Robustness of bootstrap in instrumental variable regression*

Lorenzo Camponovo†
School of Economics and Political Science
University of St.Gallen

Taisuke Otsu‡
Department of Economics
London School of Economics

December 18, 2013

Abstract

This paper studies robustness of bootstrap inference methods for instrumental variable (IV) regression models. We consider test statistics for parameter hypotheses based on the IV estimator and generalized method of trimmed moments (GMTM) estimator introduced by Čížek (2008, 2009), and compare the pairs and implied probability bootstrap approximations for these statistics by applying the finite sample breakdown point theory. In particular, we study limiting behaviors of the bootstrap quantiles when the values of outliers diverge to infinity but the sample size is held fixed. The outliers are defined as anomalous observations that can arbitrarily change the value of the statistic of interest. We analyze both just- and over-identified cases and discuss implications of the breakdown point analysis to the size and power properties of bootstrap tests. We conclude that the implied probability bootstrap test using the statistic based on the GMTM estimator shows desirable robustness properties. Simulation studies endorse this conclusion. An empirical example based on Romer's (1993) study on the effect of openness of countries to inflation rates is presented. Several extensions including the analysis for the residual bootstrap are provided.

1 Introduction

Instrumental variable (IV) regression is one of the most widely used methods in empirical economic analysis. There are numerous empirical examples and theoretical studies on IV regression. To investigate its theoretical properties, it is common to invoke the framework of the generalized method of moments (GMM), which provides a unified approach for statistical inference in econometric models specified by moment conditions. However, recent research indicates that there are considerable problems with the

*We would like to thank two anonymous referees and the seminar participants at the Info-Metrics conference 2010 at American University, Kyoto, Seoul National, and Yale for helpful comments. This research was done when the first author was visiting Yale University, financed by a fellowship of the Swiss National Science Foundation (SNSF).

†*Mailing address*: Bodanstrasse 9, 9000 St.Gallen, Switzerland. *E-mail address*: lorenzo.camponovo@unisg.ch.

‡*Mailing address*: Houghton Street, London, WC2A 2AE, UK. *E-mail address*: t.otsu@lse.ac.uk.

conventional IV regression technique particularly in its finite sample performance, and that approximations based on the asymptotic theory may yield poor results (see, e.g., special issues of the *Journal of Business & Economic Statistics*, volumes 14 and 20).

A common way to refine the approximations for the distributions of the IV regression estimators and related test statistics is to employ a bootstrap method. In the IV regression context, there are at least two approaches to conduct bootstrap approximation: the pairs bootstrap and implied probability bootstrap. The pairs bootstrap introduced by Freedman (1981) draws resamples from the original sample with equal weights and uses quantiles of the resampled statistics to approximate the distribution of the original statistic of interest. When the number of instruments exceeds the number of parameters (called over-identification), it is reasonable to impose the over-identified moment conditions to bootstrap resamples. Hall and Horowitz (1996) suggested to use the pairs bootstrap with recentered moment conditions and established a higher-order refinement result of the bootstrap inference. On the other hand, the implied probability bootstrap, proposed by Brown and Newey (2002), draws resamples with unequal weights defined by the so-called implied probabilities from the moment conditions, and uses quantiles of the resampled statistics based on the moment conditions without recentering (see also Hall and Presnell, 1999). Brown and Newey (2002) argued that the implied probability bootstrap also provides a higher-order refinement over the first-order asymptotic approximation.

Recently, Camponovo and Otsu (2012) introduced an alternative viewpoint to evaluate bootstrap methods based on the (finite sample) breakdown point theory. The breakdown point is a measure of the global reliability of a statistic that describes up to which fraction of outliers the statistic still provides reliable information (see, e.g., Hampel, 1971, and Donoho and Huber, 1983). Camponovo and Otsu (2012) extended the breakdown point theory for bootstrap quantiles (Singh, 1998) to the over-identified GMM setting and investigated robustness properties of the pairs and implied probability bootstrap methods.

The purpose of this paper is to refine the breakdown point analysis of Camponovo and Otsu (2012) by focusing on the IV regression models. In contrast to Camponovo and Otsu (2012), who focused on developing a basic framework for breakdown point analysis and considered somewhat artificial examples such as the trimmed mean with prior information, this paper focuses on the IV regression which is one of the most popular econometric models. We consider test statistics for parameter hypotheses based on the IV estimator and generalized method of trimmed moments (GMTM) estimator introduced by Čížek (2008, 2009), and compare the pairs and implied probability bootstrap approximations for these statistics by applying the finite sample breakdown point theory. In particular, we study limiting behaviors of the bootstrap quantiles when the values of outliers diverge to infinity but the sample size is held fixed. As in Singh (1998), Camponovo, Scaillet and Trojani (2012a), and Camponovo and Otsu (2012), we define the outliers as anomalous observations that can arbitrarily change the value of the statistic of interest.[1]

[1]We consider both representative and nonrepresentative outliers. Representative outliers are observations that have been correctly recorded and that cannot be assumed to be unique. Nonrepresentative outliers are instead observations

Although this may not be a popular way to define outliers in the literature,[2] our definition is useful for studying robustness of resampling methods. Our breakdown point analysis indicates that the implied probability bootstrap quantiles stay finite in a wider range than the pairs bootstrap quantiles when the values of outliers diverge. This does not necessarily have desirable implications on the size and power properties of the implied probability bootstrap tests because the original statistic based on the IV estimator may diverge as well. We also find that the implied probability bootstrap for the statistic based on the GMTM estimator shows desirable robustness properties. This finding is illustrated by striking simulation evidences. We also provide an empirical example based on Romer's (1993) study on the effect of openness of countries to inflation rates, where the data contain extremely high inflation rates of some Latin American countries.

There is a vast literature on the breakdown point theory in robust statistics (see, e.g., Hampel *et al.*, 1986, Rousseeuw, 1997, Rousseeuw and Leroy, 2003, and Maronna, Martin and Yohai, 2006). The next section presents a brief review on the literature of the breakdown point analysis in the context of resampling procedures. On the other hand, the literature of robustness study in the IV regression or GMM context is relatively thin and is currently under development. Ronchetti and Trojani (2001) extended robust estimation methods for just-identified estimating equations to the over-identified GMM setup. Gagliardini, Trojani and Urga (2005) proposed a robust GMM test for structural breaks. Čížek (2008) introduced a general trimmed estimation approach for nonlinear and limited dependent variable models, and Čížek (2008, 2009) extended this approach to the GMM context and proposed the GMTM estimator. Hill and Renault (2010) proposed a GMM estimator with asymptotically vanishing tail trimming for robust estimation of dynamic moment condition models. Kitamura, Otsu and Evdokimov (2013) studied local robustness of point estimators for moment condition models against perturbations controlled by the Hellinger distance. This paper studies global robustness of bootstrap methods in IV regression models.

In particular, in our global robustness analysis we focus on the GMTM estimator mainly for two reasons. First, the GMTM estimator is a global robust estimator characterized by a nontrivial breakdown point larger than 0. Second, the application of bootstrap approaches to the GMTM estimator allows to easily derive and clarify the global robustness properties of different bootstrap procedures. Unfortunately, a drawback of fixed-trimming estimators is that they may be biased both in finite sample and asymptotically. To overcome this problem (at least asymptotically), a possible solution consists of considering negligible trimming procedures, as in Hill and Renault (2010), Hill and Aguilar (2012) and Hill (2013). Note that the local robust approach introduced in Ronchetti and Trojani (2001) avoids also finite sample bias. However, since we are analyzing the global robustness properties of bootstrap procedures, it is more appropriate to focus on the global robust GMTM estimator.

The rest of the paper is organized as follows. In Section 2, we introduce basic concepts and the idea

whose data values are incorrect or unique in some sense, see Chambers (1986) for more details.

[2]Outliers are often defined as anomalous observations that are far away from the bulk of the data, or more generally, from the pattern set by the majority of the data (see, e.g., Hampel *et al.*, 1986).

for our breakdown point analysis and provide a brief literature review. Section 3 studies a just-identified case, which can be a benchmark for our breakdown point analysis. Section 4 generalizes the analysis in Section 3 to an over-identified model. Section 5 discusses extensions of our breakdown point analysis to different settings. Section 6 provides an empirical example. Section 7 concludes.

2 Breakdown point analysis for bootstrap: basic idea and literature

We first introduce basic concepts for our breakdown point analysis on bootstrap methods. Let $\mathcal{W}_n = \{W_i\}_{i=1}^n$ be an observed sample of size n and $S_n = S_n(\mathcal{W}_n)$ be a statistic of interest. Based on Donoho and Huber (1983), we define the (finite-sample) breakdown point of S_n as

$$\epsilon_n(S_n, \mathcal{W}_n) = \min_{1 \leq k \leq n} \left\{ \frac{k}{n} : \sup \| S_n(\mathcal{W}_{n,k}) - S_n(\mathcal{W}_n) \| = +\infty \right\}, \tag{1}$$

where the supremum is taken over all possible samples $\mathcal{W}_{n,k}$ of size n which are obtained by replacing k observations in $\mathcal{W}_n$ with arbitrary values, and $\|\cdot\|$ is the Euclidean norm. In words, the breakdown point measures the smallest fraction of contamination that can arbitrarily change the value of the statistic.[3] As emphasized in Donoho and Huber (1983), the breakdown point usually does not depend on the values of $\mathcal{W}_n$. For example, let us consider the observations $\mathcal{W}_n$ of size $n = 20$ with $W_i \in \mathbb{R}$. The sample mean $\bar{W} = \frac{1}{20}\sum_{i=1}^{20} W_i$ has a breakdown point of $\frac{1}{20}$. Let $W_{(1)} \leq \ldots \leq W_{(20)}$ be the ordered observations. The 10% trimmed mean $\tilde{W} = \frac{1}{18}\sum_{i=2}^{19} W_{(i)}$ (i.e., trim the smallest and largest observations) has a breakdown point of $\frac{1}{10}$. Note that the sample size n is held fixed.

The breakdown point analysis for the conventional bootstrap is introduced by Singh (1998). To explain the basic idea, Singh (1998) considered the bootstrap approximation for the distribution of the trimmed mean $\tilde{W}$. Note that $\tilde{W}$ is always free from the largest observation $W_{(20)}$, which is treated as an outlier. On the other hand, the bootstrap analog $\tilde{W}^{\#}$ of $\tilde{W}$ using the bootstrap resample from the empirical distribution of $\mathcal{W}_n$ is not necessarily free from $X_{(20)}$ because the bootstrap resample may contain $X_{(20)}$ more than once. Letting $B(n, q)$ be a binomial random variable with n trials and probability q, the probability that $\tilde{W}^{\#}$ is free from $X_{(20)}$ is written as

$$p^{\#} = P\left(B\left(20, \frac{1}{20}\right) \leq 1 \right) \approx 0.736.$$

Therefore, if $X_{(20)} \to +\infty$, then $100(1 - p_{\#})$ % of resamples of $\tilde{W}^{\#}$ will diverge to $+\infty$. In other words, the bootstrap t-th quantile of $\tilde{W}^{\#}$ will diverge to $+\infty$ for all $t > p_{\#}$. Note that the sample size $n = 20$ is held fixed for this analysis. Instead we analyze the limiting behavior of the bootstrap quantiles when the value of the outlier diverges, i.e., $X_{(20)} \to +\infty$. In this sense, the breakdown point analysis for

[3] There are other definitions of the breakdown point (e.g., Hampel *et al.*, 1986). Following Singh's (1998) seminal paper on breakdown point analysis for the bootstrap, we adopt Donoho and Huber's (1983) definition of the finite sample breakdown point in (1).

the bootstrap is very different from the conventional asymptotic analysis which focuses on the case of $n \to +\infty$.

There is a rich literature on breakdown point analysis in robust statistics (see, e.g., Hampel *et al.*, 1986, Rousseeuw, 1997, Rousseeuw and Leroy, 2003, and Maronna, Martin and Yohai, 2006). This paper is considered as an extension of previous research on breakdown point analysis for the bootstrap. Since the seminal work in Singh (1998), many studies have analyzed breakdown point properties of different resampling methods in different setups, such as Salibian-Barrera and Zamar (2002), Salibian-Barrera, Van Aelst and Willems (2007), and Camponovo, Scaillet and Trojani (2012a). In a recent study, Camponovo and Otsu (2012) extended the breakdown point analysis of the conventional bootstrap to the implied probability bootstrap introduced by Brown and Newey (2002). In particular, Camponovo and Otsu (2012) argued that the implied probability bootstrap is more robust than the conventional bootstrap when the implied probabilities of outliers become smaller than the uniform weight. In this case, as the values of outliers diverge to infinity, the implied probability bootstrap quantiles are well defined for a wider range than the conventional bootstrap quantiles. This paper extends the results of Camponovo and Otsu (2012) to IV regression models and derive more detailed results.

Also, our breakdown point analysis for bootstrap quantiles provides useful implications on the size and power properties of the bootstrap tests and confidence intervals. In this sense, this paper contributes to the literature of breakdown point analysis of statistical tests (see, e.g., Ylvisaker, 1977, He, Simpson and Portnoy, 1990, and Markatou and He, 1994).

Finally, the breakdown point properties of bootstrap quantiles crucially depend on the breakdown point in (1) of the statistic of interest. In IV regression models, conventional statistics, such as the t-statistic based on the IV estimator, have a trivial breakdown point $\frac{1}{n}$, i.e., a single outlier can arbitrarily change the values of the statistic (see, e.g., Krasker and Welsch, 1985). To provide more robust test statistics, recent research proposed various trimming procedures. The trimming approach has been largely applied in linear regression models with exogenous regressors. Important examples of high breakdown point robust estimators for linear regression include: the least trimmed squares estimator (Rousseeuw, 1985), the least trimmed absolute deviations estimator (Basset, 1991), and the maximum trimmed likelihood estimator (Neykov and Neitchev, 1990, and Hadi and Luceno, 1997). Čížek (2008) proposed a general trimmed estimation approach for nonlinear and limited dependent variable models. Also, Čížek (2008, 2009) extended the general trimming approach to the GMM context and proposed the GMTM estimator. Furthermore, Hill and Renault (2010) and Hill and Aguilar (2012) introduced tail trimming estimators in the GMM context, where the effect of trimming is asymptotically negligible as $n \to +\infty$. In this paper, we study robustness of bootstrap quantiles of statistics based on both the conventional IV and Čížek's (2008, 2009) GMTM estimators. It will be shown that the different bootstrap quantiles combined with the statistics based on these estimators show very different breakdown point properties.

