alpha$ is asymptotically uniformly $\rho$-equicontinuous and $(\rho, \mu)$-equiintegrable in probability.*

*If, moreover, $X_\alpha \rightsquigarrow X$, then almost all paths $t \mapsto X(t, \omega)$ are uniformly $\rho$-continuous and $(\rho, \mu)$-integrable; and the semimetric $\rho$ can without loss of generality be taken equal to any semimetric $\rho$ for which this is true and $(T, \rho)$ is totally bounded.*

*Remark.* The condition on the supremum "$0 < \rho(s, t)$" is to allow for the point masses in $\mu$ and plateaus in $X_\alpha$. In (4), this condition corresponds to the subtraction of "$x$."

*Proof.* We prove (ii) $\Rightarrow$ (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii), and then the addendum.

(ii) $\Rightarrow$ (i). Fix $\varepsilon, \eta > 0$ and take the given partiton $T_1, \ldots, T_k$. Pick one $t_i$ from each $T_i$. Then, $\|X_\alpha\|_T \le \max_i |X_\alpha(t_i)| + \varepsilon$ with inner probability at least $1 - \eta$. Since the maximum of finitely many tight nets of real variables is tight and $\|X_\alpha\|_\mu$ is assumed to be tight, it follows that the net $\|X_\alpha\|_{\mathbb{L}_\mu}$ is asymptotically tight in $\mathbb{R}$.

Fix $\zeta > 0$ and take $\varepsilon_m \searrow 0$. Let $M$ be a constant such that $\limsup P^*(\|X_\alpha\|_{\mathbb{L}_\mu} > M) < \zeta$. Taking $(\varepsilon, \eta)$ in (4) to be $(\varepsilon_m, 2^{-m}\zeta)$, we obtain for each $m$ a measurable partition $T = \bigcup_{i=1}^k T_i$ (suppressing the dependence on $m$). For each $T_i$, enumerate all of the finitely many values $0 = a_{i,0} \le a_{i,1} \le \cdots \le a_{i,p} \le M$ such that

$$\int_{T_i}(a_{i,j} - a_{i,j})|d\mu| \le \frac{\varepsilon_m}{k} \quad \text{for} \quad j = 1, \ldots, p \qquad \text{and} \qquad \int_{T_i} a_{i,p}|d\mu| \le M.$$

Since $\mu$ is not necessarily finite on the whole $T$, it can be that on some partition $T_i$ the only possible choice of $a_{i,j}$ is 0. Let $z_1, \ldots, z_q$ be the finite exhaustion of all functions in $\mathbb{L}_\mu$ that are constant on each $T_i$ and take values only on

$$0, \pm\varepsilon_m, \ldots, \pm\lfloor M/\varepsilon_m\rfloor\varepsilon_m, \quad \pm a_{1,1}, \ldots, \pm a_{1,p}, \quad \ldots, \quad \pm a_{k,1}, \ldots, \pm a_{k,p}.$$

Again, it can be that on a partition with $\int_{T_i}|d\mu| = \infty$, the only value that $z_i$ can take is 0. Let $K_m$ be the union of $q$ closed balls of radius $2\varepsilon_m$ around each $z_i$. Then, since $\inf_j \int_{T_i}|X_\alpha - a_{i,j}||d\mu| \le \frac{\varepsilon_m}{k} + \inf_x \int_{T_i}|X_\alpha - x||d\mu|$, the following three conditions

$$\|X_\alpha\|_T \le M, \quad \sup_i \sup_{s,t\in T_i}|X_\alpha(s) - X_\alpha(t)| \le \varepsilon_m, \quad \sum_i \inf_x \int_{T_i}|X_\alpha - x||d\mu| \le \varepsilon_m$$

imply that $X_\alpha \in K_m$. This holds for each $m$.

Let $K = \bigcap_{m=1}^\infty K_m$, which is closed, totally bounded, and therefore compact. Moreover, we argue that for every $\delta > 0$ there exists $m$ with $K^\delta \supset \bigcap_{j=1}^m K_j$. Suppose not. Then there is a sequence $z_m$ not in $K^\delta$, but with $z_m \in \bigcap_{j=1}^m K_j$ for every $m$. This has a subsequence contained in only one of the closed balls constituting $K_1$, and a further subsequence contained in only one of the balls constituting $K_2$, and so on. The diagonal sequence of such subsequences would eventually be contained in a ball of radius $2\varepsilon_m$ for every $m$. Therefore, it is Cauchy and its limit should be in $K$, which is a contradiction to the supposition $d(z_m, K) \geq \delta$ for every $m$.

Thus, we conclude that if $X_\alpha$ is not in $K^\delta$, then it is not in $\bigcap_{j=1}^m K_j$ for some $m$. Therefore,

$$\begin{aligned}
P^*(X_\alpha \notin K^\delta) &\leq P^*\left(X_\alpha \notin \bigcap_{j=1}^m K_j\right) \\
&\leq P^*(\|X_\alpha\|_{\mathbb{L}_\mu} > M) \\
&\quad + \sum_{j=1}^m P^*\left(\left[\sup_i \sup_{s,t\in T_i} |X_\alpha(s) - X_\alpha(t)|\right] \vee \sum_i \inf_x \int_{T_i} |X_\alpha - x||d\mu| > \varepsilon_j\right) \\
&\leq \zeta + \sum_{j=1}^m \zeta 2^{-j} < 2\zeta.
\end{aligned}$$

Hence, we obtain $\limsup_\alpha P^*(X_\alpha \notin K^\delta) < 2\zeta$, as asserted.

(i) $\Rightarrow$ (iii). If $X_\alpha$ is asymptotically tight, then so is each coordinate projection. Therefore, $X_\alpha(t)$ is asymptotically tight in $\mathbb{R}$ for every $t \in T$.

Let $K_1 \subset K_2 \subset \cdots$ be a sequence of compact sets such that $\liminf P_*(X_\alpha \in K_m^\varepsilon) \geq 1 - 1/m$ for every $\varepsilon > 0$. Define a semimetric $d$ on $T$ induced by $z$ by

$$d(s,t;z) := |z(s) - z(t)| \vee \int_T |z| \mathbb{1}\{z(s) \wedge z(t) \leq z \leq z(s) \vee z(t)\} \mathbb{1}\{z(s) \neq z(t)\}|d\mu|.$$

Observe that $d(s,s;z) = 0$ and that $d$ is measurable with respect to $\mu$.[18] Now for every $m$, define a semimetric $\rho_m$ on $T$ by

$$\rho_m(s,t) := \sup_{z\in K_m} d(s,t;z).$$

We argue that $(T, \rho_m)$ is totally bounded. For $\eta > 0$, cover $K_m$ by finitely many balls of radius $\eta$ centered at $z_1, \ldots, z_k$. Consider the partition of $\mathbb{R}^{2k}$ into cubes of edge length $\eta$. For each cube, if there exists $t \in T$ such that the following $2k$-tuple is in the cube,

$$\begin{aligned}
r(t) := \Bigg(z_1(t),\ &\int_T z_1 \mathbb{1}\{0 \wedge z_1(t) \leq z_1 \leq 0 \vee z_1(t)\}|d\mu|, \quad \ldots, \\
&\qquad z_k(t),\ \int_T z_k \mathbb{1}\{0 \wedge z_k(t) \leq z_k \leq 0 \vee z_k(t)\}|d\mu|\Bigg),
\end{aligned}$$

then pick one such $t$. Since $\|z_j\|_{\mathbb{L}_\mu}$ is finite for every $j$ (i.e., the diameter of $T$ measured by each $d(\cdot,\cdot;z_j)$ is finite), this gives finitely many points $t_1, \ldots, t_p$. Notice that the balls

[18] $T$ is not necessarily complete with respect to $d$.

$\{t : \rho_m(t, t_i) < 3\eta\}$ cover $T$, that is, $t$ is in the ball around $t_i$ for which $r(t)$ and $r(t_i)$ are in the same cube; this follows because $\rho_m(t, t_i)$ can be bounded by

$$2 \sup_{z \in K_m} \inf_j \|z - z_j\|_{\mathbb{L}_\mu} + \sup_j d(t, t_i; z_j) < 3\eta.$$

The first term is the error of approximating $z(t)$ and $z(t_i)$ by $z_j(t)$ and $z_j(t_i)$; the second is the distance of $t$ and $t_i$ measured by $d(\cdot, \cdot; z_j)$.

Define the semimetric $\rho$ by

$$\rho(s, t) := \sum_{m=1}^{\infty} 2^{-m} \big(\rho_m(s, t) \wedge 1\big).$$

We show that $(T, \rho)$ is still totally bounded. For $\eta > 0$ take $m$ such that $2^{-m} < \eta$. Since $T$ is totally bounded in $\rho_m$, we may cover $T$ with finitely many $\rho_m$-balls of radius $\eta$. Denote by $t_1, \dots, t_p$ the centers of such a cover. Since $K_m$ is nested, we have $\rho_1 \le \rho_2 \le \cdots$. Since we also have $\rho_m(t, t_i) < \eta$, for every $t$ there exists $t_i$ such that $\rho(t, t_i) \le \sum_{k=1}^{m} 2^{-k} \rho_k(t, t_i) + 2^{-m} < 2\eta$. Therefore, $(T, \rho)$ is totally bounded.

By definition we have $d(s, t; z) \le \rho_m(s, t)$ for every $z \in K_m$ and that $\rho_m(s, t) \wedge 1 \le 2^m \rho(s, t)$. And if $\|z_0 - z\|_{\mathbb{L}_\mu} < \varepsilon$ for $z \in K_m$, then $d(s, t; z_0) < 2\varepsilon + d(s, t; z)$ for every pair $(s, t)$. Hence, we conclude that

$$K_m^\varepsilon \subset \left\{ z : \sup_{\rho(s,t) < 2^{-m}\varepsilon} d(s, t; z) \le 3\varepsilon \right\}.$$

Therefore, for $\delta < 2^{-m}\varepsilon$,

$$\begin{aligned}
&\liminf_\alpha P_* \left( \sup_{\rho(s,t)<\delta} d(s, t; X_\alpha) \le 3\varepsilon \right) \\
&\qquad \ge \liminf_\alpha P_* \left( \sup_{t \in T} \left[ \sup_{\rho(s,t)<\delta} |X_\alpha(s) - X_\alpha(t)| \vee \int_{0<\rho(s,t)<\delta} |X_\alpha(s)||d\mu| \right] \le 3\varepsilon \right) \\
&\qquad \ge 1 - \frac{1}{m}.
\end{aligned}$$

(iii) $\Rightarrow$ (ii). For $\varepsilon, \eta > 0$ and correspondingly taken $\delta > 0$, one may construct the finite partition of $T$, denoted by $\{T_i^\varepsilon\}$, as follows. Since $T$ is totally bounded, it can be covered with finitely many balls of radius $\delta$; let $t_1, \dots, t_K$ be their centers. Disjointify the balls to obtain $\{T_i^\varepsilon\}$. If $\int_{\{t_i\}} |X_\alpha||d\mu| > 0$, then further separate the partition $T_i^\varepsilon$ into $\{t_i\}$ and $T_i^\varepsilon \setminus \{t_i\}$ (then they both have the same center).

There are three types of components in the partition: (a) singleton components of mass points of $\mu$, (b) components with $|\mu|(T_i^\varepsilon) = \infty$, and (c) components with $|\mu|(T_i^\varepsilon) < \infty$. The size of (a) is controlled by construction, so we are to control (b) and (c) one by one. Clearly,

$$\sup_{s,t \in T_i^\varepsilon} |X_\alpha(s) - X_\alpha(t)| \le 2 \sup_{\rho(s,t_i)<\delta} |X_\alpha(s) - X_\alpha(t_i)| \le 2\varepsilon. \tag{5}$$

Denote by $i_\infty$ the index for which $|\mu|(T_{i_\infty}^\varepsilon) = \infty$. We argue that $\sum_{i_\infty} \int_{T_{i_\infty}^\varepsilon} |X_\alpha||d\mu|$ can be arbitrarily small (with inner probability at least $1 - \eta$) for sufficiently small $\varepsilon$. By the construction of the partition, $\sup_{s\in T^\varepsilon_{i_\infty}} |X_\alpha(s)| \le 2\varepsilon$.[19] Thus, $\sum_{i_\infty}\int_{T^\varepsilon_{i_\infty}} |X_\alpha||d\mu| \le \int_T |X_\alpha|\mathbb{1}\{|X_\alpha|\le 2\varepsilon\}|d\mu|$. Since $T$ is totally bounded by the given semimetric, $\int_T |X_\alpha||d\mu|$ is bounded by $K\varepsilon$ with inner probability at least $1-\eta$ (proving asymptotic tightness of $\|X_\alpha\|_\mu$), and hence the previous integral must be arbitrarily small for small $\varepsilon$. Now we turn to (c). Let $\varepsilon'$ be such that

$$\limsup_\alpha P^*\left(\int_T |X_\alpha|\mathbb{1}\{|X_\alpha|\le 3\varepsilon'\}|d\mu| > \varepsilon\right) < 1-\eta. \tag{6}$$