3 Just-identified case

3.1 Setup

Let $\{W_i\}_{i=1}^n = \{Y_i, X_i, Z_i\}_{i=1}^n$ be an iid random sample of size n from $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^k$, where $k \geq p$, and each variable has finite variance. We consider the linear model

$$Y_i \quad = \quad X_i'\theta_0 + U_i,$$

for $i = 1, \ldots, n$, where $\theta_0 \in \mathbb{R}^p$ is a vector of unknown parameters and U_i is an error term. We suspect that the regressors X_i have endogeneity (i.e., $E[X_i U_i] \neq 0$) and the OLS estimator cannot consistently estimate the parameter of interest θ_0. In such a situation, it is common to introduce instrumental variables Z_i, which are orthogonal to the error term U_i. Based on the orthogonality, our estimation problem for θ_0 reduces to the one from the moment condition model

$$E\left[g\left(W_i, \theta_0\right)\right] = E\left[Z_i\left(Y_i - X_i'\theta_0\right)\right] = 0. \tag{2}$$

When the number of instruments equals the number of regressors (i.e., $k = p$), the model is called just-identified. When the number of instruments exceeds the number of regressors (i.e., $k > p$), the model is called over-identified. This section focuses on the case of $k = p = 1$, i.e., the model is just-identified and there is only one regressor. In this case, the IV for θ_0 is written as

$$\hat{\theta} = \frac{\sum_{i=1}^n Z_i Y_i}{\sum_{i=1}^n Z_i X_i}.$$

Based on the definition in (1), the breakdown point of $\hat{\theta}$ is $\epsilon_n\left(\hat{\theta}, \mathcal{W}_n\right) = \frac{1}{n}$, i.e. the replacement of a single observation with an arbitrary value may imply the divergence of $\hat{\theta}$. As an example of a robust estimator, we also consider the following version of the GMTM estimator introduced by Čížek (2008, 2009):

$$\hat{\theta}^d = \arg\min_\theta \left[\frac{1}{n}\sum_{i=1}^n Z_i\left(Y_i - X_i\theta\right)\mathbb{I}\left\{r\left(W_i, \theta\right) \leq r\left(W_{[n-d]}, \theta\right)\right\}\right]^2,$$

where $\mathbb{I}\{\cdot\}$ is the indicator function, $r\left(W_i, \theta\right) = \left|Z_i\left(Y_i - X_i\theta\right)\right|^2$ is a trimming function which is ordered as $r\left(W_{[1]}, \theta\right) \leq \cdots \leq r\left(W_{[n]}, \theta\right)$, and d is an integer such that $0 \leq d \leq \frac{n}{2}$ to determine the amount of trimming.[4] In this estimator, outliers are determined by the value of $r\left(W_i, \theta\right)$ and removed from the estimating equation. If $d = 0$, there is no trimming, i.e., $\hat{\theta}^d = \hat{\theta}$. Also we can see that the breakdown

[4] In Čížek (2008, 2009), the trimming term is written as $\mathbb{I}\left\{r\left(W_i, \theta\right) \leq r\left(W_{[\lambda n]}, \theta\right)\right\}$ for $\lambda \in \left(\frac{1}{2}, 1\right]$. This expression is important to analyze the asymptotic property of the GMTM estimator as $n \to \infty$, which is characterized by λ. In our breakdown point analysis, the sample size n is held fixed. So we employ the expression $\mathbb{I}\left\{r\left(W_i, \theta\right) \leq r\left(W_{[n-d]}, \theta\right)\right\}$ using an integer d for convenience. Also note that we select $0 \leq d \leq \frac{n}{2}$ because we cannot distinguish which part of the data should be fit by the model and which part should be trimmed. In practice, the selection of d depends case by case. A descriptive analysis based on scatter plot may help to identify anomalous observations and to select appropriate values of d.

point of $\hat{\theta}^d$ is $\epsilon_n\left(\hat{\theta}^d, \mathcal{W}_n\right) = \frac{d+1}{n}$, i.e., $d+1$ outliers are necessary in order to change arbitrarily the value of $\hat{\theta}^d$.

We consider parameter hypothesis testing for the null $H_0 : \theta_0 = c$ with some given $c \in \mathbb{R}$ against the two-sided alternative $H_1 : \theta_0 \neq c$. Our breakdown point analysis can be easily extended to one-sided testing by analyzing divergence properties of test statistics to positive and negative infinity separately. Based on the point estimators introduced above, we focus on the test statistics $T_n = \sqrt{n}\left(\hat{\theta} - c\right)$ and $T_n^d = \sqrt{n}\left(\hat{\theta}^d - c\right)$. In Section 3.5.3, we consider a studentized statistic.[5] To obtain critical values of the tests, we need to find approximations to the distributions of the test statistics under the null hypothesis H_0. One way to approximate these distributions is to apply the pairs bootstrap method. The pairs bootstrap draws resamples from the observations $\{W_i\}_{i=1}^n$ with the uniform weight $1/n$, and approximates the distributions of T_n and T_n^d by their resampled statistics. Another bootstrap method is to impose the moment condition $E\left[g\left(W_i, c\right)\right] = 0$ under the null hypothesis H_0, and draw bootstrap resamples using the implied probabilities (Back and Brown, 1993),[6]

$$\pi_i = \frac{1}{n} - \frac{1}{n} \frac{\left(g\left(W_i, c\right) - \bar{g}\right)\bar{g}}{\frac{1}{n}\sum_{i=1}^n g\left(W_i, c\right)^2}, \tag{3}$$

for $i = 1, \ldots, n$, where $\bar{g} = \frac{1}{n}\sum_{i=1}^n g\left(W_i, c\right)$ (note: g is assumed to be scalar-valued in this section).[7] The second term in (3) can be interpreted as a penalty term for the deviation from H_0. If $\left|g\left(W_i, c\right)\right|$ becomes larger, then the second term tends to be negative (because $\left(g\left(W_i, c\right) - \bar{g}\right)$ and $\bar{g}$ tends to take the same sign) and the weight π_i tends to be smaller than the uniform weight $\frac{1}{n}$. Intuitively, if an outlier in the observations yields a large value of $\left|g\left(W_i, c\right)\right|$, then the implied probability bootstrap tends to draw the outlier less frequently. Thus it is reasonable to expect that the pairs and implied probability bootstrap methods have different robustness properties in the presence of outliers. The next subsection formalizes this intuition by using the finite sample breakdown point theory for resample methods.

---

[5] Note that both in Sections 3 and 4 we assume that the statistics under investigation satisfy a central limit theorem.

[6] For the breakdown point analysis below, we focus on Back and Brown's (1993) implied probability in (3) because of its tractability. Back and Brown's (1993) implied probability can be interpreted as an approximation to the Fisher information projection from the empirical distribution to the space of distributions satisfying the moment conditions. It is important to extend our analysis to other implied probabilities using different information projections based on the Boltzmann-Shannon entropy yielding the exponential tilting weights (Kitamura and Stutzer, 1997, and Imbens, Spady and Johnson, 1998) and Burg entropy yielding the empirical likelihood weights (Owen, 1988) for example. In particular, Camponovo and Otsu (2012, Section 2.1) suggested a way to extend the breakdown point analysis for the implied probability bootstrap based on generalized empirical likelihood (Newey and Smith, 2004) in a limited setup. Their approach can be applied to our setup.

[7] Our breakdown point analysis assumes that all implied probabilities are non-negative. This assumption is typically justified when the sample size is sufficiently large. However, in finite samples, it is possible to have negative implied probabilities. In the simulation study below, we adopt a shrinkage-type modification suggested by Antoine, Bonnal and Renault (2007) to avoid negative implied probabilities. More precisely, for $i = 1, \ldots, n$ we consider the non-negative implied probabilities $\tilde{\pi}_i = \frac{1}{1+\tilde{\epsilon}_n}\pi_i + \frac{\tilde{\epsilon}_n}{1+\tilde{\epsilon}_n}\frac{1}{n}$, where $\tilde{\epsilon}_n = -n\min_{1\leq i\leq n}(\pi_i, 0)$.

3.2 Breakdown point analysis

Based on the above setup, we now conduct the breakdown point analysis for the pairs and implied probability bootstrap methods. We first define outliers. To fix the idea, let $\left\|W_{(1)}\right\| \leq \cdots \leq \left\|W_{(n)}\right\|$ be the observations ordered by the Euclidean norm, and let us treat $W_{(n)}$ as an outlier. Consider the statistic $T_n = \sqrt{n}\left(\hat{\theta} - c\right)$ based on the IV estimator $\hat{\theta}$ to test $H_0 : \theta_0 = c$ against $H_1 : \theta_0 \neq c$. We assume that $W_{(n)}$ is an outlier in the following sense.

Assumption 1. $W_{(n)}$ is an outlier for the statistic T_n in the sense that

$$|T_n| \to +\infty \text{ as } \left\|W_{(n)}\right\| \to +\infty.$$

Some specific examples satisfying this assumption are provided in the end of this subsection. The choice of $W_{(n)}$ as an outlier is just for convenience. Any types of divergence or convergence in the observations causing $|T_n| \to +\infty$ can be treated as outliers, and the same analysis applies.

We now consider the pairs bootstrap. The pairs bootstrap analog of T_n is written as $T_n^{\#} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$, where $\hat{\theta}^{\#}$ is the IV estimator based on the pairs bootstrap resamples. Note that $T_n^{\#}$ depends on $\hat{\theta}$, the IV estimator of the original sample. Thus, by Assumption 1, $\left|T_n^{\#}\right|$ diverges to infinity as $\left\|W_{(n)}\right\| \to +\infty$ even if the pairs bootstrap resample to compute $\hat{\theta}^{\#}$ does not contain $W_{(n)}$. Also, if the resample contains the outlier $W_{(n)}$ possibly multiple times, then $\left|T_n^{\#}\right|$ may diverge or become indeterminate as $\left\|W_{(n)}\right\| \to +\infty$. Based on these results, we can at least say that $\left|T_n^{\#}\right|$ diverges to infinity as $\left\|W_{(n)}\right\| \to +\infty$ when the resample to compute $\hat{\theta}^{\#}$ does not contain $W_{(n)}$ (because $\hat{\theta}$ diverges but $\hat{\theta}^{\#}$ does not as $\left\|W_{(n)}\right\| \to +\infty$). The probability for this event in the pairs bootstrap resampling is obtained as

$$p^{\#} = P\left(B\left(n, \frac{1}{n}\right) = 0\right),$$

where $B(n, q)$ is a binomial random variable with n trials and probability q. Therefore, at least $100p^{\#}\%$ of resamples of $\left|T_n^{\#}\right|$ will diverge to $+\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. In other words, the t-th bootstrap quantile $Q_t^{\#}$ of $\left|T_n^{\#}\right|$ will diverge to $+\infty$ for all $t > 1 - p^{\#}$.

We next consider the implied probability bootstrap. We impose the following additional assumption.

Assumption 2. Assume that

$$\left|g\left(W_{(n)}, c\right)\right| = \left|Z_{(n)}\left(Y_{(n)} - X_{(n)}c\right)\right| \to +\infty \text{ as } \left\|W_{(n)}\right\| \to +\infty,$$

This assumption is very mild. For example, if one of the elements in $\left(Y_{(n)}, X_{(n)}, Z_{(n)}\right)$ diverges, then this assumption is satisfied (unless $Y_{(n)} - X_{(n)}c = 0$ or $Z_{(n)} = 0$). Under this assumption, the implied probability in (3) for the observation $W_{(n)}$ satisfies

$$\pi_{(n)} = \frac{1}{n} - \frac{1}{n}\frac{\left(1 - \frac{1}{n} - \frac{\bar{g}_-}{g\left(W_{(n)}, c\right)}\right)\left(\frac{\bar{g}_-}{g\left(W_{(n)}, c\right)} + \frac{1}{n}\right)}{\frac{\bar{v}_-}{g\left(W_{(n)}, c\right)^2} + \frac{1}{n}} \to \frac{1}{n^2}, \tag{4}$$

as $\left\|W_{(n)}\right\| \to +\infty$, where $\bar{g}_- = \frac{1}{n}\sum_{i=1}^{n-1} g\left(W_{(i)}, c\right)$ and $\bar{v}_- = \frac{1}{n}\sum_{i=1}^{n-1} g\left(W_{(i)}, c\right)^2$. In contrast to the pairs bootstrap which draws the outlier $W_{(n)}$ with probability $\frac{1}{n}$, the implied probability bootstrap draws the outlier with smaller probability $\frac{1}{n^2}$ as $\left\|W_{(n)}\right\| \to +\infty$. The implied probability bootstrap counterpart of T_n is written as $T_n^* = \sqrt{n}\left(\hat{\theta}^* - c\right)$. Note that T_n^* is centered around the hypothetical value c instead of the estimator $\hat{\theta}$. This is due to the fact that the implied probability bootstrap resamples are drawn from the multinomial distribution satisfying $\sum_{i=1}^{n} \pi_i g\left(W_i, c\right) = 0$. Thus, $|T_n^*|$ will diverge as $\left\|W_{(n)}\right\| \to +\infty$ only when the resample to compute $\hat{\theta}^*$ contains the outlier $W_{(n)}$. From (4), the probability that the implied probability bootstrap statistic T_n^* is free from the outlier $W_{(n)}$ converges to

$$p^* = P\left(B\left(n, \frac{1}{n^2}\right) = 0\right),$$

as $\left\|W_{(n)}\right\| \to +\infty$. Therefore, under Assumptions 1 and 2, $100\left(1 - p^*\right)\%$ of resamples of $|T_n^*|$ will diverge to $+\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. In other words, the t-th bootstrap quantile Q_t^* of $|T_n^*|$ will diverge to $+\infty$ for all $t > p^*$. We summarize these findings on the pairs and implied bootstrap methods in the following proposition.

Proposition 1. *Consider the setup of this section.*

(i) *Under Assumption 1, the pairs bootstrap analog $T_n^{\#} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ always contains the outlier $W_{(n)}$, and the pairs bootstrap quantile $Q_t^{\#}$ from the resamples of $\left|T_n^{\#}\right|$ diverges to $+\infty$ for all $t > 1 - p^{\#}$ as $\left\|W_{(n)}\right\| \to +\infty$.*

(ii) *Under Assumptions 1 and 2, the implied probability bootstrap analog $T_n^* = \sqrt{n}\left(\hat{\theta}^* - c\right)$ contains the outlier $W_{(n)}$ with probability $1 - p^*$, and the implied probability bootstrap quantile Q_t^* from the resamples of $|T_n^*|$ diverges to $+\infty$ for all $t > p^*$ as $\left\|W_{(n)}\right\| \to +\infty$.*

For illustration, we present the values of $p^{\#}$ and p^* for different sample sizes.

Table A: Values of $p^{\#}$ and p^*

n	10	20	50	100	500	1000
$p^{\#}$	0.349	0.358	0.364	0.366	0.368	0.368
p^*	0.904	0.951	0.980	0.990	0.998	0.999

For example, when $n = 50$, divergence of a single outlier implies divergence of $100p^{\#} = 36.4\%$ of the pairs bootstrap resamples of $T_n^{\#}$. On the other hand, divergence of a single outlier implies divergence of $100\left(1 - p^*\right) = 2\%$ of the implied probability bootstrap resamples of T_n^*. As far as $n > 3$, it holds $p^* > 1 - p^{\#}$ and the implied probability bootstrap provides finite quantiles for a wider range than the pairs bootstrap in the case of $\left\|W_{(n)}\right\| \to +\infty$. However, this does not necessarily mean that the implied probability bootstrap test has desirable size or power properties in the presence of outliers because the original statistic $T_n = \sqrt{n}\left(\hat{\theta} - c\right)$ diverges to infinity as $\left\|W_{(n)}\right\| \to +\infty$. See Section 3.3 for a detailed discussion.

We now consider the statistic $T_n^d = \sqrt{n}\left(\hat{\theta}^d - c\right)$ with $d \geq 1$ based on the GMTM estimator $\hat{\theta}^d$. Note that under Assumption 2, $r\left(W_{(n)}, c\right) = \left|Z_{(n)}\left(Y_{(n)} - X_{(n)}c\right)\right|^2 \to +\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. Thus, the outlier $W_{(n)}$ will be trimmed, and $\hat{\theta}^d$ and T_n^d are bounded as $\left\|W_{(n)}\right\| \to +\infty$. On the other hand, the bootstrap counterparts $T_n^{d\#} = \sqrt{n}\left(\hat{\theta}^{d\#} - \hat{\theta}^d\right)$ and $T_n^{d*} = \sqrt{n}\left(\hat{\theta}^{d*} - c\right)$ diverge if the resamples contain the outlier $W_{(n)}$ more than d times. The probability that the pairs bootstrap resample to compute $\hat{\theta}^{d\#}$ contains the outlier $W_{(n)}$ less than or equal to d times is

$$p^{d\#} = P\left(B\left(n, \frac{1}{n}\right) \leq d\right).$$

Also, from (4), the probability that the implied probability bootstrap resample to compute $\hat{\theta}^{d*}$ contains the outlier $W_{(n)}$ less than or equal to d times converges to

$$p^{d*} = P\left(B\left(n, \frac{1}{n^2}\right) \leq d\right),$$

as $\left\|W_{(n)}\right\| \to +\infty$. Therefore, $100\left(1 - p^{d\#}\right)\%$ of resamples of $\left|T_n^{d\#}\right|$ will diverge to $+\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. In other words, the t-th bootstrap quantile $Q_t^{d\#}$ of $\left|T_n^{d\#}\right|$ will diverge to $+\infty$ for all $t > p^{d\#}$. Similarly, the t-th bootstrap quantile Q_t^{d*} of $\left|T_n^{d*}\right|$ will diverge to $+\infty$ for all $t > p^{d*}$. These findings are summarized as follows.

Proposition 2. *Consider the setup of this section.*

(i) *Under Assumptions 1 and 2, the pairs bootstrap analog $T_n^{d\#} = \sqrt{n}\left(\hat{\theta}^{d\#} - \hat{\theta}^d\right)$ contains the outlier $W_{(n)}$ with probability $1 - p^{d\#}$, and the pairs bootstrap quantile $Q_t^{d\#}$ from the resamples of $\left|T_n^{d\#}\right|$ diverges to $+\infty$ for all $t > p^{d\#}$ as $\left\|W_{(n)}\right\| \to +\infty$.*

(ii) *Under Assumptions 1 and 2, the implied probability bootstrap analog $T_n^{d*} = \sqrt{n}\left(\hat{\theta}^{d*} - c\right)$ contains the outlier $W_{(n)}$ with probability $1 - p^{d*}$, and the implied probability bootstrap quantile Q_t^{d*} from the resamples of $\left|T_n^{d*}\right|$ diverges to $+\infty$ for all $t > p^{d*}$ as $\left\|W_{(n)}\right\| \to +\infty$.*

For illustration, we present the values of $p^{d\#}$ and p^{d*} for $n = 50$ and $0 \leq d \leq 5$.

Table B: Values of $p^{d\#}$ and p^{d*}

$n = 50$	$d = 0$	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
$p^{d\#}$	0.364	0.736	0.930	0.982	0.997	0.999
p^{d*}	0.980	1.000	1.000	1.000	1.000	1.000

As expected, from Table B we can observe that both $p^{d\#}$ and p^{d*} increase as d increases. However, p^{d*} is always larger than $p^{d\#}$. Therefore, the implied probability bootstrap is more robust than the pairs bootstrap. It is also interesting to note that in presence of a single outlier the GMTM estimator $\hat{\theta}^d$ is stable for $d = 1$, while on the other hand $1 - 0.736 = 0.264$ of the bootstrap samples will be extremely large. This result confirms that to robustify the pairs bootstrap it is not enough to apply only robust estimators (see also Singh (1998) for a similar example).

Note that as the number of trimmed observations d increases, both $p^{d\#}$ and p^{d*} increase. Thus, the bootstrap quantiles of the statistic T_n^d based on the GMTM estimator stay finite for wider ranges than the ones of T_n in the case of $\|W_{(n)}\| \to +\infty$. Also, since $p^{d*} > p^{d\#}$, the implied probability bootstrap quantile Q_t^{d*} stays finite for a wider range than the pairs bootstrap quantile $Q_t^{d\#}$. In contrast to T_n, the statistic T_n^d will be free from $W_{(n)}$ as $\|W_{(n)}\| \to +\infty$. Thus, the robustness of the implied probability bootstrap quantile in the above sense has desirable implications on the size and power properties of the implied probability bootstrap test. We discuss this point in Section 3.3 with a striking simulation evidence in Section 3.4.

Finally, we discuss some examples that satisfy Assumption 1. The breakdown point results above apply as far as this high-level assumption is satisfied. However, it is insightful to inspect some specific types of outliers satisfying Assumption 1 in the IV regression context. In the case of $k = p = 1$, the outlier $W_{(n)}$ contains three elements $\left(Y_{(n)}, X_{(n)}, Z_{(n)}\right)$. For illustration, we consider the following cases.

Table C: Limits of T_n as $\|W_{(n)}\| \to +\infty$

Case	Diverge	Bounded	Limit of $\lvert T_n \rvert$
1	Z	$X,\ Y$	bounded
2	X	$Y,\ Z$	bounded
3	Y	$X,\ Z$	$+\infty$

In Table C (and also Tables D, E, and F below), X, Y, and Z mean $\left|X_{(n)}\right|$, $\left|Y_{(n)}\right|$, and $\left|Z_{(n)}\right|$, respectively. For example, the second row for Case 1 means that as $\left|Z_{(n)}\right| \to +\infty$, but $\left|X_{(n)}\right|$ and $\left|Y_{(n)}\right|$ are bounded, then $|T_n|$ is bounded, while the fourth row for Case 3 means that as $\left|Y_{(n)}\right| \to +\infty$, but $\left|X_{(n)}\right|$ and $\left|Z_{(n)}\right|$ are bounded, then $|T_n|$ diverges to infinity. From this table, we can regard Case 3 as an example of the outlier for T_n. Obviously, there are various other types of outliers. Let $A = \sum_{i=1}^{n-1} Z_{(i)} Y_{(i)}$ and $B = \sum_{i=1}^{n} Z_{(i)} X_{(i)}$. An inspection of

$$\hat{\theta} = \frac{\sum_{i=1}^{n} Z_i Y_i}{\sum_{i=1}^{n} Z_i X_i} = \frac{A}{B} + \frac{Z_{(n)} Y_{(n)}}{B},$$

reveals that $|T_n| = \sqrt{n}\left|\hat{\theta} - c\right|$ diverges to infinity as $\|W_{(n)}\| \to +\infty$ when $B \to 0$ or $\left|\frac{Z_{(n)} Y_{(n)}}{B}\right| \to \infty$. The situation of $B \to 0$ is somewhat unrealistic but may be caused by very weak instruments. The situation of $\left|\frac{Z_{(n)} Y_{(n)}}{B}\right| \to \infty$ can occur when $Y_{(n)}$ or both $Y_{(n)}$ and $Z_{(n)}$ diverges.

3.3 Implications on size and power properties

The breakdown point analysis in the previous subsection has important implications on the size and power properties of the bootstrap tests. Suppose Assumptions 1 and 2 hold true. First, consider the statistic $T_n = \sqrt{n}\left(\hat{\theta} - c\right)$ based on the IV estimator. A key observation is that

$$\begin{aligned} T_n &= \sqrt{n}\left(\hat{\theta} - c\right) \text{ always contains } W_{(n)}, \\ T_n^{\#} &= \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right) \text{ always contains } W_{(n)}, \\ T_n^{*} &= \sqrt{n}\left(\hat{\theta}^{*} - c\right) \text{ contains } W_{(n)} \text{ with probability } p^{*} \text{ (as } \left\|W_{(n)}\right\| \to +\infty). \end{aligned}$$

As shown in Proposition 1 (i), the distribution of the pairs bootstrap statistic $T_n^{\#} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ is heavily influenced by the presence of outliers. In particular, the pairs bootstrap quantiles of $\left|T_n^{\#}\right|$ tend to be extremely large in the presence of outliers. This yields extremely large critical values for the pairs bootstrap test (or equivalently, extremely wide confidence intervals). Thus, we tend to accept the null hypothesis very frequently, and both the size and power of the pairs bootstrap test tend to be close to 0. On the other hand, Proposition 1 (ii) says that the implied probability bootstrap quantiles of $T_n^{*} = \sqrt{n}\left(\hat{\theta}^{*} - c\right)$ are more robust to the presence of outliers. This yields relatively small and stable critical values for the implied probability bootstrap test. However, this stability of the bootstrap quantiles is not necessarily desirable for the test. By Assumption 1, the value of the original statistic $|T_n|$ tends to be large as $\left\|W_{(n)}\right\|$ increases. Thus, relatively small and stable critical values by the implied probability bootstrap yield very frequent rejections of the null hypothesis, and both the size and power of the implied probability bootstrap test tend to be close to 1.

Next, consider the statistic $T_n^d = \sqrt{n}\left(\hat{\theta}^d - c\right)$ with $d \geq 1$ based on the GMTM estimator. Again a key observation is that as $\left\|W_{(n)}\right\| \to \infty$,

$$\begin{aligned} T_n^d &= \sqrt{n}\left(\hat{\theta}^d - c\right) \text{ never contains } W_{(n)}, \\ T_n^{d\#} &= \sqrt{n}\left(\hat{\theta}^{d\#} - \hat{\theta}^d\right) \text{ contains } W_{(n)} \text{ with probability } p^{d\#}, \\ T_n^{d*} &= \sqrt{n}\left(\hat{\theta}^{d*} - c\right) \text{ contains } W_{(n)} \text{ with probability } p^{d*}. \end{aligned}$$

As shown in Proposition 2, the bootstrap quantiles of the pairs bootstrap statistic $T_n^{d\#} = \sqrt{n}\left(\hat{\theta}^{d\#} - \hat{\theta}^d\right)$ are more robust than those of the statistic $T_n^{\#}$. However, from $p^{d\#} > p^{d*}$, the bootstrap quantiles of $T_n^{d\#}$ tend to be larger than those of T_n^{d*}, and the pairs bootstrap test tends to accept the null hypothesis more often than the implied probability bootstrap test. In contrast to T_n, the statistic T_n^d never contains $W_{(n)}$. Therefore we can expect that the implied probability bootstrap is more accurate than the pairs bootstrap to approximate the distribution of T_n^d, and shows better size and power properties.

In summary, to test the null hypothesis H_0 in the presence of outliers, we recommend the use of the statistic T_n^d based on Čížek's (2008, 2009) GMTM estimator combined with the implied probability bootstrap. In the next subsection, we provide a striking simulation evidence to endorse our recommendation.

3.4 Simulation

In this subsection, we conduct a simulation study to evaluate the performance of the bootstrap methods in the presence of outliers. We consider iid samples $\{W_i\}_{i=1}^n = \{Y_i, X_i, Z_i\}_{i=1}^n$ of sizes $n = 50$ and 100 generated from $Y_i = X_i\theta_0 + U_i$ and $X_i = Z_i\pi_0 + V_i$, where $Z_i \sim N(1,1)$, $\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right)$, and $\pi_0 = 0.8$.[8] The true parameter value is set as $\theta_0 = 2$. We are interested in testing the null hypothesis $H_0 : \theta_0 = 2$ against the alternative $H_1 : \theta_0 \neq 2$. For each scenario, the number of bootstrap replications is 399 for each Monte Carlo sample and the number of Monte Carlo replications is $10,000$.

To study robustness of the bootstrap methods for approximating the distributions of the test statistics T_n and T_n^d with $d = 1$, we consider two situations: (i) $(W_1, \ldots, W_n)$ are generated from the above model (No outlier), and (ii) $\left(\tilde{W}_1, \ldots, \tilde{W}_n\right)$ with $\tilde{W}_{(i)} = W_{(i)}$ for $i = 1, \ldots, n-1$ and $\tilde{W}_{(n)} = \left(\tilde{Y}_{(n)}, \tilde{X}_{(n)}, \tilde{Z}_{(n)}\right) = \left(CY_{max}, X_{(n)}, Z_{(n)}\right)$ with $C = 5, 10, 20$ and $Y_{max} = \max\{Y_1, \ldots, Y_n\}$ (Outlier in Y). This specification of the outlier corresponds to Case 3 in Table C. Proposition 1 says that as $\left\|W_{(n)}\right\| \to \infty$, the pairs bootstrap t-th quantile $Q_t^{\#}$ from the resamples $\left|T_n^{\#}\right|$ of the statistic T_n will diverge to $+\infty$ for all $t > 1 - P\left(B\left(n, \frac{1}{n}\right) = 0\right)$, and the implied probability bootstrap t-th quantile Q_t^* from the resamples $|T_n^*|$ of T_n will diverge to $+\infty$ for all $t > P\left(B\left(n, \frac{1}{n^2}\right) = 0\right)$. Also by Proposition 2, the pairs bootstrap t-th quantile from the resamples $\left|T_n^{d\#}\right|$ of T_n^d will diverge to $+\infty$ for all $t > P\left(B\left(n, \frac{1}{n}\right) \leq 1\right)$, and the implied probability bootstrap t-th quantile from the resamples $\left|T_n^{d*}\right|$ of T_n^d will diverge to $+\infty$ for all $t > P\left(B\left(n, \frac{1}{n^2}\right) \leq 1\right)$.

Table 1 reports the Monte Carlo medians of the pairs bootstrap quantiles $Q_t^{\#}$ and $Q_t^{d\#}$ and implied probability bootstrap quantiles Q_t^* and Q_t^{d*} for $|T_n|$ and $\left|T_n^d\right|$, respectively. We set $t = 0.95$, and we report also empirical coverages of bootstrap confidence intervals. First we consider the statistic $|T_n|$. In absence of the outlier, both bootstrap methods quantiles are accurate to approximate the true quantiles. For instance, in the case of $n = 50$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 1.7972 and 1.7882, respectively, while the true quantile is 1.8210. Furthermore, also the empirical coverages are quite close to the nominal coverage probability 0.95. In contrast, in the presence of the outlier, the pairs bootstrap quantiles are extremely large, while the implied probability bootstrap quantiles tend to be close to the true quantiles without the outlier. For instance, in the case of $n = 100$ and $C = 20$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 23.3004 and 2.4642, respectively. It is important to note that in the presence of the outlier, the true quantiles are extremely large. For instance, for $n = 100$ and $C = 20$, the true quantile is 29.7983. In the presence of the outlier, both bootstrap methods do not provide accurate approximations to the true quantiles. Finally, also the empirical coverages are highly distorted. As expected, the pairs bootstrap

[8] We also tried the case of $n = 200$, but the results are similar to those of the case of $n = 100$. Unreported Monte Carlo results for $n \geq 1000$ show that the impact of the single outlier is less pronounced.

empirical coverages are close to 1, while implied probability bootstrap coverages are extremely small.

We next consider the statistic T_n^d based on the GMTM estimator. Again, in absence of the outlier, both pairs bootstrap and implied probability bootstrap quantiles are accurate to approximate the true quantiles. For instance, in the case of $n = 100$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 1.7457 and 1.7212, respectively, while the true quantile is 1.7182. Furthermore, also the empirical coverages remain very close to the nominal coverage probability. On the other hand, in the presence of the outlier, the pairs bootstrap quantiles tend to be large. This is due to the fact that the pairs bootstrap resamples often contain the outlier more than once. For instance, in the case of $n = 100$ and $C = 20$, the Monte Carlo median of the pairs bootstrap quantile is 23.0146. In contrast, even in the presence of the outlier, the implied probability bootstrap accurately approximate the true quantiles. For instance, in the case of $n = 100$ and $C = 20$, the Monte Carlo median of the implied probability bootstrap quantile is 1.7807, while the true quantile is 1.7827. As expected, the pairs bootstrap empirical coverages are close to 1. On the other hand, the implied probability empirical coverages remain very close to the nominal coverage probability. [9]

Furthermore, we study the size and the power properties of the bootstrap tests. In particular, we consider two situations: (i) $(W_1, \ldots, W_n)$ are generated from the above model (No outlier) with $\theta_0 \in [2, 2.3]$, and (ii) $\left(\tilde{W}_1, \ldots, \tilde{W}_n\right)$ with $\tilde{W}_{(i)} = W_{(i)}$ for $i = 1, \ldots, 99$ and $\tilde{W}_{(100)} = \left(\tilde{Y}_{(100)}, \tilde{X}_{(100)}, \tilde{Z}_{(100)}\right) = \left(10Y_{max}, X_{(100)}, Z_{(100)}\right)$ with $Y_{max} = \max\{Y_1, \ldots, Y_{100}\}$ (Outlier in Y). Using the bootstrap methods, we test the null hypothesis $H_0 : \theta_0 = 2$ under different parameter values of $\theta_0 \in [2, 2.3]$.

Table 6 reports the rejection frequencies of the null hypothesis under different parameter values $\theta_0 \in [2, 2.3]$. First we consider the statistic T_n. In absence of the outlier, for $\theta_0 = 2$, the rejection frequencies of both bootstrap methods are quite close to the nominal level 0.05. The power of the bootstrap tests increases as the value of θ_0 increases to 2.3. For instance, at $\theta_0 = 2.3$, the rejection frequencies are larger than 80% for both bootstrap tests. In the presence of the outlier, both size and power of the tests are dramatically distorted. In particular, the rejection frequencies of the pairs bootstrap are always smaller than 15% even when $\theta_0 = 2.3$ (very low power). In contrast, the rejection frequencies of the implied probability bootstrap are always larger than 60% even when $\theta_0 = 2$ (severe size distortion). These results endorse our findings in Section 3.3. In the presence of the outlier, the pairs bootstrap critical values tend to be very large, and both size and power of the pairs bootstrap test tend to collapse to 0. On the other hand, in the presence of the outlier, the implied probability bootstrap critical values tend to be very low and, both size and power of the implied probability bootstrap test tend to collapse to 1.

Finally, we consider the statistic T_n^d. In absence of the outlier, the rejection frequencies are very similar to those obtained for the statistic T_n. In particular, when $\theta_0 = 2$, the rejection frequencies of

[9] In this context, to compare the robustness properties of the implied probability bootstrap with other robust resampling methods, we consider also the winsorization approach described in Singh (1998). Unreported Monte Carlo results confirm the accuracy of both methods and highlight some analogies of the two procedures.

both bootstrap tests are very close to the nominal level 0.05. Furthermore, for $\theta_0 \neq 2$, the power of the bootstrap tests increases as the value of θ_0 increases to 2.3. In the presence of the outlier, the power of the pairs bootstrap test is dramatically distorted. In particular, the rejection frequencies of the pairs bootstrap are always smaller than 10% even when $\theta_0 = 2.3$. In contrast, the presence of the outlier does not deteriorate the accuracy of the implied probability bootstrap. For $\theta_0 = 2$, the rejection frequency of the implied probability bootstrap remains very close to the nominal level 0.05. For $\theta_0 = 2.3$, the rejection frequency is still close to 80%. Therefore, we conclude that the implied probability bootstrap test using the statistic T_n^d has desirable size and power properties in the presence of the outlier.