Take the partition for this $\varepsilon'$, namely $T^{\varepsilon'}_i$, to be nested on $T^\varepsilon_i$ and pick up only the components $\{T^{\varepsilon'}_j\}$ on $|\mu|(T^\varepsilon_i) < \infty$. Note that $\{T^\varepsilon_{i_\infty}\}\cup\{T^{\varepsilon'}_j\}$ defines another finite partition of $T$. For ease of notation, denote $T = T_\infty \sqcup T'$. If there exists $s\in T^{\varepsilon'}_j$ such that $|X_\alpha(s)|\le\varepsilon'$, then by the construction of the partition $\sup_{t\in T^{\varepsilon'}_j}|X_\alpha(t)|\le 3\varepsilon'$. The contrapositive of this is also true. Thus, observing

$$\begin{aligned}&\sum_j \inf_x \int_{T^{\varepsilon'}_j} |X_\alpha - x||d\mu| \\ &\qquad\le \sum_j \inf_x \int_{T^{\varepsilon'}_j} |X_\alpha - x|\mathbb{1}\{|X_\alpha|>\varepsilon'\}|d\mu| + \int_T |X_\alpha|\mathbb{1}\{|X_\alpha|\le 3\varepsilon'\}|d\mu|,\end{aligned}$$

we may assume $\inf_{T'}|X_\alpha(s)|\ge\varepsilon'>0$ at the cost of one more $\varepsilon$. Then, we also have $\int_{T'}|d\mu| \le K\varepsilon/\varepsilon'$ since $\varepsilon'\int_{T'}|d\mu|\le\int_T|X_\alpha||d\mu|$. For the partition $T^{\varepsilon'}_j$ of $T'$, further construct a nested finite partition $T^{\varepsilon'/K}_k$. Now

$$\begin{aligned}&\sum_k \inf_x \int_{T^{\varepsilon'/K}_k} |X_\alpha - x||d\mu| \\ &\qquad\le \sum_k \int_{T^{\varepsilon'/K}_k} |d\mu| \sup_{s,t\in T^{\varepsilon'/K}_k} |X_\alpha(s)-X_\alpha(t)| \le \frac{\varepsilon'}{K}\int_{T'}|d\mu| \le \varepsilon\end{aligned} \tag{7}$$

with inner probability at least $1-\eta$. Combine (5), (6), and (7) to yield the result.

Finally, we prove the addendum. Define $K_m$ as before. Then, if $X_\alpha \rightsquigarrow X$, we have $P(X\in K_m)\ge 1-1/m$, and hence $X$ concentrates on $\bigcup_{m=1}^\infty K_m$. Since elements of $K_m$ are uniformly $\rho_m$-equicontinuous and $(\rho_m,\mu)$-equiintegrable, they are also uniformly $\rho$-equicontinuous and $(\rho,\mu)$-equiintegrable. This proves the first statement. Next, note that the set of uniformly continuous and integrable functions on a totally bounded, semimetric (denote by $d$) space is complete and separable in $\mathbb{L}_\mu$. Thus the map $X$ that takes its values in this set is tight. If, moreover, $X_\alpha\rightsquigarrow X$, then $X_\alpha$ is asymptotically tight, so the compact sets for asymptotic tightness of $X_\alpha$ can be taken to be the compact sets for tightness of $X$. If every path of $X$ is uniformly $d$-continuous and $(d,\mu)$-integrable, then these compact sets can be chosen from the space of uniformly $d$-continuous and $(d,\mu)$-integrable functions. Since a compact set is totally bounded, every one of the compact sets is necessarily uniformly $d$-equicontinuous and $(d,\mu)$-equiintegrable. This completes the proof. ■

[19] This follows because $\inf_{T^\varepsilon_{i_\infty}}|X_\alpha| = 0$ given that $\int_{T^\varepsilon_{i_\infty}}|X_\alpha||d\mu|<\infty$.

Without having to resort to the classical central limit theorem for the $L_1$ spaces, Proposition A.2 can also be proved using Theorem A.6. This (loosely) checks consistency of our theorem with known results.

*Proof of Proposition A.2 through Theorem A.6.* Being continuously differentiable, $m$ can be represented as a difference $m_1 - m_2$ of two strictly increasing and continuously differentiable functions $m_1$ and $m_2$ such that $m_1(X)$ and $m_2(X)$ has a $(2+c)$th moment for $X \sim F$. In other words, we assume without loss of generality that $m$ is strictly increasing. Since asymptotic uniform equicontinuity is classical, in light of Theorem A.6 it remains only to show that the process $X_n := \sqrt{n}(\mathbb{F}_n - F)$ is asymptotically equiintegrable in probability. By Lemma A.1, $|y|^{1+c}\tilde{F} \circ m^{-1}(y)$ is integrable. This enables us to use the semimetric

$$\begin{aligned}\rho(s,t) &:= \left( \int_{m^{-1}(s)}^{m^{-1}(t)} \big(|y|^{1+c} \vee 1\big) |\tilde{F} \circ m^{-1}(y)| dy \right)^{1/2} \\ &= \left( \int_s^t \big(|m(x)|^{1+c} \vee 1\big) |\tilde{F}(x)| d\mu(x) \right)^{1/2},\end{aligned}$$

as it makes $\mathbb{R}$ totally bounded. By the Cauchy-Schwarz inequality,

$$\left( \int_s^t |X_n| d\mu \right)^2 \leq \left( \int_s^t \big(|m(x)|^{1+c} \vee 1\big) X_n^2 d\mu \right) \left( \int_{-\infty}^{\infty} \frac{1}{|m(x)|^{1+c} \vee 1} d\mu \right).$$

By the change of variables,

$$\int_{-\infty}^{\infty} \frac{1}{|m(x)|^{1+c} \vee 1} d\mu = \int_{-\infty}^{\infty} \frac{1}{|y|^{1+c} \vee 1} dy < \infty.$$

With $\mathbb{E}[X_n^2(x)] = F(x)[1-F(x)]$, this implies that for some constant $C$,

$$\mathbb{E}\left[ \left( \int_s^t |X_n| d\mu \right)^2 \right] \leq C \int_s^t \big(|m(x)|^{1+c} \vee 1\big) F(x)[1-F(x)] d\mu \leq C\rho(s,t)^2.$$

Therefore, by Van der Vaart and Wellner (1996, Theorem 2.2.4), for any $\eta, \delta > 0$,

$$\mathbb{E}\left[ \left( \sup_{\rho(s,t) \leq \delta} \int_s^t |X_n| d\mu \right)^2 \right] \leq K \left[ \int_0^\eta \sqrt{D(\varepsilon, \rho)} d\varepsilon + \delta D(\eta, \rho) \right]^2$$

for some constant $K$ where $D(\varepsilon, \rho)$ is the packing number of $T$ with respect to $\rho$. With this choice of $\rho$, the packing number satisfies $D(\varepsilon, \rho) \approx 1/\varepsilon$. Thus, we obtain

$$\mathbb{E}\left[ \left( \sup_{\rho(s,t) \leq \delta} \int_s^t |X_n| d\mu \right)^2 \right] \leq \tilde{K} \left( 2\sqrt{\eta} + \frac{\delta}{\eta} \right)^2$$

for some $\tilde{K}$. With Markov's inequality,

$$P\left( \sup_{t \in \mathbb{R}} \int_{0 < \rho(s,t) < \delta} |X_n| d\mu(s) > \varepsilon \right) \leq \frac{1}{\varepsilon^2} \mathbb{E}\left[ \left( \sup_{\rho(s,t) \leq \delta} \int_s^t |X_n| d\mu \right)^2 \right] \leq \frac{\tilde{K}}{\varepsilon^2} \left( 2\sqrt{\eta} + \frac{\delta}{\eta} \right)^2.$$

This can be however small for any $\varepsilon$ by the choice of $\eta$ and $\delta$. ■

## A.5 Differentiability of the Inverse Map

We use the space $\mathbb{L}$ constructed in the previous section to show differentiability of the inverse map for distribution functions. Let $m$ be a strictly increasing continuous function and $\mu$ the Lebesgue-Stieltjes measure associated with $m$. For a real-valued random variable $X$ distributed as $F$, Lemma A.1 implies that $m(X)$ has a first moment if and only if $\tilde{F}$ belongs to $\mathbb{L}_\mu$ with $T$ equal to $\mathbb{R}$ and $\mathcal{T}$ appropriately chosen. Specifically, the space we work on is as follows.

**Definition.** Let $\mathbb{L}_\mu$ be the space of $\mu$-measurable functions $z$ from $\mathbb{R}$ to $\mathbb{R}$ with limits $z(\pm\infty) := \lim_{x\to\pm\infty} z(x)$ and the norm

$$\|z\|_{\mathbb{L}_\mu} := \|z\|_\infty \vee \|z\|_\mu := \left(\sup_{x\in\mathbb{R}} |z(x)|\right) \vee \left(\int_{-\infty}^{\infty} |\tilde{z}(x)| d\mu\right)$$

where

$$\tilde{z}(x) = \begin{cases} z(x) - z(-\infty) & x < 0, \\ z(x) - z(+\infty) & x \geq 0. \end{cases}$$

Denote by $\mathbb{L}_{\mu,\phi}$ the subset of $\mathbb{L}_\mu$ of monotone cadlag functions with $z(-\infty) = 0$ and $z(+\infty) = 1$. If $\mu$ is the Lebesgue measure, we omit the subscript $\mu$ and denote them by $\mathbb{L}$ and $\mathbb{L}_\phi$, and $\|\cdot\|_\mu$ by $\|\cdot\|_1$.

Next, we define the space of quantile functions for distributions that have first moments. Note that there is no supremum component in its norm.

**Definition.** Let $\mathbb{B}$ be the space of ladcag functions $z$ from $(0,1)$ to $\mathbb{R}$ with the norm

$$\|z\|_{\mathbb{B}} := \int_0^1 |z(u)| du.$$

*Remark.* The metric on quantile functions induced by $\|\cdot\|_{\mathbb{B}}$ is known as the *Wasserstein metric* in probability theory and the *Mallows distance* in statistics.

The next lemma establishes differentiability of the inverse map for distribution functions with finite first moments (or more generally, monotone functions $F$ whose modifications $\tilde{F}$ are integrable).

**Lemma A.7** (Inverse map). *Let $F \in \mathbb{L}_\phi$ be a distribution function on (an interval of) $\mathbb{R}$ that has at most finitely many jumps and is otherwise continuously differentiable with strictly positive density $f$. Then, the inverse map $\phi : \mathbb{L}_\phi \to \mathbb{B}$, $\phi(F) := Q = F^{-1}$, is Hadamard differentiable at $F$ tangentially to the set $\mathbb{L}_0$ of all continuous functions in $\mathbb{L}$. The derivative is given by*

$$\phi'_F(z) = -(z \circ Q)Q' = \begin{cases} -(z/f) \circ Q & \text{if } Q \text{ is increasing,} \\ 0 & \text{if } Q \text{ is flat.} \end{cases}$$

*Proof.* Take $z_t \to z$ in $\mathbb{L}$ and $F_t := F + tz_t \in \mathbb{L}_\phi$. We want to show that

$$\left\| \frac{\phi(F_t) - \phi(F)}{t} - \phi'_F(z) \right\|_{\mathbb{B}} = \int_0^1 \left| \frac{\phi(F_t) - \phi(F)}{t} - \phi'_F(z) \right| du \longrightarrow 0 \quad \text{as} \quad t \to 0.$$

Let $j \in \mathbb{R}$ be a point of jump of $F$. For small $\varepsilon > 0$, one can separate the integral as

$$\left(\int_0^{F(j-\varepsilon)} + \int_{F(j-\varepsilon)}^{F(j+\varepsilon)} + \int_{F(j+\varepsilon)}^{1}\right)\left|\frac{\phi(F_t)-\phi(F)}{t} - \phi'_F(z)\right|du.$$

Observe that

$$\int_{F(j-\varepsilon)}^{F(j+\varepsilon)}\left|\frac{\phi(F_t)-\phi(F)}{t} - \phi'_F(z)\right|du \le 2\varepsilon\left\|\frac{F_t-F}{t}\right\|_\infty + \left(\int_{F(j-\varepsilon)}^{F(j-)} + \int_{F(j)}^{F(j+\varepsilon)}\right)|\phi'_F(z)|du.$$

The first term equals $2\varepsilon\|z_t\|_\infty$ and can be arbitrarily small by the choice of $\varepsilon$. If $\varepsilon$ is small enought that there is no other jump in $[j-\varepsilon, j+\varepsilon]$, by Fubini's theorem,

$$\int_{F(j-\varepsilon)}^{F(j-)}|\phi'_F(z)|du = \int_{j-\varepsilon}^{j}\left|\frac{z}{f}\right|dF \le \varepsilon\|z\|_\infty,$$

which can also be arbitrarily small by the choice of $\varepsilon$. Similarly, the last integral can as well be arbitrarily small. Therefore, one can ignore any finitely many jumps of $F$; we assume hereafter that $F$ has no jump and has positive density $f$ everywhere.