3.5 Extensions

3.5.1 Multiple outliers

Proposition 1 on the breakdown point properties of the bootstrap quantiles can be extended to the case where we have $m \in \{1, \ldots, n-1\}$ outliers. To this end, we extend Assumption 1 on the outlier as follows.

Assumption 3. $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ are outliers for the statistic T_n in the sense that for each $j = 1, \ldots, m$,

$$|T_n| \to +\infty \text{ as } \left\|W_{(n-j+1)}\right\| \to +\infty.$$

We first analyze the robustness properties of the pairs bootstrap. Note that the pairs bootstrap analog $T_n^{\#} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ depends on $\hat{\theta}$, the IV estimator using the original sample. Thus, by Assumption 3, $\left|T_n^{\#}\right|$ diverges to infinity as $\left\|W_{(n-j+1)}\right\| \to +\infty$ even if the pairs bootstrap resample to compute $\hat{\theta}^{\#}$ does not contain the outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$. Also, if the resample contains the outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ possibly multiple times, then $\left|T_n^{\#}\right|$ may diverge or become indeterminate as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$. Thus, similar to the single outlier case, we can at least say that $\left|T_n^{\#}\right|$ diverges to infinity as $\left\|W_{(n-j+1)}\right\| \to +\infty$ when the resample to compute $\hat{\theta}^{\#}$ does not contain any outlier in $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$. The probability for this event is obtained as

$$p_m^{\#} = P\left(B\left(n, \frac{m}{n}\right) = 0\right).$$

Therefore, at least $100 p_m^{\#}$% of resamples of $\left|T_n^{\#}\right|$ will diverge to $+\infty$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$. In other words, the t-th bootstrap quantile $Q_t^{\#}$ of $\left|T_n^{\#}\right|$ will diverge to $+\infty$ for all $t > 1 - p_m^{\#}$.

For the implied probability bootstrap, we impose the following analog of Assumption 2.

Assumption 4. For $j = 1, \ldots, m$,

$$\begin{aligned}
\left|g\left(W_{(n)}, c\right)\right| &\to +\infty \text{ as } \left\|W_{(n)}\right\| \to +\infty, \\
\frac{g\left(W_{(n-j+1)}, c\right)}{g\left(W_{(n)}, c\right)} &\to 1 \text{ as } \left\|W_{(n)}\right\| \to +\infty \text{ and } \left\|W_{(n-j+1)}\right\| \to +\infty.
\end{aligned}$$

Under this assumption, the implied probability in (3) for the observation $W_{(n)}$ satisfies

$$\pi_{(n)} = \frac{1}{n} - \frac{1}{n} \frac{\left\{1 - \frac{1}{n}\left(\frac{g_{(n-1)}+\cdots+g_{(n-m+1)}}{g_{(n)}}\right) - \frac{\bar{g}_{m-}}{g_{(n)}}\right\}\left\{\frac{\bar{g}_{m-}}{g_{(n)}} + \frac{1}{n}\left(\frac{g_{(n)}+\cdots+g_{(n-m+1)}}{g_{(n)}}\right)\right\}}{\frac{\bar{v}_{m-}}{g_{(n)}^2} + \frac{1}{n}\left(\frac{g_{(n)}^2+\cdots+g_{(n-m+1)}^2}{g_{(n)}^2}\right)} \to \frac{m}{n^2},$$

as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for all $j = 1, \ldots, m$, where $g_{(i)} = g\left(W_{(i)}, c\right)$, $\bar{g}_{m-} = \frac{1}{n}\sum_{i=1}^{n-m} g\left(W_{(i)}, c\right)$, and $\bar{v}_{m-} = \frac{1}{n}\sum_{i=1}^{n-m} g\left(W_{(i)}, c\right)^2$. By applying the same argument, we obtain $\pi_{(n-j+1)} \to \frac{m}{n^2}$ for all $j = 1, \ldots, m$. Note that the implied probability bootstrap analog T_n^* diverges to infinity as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$ when the bootstrap resample contains at least one outlier in $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$. The probability that the implied probability bootstrap resample T_n^* is free from m outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ converges to

$$p_m^* = P\left(B\left(n, \left(\frac{m}{n}\right)^2\right) = 0\right).$$

Therefore, under Assumptions 3 and 4, $100\left(1 - p_m^*\right)\%$ of resamples of $|T_n^*|$ will diverge to $+\infty$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$. In other words, the t-th bootstrap quantile Q_t^* of $|T_n^*|$ will diverge to $+\infty$ for all $t > p_m^*$. We summarize these findings in the following proposition.

Proposition 3. *Consider the setup of this section.*

(i) *Under Assumption 3, the pairs bootstrap analog $T_n^{\#} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ always contains the outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$, and the pairs bootstrap quantile $Q_t^{\#}$ from the resamples of $\left|T_n^{\#}\right|$ diverges to $+\infty$ for all $t > 1 - p_m^{\#}$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$.*

(ii) *Under Assumptions 3 and 4, the implied probability bootstrap analog $T_n^* = \sqrt{n}\left(\hat{\theta}^* - c\right)$ contains at least one outlier $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ with probability $1-p_m^*$, and the implied probability bootstrap quantile Q_t^* from the resamples of $|T_n^*|$ diverges to $+\infty$ for all $t > p_m^*$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$.*

Similar comments to Proposition 1 apply. The implied probability bootstrap provides finite quantiles for a wider range than the pairs bootstrap as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$. As the number of outliers m increases, the probability p_m^* decreases and the range where the implied probability bootstrap quantiles stay finite becomes narrower.

3.5.2 Residual bootstrap

Besides the pairs and the implied probability bootstrap methods, there is another bootstrap approach to conduct inference in the IV regression, the residual bootstrap. Let us consider the just-identified model introduced in Section 3 with the reduced form equation for the endogenous regressor X_i,

$$\begin{aligned} Y_i &= X_i\theta_0 + U_i, \\ X_i &= Z_i\pi_0 + V_i, \end{aligned}$$

where $E[Z_i U_i] = 0$. Based on Davidson and MacKinnon (2010), we discuss two kinds of residual bootstrap methods. The first method, called the unrestricted residual bootstrap, draws resamples $\{Y_i^{UR}, X_i^{UR}\}_{i=1}^n$ from

$$\begin{aligned} Y_i^{UR} &= X_i^{UR}\hat{\theta} + \hat{U}_i^{UR}, \\ X_i^{UR} &= Z_i\hat{\pi} + \hat{V}_i^{UR}, \end{aligned}$$

where $\hat{\theta} = \frac{\sum_{i=1}^n Z_i Y_i}{\sum_{i=1}^n Z_i X_i}$ is the IV estimator, $\hat{\pi} = \frac{\sum_{i=1}^n Z_i X_i}{\sum_{i=1}^n Z_i^2}$ is the OLS estimator (without intercept), and $\left(\hat{U}_i^{UR}, \hat{V}_i^{UR}\right)$ is drawn from the empirical distribution of $\left\{\hat{U}_i, \hat{V}_i\right\}_{i=1}^n$ with $\hat{U}_i = Y_i - X_i\hat{\theta}$ and $\hat{V}_i = X_i - Z_i\hat{\pi}$. To test $H_0 : \theta_0 = c$ against $H_1 : \theta_0 \neq c$, the unrestricted residual bootstrap analog of $T_n = \sqrt{n}\left(\hat{\theta} - c\right)$ is obtained as

$$T_n^{UR} = \sqrt{n}\left(\hat{\theta}^{UR} - \hat{\theta}\right) = \sqrt{n}\frac{\sum_{i=1}^n Z_i \hat{U}_i^{UR}}{\sum_{i=1}^n Z_i X_i^{UR}},$$

where $\hat{\theta}^{UR} = \frac{\sum_{i=1}^n Z_i Y_i^{UR}}{\sum_{i=1}^n Z_i X_i^{UR}}$.

Suppose that Assumption 1 holds. Since all elements of the residual vector $\left(\hat{U}_1, \ldots, \hat{U}_n\right)$ depend on $W_{(n)}$ through $\hat{\theta}$, all elements of the resampled residuals $\left(\hat{U}_1^{UR}, \ldots, \hat{U}_n^{UR}\right)$ also depend on $W_{(n)}$. Therefore, the unrestricted residual bootstrap statistic T_n^{UR} heavily depends on the outlier $W_{(n)}$. However, since the limiting behavior of T_n^{UR} is case by case, it is not easy to derive a breakdown point property of its bootstrap quantiles.

The second method, called the restricted residual bootstrap, draws resamples $\{Y_i^{RR}, X_i^{RR}\}_{i=1}^n$ from

$$\begin{aligned} Y_i^{RR} &= X_i^{RR}c + \hat{U}_i^{RR}, \\ X_i^{RR} &= Z_i\hat{\pi} + \hat{V}_i^{RR}, \end{aligned}$$

where $\left(\hat{U}_i^{RR}, \hat{V}_i^{RR}\right)$ is drawn from the empirical distribution of $\left\{\tilde{U}_i, \hat{V}_i\right\}_{i=1}^n$ with $\tilde{U}_i = Y_i - X_i c$ and $\hat{V}_i = X_i - Z_i\hat{\pi}$. The restricted residual bootstrap analog of T_n is obtained as

$$T_n^{RR} = \sqrt{n}\left(\hat{\theta}^{RR} - c\right) = \sqrt{n}\frac{\sum_{i=1}^n Z_i \hat{U}_i^{RR}}{\sum_{i=1}^n Z_i X_i^{RR}},$$

where $\hat{\theta}^{RR} = \frac{\sum_{i=1}^n Z_i Y_i^{RR}}{\sum_{i=1}^n Z_i X_i^{RR}}$.

Unlike the unrestricted residual bootstrap statistic T_n^{UR}, the restricted residual bootstrap statistic T_n^{RR} is centered around the hypothetical value c. Also, in the residual vector $\left(\tilde{U}_{(1)}, \ldots, \tilde{U}_{(n)}\right)$, only $\tilde{U}_{(n)} = Y_{(n)} - X_{(n)}c$ may be affected from the outlier $W_{(n)}$. On the other hand, all elements of the residual vector $\left(\hat{V}_1, \ldots, \hat{V}_n\right)$ may be influenced by the outlier $W_{(n)}$ through $\hat{\pi}$. It is not easy to derive a general breakdown point result when all elements of $\left(\hat{V}_1, \ldots, \hat{V}_n\right)$ diverge. To proceed, we consider the following special case, which corresponds to Case 3 in Table C.

Assumption 5. The divergence $\|W_{(n)}\| \to +\infty$ means that $|Y_{(n)}| \to +\infty$ but $|X_{(n)}|$ and $|Z_{(n)}|$ are bounded.

This assumption ensures that the residual vector $\left(\hat{V}_1, \ldots, \hat{V}_n\right)$ is not influenced by the outlier $W_{(n)}$ but the residual $\tilde{U}_{(n)}$ is. Under Assumption 5, the restricted residual bootstrap statistic T_n^{RR} diverges to infinity as $\|W_{(n)}\| \to +\infty$ only when the resample contains $\tilde{U}_{(n)}$. The probability that T_n^{RR} is free from $\tilde{U}_{(n)}$ is given by $p^{RR} = P\left(B\left(n, \frac{1}{n}\right) = 0\right)$. Therefore, $100\left(1 - p^{RR}\right)\%$ of resamples of $\left|T_n^{RR}\right|$ will diverge to $+\infty$ as $\|W_{(n)}\| \to +\infty$. In other words, the t-th bootstrap quantile Q_t^{RR} of $\left|T_n^{RR}\right|$ will diverge to $+\infty$ for all $t > p^{RR}$ as $\|W_{(n)}\| \to +\infty$. We summarize these findings in the following proposition.

Proposition 4. *Consider the setup of this section.*

(i) *Under Assumption 1, the unrestricted residual bootstrap statistic $\left|T_n^{UR}\right|$ of T_n always contains the outlier $W_{(n)}$.*

(ii) *Under Assumption 5, the restricted residual bootstrap analog $T_n^{RR} = \sqrt{n}\left(\hat{\theta}^{RR} - c\right)$ contains the outlier $W_{(n)}$ with probability $1 - p^{RR}$, and the restricted residual bootstrap quantile Q_t^{RR} from the resamples $\left|T_n^{RR}\right|$ of T_n will diverge to $+\infty$ for all $t > p^{RR}$.*

In Proposition 4, we observe that the unrestricted residual bootstrap statistic T_n^{UR} is heavily influenced by the outlier $W_{(n)}$. In contrast, under Assumption 5, the restricted residual bootstrap statistic T_n^{RR} is more robust to the outlier $W_{(n)}$. However, since $p^{RR} = p^{\#} = P\left(B\left(n, \frac{1}{n}\right) = 0\right)$ is relatively smaller than $p^* = P\left(B\left(n, \frac{1}{n^2}\right) = 0\right)$ (see Table A), the presence of a single outlier can imply divergence of a larger proportion of the resampled statistics.

We conduct a simulation study to evaluate the performance of the residual bootstrap methods in the presence of outliers. For comparison, we consider the same simulation setting in Section 3.4. Table 2 reports the Monte Carlo medians of the unrestricted and restricted residual bootstrap quantiles Q_t^{UR} and Q_t^{RR}, respectively. We set $t = 0.95$, and we report also empirical coverages of bootstrap confidence intervals. First we consider the statistic T_n. In absence of the outlier, the Monte Carlo medians of both residual bootstrap quantiles are very close to the true quantiles. Furthermore, also the empirical coverages are very close to the nominal coverage probability 0.95. For instance, in the case of $n = 100$, the Monte Carlo medians of the unrestricted and restricted residual bootstrap quantiles are 1.7638 and 1.7610, respectively, while the true quantile is 1.7713. In contrast, in the presence of the outlier, the quantiles of both bootstrap methods (and also the empirical coverages) end to be quite large. For instance, in the case of $n = 100$ and $C = 20$, the Monte Carlo medians of the unrestricted and restricted residual bootstrap quantiles are 31.3310 and 31.2815, respectively. Similar findings arise also for residual bootstrap methods applied to the statistic T_n^d. In particular, we can observe than in absence of the outlier both residual bootstrap imply bootstrap quantiles very close to the true quantiles. On the other hand, in the presence of the outlier both bootstrap quantiles and empirical coverages are extremely large and very far from the true values. These findings confirm our theoretical results. The residual bootstrap quantiles are heavily influenced by the outlier and tend to be very large.

We also analyze the size and power properties of the unrestricted and restricted residual bootstrap tests. Again we consider the same settings introduced in Section 3.4. Table 7 reports the rejection frequencies of the null hypothesis for different values of $\theta_0 \in [2, 2.3]$. Without the outlier, in the case of $\theta_0 = 2$, the rejection frequencies of both residual bootstrap tests are very close to the nominal level 0.05. The power of the bootstrap tests increases as the value of θ_0 increases to 2.3. For instance, at $\theta_0 = 2.3$, the rejection frequencies are larger than 85% for both bootstrap tests. In the presence of the outlier, the size and power of the bootstrap tests are dramatically distorted. In particular, the rejection frequencies of both residual bootstrap methods are always smaller than 10% even when $\theta_0 = 2.3$ (very low power). In the presence of the outlier, both residual bootstrap critical values tend to be very large, and both size and power of the bootstrap tests tend to collapse to 0.

3.5.3 Studentized statistic

So far we consider the nonstudentized statistic T_n. Our breakdown point analysis for the bootstrap can be extended to the studentized $t_n = \frac{\hat{\theta}-c}{\hat{\sigma}}$, where $\hat{\sigma} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2\right)\left(\sum_{i=1}^n Z_i^2\right) / \left(\sum_{i=1}^n Z_i X_i\right)^2}$ is the standard error of $\hat{\theta}$, and $\hat{U}_i = Y_i - X_i\hat{\theta}$ is the residual.[10] In this case, we modify Assumption 1 as follows.

Assumption 6. $W_{(n)}$ is an outlier for the statistic t_n in the sense that

$$|t_n| \to +\infty \text{ as } \left\|W_{(n)}\right\| \to +\infty.$$

Assumption 6 characterizes the outliers for the statistic t_n, which may be different from those for T_n defined in Assumption 1. Below we provide some specific example that satisfy this assumption.