For every $\varepsilon > 0$ there exists a large number $M$ such that $F(-M) < \varepsilon$ and $1-F(M) < \varepsilon$. Write

$$\begin{aligned}\left\|\frac{\phi(F_t)-\phi(F)}{t} - \phi'_F(z)\right\|_{\mathbb{B}} \le& \int_{F(-M)+\varepsilon}^{F(M)-\varepsilon}\left|\frac{\phi(F_t)-\phi(F)}{t} - \phi'_F(z)\right|du \\ &+ \left(\int_0^{2\varepsilon} + \int_{1-2\varepsilon}^{1}\right)\left|\frac{\phi(F_t)-\phi(F)}{t} - \phi'_F(z)\right|du.\end{aligned}$$

By Van der Vaart and Wellner (1996, Theorem 3.9.23 (i)), uniform convergence of the integrand holds on $[F(-M)+\varepsilon, F(M)-\varepsilon]$. Since the first integral is bounded by

$$\sup_{u\in[F(-M)+\varepsilon, F(M)-\varepsilon]}\left|\frac{\phi(F_t)-\phi(F)}{t} - \phi'_F(z)\right|,$$

it vanishes as $t \to 0$.

We now turn to the second integral. The triangle inequality bounds the integral by

$$\int_0^{2\varepsilon}\left|\frac{\phi(F_t)-\phi(F)}{t}\right|du + \int_0^{2\varepsilon}|\phi'_F(z)|du.$$

Since $F$ and $F_t$ are nondecreasing, by Fubini's theorem,

$$\begin{aligned}\int_0^{2\varepsilon}\left|\frac{\phi(F_t)-\phi(F)}{t}\right|du = \frac{1}{|t|}\int_0^{2\varepsilon}\left|F_t^{-1} - F^{-1}\right|du &\le \frac{1}{|t|}\int_{-\infty}^{F^{-1}(2\varepsilon+\|tz_t\|_\infty)}|tz_t|dx \\ &\le \|z_t - z\|_{\mathbb{L}_1} + \int_{-\infty}^{F^{-1}(2\varepsilon+t\|z_t\|_\infty)}|z|dx.\end{aligned}$$

The first term goes to 0 and the second term can be arbitrarily small by the choice of $\varepsilon$. Finally, by the change of variables,

$$\int_0^{2\varepsilon}|\phi'_F(z)|du = \int_{-\infty}^{F^{-1}(2\varepsilon)}\left|\frac{z}{f}\right|dF = \int_{-\infty}^{F^{-1}(2\varepsilon)}|z|dx.$$

This quantity can be arbitrarily small. Similarly, the integral from $1-2\varepsilon$ to 1 can be shown to converge to 0. This completes the proof. ∎

*Remark.* A probability distribution $F$ has a $p$th moment if and only if its quantile function $Q$ is in $L_p$ (Lemma A.1). This may spur speculations that if $F$ has a $p$th moment, then the map $F \mapsto Q$ may be differentiable for $Q \in L_p$. However, we have not been able to prove that this is the case (although the Glivenko-Cantelli type results do hold; see Addendum A.22). The success of Lemma A.7 hinges upon the fact that Fubini's theorem is compatible with the $L_1$ norm. Nevertheless, it is possible to extend differentiability of inverse maps to subsume $p$th moments, or even to more general transformations, by regarding the range space to be $L_1$ as done below.

Now we extend the result to subsume transformations. Consider, for example, the second moment. Observe that the second moment of $X$ can be thought of as the first moment of $Z := X^2$. In other words,

$$\int_{-\infty}^{\infty} x^2 dF(x) = \int_0^{\infty} z dF(\sqrt{z}) + \int_{-\infty}^{0} -z dF(-\sqrt{-z}) = \int_0^{\infty} z d\big(F(\sqrt{z}) + F(-\sqrt{z})\big).$$

Here, $F(\sqrt{\cdot}) + F(-\sqrt{\cdot})$ is the distribution function of the random variable $Z$.[20] This is in line with the informal mental exercise that if one inverts the $p$th power of an inverse function, $(F^{-1})^p$, one obtains the composition of the original function and the $1/p$th power, $F \circ (\cdot)^{1/p}$. Thus, one expects that this composition, $F \circ (\cdot)^{1/p}$, is in $\mathbb{L}$ whenever $F$ is in $\mathbb{L}_{|x|^p}$. Then, Hadamard differentiability of the map $F \mapsto (F^{-1})^p$ may follow by the chain rule on $F \mapsto F \circ (\cdot)^{1/p} \mapsto [F \circ (\cdot)^{1/p}]^{-1} = (F^{-1})^p$. For this, one only needs Hadamard differentiability (or anything stronger) of the first map, $F \mapsto F \circ (\cdot)^{1/p}$.

*Remark.* As remarked above, a subtle but important distinction is that the last element in this chain $(F^{-1})^p$ should be seen as itself belonging to $\mathbb{B}$ (the $L_1$ space), but not as $F^{-1}$ belonging to the $L_p$ space.

More generally, for a monotone function $m$, we exploit the relationship $m(F^{-1}) = (F \circ m^{-1})^{-1}$. The requirement of monotonicity of $m$ is almost innocuous since any function of locally bounded variation can be represented by a difference of two increasing functions, so the lemma extends naturally to more general transformation.

**Lemma A.8.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a strictly increasing continuous function and $\mu$ be the associated Lebesgue-Stieltjes measure. Then, the map $\psi : \mathbb{L}_\mu \to \mathbb{L}$, $\psi(F) := F \circ m^{-1}$, is uniformly Fréchet differentiable with rate function $q \equiv 0$.*[21] *The derivative is given by $\psi'_F(z) := z \circ m^{-1}$.*

*Proof.* Observe that $\psi(F + z) - \psi(F) = (F + z)(m^{-1}) - F(m^{-1}) = z(m^{-1})$. Therefore, $\psi(F + z) - \psi(F) - \phi'_F(z) = 0$. ■

This lemma is obvious since the map $F \mapsto F \circ m^{-1}$ is by itself linear. Now differentiability of the inverse map for general transformations follows by the chain rule.

**Theorem A.9** (Transformed inverse map)**.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function and $\mu$ be the associated Lebesgue-Stieltjes measure. Let $F \in \mathbb{L}_{\mu,\phi}$ be a distribution function on (an interval of) $\mathbb{R}$ that has at most finitely many jumps and*

[20] After a minor fix for right-continuity.

[21] A map $\psi : \mathbb{L} \to \mathbb{B}$ is *uniformly Fréchet differentiable with rate function $q$* if there exists a continuous linear map $\psi'_F : \mathbb{L} \to \mathbb{B}$ such that $\|\psi(F + z) - \psi(F) - \psi'_F(z)\|_{\mathbb{B}} = O(q(\|z\|_{\mathbb{L}}))$ uniformly over $F \in \mathbb{L}$ as $z \to 0$ and $q$ is monotone with $q(t) = o(t)$.

*is otherwise continuously differentiable with strictly positive density $f$. Then, the map $\phi \circ \psi : \mathbb{L}_{\mu,\phi} \to \mathbb{B}$, $\phi \circ \psi(F) := m(Q)$, is Hadamard differentiable at $F$ tangentially to the set $\mathbb{L}_{\mu,0}$ of all continuous functions in $\mathbb{L}_\mu$. The derivative is given by*

$$(\phi \circ \psi)'_F(z) := -(m'z/f) \circ Q.$$

*Proof.* Since $m$ is continuously differentiable and hence of locally bounded variation, one can set $m(x) = m_1(x) - m_2(x)$, where $m_1$ and $m_2$ are both increasing and continuously differentiable functions. Moreover, $m_1$ and $m_2$ can be taken such that they are strictly increasing, and for their corresponding Lebesgue-Stieltjes measures $\mu_1$ and $\mu_2$, $F$ belongs to both $\mathbb{L}_{\mu_1,\phi}$ and $\mathbb{L}_{\mu_2,\phi}$.[22] Since the derivative formula is linear in $m'$, it suffices to show that the claim holds for each of $m_1$ and $m_2$ separately. In other words, we can assume without loss of generality that $m$ is strictly increasing.

Now observe that $z$ is in $\mathbb{L}_\mu$ (or $\mathbb{L}_{\mu,0}$) if and only if $z \circ m^{-1}$ is in $\mathbb{L}$ (or $\mathbb{L}_0$). The assertion then follows by the chain rule (Van der Vaart and Wellner, 1996, Lemma 3.9.3) applied to Lemmas A.7 and A.8. ∎

The main conclusion of this section is summarized as follows.

**Proposition A.10.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function. For a distribution function $F$ on (an interval of) $\mathbb{R}$ that has at most finitely many jumps and is otherwise continuously differentiable with strictly positive density $f$ such that $m(X)$ has a $(2+c)$th moment for $X \sim F$ and some $c > 0$, the process $\sqrt{n}(m(\mathbb{Q}_n) - m(Q))$ converges weakly in $\mathbb{B}$ to a Gaussian process with mean zero and covariance function $\operatorname{Cov}(s,t) = m'(Q(s))Q'(s)m'(Q(t))Q'(t)(s \wedge t - st)$.*

*Proof.* This follows in combination of Proposition A.2 and Theorem A.9. ∎

## A.6 Differentiability of $L$-Statistics

Given the weak convergence of the integral of empirical quantiles, we investigate whether this is the case for the integral with respect to a random measure. Or precisely, we seek conditions under which the integral of a stochastic process with respect to another stochastic process converges weakly. This is the extension of the results on Wilcoxon statistics (Van der Vaart and Wellner, 1996, Section 3.9.4.1) to allow for an unbounded integrand. Then, the general $L$-statistics result follows with Proposition A.10.

The spaces we work on in this section are given as follows.

**Definition.** Let $Q \in \mathbb{B}$ and define the space $\mathbb{L}_Q$ of functions $\kappa : (0,1) \to \mathbb{R}$ with the norm

$$\|\kappa\|_{\mathbb{L}_Q} := \|\kappa\|_{Q,\infty} \vee \|\kappa\|_Q := \Big(\sup_{u\in(0,1)} |(|Q| \vee 1)(u)\kappa(u)|\Big) \vee \Big(\int_0^1 |\kappa(u)||dQ(u)|\Big).$$

Let $\mathbb{L}_{Q,M}$ be the subset of $\mathbb{L}_Q$ of Lipschitz functions.

First, we give a generalization of differentiability of Wilcoxon statistics.

[22] For example, take $m_1$ and $m_2$ to be the least steep functions and add a normal cdf.

**Theorem A.11** (Wilcoxon statistic)**.** *For each fixed $M$, the maps $\lambda : \mathbb{B} \times \mathbb{L}_{Q,M} \to \mathbb{R}$ and $\tilde{\lambda} : \mathbb{B} \times \mathbb{L}_{Q,M} \to L_\infty(0,1)^2$,*

$$\lambda(Q,K) := \int_0^1 QdK \qquad \textit{and} \qquad \tilde{\lambda}(Q,K) := \int_s^t QdK,$$

*are Hadamard differentiable at every $(Q,K) \in \mathbb{B} \times \mathbb{L}_{Q,M}$ uniformly over $\mathbb{L}_{Q,M}$. The derivative maps are given by*

$$\lambda'_{Q,K}(z,\kappa) := \int_0^1 Qd\kappa + \int_0^1 zdK$$

*where $\int Qd\kappa$ is defined via integration by parts if $\kappa$ is of unbounded variation.*

*Proof.* The derivative map is linear by construction; it is also continuous since

$$\begin{aligned}\left|\lambda'_{Q,K}(z_1,\kappa_1) - \lambda'_{Q,K}(z_2,\kappa_2)\right| &= \left|\int_0^1 Qd(\kappa_1-\kappa_2) + \int_0^1 (z_1-z_2)dK\right| \\ &\le \|\kappa_1-\kappa_2\|_{Q,\infty} + \|\kappa_1-\kappa_2\|_Q + M\|z_1-z_2\|_{\mathbb{B}},\end{aligned}$$

which vanishes as $\|z_1 - z_2\|_{\mathbb{B}} \to 0$ and $\|\kappa_1-\kappa_2\|_{\mathbb{L}_Q} \to 0$. Let $z_t \to z$ and $\kappa_t \to \kappa$ such that $Q_t := Q + tz_t$ is in $\mathbb{B}$ and $K_t := K + t\kappa_t$ is in $\mathbb{L}_{Q,M}$. Observe

$$\frac{\lambda(Q_t,K_t) - \lambda(Q,K)}{t} - \lambda'_{Q,K}(z_t,\kappa_t) = \int (z_t - z)d(K_t-K) + \int zd(K_t-K).$$

We want to show that this converges to zero as $t \to 0$. The first term vanishes since

$$\left|\int (z_t-z)d(K_t-K)\right| \le 2M\int |z_t - z|du = 2M\|z_t-z\|_{\mathbb{B}}.$$

Since $z$ is integrable, for every $\varepsilon > 0$ there exists a small number $\delta > 0$ such that

$$\left(\int_0^\delta + \int_{1-\delta}^1\right)|z|du + \int_\delta^{1-\delta}(|z| - (|z| \wedge \delta^{-1}))du \le \varepsilon.$$

This gives the inequality

$$\begin{aligned}\left|\int zd(K_t-K)\right| &\le \left|\int_\delta^{1-\delta}(-\delta^{-1} \vee z \wedge \delta^{-1})d(K_t-K)\right| \\ &\quad + \left|\int zd(K_t-K) - \int_\delta^{1-\delta}(-\delta^{-1} \vee z \wedge \delta^{-1})d(K_t-K)\right| \\ &\le \left|\int_\delta^{1-\delta}(-\delta^{-1} \vee z \wedge \delta^{-1})d(K_t-K)\right| + 2M\varepsilon.\end{aligned}$$

Let $\tilde{z} := -\delta^{-1} \vee z \wedge \delta^{-1}$. Since $\tilde{z}$ is ladcag on $[\delta, 1-\delta]$, there exists a partition $\delta = t_0 < t_1 < \cdots < t_m = 1-\delta$ such that $\tilde{z}$ varies less than $\varepsilon$ on each interval $(t_{i-1}, t_i]$. Let $\bar{z}$ be the piecewise constant function that equals $\tilde{z}(t_i)$ on each interval $(t_{i-1}, t_i]$. Then

$$\begin{aligned}\left|\int_\delta^{1-\delta}\tilde{z}d(K_t-K)\right| &\le 2M\sup_{u\in[\delta,1-\delta]}|\tilde{z}-\bar{z}| + |\tilde{z}(\delta)||(K_t-K)(\{\delta\})| \\ &\quad + \sum_{i=1}^m |\tilde{z}(t_i)||(K_t-K)((t_{i-1},t_i])|.\end{aligned}$$

The first term is arbitrarily small by the choice of $\varepsilon$, and the second and third terms are collectively bounded by $(2m+1)\delta^{-1}\|K_t - K\|_\infty = (2m+1)\delta^{-1}t\|\kappa_t\|_\infty$, which converges to 0 regardless of the choice of $K$.