We first consider the pairs bootstrap. Note that the pairs bootstrap statistic $t_n^{\#} = \frac{\hat{\theta}^{\#}-\hat{\theta}}{\hat{\sigma}^{\#}}$ always contains the outlier $W_{(n)}$ through $\hat{\theta}$. Thus even when the resample to compute $\hat{\theta}^{\#}$ and $\hat{\sigma}^{\#}$ does not contain the outlier $W_{(n)}$, the statistic $t_n^{\#}$ diverges if $\hat{\theta}$ diverges as $\left\|W_{(n)}\right\| \to +\infty$. If the resample to compute $\hat{\theta}^{\#}$ and $\hat{\sigma}^{\#}$ contains the outlier $W_{(n)}$, the limiting behavior of $t_n = \frac{\hat{\theta}^{\#}-\hat{\theta}}{\hat{\sigma}^{\#}}$ is case by case and may diverge or become indeterminate as $\left\|W_{(n)}\right\| \to +\infty$.

We next consider the implied probability bootstrap. Note that the implied probability bootstrap statistic $t_n^* = \frac{\hat{\theta}^*-c}{\hat{\sigma}^*}$ is centered around the hypothetical value c. Thus, $|t_n^*|$ will diverge as $\left\|W_{(n)}\right\| \to +\infty$ only when the resample contains the outlier $W_{(n)}$. Furthermore, under Assumption 2, the implied probability in (3) for the observation $W_{(n)}$ satisfies $\pi_{(n)} \to \frac{1}{n^2}$ as $\left\|W_{(n)}\right\| \to +\infty$, as established in (4). Thus, the probability that the implied probability bootstrap statistic t_n^* is free from the outlier $W_{(n)}$ converges to $p^* = P\left(B\left(n, \frac{1}{n^2}\right) = 0\right)$ as $\left\|W_{(n)}\right\| \to +\infty$. Therefore, under Assumptions 2 and 6, $100\,(1-p^*)\,\%$ of resamples of $|t_n^*|$ will diverge to $+\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. In other words, the t-th bootstrap quantile Q_t^* of $|t_n^*|$ will diverge to $+\infty$ for all $t > p^*$. We summarize these findings in the following proposition.

Proposition 5. *Consider the setup of this section.*

[10] The results in this subsection do not change even if we use the heteroskedasticity robust standard error $\hat{\sigma} = \sqrt{\left(\sum_{i=1}^n \hat{U}_i^2 Z_i^2\right) / \left(\sum_{i=1}^n Z_i X_i\right)^2}$.

(i) *Under Assumption 6, the pairs bootstrap analog* $t_n^{\#} = \frac{\hat{\theta}^{\#}-\hat{\theta}}{\hat{\sigma}^{\#}}$ *always contains the outlier* $W_{(n)}$.

(ii) *Under Assumptions 2 and 6, the implied probability bootstrap analog* $t_n^{*} = \frac{\hat{\theta}^{*}-c}{\hat{\sigma}^{*}}$ *contains the outlier* $W_{(n)}$ *with probability* $1-p^*$, *and the implied probability bootstrap quantile* Q_t^* *from the resamples of* $|t_n^*|$ *diverges to* $+\infty$ *for all* $t > p^*$ *as* $\left\|W_{(n)}\right\| \to +\infty$.

Similar comments to Proposition 1 apply here. In particular, since the pairs bootstrap statistic $t_n^{\#}$ always contains the outlier, we expect that the pairs bootstrap quantiles tend to be very large. In contrast, the implied probability bootstrap distribution is less affected from the outlier $W_{(n)}$ and its bootstrap quantiles tend to be low and stable. However, under Assumption 6, the original statistic t_n diverges as $\left\|W_{(n)}\right\| \to +\infty$, and the size and power of the implied probability bootstrap test tend to be close to 1.

We now discuss some example of anomalous observations that satisfy Assumption 6. In Table D, we consider the same cases introduced in Table C.

Table D: Limits of t_n as $\left\|W_{(n)}\right\| \to +\infty$

Case	Diverge	Bounded	Limit of $\lvert t_n \rvert$
1	Z	$X,\ Y$	bounded
2	X	$Y,\ Z$	$+\infty$
3	Y	$X,\ Z$	bounded

From this table, we can regard Case 2 as the outliers for t_n. Therefore, in the presence of the outlier in the explanatory variable, we expect that both pairs and implied probability bootstrap tests based on the statistic t_n perform poorly.

We conduct a simulation study to evaluate the performance of the bootstrap methods for t_n in the presence of outliers. We consider two situations: (i) $(W_1, \ldots, W_n)$ are generated from the data generating process introduced in Section 3.4 (No outlier), and (ii) $\left(\tilde{W}_1, \ldots, \tilde{W}_n\right)$ with $\tilde{W}_{(i)} = W_{(i)}$ for $i = 1, \ldots, n-1$ and $\tilde{W}_{(n)} = \left(\tilde{Y}_{(n)}, \tilde{X}_{(n)}, \tilde{Z}_{(n)}\right) = \left(Y_{(n)}, CX_{max}, Z_{(n)}\right)$ with $C = 5, 10, 20$ and $X_{max} = \max\{X_1, \ldots, X_n\}$ (Outlier in X). This setup for the outlier corresponds to Case 2 in Table D. Proposition 5 says that the pairs bootstrap statistic $t_n^{\#}$ always contains the outlier, while the implied probability bootstrap statistic t_n^* contains the outlier with probability $1-p^*$ as $\left\|W_{(n)}\right\| \to +\infty$.

Table 3 reports the Monte Carlo medians of the pairs and implied probability bootstrap quantiles $Q_t^{\#}$ and Q_t^*, respectively. We set $t = 0.95$, and we report also empirical coverages of bootstrap confidence intervals. In absence of the outlier, the Monte Carlo medians of both bootstrap methods are very close to the true quantiles. Furthermore, also the empirical coverages are very close to the nominal coverage probability 0.95 For instance, in the case of $n = 50$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 1.9466 and 1.9373, respectively, while the true quantile is 1.9664. In contrast, in the presence of the outlier, the pairs bootstrap quantiles tend to be large, while the implied probability bootstrap quantiles tend to be close to the true quantiles without the outlier. For instance, in the case of $n = 50$ and $C = 20$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 10.5556 and 2.3659, respectively. As expected , the pairs bootstrap empirical coverages are very close to 1, while the implied probability bootstrap empirical coverages are extremely small. These findings confirm the theoretical results in Proposition 5.

4 Over-identified case

4.1 Setup

We now extend our breakdown point analysis to an over-identified case with $k = 2$ and $p = 1$. Extensions to high dimensional cases will be briefly discussed in Section 5. In this case, the moment conditions (2) are written as

$$E\left[g\left(W_i,\theta_0\right)\right] = E\left[\begin{array}{c} g_1\left(W_i,\theta_0\right) \\ g_2\left(W_i,\theta_0\right) \end{array}\right] = E\left[\begin{array}{c} Z_{1i}\left(Y_i - X_i\theta_0\right) \\ Z_{2i}\left(Y_i - X_i\theta_0\right) \end{array}\right] = 0.$$

where $W_i = (Y_i, X_i, Z_i')' \in \mathbb{R}^4$ and $Z_i = (Z_{1i}, Z_{2i})'$. Also in this setting, we assume that all variables have finite variance. Similar to the just-identified case, we consider the parameter hypothesis testing problem $H_0 : \theta_0 = c$ against the two-sided alternative $H_1 : \theta_0 \neq c$ using the statistics $T_{n,over} = \sqrt{n}\left(\hat{\theta} - c\right)$ based on a conventional estimator $\hat{\theta}$ and $T^d_{n,over} = \sqrt{n}\left(\hat{\theta}^d - c\right)$ based on the GMTM estimator. The point estimator $\hat{\theta}$ for θ_0 is either the two-stage least square estimator

$$\hat{\theta}_{2SLS} = \left[\left(\sum_{i=1}^{n} X_i Z_i\right)'\left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^{n} X_i Z_i\right)\right]^{-1}\left(\sum_{i=1}^{n} X_i Z_i\right)'\left(\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^{n} Z_i Y_i\right),$$

or the two-step GMM estimator

$$\hat{\theta}_{GMM} = \left[\left(\sum_{i=1}^{n} X_i Z_i\right)'\left(\sum_{i=1}^{n} \hat{U}_i^2 Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^{n} X_i Z_i\right)\right]^{-1}\left(\sum_{i=1}^{n} X_i Z_i\right)'\left(\sum_{i=1}^{n} \hat{U}_i^2 Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^{n} Z_i Y_i\right),$$

where $\hat{U}_i = Y_i - X_i\hat{\theta}_{2SLS}$ is the residual from the first-step estimation using $\hat{\theta}_{2SLS}$. Since both $\hat{\theta}_{2SLS}$ and $\hat{\theta}_{GMM}$ have the same breakdown point properties, we denote $\hat{\theta}$ for either $\hat{\theta}_{2SLS}$ or $\hat{\theta}_{GMM}$ in this section. For the over-identified case, we consider the following version of the GMTM estimator

$$\hat{\theta}^d = \arg\min_{\theta}\left\|\frac{1}{n}\sum_{i=1}^{n} Z_i\left(Y_i - X_i\theta\right)\mathbb{I}\left\{r\left(W_i,\theta\right) \leq r\left(W_{[n-d]},\theta\right)\right\}\right\|,$$

where $r\left(W_i,\theta\right) = \left\|Z_i\left(Y_i - X_i\theta\right)\right\|$ is a trimming function ordered as $\left\|r\left(W_{[1]},\theta\right)\right\| \leq \cdots \leq \left\|r\left(W_{[n]},\theta\right)\right\|$, and d is an integer such that $0 \leq d < \frac{n}{2}$ to determine the amount of trimming.

Similar to the last section, we compare the breakdown point properties of the pairs and implied probability bootstrap quantiles. In this case, the moment function g is a vector. Thus Back and

Brown's (1993) implied probability for the observation W_i from the moment condition $E\left[g\left(W_i,c\right)\right]=0$ is written as

$$\pi_i=\frac{1}{n}-\frac{1}{n}\left(g\left(W_i,c\right)-\bar{g}\right)'\left[\frac{1}{n}\sum_{i=1}^{n}g\left(W_i,c\right)g\left(W_i,c\right)'\right]^{-1}\bar{g}. \tag{5}$$

Although the implied probability takes a more complicated form than the just-identified case, we can still apply the same breakdown point analysis to this setting.

4.2 Breakdown point analysis

Based on the above setup, we now conduct the breakdown point analysis for the bootstrap. To define outliers, Assumption 1 is modified as follows.

Assumption 1'. $W_{(n)}$ is an outlier for the statistic $T_{n,over}$ in the sense that

$$|T_{n,over}|\to+\infty \text{ as } \left\|W_{(n)}\right\|\to+\infty.$$

We first consider the pairs bootstrap. As in Section 3.2, the pairs bootstrap statistic $T^{\#}_{n,over}=\sqrt{n}\left(\hat{\theta}^{\#}-\hat{\theta}\right)$ always contains the outlier $W_{(n)}$ through $\hat{\theta}$. Even if the pairs bootstrap resample to compute $\hat{\theta}^{\#}$ does not contain the outlier, $T^{\#}_{n,over}$ diverges to infinity. If the resample contains the outlier $W_{(n)}$ possibly multiple times, then $\left|T^{\#}_{n,over}\right|$ may diverge or become indeterminate as $\left\|W_{(n)}\right\|\to+\infty$. Therefore, in this case we can at least say that $\left|T^{\#}_{n,over}\right|$ diverges to infinity as $\left\|W_{(n)}\right\|\to+\infty$ when the resample to compute $\hat{\theta}^{\#}$ does not contain $W_{(n)}$. The probability for this event is obtained as $p^{\#}=P\left(B\left(n,\frac{1}{n}\right)=0\right)$.

We next consider the implied probability bootstrap. In this case, we impose the following additional assumption.

Assumption 2'. Let $g_1\left(W_{(n)},c\right)=Z_{1(n)}\left(Y_{(n)}-X_{(n)}c\right)$ and $g_2\left(W_{(n)},c\right)=Z_{2(n)}\left(Y_{(n)}-X_{(n)}c\right)$. Assume that

$$\left|g_1\left(W_{(n)},c\right)\right|\to+\infty \text{ as } \left\|W_{(n)}\right\|\to+\infty,$$
$$\left|g_2\left(W_{(n)},c\right)\right|\to+\infty \text{ as } \left\|W_{(n)}\right\|\to+\infty.$$

Similar to Assumption 2, Assumption 2' is very mild. Using the results in Camponovo and Otsu (2012), we can show that under Assumption 2', the implied probability in (5) for the observation $W_{(n)}$ satisfies

$$\pi_{(n)}\to\frac{1}{n^2}+\frac{1}{n}\frac{\left(\bar{g}_{1-}-\bar{g}_{2-}\right)^2}{v_{11}+v_{22}-2v_{12}}, \tag{6}$$

as $\left\|W_{(n)}\right\|\to+\infty$, where $\bar{g}_{1-}=\frac{1}{n}\sum_{i=1}^{n-1}g_{1(i)}$, $\bar{g}_{2-}=\frac{1}{n}\sum_{i=1}^{n-1}g_{2(i)}$, $v_{11}=\frac{1}{n}\sum_{i=1}^{n-1}g^2_{1(i)}$, $v_{22}=\frac{1}{n}\sum_{i=1}^{n-1}g^2_{2(i)}$, and $v_{12}=\frac{1}{n}\sum_{t=1}^{n-1}g_{1(i)}g_{2(i)}$.[11] Unlike the just-identified case, the limit of the implied probability $\pi_{(n)}$

[11]The technical details for the derivation of the limit of $\pi_{(n)}$ are provided in equation (9) in Camponovo and Otsu (2012).

depends on the terms $\bar{g}_{1-}$, $\bar{g}_{2-}$, v_{11}, v_{22}, and v_{12}. Therefore, the implied probability bootstrap does not necessarily draw the outlier with probability smaller than $\frac{1}{n}$. Nevertheless, it should be noted that the terms $\bar{g}_{1-}$, $\bar{g}_{2-}$, v_{11}, v_{22}, and v_{12} do not contain the outlier $W_{(n)}$ and thus the second term appearing in the limit (6) tends to be small when the sample size n is large. Also we can empirically evaluate the second term in (6) and assess the difference with the uniform weight $\frac{1}{n}$.

Note that the implied probability bootstrap statistic $T^*_{n,over} = \sqrt{n}\left(\hat{\theta}^* - c\right)$ diverges to infinity only when the resample to compute $\hat{\theta}^*$ contains the outlier $W_{(n)}$. From (6), the probability that the implied probability bootstrap statistic $T^*_{n,over}$ is free from the outlier $W_{(n)}$ converges to

$$p^*_{over} = P\left(B\left(n, \frac{1}{n^2} + \frac{1}{n}\frac{(\bar{g}_{1-} - \bar{g}_{2-})^2}{v_{11} + v_{22} - 2v_{12}}\right) = 0\right),$$

as $\left\|W_{(n)}\right\| \to +\infty$. Therefore, under Assumptions 1' and 2', $100\left(1 - p^*_{over}\right)\%$ of resamples of $\left|T^*_{n,over}\right|$ will diverge to $+\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. In other words, the t-th bootstrap quantile Q^*_t of $\left|T^*_{n,over}\right|$ will diverge to $+\infty$ for all $t > p^*_{over}$. We summarize these findings in the following proposition.

Proposition 6. *Consider the setup of this section.*

(i) *Under Assumption 1', the pairs bootstrap analog* $T^{\#}_{n,over} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ *always contains the outlier* $W_{(n)}$*, and the pairs bootstrap quantile* $Q^{\#}_t$ *from the resamples of* $\left|T^{\#}_{n,over}\right|$ *diverges to* $+\infty$ *for all* $t > 1 - p^{\#}$ *as* $\left\|W_{(n)}\right\| \to +\infty$.

(ii) *Under Assumptions 1' and 2', the implied probability bootstrap analog* $T^*_{n,over} = \sqrt{n}\left(\hat{\theta}^* - c\right)$ *contains the outlier* $W_{(n)}$ *with probability* $1 - p^*_{over}$*, and the implied probability bootstrap quantile* Q^*_t *from the resamples of* $\left|T^*_{n,over}\right|$ *diverges to* $+\infty$ *for all* $t > p^*_{over}$ *as* $\left\|W_{(n)}\right\| \to +\infty$.