The proof for $\tilde{\lambda}$ is basically the same as that for $\lambda$. ∎

Next, we give conditions under which the random measure $\mathbb{K}_n$ converges in $\mathbb{L}_Q$. Not surprisingly, this convergence hinges on the tail behavior around 0 and 1. Roughly speaking, if $Q$ has a $(2+c)$th moment, then weak convergence of $\frac{X_\alpha}{u^r(1-u)^r}$ to $\frac{X}{u^r(1-u)^r}$ in $L_\infty$ for some $r > \frac{1}{2+c}$ implies weak convergence of $X_\alpha$ to $X$ in $\mathbb{L}_Q$.

**Lemma A.12.** *Let $Q : (0,1) \to \mathbb{R}$ be a quantile function whose probability measure has a $(2+c)$th moment for some $c > 0$. If for a net of processes $X_\alpha : \Omega_\alpha \to \mathbb{L}_Q$ there exists $r > \frac{1}{2+c}$ such that for every $\eta > 0$ there exists $M$ satisfying*

$$\limsup_\alpha P^*\left(\left\|\frac{X_\alpha}{u^r(1-u)^r}\right\|_\infty > M\right) < \eta,$$

*then there exists a semimetric $\rho$ on $(0,1)$ such that $(0,1)$ is totally bounded, $QX_\alpha$ is asymptotically uniformly $\rho$-equicontinuous in probability, and $X_\alpha$ is asymptotically $(\rho, Q)$-equiintegrable in probability.*

*Proof.* Assume $r < 1$ first. Define $\rho$ by[23]

$$\rho(s,t) := \int_{(s,t)} u^r(1-u)^r dQ.$$

We show that $(0,1)$ is totally bounded with respect to $\rho$. Observe that Lemma A.1 and $r > \frac{1}{2+c}$ imply $u^r(1-u)^r Q(u) \to 0$ as $u \to \{0,1\}$. Therefore, integrating by parts,

$$\rho(0,1) \leq \int_{(0,1)} u^r \wedge (1-u)^r dQ \leq |Q|\left(\frac{1}{2}\right) + \int_0^{\frac{1}{2}} u^{r-1}|Q|du + \int_{\frac{1}{2}}^1 (1-u)^{r-1}|Q|du.$$

Since $Q \in L_{2+c}$ and $u^{r-1} \wedge (1-u)^{r-1} \in L_q$ for every $q < 1/(1-r)$, in particular for $q = (2+c)/(1+c)$, this integral is finite by Hölder's inequality. This means that the diameter of $(0,1)$ is finite, concluding that $(0,1)$ is totally bounded.

Note that $|Q|$ is eventually smaller than $1/u^r(1-u)^r$ near 0 and 1, so that for every $\eta$ there exists $M$ such that

$$\limsup_\alpha P^*(\|(|Q| \vee 1)X_\alpha\|_\infty > M) \leq \limsup_\alpha P^*\left(\left\|\frac{X_\alpha}{u^r(1-u)^r}\right\|_\infty > M\right) < \eta.$$

This shows uniform equicontinuity.

Next, for every $0 < s \leq t < 1$,

$$\int_{(s,t)} |X_\alpha| dQ \leq \left(\sup_{u\in(0,1)} \frac{|X_\alpha(u)|}{u^r(1-u)^r}\right) \int_{(s,t)} u^r(1-u)^r dQ \leq \left\|\frac{X_\alpha}{u^r(1-u)^r}\right\|_\infty \rho(s,t).$$

Therefore,

$$P^*\left(\sup_{t\in(0,1)} \int_{0<\rho(s,t)<\delta} |X_\alpha| dQ(s) > \varepsilon\right) \leq P^*\left(\left\|\frac{X_\alpha}{u^r(1-u)^r}\right\|_\infty > \frac{\varepsilon}{\delta}\right).$$

[23]This semimetric is reminiscent of the condition for $\|\cdot/q\|_p$-metric convergence in Csörgő et al. (1993).

By assumption, this can be however small by the choice of $\delta$. Conclude that $X_\alpha$ is asymptotically $(\rho, Q)$-equiintegrable in probability.

Finally, if $r \geq 1$, replace every $r$ appeared in the proof by $1/2$. Then the result follows since $\big\|\frac{X_\alpha}{u^r(1-r)^r}\big\|_\infty \geq \big\|\frac{X_\alpha}{u^{1/2}(1-u)^{1/2}}\big\|_\infty$. ∎

Next, we apply this lemma to show that most "well-behaved" sample selection measures satisfy the condition. Let $X_1, \dots, X_n$ be independent continuous random variables and $X_{1,n}, \dots, X_{m,n}$ be subset of $X_1, \dots, X_n$ that are selected by some (possibly random) criterion. Then, roughly speaking, if the empirical distribution of the selected sample $X_{1,n}, \dots, X_{m,n}$ converges in $L_\infty$ to a smooth distribution, then the selection measure $\mathbb{K}_n$ defined in the text converges in $\mathbb{L}_Q$.

**Proposition A.13.** *Let $U_1, \dots, U_n$ be independent uniformly distributed random variables on $(0,1)$ and $w_{1,n}, \dots, w_{n,n}$ random variables bounded by some constant $M$ whose distribution can depend on $U_1, \dots, U_n$ and $n$. Define*

$$\mathbb{F}_n(u) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{U_i \leq u\}, \qquad \mathbb{G}_n(u) := \frac{1}{n}\sum_{i=1}^n w_{i,n}\mathbb{1}\{U_i \leq u\}.$$

*Let $I(u) := u$ and assume that $K(u) := \lim_{n\to\infty} \mathbb{E}[\mathbb{G}_n(u)]$ exists and is Lipschitz and differentiable. If $\sqrt{n}(\mathbb{G}_n - K)$ weakly converges jointly with $\sqrt{n}(\mathbb{F}_n - I)$ in $L_\infty$, then for the selection measure*

$$\mathbb{K}_n(u) := \frac{1}{n}\sum_{i=1}^n w_{i,n}\mathbb{1}\big\{0 \vee \big(nu - n\mathbb{F}_n(U_i) + 1\big) \wedge 1\big\},$$

*we have $\sqrt{n}(\mathbb{K}_n - K)$ converge weakly in $\mathbb{L}_Q$ for every quantile function $Q$ whose distribution has a $(2+c)$th moment for some $c > 0$.*

*Proof.* Assume without loss of generality that $M = 1$. Define $U_{(0)} := 0$. Let $\tilde{\mathbb{F}}_n$ and $\tilde{\mathbb{G}}_n$ be the continuous linear interpolations of $\mathbb{F}_n$ and $\mathbb{G}_n$, that is, for $U_{(i-1)} \leq u < U_{(i)}$,

$$\begin{aligned}
\tilde{\mathbb{F}}_n(u) &:= \frac{i-1}{n} + \frac{u - U_{(i-1)}}{n(U_{(i)} - U_{(i-1)})}, \\
\tilde{\mathbb{G}}_n(u) &:= \frac{1}{n}\sum_{i=1}^n w_{i,n}\mathbb{1}\{U_i \leq u\} + \frac{w_{i,n}(u - U_{(i-1)})}{n(U_{(i)} - U_{(i-1)})},
\end{aligned}$$

and for $u \geq U_{(n)}$, $\tilde{\mathbb{F}}_n(u) := 1$ and $\tilde{\mathbb{G}}_n(u) := \frac{1}{n}\sum w_{i,n}$. Observe that $\mathbb{K}_n(u) = \tilde{\mathbb{G}}_n(\tilde{\mathbb{F}}_n^{-1}(u))$. By Lemma A.14 it suffices to show that $\sqrt{n}(\tilde{\mathbb{F}}_n - I)$ and $\sqrt{n}(\tilde{\mathbb{G}}_n - K)$ converge weakly jointly in $\mathbb{L}_Q$. Note that

$$\begin{aligned}
\|\mathbb{F}_n - I\|_\infty - \frac{1}{n} \leq \|\tilde{\mathbb{F}}_n - I\|_\infty \leq \|\mathbb{F}_n - I\|_\infty + \frac{1}{n}, \\
\|\mathbb{F}_n - I\|_Q - C \leq \|\tilde{\mathbb{F}}_n - I\|_Q \leq \|\mathbb{F}_n - I\|_Q + C,
\end{aligned}$$

for $C = \int(\tilde{I} - \lfloor n\tilde{I}\rfloor/n)dQ = O(1/n)$. Thus, $\sqrt{n}(\tilde{\mathbb{F}}_n - I)$ converges weakly in $\mathbb{L}_Q$ if and only if $\sqrt{n}(\mathbb{F}_n - I)$ does, and they share the same limit. The same is true for $\tilde{\mathbb{G}}_n$ and $\mathbb{G}_n$.

The classical results imply that $\sqrt{n}(\mathbb{F}_n - I)$ converges weakly in $L_\infty$ to a Brownian bridge and

$$\left\| \frac{\sqrt{n}(\mathbb{F}_n - I)}{u^r(1-u)^r} \right\|_\infty = O_P(1)$$

for every $r < 1/2$ (Csörgő and Horváth, 1993). By Lemma A.12 it follows that $\sqrt{n}(\mathbb{F}_n - I)$ converges weakly in $\mathbb{L}_Q$. By assumption $\sqrt{n}(\mathbb{G}_n - K)$ converges weakly in $L_\infty$ jointly with $\sqrt{n}(\mathbb{F}_n - I)$, and since $M|\mathbb{F}_n - I| \geq |\mathbb{G}_n - K|$, conclude that $\sqrt{n}(\mathbb{G}_n - K)$ converges weakly in $\mathbb{L}_Q$ jointly with $\sqrt{n}(\mathbb{F}_n - I)$. ■

**Lemma A.14** (Inverse composition map). *Let $\mathbb{L}_Q$ contain the identity map $I(u) := u$. Let $\mathcal{D}$ be the subset of $\mathbb{L}_Q \times \mathbb{L}_Q$ such that every $(A, B) \in \mathcal{D}$ satisfies $A(u_1) - A(u_2) \geq B(u_1) - B(u_2) \geq 0$ for every $u_1 \geq u_2$, the range of $A$ contains $(0, 1)$, and $B$ is differentiable and Lipschitz. Let $\mathbb{L}_{Q,UC}$ be the subset of $\mathbb{L}_Q$ of uniformly continuous functions. Then, the map $\chi : \mathcal{D} \to \mathbb{L}_Q$, $\chi(A, B) := B \circ A^{-1}$, is Hadamard differentiable at $(A, B) \in \mathcal{D}$ for $A = I$ tangentially to $\mathbb{L}_Q \times \mathbb{L}_{Q,UC}$. The derivative is given by*

$$\chi'_{I,B}(a, b)(u) = b(u) + B'(u)a(u), \qquad u \in (0, 1).$$

*Proof.* For $(A, B) \in \mathcal{D}$ and $u_1 \geq u_2$, denote $v_1 := A(u_1)$ and $v_2 := A(u_2)$. By assumption we have $v_1 - v_2 \geq B(A^{-1}(v_1)) - B(A^{-1}(v_2)) \geq 0$ for every $v_1 \geq v_2$. Therefore, $B \circ A^{-1}$ is monotone and bounded by the identity map up to a constant. This implies

$$\int_{(0,1)} |\widetilde{B \circ A^{-1}}||dQ| \leq \int_{(0,1)} |\tilde{I}||dQ| < \infty$$

and $\|Q(B \circ A^{-1})\|_\infty < \infty$; it follows that $B \circ A^{-1}$ is in $\mathbb{L}_{Q,1}$.