We now consider the statistic $T^d_{n,over} = \sqrt{n}\left(\hat{\theta}^d - c\right)$ with $d \geq 1$ based on the GMTM estimator. Under Assumption 2', it holds $r\left(W_{(n)}, c\right) = \left\|Z_{(n)}\left(Y_{(n)} - X_{(n)}c\right)\right\| \to +\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. Thus, the outlier $W_{(n)}$ will be trimmed and $\hat{\theta}^d$ and $T^d_{n,over}$ are bounded as $\left\|W_{(n)}\right\| \to +\infty$. On the other hand, the bootstrap counterparts $T^{d\#}_{n,over} = \sqrt{n}\left(\hat{\theta}^{d\#} - \hat{\theta}^d\right)$ and $T^{d*}_{n,over} = \sqrt{n}\left(\hat{\theta}^{d*} - c\right)$ diverge if the resample contains the outlier $W_{(n)}$ more than d times. The probability that the pairs bootstrap resample to compute $\hat{\theta}^{d\#}$ contains the outlier $W_{(n)}$ less than or equal to d times is

$$p^{d\#} = P\left(B\left(n, \frac{1}{n}\right) \leq d\right).$$

Also, from (6), the probability that the implied probability bootstrap resample to compute $\hat{\theta}^{d*}$ contains the outlier $W_{(n)}$ less than or equal to d times converges to

$$p^{d*}_{over} = P\left(B\left(n, \frac{1}{n^2} + \frac{1}{n}\frac{(\bar{g}_{1-} - \bar{g}_{2-})^2}{v_{11} + v_{22} - 2v_{12}}\right) \leq d\right),$$

as $\left\|W_{(n)}\right\| \to +\infty$. Thus, under Assumptions 1' and 2', the t-th bootstrap quantile $Q^{d\#}_t$ of $\left|T^{d\#}_{n,over}\right|$ will diverge to $+\infty$ for all $t > p^{d\#}$, as $\left\|W_{(n)}\right\| \to +\infty$. Moreover, under Assumptions 1' and 2' the

t-th bootstrap quantile Q_t^{d*} of $\left|T_{n,over}^{d*}\right|$ will diverge to $+\infty$ for all $t > p_{over}^{d*}$, as $\left\|W_{(n)}\right\| \to +\infty$. These findings are summarized as follows.

Proposition 7. *Consider the setup of this section.*

(i) *Under Assumptions 1' and 2', the pairs bootstrap analog $T_{n,over}^{d\#} = \sqrt{n}\left(\hat{\theta}^{d\#} - \hat{\theta}^{d}\right)$ contains the outlier $W_{(n)}$ with probability $1 - p^{\#}$, and the pairs bootstrap quantile $Q_t^{d\#}$ from the resamples of $\left|T_{n,over}^{d\#}\right|$ diverges to $+\infty$ for all $t > p^{d\#}$ as $\left\|W_{(n)}\right\| \to +\infty$.*

(ii) *Under Assumptions 1' and 2', the implied probability bootstrap analog $T_{n,over}^{d*} = \sqrt{n}\left(\hat{\theta}^{d*} - c\right)$ contains the outlier $W_{(n)}$ with probability $1 - p_{over}^{d*}$, and the implied probability bootstrap quantile Q_t^{d*} from the resamples of $\left|T_{n,over}^{d*}\right|$ diverges to $+\infty$ for all $t > p_{over}^{d*}$ as $\left\|W_{(n)}\right\| \to +\infty$.*

Similar comments to Proposition 2 apply here. As the number of trimmed observations d increases, both $p^{d\#}$ and p_{over}^{d*} increase. However, for the over-identified case, it is not clear whether $p_{over}^{d*} > p^{d\#}$ or not. If $p_{over}^{d*} > p^{d\#}$, then the implied probability bootstrap quantile Q_t^{d*} becomes more robust than the pairs bootstrap quantile $Q_t^{d\#}$.

Also, similar implications on the size and power properties of the bootstrap tests apply to the over-identified case. For example, suppose that $\frac{1}{n^2} + \frac{1}{n}\frac{(\bar{g}_{1-} - \bar{g}_{2-})^2}{v_{11} + v_{22} - 2v_{12}} < \frac{1}{n}$. Then, in the presence of the outlier, the pairs bootstrap quantiles tend to be quite large. In contrast, the implied probability bootstrap quantiles are relatively small and stable. Since the statistic T_n^d never contains $W_{(n)}$, we can expect that the implied probability bootstrap provides more accurate approximations to the distribution of T_n^d and shows desirable size and power properties. These implications are confirmed by a simulation study in the next subsection.

Finally, we can discuss some example of anomalous observations that satisfy Assumption 1'. In the case of $k = 2$ and $p = 1$, $W_{(n)}$ contains four elements $\left(Y_{(n)}, X_{(n)}, Z_{1(n)}, Z_{2(n)}\right)$. For illustration, we consider the following cases.

Table E: Limits of $T_{n,over}$ as $\left\|W_{(n)}\right\| \to +\infty$

Case	Diverge	Bounded	Limit of $\lvert T_{n,over}\rvert$
1	Z_1	$Z_2,\ X,\ Y$	bounded
2	Z_2	$Z_1,\ X,\ Y$	bounded
3	X	$Z_1,\ Z_2,\ Y$	bounded
4	Y	$Z_1,\ Z_2,\ X$	$+\infty$

From this table, we can regard Case 4 as the outliers for T_n. Therefore, in the presence of outliers in the dependent variable, we recommend to use the implied probability bootstrap using the statistic T_n^d.

4.3 Simulation

We conduct a simulation study to illustrate the theoretical findings in the last subsection. Consider iid samples $\{W_i\}_{i=1}^n = \{Y_i, X_i, Z_i\}_{i=1}^n$ of sizes $n = 50$ and 100 generated from $Y_i = X_i\theta_0 + U_i$ and $X_i = Z_i'\pi_0 + V_i$, where $Z_i = \begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$, $\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right)$, and $\pi_0 = (0.8, 0.6)'$. The true parameter value is set as $\theta_0 = 2$. We are interested in testing the null hypothesis $H_0 : \theta_0 = 2$ against the alternative $H_1 : \theta_0 \neq 2$. For each scenario, the number of bootstrap replications is 399 for each Monte Carlo sample and the number of Monte Carlo replications is 10,000.

To study robustness of the bootstrap methods for the test statistics $T_{n,over}$ and $T_{n,over}^d$ with $d = 1$, we consider two situations: (i) $(W_1, \ldots, W_n)$ are generated from the above model (No outlier), and (ii) $\left(\tilde{W}_1, \ldots, \tilde{W}_n\right)$ with $\tilde{W}_{(i)} = W_{(i)}$ for $i = 1, \ldots, n-1$ and $\tilde{W}_{(n)} = \left(\tilde{Y}_{(n)}, \tilde{X}_{(n)}, \tilde{Z}_{1(n)}, \tilde{Z}_{2(n)}\right) = \left(CY_{max}, X_{(n)}, Z_{1(n)}, Z_{2(n)}\right)$ with $C = 5, 10, 20$ and $Y_{max} = \max\{Y_1, \ldots, Y_n\}$ (Outlier in Y). This specification for the outlier corresponds to Case 4 in Table E. Proposition 6 says that as $\|W_{(n)}\| \to \infty$, the pairs bootstrap t-th quantile $Q_t^{\#}$ from the resamples $\left|T_{n,over}^{\#}\right|$ of $T_{n,over}$ will diverge to $+\infty$ for all $t > 1 - P\left(B\left(n, \frac{1}{n}\right) = 0\right)$, and the implied probability bootstrap t-th quantile Q_t^* from the resamples $\left|T_{n,over}^*\right|$ of $T_{n,over}$ will diverge to $+\infty$ for all $t > P\left(B\left(n, \frac{1}{n^2} + \frac{1}{n}\frac{(\bar{g}_{1-} - \bar{g}_{2-})^2}{v_{11} + v_{22} - 2v_{12}}\right) = 0\right)$. Also by Proposition 7, the pairs bootstrap t-th quantile from the resamples $\left|T_{n,over}^{d\#}\right|$ of $T_{n,over}^d$ will diverge to $+\infty$ for all $t > P\left(B\left(n, \frac{1}{n}\right) \leq 1\right)$, and the implied probability bootstrap t-th quantile from the resamples $\left|T_{n,over}^{d*}\right|$ of $T_{n,over}^d$ will diverge to $+\infty$ for all $t > P\left(B\left(n, \frac{1}{n^2} + \frac{1}{n}\frac{(\bar{g}_{1-} - \bar{g}_{2-})^2}{v_{11} + v_{22} - 2v_{12}}\right) \leq 1\right)$.

Table 4 reports the Monte Carlo medians of the pairs bootstrap quantiles $Q_t^{\#}$ and $Q_t^{d\#}$ and implied probability bootstrap quantiles Q_t^* and Q_t^{d*} for $|T_{n,over}|$ and $\left|T_{n,over}^d\right|$, respectively. We set $t = 0.95$, and we report also empirical coverages of bootstrap confidence intervals. First we consider the statistic $|T_{n,over}|$. In absence of the outlier, the Monte Carlo medians of both bootstrap quantiles are very close to the true quantiles. For instance, in the case of $n = 100$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 1.5411 and 1.5207, respectively, while the true quantile is 1.5488. Also the empirical coverages are very close to the nominal coverage probability 0.95. In contrast, in the presence of the outlier, the pairs bootstrap quantiles are quite large, while the implied probability bootstrap quantiles tend to be close to the true quantiles without the outlier. For instance, in the case of $n = 100$ and $C = 20$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 21.7734 and 1.9692, respectively. These results suggest that in this setting the term $\frac{1}{n^2} + \frac{1}{n}\frac{(\bar{g}_{1-} - \bar{g}_{2-})^2}{v_{11} + v_{22} - 2v_{12}}$ is quite small, and the implied probability bootstrap quantiles are less influenced from the outlier than the pairs bootstrap quantiles. Finally, as expected the pairs bootstrap empirical coverages are close to 1, while the implied probability bootstrap coverages are extremely small.

We next consider the statistic $\left|T_{n,over}^d\right|$. In absence of the outlier, the Monte Carlo medians of both pairs and implied probability bootstrap quantiles are very close to the true quantiles. For instance, in the case of $n = 100$, the Monte Carlo medians of the pairs and implied probability bootstrap quantiles are 1.5292 and 1.4921, respectively, while the true quantile is 1.5031. In contrast, in the presence of the outlier, the pairs bootstrap quantiles (and also the empirical coverages) tend to be very large. For instance, in the case of $n = 100$ and $C = 20$, the Monte Carlo median of the pairs bootstrap quantile is 21.2601. On the other hand, since the implied probability bootstrap is less affected from the outlier, the implied probability bootstrap quantiles (and also the empirical coverages) remain close to the true quantiles. For instance, in the case of $n = 100$ and $C = 20$, the Monte Carlo median of the implied probability bootstrap quantile is 1.5201, while the true quantile is 1.5524.

We also study the size and the power properties of the bootstrap tests. In particular, we consider two situations: (i) $(W_1, \ldots, W_n)$ are generated from the above model with the parameter value $\theta_0 \in [2, 2.3]$ (No outlier), and (ii) $\left(\tilde{W}_1, \ldots, \tilde{W}_n\right)$ with $\tilde{W}_{(i)} = W_{(i)}$ for $i = 1, \ldots, 99$ and $\tilde{W}_{(100)} = \left(\tilde{Y}_{(100)}, \tilde{X}_{(100)}, \tilde{Z}_{1(100)}, \tilde{Z}_{2(100)}\right) = \left(10Y_{max}, X_{(100)}, Z_{1(100)}, Z_{2(100)}\right)$ with $Y_{max} = \max\{Y_1, \ldots, Y_{100}\}$ (Outlier in Y). We test the null hypothesis $H_0 : \theta_0 = 2$ under different parameter values of $\theta_0 \in [2, 2.3]$.

Table 8 reports the rejection frequencies of the null hypothesis under different parameter values $\theta_0 \in [2, 2.3]$. First we consider the statistic $T_{n,over}$. In absence of the outlier, for $\theta_0 = 2$, the rejection frequencies of both bootstrap tests are quite close to the nominal level 0.05. The power of the bootstrap tests increases as the value of θ_0 increases to 2.3. For instance, at $\theta_0 = 2.3$, the rejection frequencies are larger than 85% for both bootstrap tests. In the presence of the outlier, both size and power of the tests are dramatically distorted. In particular, the rejection frequencies of the pairs bootstrap test are always smaller than 15% even when $\theta_0 = 2.3$ (very low power). In contrast, the rejection frequencies of the implied probability bootstrap test are always larger than 65% even when $\theta_0 = 2$ (severe size distortion). These are similar to the results obtained in Section 3.4 for the just-identified case. Indeed, in the presence of the outlier, both size and power of the pairs bootstrap test tend to collapse to 0. On the other hand, in the presence of the outlier, both size and power of the implied probability bootstrap test tend to collapse to 1.

Finally, we consider the statistic $T^d_{n,over}$. In absence of the outlier, the rejection frequencies are very similar to those obtained for the statistic $T_{n,over}$. In particular, when $\theta_0 = 2$, the rejection frequencies are very close to the nominal level 0.05. Furthermore, for $\theta_0 \neq 2$, the rejection frequencies increase (as θ_0 increases to 2.3) and are larger than 85% when $\theta_0 = 2.3$. In the presence of the outlier, we observe that both size and power of the pairs bootstrap test are again dramatically distorted. In particular, the rejection frequencies of the pairs bootstrap test are always smaller than 15% even when $\theta_0 = 2.3$ (very low power). In contrast, the presence of the outlier does not deteriorate the accuracy of the implied probability bootstrap. For $\theta_0 = 2$, the rejection frequency of the implied probability bootstrap is very close to the nominal level 0.05. For $\theta_0 = 2.3$, the rejection frequency is larger than 85%. Therefore, similar to the just-identified case, we conclude that the implied probability bootstrap test using the statistic T^d_n has desirable size and power properties in the presence of the outlier.

4.4 Extensions

4.4.1 Multiple outliers

Proposition 6 may be extended to the case where we have $m \in \{1, \ldots, n-1\}$ outliers. To this end, we modify Assumption 3 as follows.

Assumption 3'. $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ are outliers for the statistic $T_{n,over}$ in the sense that for each $j = 1, \ldots, m$,

$$|T_{n,over}| \to +\infty \text{ as } \left\|W_{(n-j+1)}\right\| \to +\infty.$$

The pairs bootstrap statistic $T^{\#}_{n,over} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ always contains the outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ through $\hat{\theta}$. Thus, by Assumption 3', $\left|T^{\#}_{n,over}\right|$ diverges to infinity as $\left\|W_{(n-j+1)}\right\| \to +\infty$ even if the pairs bootstrap resample to compute $\hat{\theta}^{\#}$ does not contain the outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$. If the resample contains the outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ possibly multiple times, then $\left|T^{\#}_{n,over}\right|$ may diverge or become indeterminate as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$. Thus, similar to the single outlier case, we can at least say that $\left|T^{\#}_{n,over}\right|$ diverges to infinity as $\left\|W_{(n-j+1)}\right\| \to +\infty$ when the resample to compute $\hat{\theta}^{\#}$ does not contain any outlier in $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$. The probability for this event is obtained as

$$p^{\#}_{m} = P\left(B\left(n, \frac{m}{n}\right) = 0\right).$$

Therefore, at least $100 p^{\#}_{m}\%$ of resamples of $\left|T^{\#}_{n,over}\right|$ will diverge to $+\infty$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$, $j = 1, \ldots, m$. In other words, the t-th bootstrap quantile $Q^{\#}_{t}$ of $\left|T^{\#}_{n,over}\right|$ will diverge to $+\infty$ for all $t > 1 - p^{\#}_{m}$.

For the implied probability bootstrap, we impose the following additional assumption.