Let $a_t \to a$ and $b_t \to b$ in $\mathbb{L}_{\nu_2}$ and $(A_t, B_t) := (I + ta_t, B + tb_t) \in \mathcal{D}$. We want to show that

$$\left\| \frac{B_t \circ A_t^{-1} - B \circ I^{-1}}{t} - b - B'a \right\|_{\mathbb{L}_Q} \longrightarrow 0 \qquad \text{as} \qquad t \to 0.$$

That $\| \cdot \|_{Q,\infty} \to 0$ follows by applying Van der Vaart and Wellner (1996, Lemma 3.9.27) to $(A^{-1}, QB)$ as elements in $L_\infty$. Thus, it remains to show that $\| \cdot \|_Q \to 0$. In the assumed inequality, substitute $(u_1, u_2)$ by $(u, A_t^{-1}(u))$ to find that

$$|A_t(u) - u| \geq |B_t(A_t^{-1}(u)) - B_t(u)| \geq 0.$$

Therefore, the following inequality holds pointwise:

$$|B_t \circ A_t^{-1} - B| \leq |B_t \circ A_t^{-1} - B_t| + |B_t - B| \leq |A_t - I| + |B_t - B| = |ta_t| + |tb_t|.$$

For $\varepsilon > 0$, write $\| \cdot \|_Q$ as

$$\left( \int_0^\varepsilon + \int_\varepsilon^{1-\varepsilon} + \int_{1-\varepsilon}^1 \right) \left| \frac{B_t \circ A_t^{-1} - B}{t} - b - B'a \right| |dQ|.$$

For any fixed $\varepsilon > 0$ the middle term vanishes as $t \to 0$ since $\| \cdot \|_{Q,\infty} \to 0$. It remains to show that the first term can be however small by the choice of $\varepsilon$ since then by symmetry the third term must likewise be ignorable. Using the inequality obtained above, write

$$\int_0^\varepsilon \left| \frac{B_t \circ A_t^{-1} - B}{t} - b - B'a \right| |dQ| \leq \int_0^\varepsilon \Big( |a_t| + |b_t| + |b| + |B'a| \Big) |dQ|.$$

Since $\|a_t - a\|_Q \to 0$ and $\|b_t - b\|_Q \to 0$, this integral should be arbitrarily small by the choice of $\varepsilon$, as desired. ■

Now we are ready to give the main conclusion of this paper.

**Proposition A.15** ($L$-statistic)**.** *Let $m_1, m_2 : \mathbb{R} \to \mathbb{R}$ be continuously differentiable functions and $F : \mathbb{R}^2 \to [0,1]$ be a distribution function on (a rectangular of) $\mathbb{R}^2$ with marginal distributions $(F_1, F_2)$ that have at most finitely many jumps and are otherwise continuously differentiable with strictly positive marginal densities $(f_1, f_2)$ such that $m_1(X_1)$ and $m_2(X_2)$, $(X_1, X_2) \sim F$, have $(2+c)$th moments for some $c > 0$. Along with i.i.d. random variables $X_{1,1}, \dots, X_{n,1}$ and $X_{1,2}, \dots, X_{n,2}$, let $w_{1,n,1}, \dots, w_{n,n,1}$ and $w_{1,n,2}, \dots, w_{n,n,2}$ be random variables bounded by a constant $M$ whose distributions can depend on $n$ and all of $X_{1,1}, \dots, X_{n,1}$ and $X_{1,2}, \dots, X_{n,2}$ such that the empirical distributions of $X_{i,1}$, $X_{i,2}$, $w_{i,n,1}X_{i,1}$, and $w_{i,n,2}X_{i,2}$ converge uniformly jointly to continuously differentiable distribution functions. Then, the normalized $L$-statistics*

$$\sqrt{n}\begin{pmatrix} \mathbb{E}_n[m_1(X_{i,1})w_{i,n,1}] - \mathbb{E}[m_1(X_{i,1})w_{i,n,1}] \\ \mathbb{E}_n[m_2(X_{i,2})w_{i,n,2}] - \mathbb{E}[m_2(X_{i,2})w_{i,n,2}] \end{pmatrix} = \sqrt{n}\begin{pmatrix} \int_0^1 m_1(\mathbb{Q}_{n,1})d\mathbb{K}_{n,1} - \int_0^1 m_1(Q_1)dK_1 \\ \int_0^1 m_2(\mathbb{Q}_{n,2})d\mathbb{K}_{n,2} - \int_0^1 m_2(Q_2)dK_2 \end{pmatrix}$$

*where*

$$\mathbb{K}_{n,j}(u) := \frac{1}{n}\sum_{i=1}^{n} w_{i,n,j}\mathbb{1}\big\{0 \vee \big(nu - n\mathbb{F}_{n,j}(X_i) + 1\big) \wedge 1\big\},$$
$$K_j(u) := \lim_{n\to\infty} \mathbb{E}[w_{i,n,j} \mid F_j(X_{i,j}) \leq u],$$

*converge weakly in $\mathbb{R}^2$ to a normal vector $(\xi_1, \xi_2)$ with mean zero and (co)variance*

$$\begin{aligned}\mathrm{Cov}(\xi_j, \xi_k) = \int_0^1 \int_0^1 & m_j'(Q_j(s))Q_j'(s)m_k'(Q_k(t))Q_k'(t) \times \\ & \Big([F_{jk}^Q(s,t) - st] + [K_{jk}(s,t)F_{jk}^Q(s,t) - stK_j(s)K_k(t)] \\ & - K_j(s)[F_{jk}^Q(s,t) - st] - K_k(t)[F_{jk}^Q(s,t) - st]\Big)dsdt,\end{aligned}$$

*where $F_{jk}^Q(s,t) := \Pr(X_{i,j} \leq Q_j(s), X_{i,k} \leq Q_k(t))$ and $K_{jk}(s,t) := \lim_{n\to\infty} \mathbb{E}[w_{i,n,j}w_{i,n,k} \mid X_{i,j} \leq Q_j(s), X_{i,k} \leq Q_k(t)]$. If $F$ has no jumps, this is equal to*

$$\begin{aligned}\mathrm{Cov}(\xi_j, \xi_k) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & \Big([1 - K_j^F(x) - K_k^F(y)][F_{jk}(x,y) - F_j(x)F_k(y)] \\ & + [K_{jk}^F(x,y)F_{jk}(x,y) - K_j^F(x)K_k^F(y)F_j(x)F_k(y)]\Big)dm_j(x)dm_k(y),\end{aligned}$$

*where $F_{jk}(x,y) := \Pr(X_{i,j} \leq x, X_{i,k} \leq y)$ and $K_{jk}^F(x,y) := \lim_{n\to\infty} \mathbb{E}[w_{i,n,j}w_{i,n,k} \mid X_{i,j} \leq x, X_{i,k} \leq y]$. Given $m_j$ and $m_k$ known, this can be consistently estimated by its sample analogue*

$$\begin{aligned}\widehat{\mathrm{Cov}(\xi_j, \xi_k)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} & \Big([1 - \mathbb{K}_{n,j}^F(x) - \mathbb{K}_{n,k}^F(y)][\mathbb{F}_{n,jk}(x,y) - \mathbb{F}_{n,j}(x)\mathbb{F}_{n,k}(y)] \\ & + [\mathbb{K}_{n,jk}^F(x,y)\mathbb{F}_{n,jk}(x,y) - \mathbb{K}_{n,j}^F(x)\mathbb{K}_{n,k}^F(y)\mathbb{F}_{n,j}(x)\mathbb{F}_{n,k}(y)]\Big)dm_j(x)dm_k(y),\end{aligned}$$

*where* $\mathbb{F}_{n,jk}(x,y) := \mathbb{E}_n[\mathbb{1}\{X_{i,j} \le x, X_{i,k} \le y\}]$ *and* $\mathbb{K}^F_{n,jk}(x,y) := \mathbb{E}_n[w_{i,n,j}w_{i,n,k} \mid X_{i,j} \le x, X_{i,k} \le y]$.

*Proof.* Weak convergence follows by Propositions A.10 and A.13 and Theorem A.11. The derivative formulas give us

$$\begin{aligned}\operatorname{Cov}(\xi_j,\xi_k) &= \int_0^1\int_0^1 m_j'(Q_j(s))Q_j'(s)m_k'(Q_k(t))Q_k'(t)[F^Q_{ik}(s,t)-st]dsdt\\&+\int_0^1\int_0^1 m_j'(Q_j(s))Q_j'(s)m_k'(Q_k(t))Q_k'(t)[K_{jk}(s,t)F^Q_{jk}(s,t)-stK_j(s)K_k(t)]dsdt\\&-\int_0^1\int_0^1 m_j'(Q_j(s))Q_j'(s)m_k'(Q_k(t))Q_k'(t)K_j(s)[F^Q_{jk}(s,t)-st]dsdt\\&-\int_0^1\int_0^1 m_j'(Q_j(s))Q_j'(s)m_k'(Q_k(t))Q_k'(t)K_k(t)[F^Q_{jk}(s,t)-st]dsdt,\end{aligned}$$

where $F^Q_{jk}(s,t) := \Pr(X_{i,j} \le Q_j(s), X_{i,k} \le Q_k(t))$ and $K_{jk}(s,t) := \lim_{n\to\infty}\mathbb{E}[w_{i,n,j}w_{i,n,k} \mid X_{i,j} \le Q_j(s), X_{i,k} \le Q_k(t)]$. Apply the change of variables $x = Q_j(s)$ and $y = Q_k(t)$ to obtain the second formula. Consistency of the sample analogue estimator follows by uniform convergence of $\mathbb{K}^F_{n,j}$ and $\mathbb{K}^F_{n,k}$ and Addendum A.21. ∎

## A.7 Validity of Nonparametric Bootstrap

In this section, we establish the validity of nonparametric bootstrap. Since we have proved the Hadamard differentiability of maps involved, it remains to show the weak convergence of bootstrap processes conditional on the original observations.

The bootstrap process is given by

$$\begin{aligned}\hat{\mathbb{Z}}_n(x) := \sqrt{n}(\hat{\mathbb{F}}_n - \mathbb{F}_n)(x) &:= \frac{1}{\sqrt{n}}\sum_{i=1}^n (M_{ni}-1)\mathbb{1}\{X_i \le x\}\\&= \frac{1}{\sqrt{n}}\sum_{i=1}^n (M_{ni}-1)(\mathbb{1}\{X_i \le x\} - F(x))\end{aligned}$$

where $M_{ni}$ is the number of times $X_i$ is drawn in the bootstrap sample. We want to prove that this process converges weakly to the same limit as $\mathbb{Z}_n := \sqrt{n}(\mathbb{F}_n - F)$ conditional on the observations $X_i$. The strategy is very similar to the one employed in Van der Vaart and Wellner (1996, Chapter 3.6) and goes as follows: since $M_{ni}$ sums up to $n$, it is slightly dependent to each other; we replace $M_{ni}$ with *independent* Poisson random variables $\xi_i$ by Poissonization, that is, we show equivalence of weak convergence of the bootstrap process $\hat{\mathbb{Z}}_n$ and the multiplier process $\mathbb{Z}_n' := n^{-1/2}\sum \xi_i(\mathbb{1}\{X_i \le x\} - F)$ (Lemma A.16); then, we prove unconditional convergence of $\mathbb{Z}_n'$ (so randomness comes from both $X_i$ and $\xi_i$) by symmetrization, which replaces $\xi_i$ with independent Rademacher random variables $\varepsilon_i$ (Lemma A.17); next, we show conditional convergence of $\mathbb{Z}_n'$ conditional on $\mathbb{Z}_n$ (so randomness comes only from $\xi_i$) by discreting $\mathbb{Z}_n'$ (Lemma A.18).

We observe that many results in Van der Vaart and Wellner (1996, Chapters 2.9, 3.6, and A.1) translate directly to our norm $\mathbb{L}_\mu$; therefore, we will not reproduce the entire discourse but rather note when this is the case and prove results that require modification to accommodate our norm. Additionally, in order to show validity of bootstrap for $L$-statistics, we also need to show conditional weak convergence of the bootstrap sample selection process $\sqrt{n}(\hat{\mathbb{K}}_n - \mathbb{K}_n)$. We restrict attention to the case of sample selection based on sample quantiles and show validity by writing the sample selection process as a function of the sum of i.i.d. random variables (Lemma A.19).

We first prove the key lemma used in Poissonization, the counterpart of Van der Vaart and Wellner (1996, Lemma 3.6.16).

**Lemma A.16.** *For each $n$, let $(W_{n1}, \dots, W_{nn})$ be an exchangeable nonnegative random vector independent of $X_1, X_2, \dots$ such that $\sum_{i=1}^n W_{ni} = 1$ and $\max_{1\le i\le n} |W_{ni}|$ converges to zero in probability. Then, for every $\varepsilon > 0$, as $n \to \infty$,*

$$\Pr_W\left(\left\|\sum_{i=1}^n W_{ni}\big(\mathbb{1}\{X_i \le x\} - F(x)\big)\right\|_\mu^* > \varepsilon\right) \xrightarrow{\text{as}*} 0.$$

*Proof.* Assume without loss of generality that $\mu$ is a positive measure, that is, $m$ is increasing (we may do so since $m$ is of locally bounded variation). Since Van der Vaart and Wellner (1996, Lemma 3.6.7) can be restated with our norm $\|\cdot\|_{\mathbb{L}_\mu}$, the proof of this lemma is almost identical to that of Van der Vaart and Wellner (1996, Lemma 3.6.16) modulo the norm. Essentially, the only part that requires modification is the boundedness of $n^{-1}\sum_{i=1}^n \|\mathbb{1}\{X_i \le x\} - F(x)\|_\mu^r$. Note that

$$|\mathbb{1}\{X_i \le x\} - F(x)| \le |\mathbb{1}\{X_i \le x\} - \mathbb{1}\{0 \le x\}| + |\tilde{F}(x)|.$$

Therefore,

$$\|\mathbb{1}\{X_i \le x\} - F(x)\|_\mu \le |m(X_i) - m(0)| + \|\tilde{F}\|_\mu.$$

Find that

$$\frac{1}{n}\sum_{i=1}^n \|\mathbb{1}\{X_i \le x\} - F(x)\|_\mu^r \le \frac{1}{n}\sum_{i=1}^n |m(X_i) - m(0)|^r + \|\tilde{F}\|_\mu^r,$$

which converges almost surely to $\mathbb{E}[|m(X_i) - m(0)|^r] + \|\tilde{F}\|_\mu^r$, which is finite. ∎

Given this lemma, we may infer by the same arguments as Van der Vaart and Wellner (1996, Theorem 3.6.1) that conditional weak convergence of the bootstrap process $\hat{\mathbb{Z}}_n$ follows from conditional weak convergence of the multiplier process $\mathbb{Z}_n'$. Before moving on to the conditional convergence, however, we need to show the unconditional convergence of the multiplier process $\mathbb{Z}_n'$ in our norm. For a random variable $\xi$, we use the notation

$$\|\xi\|_{2,1} := \int_0^\infty \sqrt{\Pr(|\xi| > x)}dx.$$

That $\|\xi\|_{2,1} < \infty$ means that $\xi$ has slightly more than a variance. The following is a modification of Van der Vaart and Wellner (1996, Theorem 2.9.2).

**Lemma A.17.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a function of locally bounded variation and $\mu$ the Lebesgue-Stieltjes measure associated with $m$. Let $\xi_1, \dots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \dots, X_n$. For a probability distribution $F$ on $\mathbb{R}$ such that $m(X)$ has a $(2+c)$th moment for $X \sim F$ and some $c > 0$, the process $\mathbb{Z}_n'(x) := n^{-1/2}\sum_{i=1}^n \xi_i[\mathbb{1}\{X_i \le x\} - F(x)]$ converges weakly to a tight limit process in $\mathbb{L}_\mu$ if and only if $\mathbb{Z}_n := n^{-1/2}\sum_{i=1}^n[\mathbb{1}\{X_i \le x\} - F(x)]$ does. In that case, they share the same limit processes.*

*Proof.* By Proposition A.2 and Van der Vaart and Wellner (1996, Theorem 2.9.2), marginal convergence and asymptotic equicontinuity of $\mathbb{Z}'_n$ are trivial. It remains to show the equivalence of asymptotic equiintegrability of $\mathbb{Z}'_n$ and $\mathbb{Z}_n$.

Note that the proofs of Van der Vaart and Wellner (1996, Lemmas 2.3.1, 2.3.6, and 2.9.1 and Propositions A.1.4 and A.1.5) do not depend on the specificity of the norm $\|\cdot\|_{\mathcal{F}}$; therefore, they continue to hold with $\|\cdot\|_{\mathbb{L}_\mu}$. Given these, Van der Vaart and Wellner (1996, Lemma 2.3.11) also holds with $\|\cdot\|_{\mathbb{L}_\mu}$ (and $\|\cdot\|_{\mathbb{L}_{\mu,\delta_n}}$). Finally, rewriting the proof of Van der Vaart and Wellner (1996, Theorem 2.9.2) in terms of $\|\cdot\|_{\mathbb{L}_\mu}$ yields the proof of this lemma. ∎

Third, we show conditional convergence of the multiplier process $\mathbb{Z}'_n$ using the above result. Van der Vaart and Wellner (1996, Theorem 2.9.6).

**Lemma A.18.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a function of locally bounded variation and $\mu$ the Lebesgue-Stieltjes measure associated with $m$. Let $\xi_1, \dots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \dots, X_n$. For a probability distribution $F$ on $\mathbb{R}$ such that $m(X)$ has a $(2+c)$th moment for $X \sim F$ and some $c > 0$, the process $\mathbb{Z}'_n(x) = n^{-1/2} \sum_{i=1}^n \xi_i [\mathbb{1}\{X_i \leq x\} - F(x)]$ satisfies*

$$\sup_{h \in \mathrm{BL}_1(\mathbb{L}_\mu)} \left| \mathbb{E}_\xi h(\mathbb{Z}'_n) - \mathbb{E} h(\mathbb{Z}) \right| \longrightarrow 0$$

*in outer probability, and the sequence $\mathbb{Z}'_n$ is asymptotically measurable.*

*Proof.* By Lemma A.17, $\mathbb{Z}'_n$ is asymptotically measurable. Assume without loss of generality that $m$ is continuous and strictly monotone (see footnote 22), and define a semimetric $\rho$ on $\mathbb{R}$ by

$$\rho(s,t) = \left( \int_s^t \left( |m(x)|^{1+c} \vee 1 \right) |\tilde{F}(x)| d\mu(x) \right)^{1/2}.$$

For $\delta > 0$, $t_1 < \cdots < t_p$ be such that $\rho(-\infty, t_1) \leq \delta$, $\rho(t_j, t_{j+1}) \leq \delta$, and $\rho(t_p, \infty) \leq \delta$. Define $\mathbb{Z}_\delta$ by

$$\mathbb{Z}_\delta(x) := \begin{cases} 0 & x < t_1 \text{ or } x \geq t_p, \\ \mathbb{Z}(t_i) & t_i \leq x \leq t_{i+1},\ i = 1, \dots, p-1. \end{cases}$$

Define $\mathbb{Z}'_{n,\delta}$ analogously. By the continuity and integrability of the limit process $\mathbb{Z}$, we have $\mathbb{Z}_\delta \to \mathbb{Z}$ in $\mathbb{L}_\mu$ almost surely as $\delta \to 0$. Therefore,

$$\sup_{h \in \mathrm{BL}_1(\mathbb{L}_\mu)} \left| \mathbb{E} h(\mathbb{Z}_\delta) - \mathbb{E} h(\mathbb{Z}) \right| \longrightarrow 0 \qquad \text{as} \qquad \delta \to 0.$$

Second, by Van der Vaart and Wellner (1996, Lemma 2.9.5),

$$\sup_{h \in \mathrm{BL}_1(\mathbb{L}_\mu)} \left| \mathbb{E}_\xi h(\mathbb{Z}'_{n,\delta}) - \mathbb{E} h(\mathbb{Z}_\delta) \right| \longrightarrow 0 \qquad \text{as} \qquad n \to \infty$$

for almost every sequence $X_1, X_2, \dots$ and fixed $\delta > 0$. Since $\mathbb{Z}_\delta$ and $\mathbb{Z}'_{n,\delta}$ take only on a finite number of values and their tail values are zero, one can replace the supremum over $\mathrm{BL}_1(\mathbb{L}_\mu)$ with a supremum over $\mathrm{BL}_1(\mathbb{R}^p)$. Observe that $\mathrm{BL}_1(\mathbb{R}^p)$ is separable with respect to the topology of uniform convergence on compact sets; this supremum is effectively over a countable set, hence measurable. Third,

$$\begin{aligned}\sup_{h\in \mathrm{BL}_1(\mathbb{L}_\mu)} \left|\mathbb{E}_\xi h(\mathbb{Z}'_{n,\delta}) - \mathbb{E}_\xi h(\mathbb{Z}'_n)\right| &\le \sup_{h\in \mathrm{BL}_1(\mathbb{L}_\mu)} \mathbb{E}_\xi \left|h(\mathbb{Z}'_{n,\delta}) - h(\mathbb{Z}'_n)\right| \\ &\le \mathbb{E}_\xi \|\mathbb{Z}'_{n,\delta} - \mathbb{Z}'_n\|^*_{\mathbb{L}_\mu} \le \mathbb{E}_\xi \|\mathbb{Z}'_n\|_{\mathbb{L}^*_{\mu,\delta}}\end{aligned}$$

This implies that the outer expectation of the LHS is bounded above by $\mathbb{E}^*\|\mathbb{Z}'_n\|_{\mathbb{L}_{\mu,\delta}}$, which vanishes as $n \to \infty$ by the modified Van der Vaart and Wellner (1996, Lemma 2.9.1) as in Lemma A.17. ■

These results show that nonparametric bootstrap works for the empirical process $\sqrt{n}(\mathbb{F}_n - F)$ and the empirical quantile process $\sqrt{n}(\mathbb{Q}_n - Q)$. However, we also need to verify validity for the sample selection process $\sqrt{n}(\mathbb{K}_n - K)$. The key in the proof is to represent $\mathbb{K}_n$ in terms of Hadamard differentiable functions of "$\mathbb{F}_n$" in Proposition A.13.

**Lemma A.19.** *Let $U_1, \dots, U_n$ be independent uniformly distributed random variables on $(0, 1)$ and $\xi_1, \dots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $U_1, \dots, U_n$. Define the bootstrapped empirical process of $U$ by*

$$\mathbb{F}'_n(u) := \frac{1}{n}\sum_{i=1}^n \xi_i \mathbb{1}\{U_i \le u\},$$

*and let $w'_{i,n}$ be the indicator of whether $U_i$ is above the $\alpha$-quantile of the bootstrapped sample, that is, $w'_{i,n} = \mathbb{1}\{U_i > \mathbb{F}'^{-1}_n(\alpha)\}$. Define*

$$\mathbb{G}'_n(u) := \frac{1}{n}\sum_{i=1}^n \xi_i w'_{i,n} \mathbb{1}\{U_i \le u\}.$$

*Then, for $F(u) = 0 \vee u \wedge 1$ and $G(u) = 0 \vee (u - \alpha) \wedge (1 - \alpha)$,*

$$\begin{aligned}\sup_{h\in BL_1(\mathbb{L}_{\nu_2})} \left|\mathbb{E}_\xi h(\sqrt{n}(\mathbb{F}'_n - F)) - \mathbb{E} h(\sqrt{n}(\mathbb{F}_n - F))\right| &\longrightarrow 0, \\ \sup_{h\in BL_1(\mathbb{L}_{\nu_2})} \left|\mathbb{E}_\xi h(\sqrt{n}(\mathbb{G}'_n - G)) - \mathbb{E} h(\sqrt{n}(\mathbb{G}_n - G))\right| &\longrightarrow 0,\end{aligned}$$

*in outer probability, and the sequences $\sqrt{n}(\mathbb{F}'_n - F)$ and $\sqrt{n}(\mathbb{G}'_n - G)$ are asymptotically measurable.*

*Proof.* Noting that $\mathbb{G}'_n(u) = 0 \vee [\mathbb{F}'_n(u) - \mathbb{F}'_n \circ \mathbb{F}'^{-1}_n(\alpha)]$, weak convergence of $\sqrt{n}(\mathbb{F}'_n - F)$ and $\sqrt{n}(\mathbb{G}'_n - G)$ follows from Lemmas A.14 and A.18. ■

*Remark.* Note that Lemma A.19 immediately implies the validity of bootstrap for any selection of samples based on a finite number of empirical quantiles.

Now we show the validity of nonparametric bootstrap when sample selection is based on empirical quantiles.

**Proposition A.20** (Validity of nonparametric bootstrap)**.** *In the assumptions stated in Proposition A.15, assume further that $w_{i,n,j}$ represents sample selection based on a fixed number of empirical quantiles.*[24] *Then, the joint distribution of $(\hat\beta_1, \dots, \hat\beta_d)$ can be consistently estimated by nonparametric bootstrap. The algorithm is as follows. Here, $X_i$ denotes a vector $(X_{i,1}, \dots, X_{i,d})$.*

[24] The assumption on convergence must be extended (from bivariate) to joint over all processes involved.

*i. Bootstrap $n$ (or fewer) random observations from $X_1, \ldots, X_n$ with replacement.*

*ii. Compute the statistics $(\hat{\beta}_1^*, \ldots, \hat{\beta}_d^*)$ for the bootstrapped sample.*

*iii. Repeat the above steps $S$ times.*

*iv. Use the empirical distribution of $(\hat{\beta}_1^*, \ldots, \hat{\beta}_d^*)$ as the approximation to the theoretical asymptotic distribution of $(\hat{\beta}_1, \ldots, \hat{\beta}_d)$.*

*Proof.* With the remark below Lemma A.16, the proposition follows from Lemmas A.18 and A.19 and Van der Vaart and Wellner (1996, Theorem 3.9.11). ■

## A.8 Glivenko-Cantelli Type Results

We state some of the Glivenko-Cantelli type results for distribution functions and quantile functions when we have additional information about their expectations. The results are generally stronger than the classical Glivenko-Cantelli theorems. It is noteworthy that the Glivenko-Cantelli theorem of quantile functions holds for the $L_p$ spaces (see Appendix A.5 for discussion). Related results are found in Parzen (1980) and Csörgő and Horváth (1993).

The first addendum provides a stronger Glivenko-Cantelli result for the distribution function when the underlying distribution has a finite expectation when transformed by a function $m$.