Assumption 4'. For $j = 1, \ldots, m$ and $l = 1, 2$,

$$\begin{aligned} \left|g_l\left(W_{(n)}, c\right)\right| &\to +\infty \text{ as } \left\|W_{(n)}\right\| \to +\infty, \\ \frac{g_l\left(W_{(n-j+1)}, c\right)}{g_l\left(W_{(n)}, c\right)} &\to 1 \text{ as } \left\|W_{(n)}\right\| \to +\infty \text{ and } \left\|W_{(n-j+1)}\right\| \to +\infty. \end{aligned}$$

Using the results in Camponovo and Otsu (2012), under Assumption 4', we can show that the implied probability in (5) for the observation $W_{(n)}$ satisfies

$$\pi_{(n)} \to \frac{m}{n^2} + \frac{1}{n} \frac{\left(\bar{g}_{1-,m} - \bar{g}_{2-,k}\right)^2}{v_{11,m} + v_{22,m} - 2v_{12,m}},$$

as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for all $j = 1, \ldots, m$, where $\bar{g}_{1-,m} = \frac{1}{n}\sum_{i=1}^{n-m} g_{1(i)}$, $\bar{g}_{2-,m} = \frac{1}{n}\sum_{i=1}^{n-m} g_{2(i)}$, $v_{11,m} = \frac{1}{n}\sum_{t=1}^{n-m} g^2_{1(m)}$, $v_{22,m} = \frac{1}{n}\sum_{i=1}^{n-m} g^2_{2(i)}$, and $v_{12,m} = \frac{1}{n}\sum_{i=1}^{n-m} g_{1(i)} g_{2(i)}$. By applying the same argument, we obtain $\pi_{(n-j+1)} \to \frac{m}{n^2} + \frac{1}{n}\frac{\left(\bar{g}_{1-,m} - \bar{g}_{2-,k}\right)^2}{v_{11,m} + v_{22,m} - 2v_{12,m}}$, for all $j = 1, \ldots, m$. Note that the implied probability bootstrap statistic $T^{*}_{n,over}$ diverges to infinity when the resample contains at least one outlier from $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$. The probability that the implied probability bootstrap resample $T^{*}_{n,over}$ of $T_{n,over}$ is free from m outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ converges to

$$p^{*}_{m,over} = P\left(B\left(n, \left(\frac{m}{n}\right)^2 + \frac{m}{n}\frac{\left(\bar{g}_{1-,m} - \bar{g}_{2-,m}\right)^2}{v_{11,m} + v_{22,m} - 2v_{12,m}}\right) = 0\right),$$

as $\left\|W_{(n-j+1)}\right\| \to +\infty$ for $j = 1, \ldots, m$. Therefore, under Assumptions 3' and 4', $100\left(1 - p^*_{m,over}\right)\%$ of resamples of $\left|T^*_{n,over}\right|$ will diverge to $+\infty$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$, $j = 1, \ldots, m$. In other words, the t-th bootstrap quantile Q^*_t of $\left|T^*_{n,over}\right|$ will diverge to $+\infty$ for all $t > p^*_{m,over}$. We summarize these findings in the following proposition.

Proposition 8. *Consider the setup of this section.*

(i) *Under Assumption 3', the pairs bootstrap analog $T^{\#}_{n,over} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ always contains the outliers $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$, and the pairs bootstrap quantile $Q^{\#}_t$ from the resamples of $\left|T^{\#}_{n,over}\right|$ diverges to $+\infty$ for all $t > 1 - p^{\#}_m$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$, $j = 1, \ldots, m$.*

(ii) *Under Assumptions 3' and 4', the implied probability bootstrap analog $T^*_{n,over} = \sqrt{n}\left(\hat{\theta}^* - c\right)$ contains the outlier $\left(W_{(n-m+1)}, \ldots, W_{(n)}\right)$ with probability $1 - p^*_{m,over}$, and the implied probability bootstrap quantile Q^*_t from the resamples of $\left|T^*_{n,over}\right|$ diverges to $+\infty$ for all $t > p^*_{m,over}$ as $\left\|W_{(n-j+1)}\right\| \to +\infty$, $j = 1, \ldots, m$.*

Similar comments to Proposition 3 apply. The implied probability bootstrap provides finite quantiles for a wider range than the pairs bootstrap if $\left(\frac{m}{n}\right)^2 + \frac{m}{n}\frac{(\bar{g}_{1-,m} - \bar{g}_{2-,m})^2}{v_{11,m} + v_{22,m} - 2v_{12,m}} < \frac{m}{n}$. As the number of outliers m increases, the probability $p^*_{m,over}$ decreases and the range where the implied probability bootstrap quantiles stay finite becomes narrower.

4.4.2 Studentized statistic

Our breakdown point analysis for the bootstrap can be extended to the studentized statistic $t_{n,over} = \frac{\hat{\theta} - c}{\hat{\sigma}}$, where $\hat{\sigma}$ is the standard error of $\hat{\theta}$.[12] The outlier in this context is defined as follows.

Assumption 5'. $W_{(n)}$ is an outlier for the statistic $t_{n,over}$ in the sense that

$$|t_{n,over}| \to +\infty \text{ as } \left\|W_{(n)}\right\| \to +\infty.$$

Let us consider the pairs bootstrap. Note that the pairs bootstrap statistic $t^{\#}_{n,over} = \frac{\hat{\theta}^{\#} - \hat{\theta}}{\hat{\sigma}^{\#}}$ always contains the outlier $W_{(n)}$ through $\hat{\theta}$. Similar to the studentized statistic t_n analyzed in Section 3.5.3, even when the resample to compute $\hat{\theta}^{\#}$ and $\hat{\sigma}^{\#}$ does not contain the outlier, the statistic $t^{\#}_{n,over}$ diverges if $\hat{\theta}$ diverges as $\left\|W_{(n)}\right\| \to +\infty$. If the resample to compute $\hat{\theta}^{\#}$ and $\hat{\sigma}^{\#}$ contains the outlier $W_{(n)}$, the limiting behavior of $t^{\#}_{n,over}$ is case by case and may diverge or become indeterminate as $\left\|W_{(n)}\right\| \to +\infty$.

We now consider the implied probability bootstrap. Assumption 2' guarantees that the implied probability in (5) for the observation $W_{(n)}$ satisfies $\pi_{(n)} \to \frac{1}{n^2} + \frac{1}{n}\frac{(\bar{g}_{1-} - \bar{g}_{2-})^2}{v_{11} + v_{22} - 2v_{12}}$ as $\left\|W_{(n)}\right\| \to +\infty$. In

[12] More precisely, for the two-stage least square estimator $\hat{\theta}_{2SLS}$ and the two-step GMM estimator $\hat{\theta}_{GMM}$ we consider the standard error $\hat{\sigma}_{2SLS} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^n \hat{U}_i^2\right)\left[\left(\sum_{i=1}^n X_i Z_i\right)'\left(\sum_{i=1}^n Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^n X_i Z_i\right)\right]^{-1}}$ and $\hat{\sigma}_{GMM} = \sqrt{\left[\left(\sum_{i=1}^n X_i Z_i\right)'\left(\sum_{i=1}^n \tilde{U}_i^2 Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^n X_i Z_i\right)\right]^{-1}}$, respectively.

this case, the statistic $\left|t^*_{n,over}\right|$ diverges as $\left\|W_{(n)}\right\| \to +\infty$ only when the resample to compute $\hat{\theta}^*$ and $\hat{\sigma}^*$ contains the outlier $W_{(n)}$. From (6), the probability that the implied probability bootstrap statistic $t^*_{n,over}$ is free from the outlier $W_{(n)}$ converges to

$$p^*_{over} = P\left(B\left(n, \frac{1}{n^2} + \frac{1}{n}\frac{\left(\bar{g}_{1-} - \bar{g}_{2-}\right)^2}{v_{11} + v_{22} - 2v_{12}}\right) = 0\right),$$

as $\left\|W_{(n)}\right\| \to +\infty$. Therefore, under Assumptions 2' and 5', $100\left(1 - p^*_{over}\right)\%$ of resamples of $\left|t^*_{n,over}\right|$ will diverge to $+\infty$ as $\left\|W_{(n)}\right\| \to +\infty$. In other words, the t-th bootstrap quantile Q^*_t of $\left|t^*_{n,over}\right|$ will diverge to $+\infty$ for all $t > p^*_{over}$. We summarize these findings in the following proposition.

Proposition 9. *Consider the setup of this section.*

(i) *Under Assumption 5', the pairs bootstrap analog $t^{\#}_{n,over} = \sqrt{n}\left(\hat{\theta}^{\#} - \hat{\theta}\right)$ always contains the outlier $W_{(n)}$.*

(ii) *Under Assumptions 2' and 5', the implied probability bootstrap analog $t^*_{n,over} = \sqrt{n}\left(\hat{\theta}^* - c\right)$ contains the outlier $W_{(n)}$ with probability $1 - p^*_{over}$, and the implied probability bootstrap quantile Q^*_t from the resamples of $\left|t^*_{n,over}\right|$ diverges to $+\infty$ for all $t > p^*_{over}$ as $\left\|W_{(n)}\right\| \to +\infty$.*

We discuss some example of anomalous observations that satisfy Assumption 5'. In Table F, we consider the same cases introduced in Table E.

Table F: Limits of $t_{n,over}$ as $\left\|W_{(n)}\right\| \to +\infty$

Case	Diverge	Bounded	Limit of $\lvert t_{n,over}\rvert$
1	Z_1	$Z_2,\ X,\ Y$	bounded
2	Z_2	$Z_1,\ X,\ Y$	bounded
3	X	$Z_1,\ Z_2,\ Y$	$+\infty$
4	Y	$Z_1,\ Z_2,\ X$	bounded

From this table, we can regard Case 3 as an example of the outlier for $t_{n,over}$. In the presence of an outlier in the explanatory variable, we expect that both pairs and implied probability bootstrap tests based on the statistic $t_{n,over}$ perform poorly.

We conduct a simulation study to investigate robustness of the bootstrap methods for the test statistic $t_{n,over}$. We consider two situations: (i) $(W_1, \ldots, W_n)$ are generated using the data generating mechanism introduced in Section 4.3 (No outlier), and (ii) $\left(\tilde{W}_1, \ldots, \tilde{W}_n\right)$ with $\tilde{W}_{(i)} = W_{(i)}$ for $i = 1, \ldots, n-1$ and $\tilde{W}_{(n)} = \left(\tilde{Y}_{(n)}, \tilde{X}_{(n)}, \tilde{Z}_{1(n)}, \tilde{Z}_{2(n)}\right) = \left(Y_{max}, CX_{(n)}, Z_{1(n)}, Z_{2(n)}\right)$ with $C = 5, 10, 20$ and $X_{max} = \max\{X_1, \ldots, X_n\}$ (Outlier in X). This specification of the outlier corresponds to Case 3 in Table F. Proposition 9 says that the pairs bootstrap statistic $t^{\#}_{n,over}$ always contains the outlier, while the implied probability bootstrap statistic $t^*_{n,over}$ contains the outlier with probability $1 - p^*_{m,over}$ as $\left\|W_{(n)}\right\| \to +\infty$.

Table 5 reports the Monte Carlo medians of the pairs and implied probability bootstrap quantiles $Q_t^{\#}$ and Q_t^*, respectively. We set $t = 0.95$, and we report also empirical coverages of bootstrap confidence intervals. In absence of the outlier, the Monte Carlo medians of both bootstrap methods are very close to the true quantiles. Also the empirical coverages are very close to the nominal coverage probability 0.95. For instance, in the case of $n = 50$, the pairs and implied probability bootstrap quantiles are 1.9541 and 1.9454, respectively, while the true quantile is 1.9766. In contrast, in the presence of the outlier, the pairs bootstrap quantiles are quite large, while the implied probability bootstrap quantiles tend to be close to the true quantiles without the outlier. For instance, in the case of $n = 50$ and $C = 20$, the pairs and implied probability bootstrap quantiles are 14.2579 and 2.4643, respectively. These findings confirm the theoretical results in Proposition 9. The implied probability bootstrap is less influenced from the outlier than the pairs bootstrap.

5 Discussions

5.1 Over-identifying restriction test

Another important issue for over-identified IV regression models is to check the validity of the instruments (i.e., test $H_0 : E\left[Z_i\left(Y_i - X_i'\theta\right)\right] = 0$ for some θ against $H_1 : E\left[Z_i\left(Y_i - X_i'\theta\right)\right] \neq 0$ for any θ). This problem is called the over-identifying restriction test. In the GMM context, the over-identifying restriction test statistic (so-called Hansen's J-statistic) is defined as

$$J_n = \left(\sum_{i=1}^n Z_i\left(Y_i - X_i'\hat{\theta}_{GMM}\right)\right)'\left(\sum_{i=1}^n \hat{U}_i^2 Z_i Z_i'\right)^{-1}\left(\sum_{i=1}^n Z_i\left(Y_i - X_i'\hat{\theta}_{GMM}\right)\right),$$

where $\hat{U}_i = Y_i - X_i'\hat{\theta}_{2SLS}$. Hall and Horowitz (1996) and Brown and Newey (2002) demonstrated higher order refinements of the pairs bootstrap with recentered moments and implied probability bootstrap, respectively, over the first-order asymptotic approximation based on the χ^2 distribution. Our breakdown point analysis presented in the last section can be extended to this statistic. To this end, we first modify Assumption 1' to define the outlier as the one causing divergence of J_n. We then characterize limiting behaviors of the implied probability associated with the outlier and derive breakdown point properties of the bootstrap quantiles.

5.2 Moment function with higher dimension

Our breakdown point analysis can be extended to the case of higher dimensional moment functions with $k > 2$. The main issue is to compute the limit of the implied probability $\pi_{(n)}$ defined in (5). As pointed out in Camponovo and Otsu (2012), if each element of $g\left(W_{(n)}, c\right)$ takes a different limit as $\left\|W_{(n)}\right\| \to$

$+\infty$, it is necessary to evaluate explicitly the limit of the inverse $\left[\frac{1}{n}\sum_{i=1}^{n} g\left(W_i, c\right) g\left(W_i, c\right)'\right]^{-1}$ appearing in $\pi_{(n)}$. Consequently, the result may become more complicated and less intuitive. To obtain a comprehensible result, it would be reasonable to consider the case where all elements of $g\left(X_{(n)}, c\right)$ take only two limiting values. In this case, we can split $g\left(X_{(n)}, c\right)$ into two sub-vectors and apply the partitioned matrix inverse formula for $\left[\frac{1}{n}\sum_{i=1}^{n} g\left(X_i, c\right) g\left(X_i, c\right)'\right]^{-1}$ to derive the limit of the implied probability $\pi_{(n)}$.

5.3 Time series data

For time series data, the bootstrap methods discussed in this paper need to be modified to reflect dependence of the data generating process. Combining the ideas of Kitamura (1997) and Brown and Newey (2002), Allen, Gregory and Shimotsu (2011) proposed an extension of the implied probability bootstrap to a time series context by using block averages of moment functions. We expect that the breakdown point analysis of this paper can be adapted to such a modified bootstrap method (see Camponovo, Scaillet and Trojani, 2012b, for the breakdown point analysis of resampling methods in time series data).

6 Empirical example

In this section, we illustrate our breakdown point analysis for bootstrap methods by an empirical example. We consider the following regression model studied by Romer (1993):

$$Inf_i = \alpha_0 + \beta_0 \cdot Open_i + U_i,$$

for $i = 1, \ldots, n$, where Inf_i is country i's average annual inflation rate and $Open_i$, a proxy variable for openness, is country i's share of imports in the GDP. Romer (1993) employed this model to investigate whether more open economies tend to have lower inflation rates. To deal with endogeneity in the openness variable, Romer (1993) used the logarithm of country i's land area $Land_i$ as an instrumental variable. In this case, the IV estimator of β_0 is written as

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Z_i Y_i}{\sum_{i=1}^{n} Z_i X_i},$$

where $Y_i = Inf_i - n^{-1}\sum_{i=1}^{n} Inf_i$, $X_i = Open_i - n^{-1}\sum_{i=1}^{n} Open_i$, and $Z_i = Land_i - n^{-1}\sum_{i=1}^{n} Land_i$.

As emphasized in Desbordes and Verardi (2012), Romer's (1993) dataset may contain anomalous observations related to extremely high inflation rates of some Latin American countries in the 1980's. A scatter plot for openness and inflation rates in Figure 1 endorses this concern. Particular anomalous observations are (i) Bolivia ($Inf_i = 206.7$), (ii) Argentina ($Inf_i = 117.0$), and (iii) Singapore ($Open_i = 163.8$).