**Addendum A.21.** *Let $m : \mathbb{R} \to \mathbb{R}$ be a function of locally bounded variation and $\mu$ the associated Lebesgue-Stieltjes measure. For a probability measure $F$ on $\mathbb{R}$ such that $m(X)$ has a finite expectation for $X \sim F$, we have*

$$\big\|m(t)\mathbb{F}_n(t) - m(t)F(t)\big\|_\infty \xrightarrow{\text{as}*} 0, \qquad \bigg\|\int_{[s,t]} |m| d\mathbb{F}_n - \int_{[s,t]} |m| dF\bigg\|_\infty \xrightarrow{\text{as}*} 0,$$

$$\bigg\|\int_{[s,t]} |\tilde{\mathbb{F}}_n||d\mu| - \int_{[s,t]} |\tilde{F}||d\mu|\bigg\|_\infty \xrightarrow{\text{as}} 0, \qquad \int_{\mathbb{R}} |\mathbb{F}_n - F||d\mu| \xrightarrow{\text{as}} 0,$$

*where the supremum is taken respectively over $t \in \mathbb{R}$, $(s,t) \in \overline{\mathbb{R}}^2$, and $(s,t) \in \overline{\mathbb{R}}^2$.*

*Proof.* Note first that since a function of bounded variation can be represented as a difference of two increasing functions, we may assume without loss of generality that $m$ is increasing (and hence $\mu$ is a positive measure). Moreover, we may also assume $m(0) = 0$ for otherwise the residual terms $m(0)(\mathbb{F}_n - F)$ of the first two quantities vanish by the classical Glivenko-Cantelli theorem.

In view of Van der Vaart and Wellner (1996, Theorem 2.4.1), to prove the first two claims it suffices to show that the classes of functions,

$$\mathcal{F} = \Big\{f_t : \mathbb{R} \to \mathbb{R} : t \in \overline{\mathbb{R}},\ f_t(x) = m(t)\mathbb{1}\{x \le t\}\Big\},$$
$$\mathcal{G} = \Big\{g_{s,t} : \mathbb{R} \to \mathbb{R} : s, t \in \overline{\mathbb{R}},\ g_{s,t}(x) = |m(x)|\mathbb{1}\{s \le x \le t\}\Big\},$$

have finite bracketing numbers with respect to $L_1(P)$, i.e., $N_{[]}(\varepsilon, \mathcal{F}, L_1(F)) < \infty$ and $N_{[]}(\varepsilon, \mathcal{G}, L_1(F)) < \infty$ for every $\varepsilon > 0$. For $\mathcal{F}$ take $-\infty = t_0 < t_1 < \cdots < t_m = \infty$

such that $\left|\int (f_{t_{i+1}} - f_{t_i}) dF\right| < \varepsilon$ for each $i$ and consider the brackets $\{f_{t_i}\}$.[25] Such a partition is finite since $m(X)$ has a finite first moment and by Lemma A.1. For $\mathcal{G}$ take $-\infty = t_0 < t_1 < \cdots < t_m = \infty$ such that $\left|\int_{(-\infty,t_{i+1}]} |m| dF - \int_{(-\infty,t_i]} |m| dF\right| < \varepsilon$ for each $i$, then consider the brackets $\{g_{s,t}\}$ for every pair $s,t \in \{t_0,\dots,t_m\}$.[26] Such a partition is finite by the assumption that $m(X)$ has a finite first moment.

To prove the third claim, observe that by integration by parts,

$$\int_{[s,t]} |m| d\mathbb{F}_n = \int_{[s,t]} |m| d\tilde{\mathbb{F}}_n = \Big[|m|\tilde{\mathbb{F}}_n\Big]_s^t + \int_{[s,t]} |\tilde{\mathbb{F}}_n| d\mu.$$

Then the claim follows in observation of the following triangle inequality and the two claims proved so far,[27]

$$\begin{aligned}\left\|\int_{[s,t]} |\tilde{\mathbb{F}}_n| d\mu - \int_{[s,t]} |\tilde{F}| d\mu\right\|_\infty &\le 2\big\|m(t)\mathbb{F}_n(t) - m(t)F(t)\big\|_\infty \\ &\quad + \left\|\int_{[s,t]} |m| d\mathbb{F}_n - \int_{[s,t]} |m| dF\right\|_\infty.\end{aligned}$$

To prove the last claim, observe that Lemma A.1 and the preceding claim imply that for $\varepsilon > 0$ there exists $M < \infty$ such that

$$\left(\int_{(-\infty,-M]} + \int_{[M,\infty)}\right) |\tilde{\mathbb{F}}_n| d\mu + \left(\int_{(-\infty,-M]} + \int_{[M,\infty)}\right) |\tilde{F}| d\mu < \varepsilon$$

with probability tending to 1. By the triangle inequality,

$$\int_{\mathbb{R}} \big|\tilde{\mathbb{F}}_n - \tilde{F}\big| d\mu \le \int_{(-M,M)} \big|\tilde{\mathbb{F}}_n - \tilde{F}\big| d\mu + \varepsilon \le \|\mathbb{F}_n - F\|_\infty \mu((-M,M)) + \varepsilon.$$

Then the assertion follows by the Glivenko-Cantelli theorem. ■

We next provide the Glivenko-Cantelli results for the quantile functions. Two points are interesting. First, despite the fact that we were unable to prove Hadamard differentiability of $F \mapsto Q$ as a map from $\mathbb{L}_p$ to $L_p$, the Glivenko-Cantelli result still holds for the $L_p$ norm. Second, the addendum gives the "sup norm" for quantile functions. Although the quantile function for an unbounded random variable is unbounded, we still have a reasonable pseudo-uniform convergence when $Q$ is continuous.

**Addendum A.22.** *Let $F$ be a probability distribution on $\mathbb{R}$ with a pth moment for $p > 0$ and $Q := F^{-1}$. Then*

$$\left(\int_0^1 |\mathbb{Q}_n - Q|^p du\right)^{1/p} \xrightarrow{\text{as}} 0.$$

*Moreover, if and only if $Q$ is continuous, we have*

$$\big\|u^{1/p}(1-u)^{1/p}(\mathbb{Q}_n - Q)\big\|_\infty \xrightarrow{\text{as*}} 0,$$

*where the supremum is taken over $u \in (0,1)$.*

[25] If $F$ has a probability mass at $t$, then for small $\varepsilon$ take, instead of $f_t$, $\tilde{f}_{t,c}(x) = m(t)[c\mathbb{1}\{x \le t\} + (1-c)\mathbb{1}\{x < t\}]$ for appropriately chosen $c$.

[26] Again, if $F$ has a mass, similar adjustments are required.

[27] Measurability of the sup on the LHS follows by continuity of the Lebesgue integrals.

*Proof.* By the strong law of large numbers on $Y = |X|^p$, $X \sim F$, we have

$$\int_{-\infty}^{\infty} |x|^p d\mathbb{F}_n - \int_{-\infty}^{\infty} |x|^p dF \xrightarrow{\text{as}} 0.$$

Applying the change of variables,

$$\int_0^1 |\mathbb{Q}_n|^p du - \int_0^1 |Q|^p du \xrightarrow{\text{as}} 0.$$

In view of this (and since $Q$ is in $L_p$ by Lemma A.1), for $\varepsilon > 0$ one can take $\delta > 0$ such that

$$\Big(\int_0^{2\delta} + \int_{1-2\delta}^1\Big)|\mathbb{Q}_n|^p du < \varepsilon, \qquad \Big(\int_0^{2\delta} + \int_{1-2\delta}^1\Big)|Q|^p du < \varepsilon,$$

with probability tending to 1. Combination with the triangle inequality allows us to bound the target as

$$\int_0^1 |\mathbb{Q}_n - Q|^p du \le \int_{2\delta}^{1-2\delta} |\mathbb{Q}_n - Q|^p du + 2\varepsilon.$$

Moreover, observe

$$\int_{2\delta}^{1-2\delta} |\mathbb{Q}_n - Q|^p du \le \varepsilon^p + \int_{2\delta}^{1-2\delta} |\mathbb{Q}_n - Q|^p \mathbb{1}\big\{|\mathbb{Q}_n - Q| > \varepsilon\big\} du.$$

We want to bound the last integrand by a multiple of $|\mathbb{Q}_n - Q|$ to eliminate the $p$th power, since then we may further apply Fubini's theorem to eliminate the inverse. Toward this goal, we aim to use the following inequality: for $p > 0$ and $M \ge 0$,

$$|x|^p \mathbb{1}\{|x| > \varepsilon\} \le \Big(\varepsilon^{p-1} \vee M^{p-1}\Big)|x| \quad \text{for every } |x| \le M.$$

But to do that, we need to find the bound $M$ on $|\mathbb{Q}_n - Q|$ over $(2\delta, 1 - 2\delta)$ that does not depend on $n$.

By the classical Glivenko-Cantelli theorem, we may take $n$ large enough so that $\|\mathbb{F}_n - F\|_\infty < \delta$ with outer probability at least $1 - \varepsilon$. This implies $F - \delta \le \mathbb{F}_n \le F + \delta$, and since $F$ and $\mathbb{F}_n$ are nondecreasing,

$$(F - \delta)^{-1} \le \mathbb{F}_n^{-1} = \mathbb{Q}_n \le (F + \delta)^{-1}.$$

This yields the following bounds on $\mathbb{Q}_n$ over the region of integration.

$$\mathbb{Q}_n(1 - 2\delta) \le (F + \delta)^{-1}(1 - 2\delta) = Q(1 - \delta),$$
$$\mathbb{Q}_n(2\delta) \ge (F - \delta)^{-1}(2\delta) = Q(\delta).$$

Note that nondecreasingness of $F$ implies analogous inequalities for $Q$ itself, namely, $Q(1 - 2\delta) \le Q(1 - \delta)$ and $Q(2\delta) \ge Q(\delta)$. Therefore, the difference $|\mathbb{Q}_n - Q|$ is bounded by $M := Q(1 - \delta) - Q(\delta)$ over the region of integration.

Given this, we can successfully bound the last integral by

$$\Big(\varepsilon^{p-1} \vee M^{p-1}\Big) \int_{2\delta}^{1-2\delta} |\mathbb{Q}_n - Q| du.$$

Finally, invoke Fubini's theorem to find

$$\int_{2\delta}^{1-2\delta} |\mathbb{Q}_n - Q| du = \int_{2\delta}^{1-2\delta} \int_{Q\wedge\mathbb{Q}_n}^{Q\vee\mathbb{Q}_n} dx du \leq \int_{Q(\delta)}^{Q(1-\delta)} |\mathbb{F}_n - F| dx \leq M\|\mathbb{F}_n - F\|_\infty,$$

which vanishes outer almost surely.

Next, we prove the second claim. Since $Q \in L_p$, $\mathbb{Q}_n \in L_p$, and $\mathbb{Q}_n$ converges to $Q$ in $L_p$, for every $\varepsilon > 0$ there exists $\delta$ such that

$$\left(\int_0^\delta + \int_{1-\delta}^1\right)\left(|\mathbb{Q}_n|^p + |Q|^p\right) du < \varepsilon$$

with probability tending to 1. When this inequality holds, we argue that $|\mathbb{Q}_n|^p$ and $|Q|^p$ never exceed the function $\varepsilon/[u(1-u)]$ on $u \in (0,\delta) \cup (1-\delta, 1)$. Suppose otherwise (without loss of generality consider $Q$ only) and let $u$ be the point of exceedance. Since $Q$ is monotone, the area of $|Q|^p$ must contain either the left rectangle $(0,u)\times(0,\varepsilon/[u(1-u)])$ or the right rectangle $(u,1)\times(0,\varepsilon/[u(1-u)])$. In either case, the integral of $|Q|^p$ over $(0,\delta)\cup(1-\delta,1)$ must exceed $\varepsilon$, since each rectangle has area bigger than $\varepsilon$. This is a contradiction to the assumption that the integral is less than $\varepsilon$.

This ensures that the difference $|\mathbb{Q}_n - Q|^p$ is less than $2^p\varepsilon/[u(1-u)]$ on this region. In other words, the supremum of $u(1-u)|\mathbb{Q}_n - Q|^p$ over $(0,\delta)\cup(1-\delta,1)$ is less than $2^p\varepsilon$. Meanwhile, if $Q$ is continuous, then $\mathbb{Q}_n$ converges to $Q$ pointwise on any fixed closed interval of $(0,1)$ since $\mathbb{Q}_n$ and $Q$ are monotone. Therefore, the supremum of $u(1-u)|\mathbb{Q}_n - Q|^p$ over $[\delta, 1-\delta]$ must vanish as $n \to \infty$. Combining the results, sufficiency follows.

On the other hand, suppose $Q$ is discontinuous at $u \in (0,1)$. By construction $\mathbb{Q}_n$ can only have discontinuity points on $\{1/n, \ldots, n/n\}$. But there exist infinitely many $n$ such that $u \notin \{1/n, \ldots, n/n\}$. Therefore, $|\mathbb{Q}_n - Q|$ can infinitely often be as large as half the jump height at $u$. ■

## References


Aaberge, R. (2009): "Ranking Intersecting Lorenz Curves," *Social Choice and Welfare*, 33, 235–259.

Acemoglu, D., S. Johnson, A. Kermani, J. Kwak, and T. Mitton (2016): "The Value of Connections in Turbulent Times: Evidence from the United States," *Journal of Financial Economics*, 121, 368–391.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

——— (2012): "The Colonial Origins of Comparative Development: An Empirical Investigation: Reply," *American Economic Review*, 102, 3077–3110.

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2017): "Democracy Does Cause Growth," *Journal of Political Economy*, forthcoming.

Agarwal, N., C. Banternghansa, and L. T. Bui (2010): "Toxic Exposure in America: Estimating Fetal and Infant Health Outcomes from 14 Years of TRI Reporting," *Journal of Health Economics*, 29, 557–574.

Aggarwal, C. C. (2013): *Outlier Analysis*, New York: Springer.

Alatas, V., A. Banerjee, R. Hanna, B. A. Olken, R. Purnamasari, and M. Wai-Poi (2016): "Self-Targeting: Evidence from a Field Experiment in Indonesia," *Journal of Political Economy*, 124, 371–427.

Albouy, D. Y. (2012): "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment," *American Economic Review*, 102, 3059–3076.

Anderson, M. L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563.

Arora, S. and K. Jain (2006): "Testing for Generalized Lorenz Dominance," *Statistical Methods and Applications*, 15, 75–88.

Auerbach, A. J., K. A. Hassett, and S. D. Oliner (1994): "Reassessing the Social Returns to Equipment Investment," *Quarterly Journal of Economics*, 109, 789–802.

Augsburg, B., R. D. Haas, H. Harmgart, and C. Meghir (2015): "The Impacts of Microcredit: Evidence from Bosnia and Herzegovina," *American Economic Journal: Applied Economics*, 7, 183–203.

Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan (2015a): "The Miracle of Microfinance? Evidence from a Randomized Evaluation," *American Economic Journal: Applied Economics*, 7, 22–53.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Parienté, J. Shapiro, B. Thuysbaert, and C. Udry (2015b): "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries," *Science*, 348, 772–788.

Banerjee, A., E. Duflo, and R. Hornbeck (2014): "(Measured) Profit is Not Welfare: Evidence from an Experiment on Bundling Microcredit and Insurance," Working paper.

Barrett, G. F., S. G. Donald, and D. Bhattacharya (2014): "Consistent Nonparametric Tests for Lorenz Dominance," *Journal of Business & Economic Statistics*, 32, 1–13.

Belloni, A. and V. Chernozhukov (2011): "$\ell_1$-Penalized Quantile Regression in High-Dimensional Sparse Models," *Annals of Statistics*, 39, 82–130.

Benjamini, Y. and Y. Hochberg (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society, Series B*, 57, 289–300.

Benjamini, Y. and D. Yekutieli (2001): "The Control of the False Discovery Rate in Multiple Testing under Dependency," *Annals of Statistics*, 29, 1165–1188.

Berg, A. and D. Politis (2009): "CDF and Survival Function Estimation with Infinite-Order Kernels," *Electronic Journal of Statistics*, 3, 1436–1454.

Beutner, E. and H. Zähle (2010): "A Modified Functional Delta Method and its Application to the Estimation of Risk Functionals," *Journal of Multivariate Analysis*, 101, 2452–2463.

Bhattacharya, D. (2005): "Asymptotic Inference from Multi-Stage Samples," *Journal of Econometrics*, 126, 145–171.

Bishop, J. A., S. Chakraborti, and P. D. Thistle (1988): "Large Sample Tests for Absolute Lorenz Dominance," *Economics Letters*, 26, 291–294.

Bishop, J. A., J. P. Formby, and W. J. Smith (1991): "Lorenz Dominance and Welfare: Changes in the U.S. Distribution of Income, 1967–1986," *Review of Economics and Statistics*, 73, 134–139.

——— (1993): "International Comparisons of Welfare and Poverty: Dominance Orderings for Ten Countries," *Canadian Journal of Economics / Revue canadienne d'Economique*, 26, 707–726.

Casey, K., R. Glennerster, and E. Miguel (2012): "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan," *Quarterly Journal of Economics*, 127, 1755–1812.

Chatterjee, S. and A. S. Hadi (1988): *Sensitivity Analysis in Linear Regression*, New York: John Wiley & Sons, Inc.

Chernozhukov, V. (2005): "Extremal Quantile Regression," *Annals of Statistics*, 33, 806–839.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010): "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on Counterfactual Distributions," *Econometrica*, 81, 2205–2268.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chhaochharia, V. and Y. Grinstein (2009): "CEO Compensation and Board Structure," *Journal of Finance*, 64, 231–261.

Cox, C., M. F. Schneider, and A. Muñoz (2013): "Quantiles of Residual Survival," in *Risk Assessment and Evaluation of Predictions*, ed. by M.-L. T. Lee, M. Gail, R. Pfeiffer, G. Satten, T. Cai, and A. Gandy, New York: Springer, chap. 6, 87–103.

Csörgő, M. (1983): *Quantile Processes with Statistical Applications*, Philadelphia: Society for Industrial and Applied Mathematics.

Csörgő, M., S. Csörgő, L. Horváth, and D. M. Mason (1986): "Weighted Empirical and Quantile Processes," *Annals of Probability*, 14, 31–85.

Csörgő, M. and L. Horváth (1988): "On the Distributions of $L_p$ Norms of Weighted Uniform Empirical and Quantile Processes," *Annals of Probability*, 16, 142–161.

——— (1990): "On the Distributions of $L_p$ Norms of Weighted Quantile Processes," *Annales de l'Institut Henri Poincaré, Section B, Calcul des Probabilités et statistiques*, 26, 65–85.

——— (1993): *Weighted Approximations in Probability and Statistics*, Chichester: John Wiley & Sons Ltd.

Csörgő, M., L. Horváth, and Q.-M. Shao (1993): "Convergence of Integrals of Uniform Empirical and Quantile Processes," *Stochastic Processes and their Applications*, 45, 283–294.

Dasgupta, P., A. Sen, and D. Starrett (1973): "Notes on the Measurement of Inequality," *Journal of Economic Theory*, 6, 180–187.

de Haan, L. and A. Ferreira (2006): *Extreme Value Theory: An Introduction*, New York: Springer.

de Long, J. B. and L. H. Summers (1991): "Equipment Investment and Economic Growth," *Quarterly Journal of Economics*, 106, 445–502.

——— (1994): "Equipment Investment and Economic Growth: Reply," *Quarterly Journal of Economics*, 109, 803–807.

de Long, J. B., L. H. Summers, and A. B. Abel (1992): "Equipment Investment and Economic Growth: How Strong is the Nexus?" *Brookings Papers on Economic Activity*, 1992, 157–211.

Dehling, H., O. Durieu, and M. Tusche (2014): “Approximating Class Approach for Empirical Processes of Dependent Sequences Indexed by Functions,” *Bernoulli*, 20, 1372–1403.

Dehling, H., T. Mikosch, and M. Sørensen, eds. (2002): *Empirical Process Techniques for Dependent Data*, New York: Springer.

del Barrio, E., E. Giné, and C. Matrán (1999): “Central Limit Theorems for the Wasserstein Distance between the Empirical and the True Distributions,” *Annals of Probability*, 27, 1009–1071.

Demont, T. (2013): “Poverty, Access to Credit and Absorption of Weather Shocks: Evidence from Indian Self-Help Groups,” Working Paper.

Drees, H. and H. Rootzén (2010): “Limit Theorems for Empirical Processes of Cluster Functionals,” *Annals of Statistics*, 38, 2145–2186.

——— (2016): “Correction Note: Limit Theorems for Empirical Processes of Cluster Functionals,” *Annals of Statistics*, 44, 1360–1361.

Dudley, R. M. (1997): “Empirical Processes and $p$-Variation,” in *Festschrift for Lucien LeCam*, ed. by D. Pollard, E. Torgersen, and G. L. Yang, New York: Springer, chap. 13, 219–233.

——— (2014): *Uniform Central Limit Theorems*, Cambridge: Cambridge University Press, second ed.

Einmahl, J. H. (1992): “Limit Theorems for Tail Processes with Application to Intermediate Quantile Estimation,” *Journal of Statistical Planning and Inference*, 32, 137–145.

Embrechts, P. and M. Hofert (2013): “A Note on Generalized Inverses,” *Mathematical Methods of Operations Research*, 77, 423–432.

Fabrizio, K. R., N. L. Rose, and C. D. Wolfram (2007): “Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency,” *American Economic Review*, 97, 1250–1277.

Falkner, N. and G. Teschl (2012): “On the Substitution Rule for Lebesgue-Stieltjes Integrals,” *Expositiones Mathematicae*, 30, 412–418.

Fellman, J. (2002): “The Redistributive Effect of Tax Policies,” *Sankhyā: The Indian Journal of Statistics*, 64, 90–100.

Feng, C., H. Wang, X. M. Tu, and J. Kowalski (2012): “A Note on Generalized Inverses of Distribution Function and Quantile Transformation,” *Applied Mathematics*, 3, 2098–2100.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): “Unconditional Quantile Regressions,” *Econometrica*, 77, 953–973.

Gini, C. (1912): *Variabilità e Mutabilità, contributo allo studio delle distribuzioni e delle relazioni statistiche*, Bologna: Tipografia di Paolo Cuppin.

Gujarati, D. N. and D. C. Porter (2009): *Basic Econometrics*, New York: McGraw-Hill, Inc., fifth ed.

Guthrie, K., J. Sokolowsky, and K.-M. Wan (2012): "CEO Compensation and Board Structure Revisited," *Journal of Finance*, 67, 1149–1168.

Hall, P., R. C. L. Wolff, and Q. Yao (1999): "Methods for Estimating a Conditional Distribution Function," *Journal of the American Statistical Association*, 94, 154–163.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986): *Robust Statistics: The Approach Based on Influence Functions*, New York: John Wiley & Sons, Inc.

Hawkins, D. M. (1980): *Identification of Outliers*, Dordrecht: Springer.

Heckman, J. J., J. Smith, and N. Clements (1997): "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535.

Herndon, T., M. Ash, and R. Pollin (2014): "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff," *Cambridge Journal of Economics*, 38, 257–279.

Huber, P. J. and E. M. Ronchetti (2009): *Robust Statistics*, New Jersey: John Wiley & Sons, Inc., second ed.

Jayachandran, S. (2006): "Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries," *Journal of Political Economy*, 114, 538–575.

Kaji, T. (2017): "Assessing Outcome-Dependent Heterogeneity in Treatment Effects," Working paper.

Kaji, T. and H. Kang (2017): "Simple Rule for Controlling Tail Risk Measures with Estimation Error," Working paper.

Koenker, R. (2005): *Quantile Regression*, New York: Cambridge University Press.

Koenker, R. and G. Bassett, Jr (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Koenker, R. and Z. Xiao (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

Kosorok, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*, New York: Springer.

Koul, H. L. (2002): *Weighted Empirical Processes in Dynamic Nonlinear Models*, New York: Springer, second ed.

Krasker, W. S. (1980): "Estimation in Linear Regression Models with Disparate Data Points," *Econometrica*, 48, 1333–1346.

LAMBERT, P. J. (2001): *The Distribution and Redistribution of Income*, Manchester: Manchester University Press, third ed.

LAPLACE, P.-S. (1812): *Théorie Analytique des Probabilités*, Paris: Courcier.

LEDOUX, M. AND M. TALAGRAND (1991): *Probability in Banach Spaces: Isoperimetry and Processes*, Berlin Heidelberg: Springer-Verlag.

LEHMANN, E. L. AND J. P. ROMANO (2005): *Testing Statistical Hypotheses*, New York: Springer, third ed.

LORENZ, M. O. (1905): "Methods of Measuring the Concentration of Wealth," *Publications of the American Statistical Association*, 9, 209–219.

MARONNA, R. A., R. D. MARTIN, AND V. J. YOHAI (2006): *Robust Statistics: Theory and Methods*, Chichester: John Wiley & Sons Ltd.

MORELLI, S., T. SMEEDING, AND J. THOMPSON (2015): "Post-1970 Trends in Within-Country Inequality and Poverty: Rich and Middle-Income Countries," in *Handbook of Income Distribution*, ed. by A. B. Atkinson and F. Bourguignon, Oxford: North-Holland, vol. 2A, chap. 8, 593–696.

PARZEN, E. (1980): "Quantile Functions, Convergence in Quantile, and Extreme Value Distribution Theory," *Technical Report*, B–3, Statistical Inst., Texas A & M Univ., College Station, TX.

REINHART, C. M. AND K. S. ROGOFF (2010): "Growth in a Time of Debt," *American Economic Review: Papers & Proceedings*, 100, 573–578.

ROMANO, J. P., A. M. SHAIKH, AND M. WOLF (2008): "Control of the False Discovery Rate under Dependence using the Bootstrap and Subsampling," *Test*, 17, 417–442.

——— (2010): "Hypothesis Testing in Econometrics," *Annual Review of Economics*, 2, 75–104.

ROOTZÉN, H. (2009): "Weak Convergence of the Tail Empirical Process for Dependent Sequences," *Stochastic Processes and their Applications*, 119, 468–490.

SHORACK, G. R. AND J. A. WELLNER (1986): *Empirical Processes with Applications to Statistics*, New York: John Wiley & Sons, Inc.

SHORROCKS, A. F. (1983): "Ranking Income Distributions," *Economica*, 50, 3–17.

TAROZZI, A., J. DESAI, AND K. JOHNSON (2015): "The Impacts of Microcredit: Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 7, 54–89.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge: Cambridge University Press.

VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer.

Yekutieli, D. (2007): "False Discovery Rate Control for Non-Positively Regression Dependent Test Statistics," *Journal of Statistical Planning and Inference*, 138, 405–415.

Young, A. (2017): "Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," Working paper.