Using our breakdown point analysis, we can determine which of these observations may dramatically influence $\hat{\beta}$ and the bootstrap inference. Since $\hat{\beta}$ remains bounded for anomalous observations in the explanatory variable X_i (see Case 2 in Table C), it turns out that Singapore does not have a large impact on $\hat{\beta}$ and the bootstrap inference. In contrast, since $\hat{\beta}$ is not bounded for anomalous observations in the dependent variable Y_i (see Case 3 in Table C), it turns out that Bolivia and Argentina may dramatically influence $\hat{\beta}$ and the bootstrap inference. Based on this preliminary analysis, we also consider the GMTM estimator $\hat{\beta}_{GMTM}$ with trimming two observations ($d = 2$):

$$\hat{\beta}_{GMTM} = \arg\min_{\beta} \left(\frac{1}{n} \sum_{i=1}^{n} Z_i \left(Y_i - X_i \beta \right) \mathbb{I} \left\{ r_i \left(\beta \right) \leq r_{(n-2)} \left(\beta \right) \right\} \right)^2 ,$$

where $r_i(\beta) = [Z_i(Y_i - X_i\beta)]^2$ and $r_{(1)}(\beta) \leq \cdots \leq r_{(n)}(\beta)$.

Table 9 reports the pairs and implied probability bootstrap confidence intervals based on $\hat{\beta}$ and $\hat{\beta}_{GMTM}$. We observe that both bootstrap methods tend to reject the null hypothesis $H_0 : \beta_0 = 0$ (the pairs bootstrap rejects H_0 at the 5% significance level, while the implied probability bootstrap rejects H_0 at the 1% significance level). However, it should be noted that the implied probability bootstrap confidence intervals are much shorter than the pairs bootstrap confidence intervals. For instance, the length of the 99% implied probability bootstrap confidence interval based on $\hat{\beta}$ is 0.5350, while for the pairs bootstrap the length is 0.9286. By our breakdown point analysis, these results can be explained as follows. For the anomalous observations, the implied probabilities are 0 for Bolivia and 0.0013 for Argentina. On the other hand, the pairs bootstrap draws those observations with probability $\frac{1}{114} = 0.0088$. Therefore, the pairs bootstrap analogs of $\hat{\beta}$ contain more frequently the anomalous observations than the implied probability bootstrap analogs of $\hat{\beta}$, and the pairs bootstrap confidence interval tends to be longer than the implied probability bootstrap confidence interval.

In Table 9, we observe that both bootstrap confidence intervals based on $\hat{\beta}_{GMTM}$ tend to be less significant against the null hypothesis $H_0 : \beta_0 = 0$ (the pairs bootstrap does not reject H_0 at the 10% significance level, while the implied probability bootstrap rejects H_0 at the 5% significance level). These results are in line with Desbordes and Verardi (2012), who also found that confidence intervals based on robust estimators tend to be less significant against H_0. Also the point estimates are very different ($\hat{\beta} = -0.3329$ and $\hat{\beta}_{GMTM} = -0.1716$) and $\hat{\beta}_{GMTM}$ is closer to 0.[13] Similar to the confidence intervals based on $\hat{\beta}$, the implied probability bootstrap confidence intervals based on $\hat{\beta}_{GMTM}$ are much shorter than the pairs bootstrap confidence intervals based on $\hat{\beta}_{GMTM}$. For instance, the length of the 99% implied probability bootstrap confidence interval is 0.4230, while for the pairs bootstrap the length is 0.8002. Again this difference in the lengths of the bootstrap confidence intervals can be explained by the fact that the pairs bootstrap analogs of $\hat{\beta}_{GMTM}$ contain more frequently the anomalous observations than the implied probability bootstrap analogs of $\hat{\beta}_{GMTM}$ even though the estimator $\hat{\beta}_{GMTM}$ is robust to the anomalous observations.

In this empirical example, based on our breakdown point analysis, if the researcher wishes to treat

[13] We also considered trimming more than two observations but the point estimates $\hat{\beta}_{GMTM}$ remain quite close to -0.1716.

the observations in Bolivia and Argentina as outliers, we recommend to use the GMTM estimator $\hat{\beta}_{GMTM}$ and implied probability bootstrap confidence interval.

7 Conclusion

This paper studies robustness of the pairs and implied probability bootstrap inference methods for instrumental variable regression models. In particular, we analyze the breakdown point properties of the quantiles of those bootstrap methods for robust and non-robust test statistics for parameter hypotheses. Simulation studies illustrate the theoretical findings. Our breakdown point analysis can be an informative guideline for applied researchers to decide which bootstrap method should be applied under existence of outliers. It is important to extend our analysis to dependent data setups, where different bootstrap methods, such as block bootstrap, need to be employed. Also, it is interesting to analyze the breakdown point properties for other implied probabilities, such as the exponential tilting weights obtained from the information projection by the Boltzmann-Shannon entropy.

References

[1] Allen, J., Gregory, A. W. and K. Shimotsu (2011) Empirical likelihood block bootstrapping, *Journal of Econometrics*, 161, 110-121.

[2] Antoine, B., Bonnal, H. and E. Renault (2007) On the efficient use of the informational content of estimating equations: implied probabilities and Euclidean likelihood, *Journal of Econometrics*, 138, 461-487.

[3] Back, K. and D. P. Brown (1993) Implied probabilities in GMM estimators, *Econometrica*, 61, 507-517.

[4] Bassett, G. W. (1991) Equivariant, monotonic, 50% breakdown estimators, *American Statistician*, 45, 135-137.

[5] Brown, B. W. and W. K. Newey (2002) Generalized method of moments, efficient bootstrapping, and improved inference, *Journal of Business and Economic Statistics*, 20, 971-975.

[6] Camponovo, L. and T. Otsu (2012) Breakdown point theory for implied probability bootstrap, *Econometrics Journal*, 15, 32-55.

[7] Camponovo, L., Scaillet, O. and F. Trojani (2012a) Robust subsampling, *Journal of Econometrics*, 167, 197-210.

[8] Camponovo, L., Scaillet, O. and F. Trojani (2012b) Robust resampling methods for time series, Working paper.

[9] Chambers, R.L. (1986) Outlier robust finite population estimation, *Journal of the American Statistical Association*, 81, 1063-1069.

[10] Čížek's, P. (2008) General trimmed estimation: Robust approach to nonlinear and limited dependent variable models, *Econometric Theory*, 24, 1500-1529.

[11] Čížek's, P. (2009) Generalized method of trimmed moments, Working paper.

[12] Davidson, R. and J. G. MacKinnon (2010) Wild bootstrap tests for IV regression, *Journal of Business and Economic Statistics*, 28, 128-144.

[13] Desbordes, R. and V. Verardi (2012) A robust instrumental variables estimator, forthcoming in *Stata Journal*.

[14] Donoho, D. L. and P. J. Huber (1983) The notion of breakdown point, in *A Festschrift for Erich L. Lehmann* by Bickel, P. J., Doksum, K. A. and J. L. Hodges Jr. (eds.), 157-184, Wadsworth, Belmont, California.

[15] Freedman, D. A. (1981) Bootstrapping regression models, *Annals of Statistics*, 9, 1218-1228.

[16] Gagliardini, P., Trojani, F. and G. Urga (2005) Robust GMM tests for structural breaks, *Journal of Econometrics*, 129, 139-182.

[17] Hadi, A. and A. Luceno (1997) Maximum trimmed likelihood estimators: a unified approach, examples and algorithms, *Computational Statistics and Data Analysis*, 25, 251-272.

[18] Hall, P. and J. L. Horowitz (1996) Bootstrap critical values for tests based on generalized- method-of-moment estimators, *Econometrica*, 64, 891-916.

[19] Hall, P. and B. Presnell (1999) Intentionally biased bootstrap methods, *Journal of the Royal Statistical Society*, B, 61, 143-158.

[20] Hampel, F. R. (1971) A general qualitative definition of robustness, *Annals of Mathematical Statistics*, 42, 1887-1896.

[21] Hampel, F., Ronchetti, E., Rousseeuw, P. and W. Stahel (1986) *Robust statistics: The approach based on influence functions*, Wiley, New York.

[22] He, X., Simpson, D. G. and S. L. Portnoy (1990) Breakdown robustness of tests, *Journal of the American Statistical Association*, 85, 446-452.

[23] Hill, J.B. (2013). Least tail-trimmed squares for infinite variance autoregressions. Journal of Time Series Analysis 34, 168-186.

[24] Hill, J.B. and M. Aguilar (2012) Moment condition tests for heavy tailed time series, forthcoming in *Journal of Econometrics*, Annals Issue on Extreme Value theory.

[25] Hill, J. P. and E. Renault (2010) Generalized method of moments with tail trimming, Working paper.

[26] Imbens, G. W., Spady, R. H. and P. Johnson (1998) Information theoretic approaches to inference in moment condition models, *Econometrica*, 66, 333-357.

[27] Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent process, *Annals of Statistics*, 25, 2084-2102.

[28] Kitamura, Y., Otsu, T. and K. Evdokimov (2013) Robustness, infinitesimal neighborhoods, and moment restrictions, forthcoming in *Econometrica*.

[29] Kitamura, Y. and M. Stutzer (1997) An information-theoretic alternative to generalized method of moments estimation, *Econometrica*, 65, 861-874.

[30] Krasker, W. S. and R. E. Welsch (1985) Resistant estimation for simultaneous-equations models using weighted instrumental variables, *Econometrica* 53, 1475-1488.

[31] Markatou, M. and X. He (1994) Bounded influence and high breakdown point testing procedures in linear models, *Journal of the American Statistical Association*, 89, 543-549.

[32] Maronna, R.A., Martin, R.D. and V.J. Yohai (2006) *Robust Statistics. Theory and Methods*, Wiley, New York.

[33] Newey, W. K. and R. J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, 72, 219-255.

[34] Neykov, N. M. and P. N. Neytchev (1990) A robust alternative of the maximum likelihood estimator, Short communications of *COMPSTAT*, 99-100.

[35] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

[36] Romer, D. (1993) Openness and inflation: theory and evidence, *The Quarterly Journal of Economics*, 108, 869-903.

[37] Ronchetti, E. and F. Trojani (2001) Robust inference with GMM estimators, *Journal of Econometrics*, 101, 37-69.

[38] Rousseeuw, P. J. (1985) Multivariate estimation with high breakdown point, in W. Grossman, G. Pflug, I. Vincze, and W. Wertz (eds.) *Mathematical statistics and applications*, Vol. B, Dordrecht, Netherlands: Reidel, 283-297.

[39] Rousseeuw, P. J. (1997) Introduction to positive-breakdown methods, in G. S. Maddala and C. R. Rao (eds.) *Handbook of statistics*, Vol. 15: Robust inference. Amsterdam: Elsevier, 101-121.

[40] Rousseeuw, P. J. and A. M. Leroy (2003) *Robust Regression and Outlier Detection*, Wiley, New York.

[41] Salibian-Barrera, M., Van Aelst, S. and G. Willems (2007) Fast and robust bootstrap, *Statistical Methods and Applications*, 17, 41-71.

[42] Salibian-Barrera, M. and R. Zamar (2002) Bootstrapping robust estimates of regression, *Annals of Statistics*, 30, 556-582.

[43] Singh, K. (1998) Breakdown theory for bootstrap quantiles, *Annals of Statistics*, 26, 1719-1732.

[44] Ylvisaker, D. (1977) Test resistance, *Journal of the American Statistical Association*, 72, 551-556.

		No Outlier	Outlier		
$\lvert T_n \rvert$	n		$Y_{(n)} = 5Y_{max}$	$Y_{(n)} = 10Y_{max}$	$Y_{(n)} = 20Y_{max}$
True	50	1.8210	9.3181	19.1922	39.0682
	100	1.7713	7.2143	14.6622	29.7983
Pairs	50	1.7972 (0.9491)	7.9834 (0.9937)	16.1951 (0.9964)	32.5251 (0.9987)
	100	1.7537 (0.9481)	5.8554 (0.9896)	11.7248 (0.9967)	23.3004 (0.9987)
Implied	50	1.7882 (0.9450)	2.6851 (0.5016)	2.7357 (0.3403)	2.6658 (0.1919)
	100	1.7497 (0.9461)	2.4731 (0.5425)	2.5712 (0.3765)	2.4642 (0.2191)
		No Outlier	Outlier		
$\lvert T_n^d \rvert$	n		$Y_{(n)} = 5Y_{max}$	$Y_{(n)} = 10Y_{max}$	$Y_{(n)} = 20Y_{max}$
True	50	1.7388	1.8400	1.8400	1.8400
	100	1.7182	1.7827	1.7827	1.7827
Pairs	50	1.7845 (0.9592)	7.6376 (0.9961)	15.5285 (0.9968)	31.1589 (0.9980)
	100	1.7457 (0.9591)	5.8030 (0.9952)	11.7246 (0.9976)	23.0146 (0.9988)
Implied	50	1.7473 (0.9534)	1.8857 (0.9619)	1.8327 (0.9590)	1.8005 (0.9574)
	100	1.7212 (0.9531)	1.8693 (0.9608)	1.8163 (0.9587)	1.7807 (0.9575)

Table 1: **Quantiles of the pairs and implied probability bootstrap for the just-identified case.** The rows labelled "True" report the simulated quantiles of the distribution of $\lvert T_n \rvert$, and $\left\lvert T_n^d \right\rvert$ based on 20,000 realizations. The rows labelled "Pairs" report the pairs bootstrap quantiles. The rows labelled "Implied" report the implied probability bootstrap quantiles. The sample sizes are $n = 50$ and 100. In brackets, we report the empirical coverages of 95% bootstrap confidence intervals.

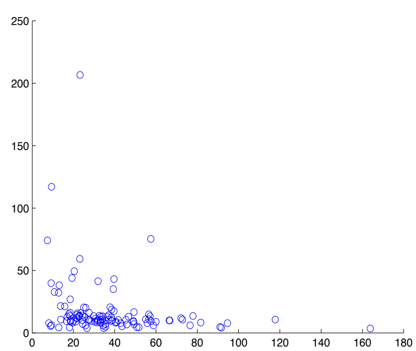

Figure 1: **Scatter plot for the empirical example.** On the $x-$Axis and $y-$Axis, are represented the proxy variable for the openess of a country and the average annual inflation rate of a country, respectively.

		No Outlier	Outlier		
$\lvert T_n \rvert$	n		$Y_{(n)} = 5Y_{max}$	$Y_{(n)} = 10Y_{max}$	$Y_{(n)} = 20Y_{max}$
True	50	1.8210	9.3181	19.1922	39.0682
	100	1.7713	7.2143	14.6622	29.7983
R. Residual	50	1.7788 (0.9471)	9.9587 (0.9725)	20.2172 (0.9781)	40.7488 (0.9802)
	100	1.7610 (0.9444)	7.7124 (0.9691)	15.5270 (0.9727)	31.2815 (0.9793)
U. Residual	50	1.7858 (0.9473)	9.9887 (0.9735)	20.2941 (0.9786)	40.9144 (0.9810)
	100	1.7638 (0.9454)	7.7183 (0.9706)	15.5622 (0.9733)	31.3310 (0.9796)
		No Outlier	Outlier		
$\lvert T_n^d \rvert$	n		$Y_{(n)} = 5Y_{max}$	$Y_{(n)} = 10Y_{max}$	$Y_{(n)} = 20Y_{max}$
True	50	1.7388	1.8400	1.8400	1.8400
	100	1.7182	1.7827	1.7827	1.7827
R. Residual	50	1.7461 (0.9582)	6.5794 (1.0000)	12.6706 (1.0000)	25.1610 (1.0000)
	100	1.7375 (0.9571)	5.1123 (1.0000)	9.4927 (1.0000)	18.8101 (1.0000)
U. Residual	50	1.7544 (0.9591)	6.6319 (1.0000)	12.7875 (1.0000)	25.6547 (1.0000)
	100	1.7408 (0.9577)	5.1947 (1.0000)	9.7543 (1.0000)	19.2149 (1.0000)

Table 2: **Quantiles of the restricted and unrestricted residual bootstrap for the just-identified case.** The rows labelled "True" report the simulated quantiles of the distribution of $|T_n|$, based on 20,000 realizations. The rows labelled "R. Residual" report the restricted residual bootstrap quantiles. The rows labelled "U. Residual" report the unrestricted residual bootstrap quantiles. The sample sizes are $n = 50$ and 100. In brackets, we report the empirical coverages of 95% bootstrap confidence intervals.

		No Outlier	Outlier		
$\lvert t_n \rvert$	n		$X_{(n)} = 5X_{max}$	$X_{(n)} = 10X_{max}$	$X_{(n)} = 20X_{max}$
True	50	1.9664	2.8926	4.0525	6.4031
	100	1.9652	2.4056	3.0520	4.3312
Pairs	50	1.9466 (0.9438)	4.5537 (1.0000)	7.1891 (1.0000)	10.5556 (1.0000)
	100	1.9281 (0.9444)	3.8653 (1.0000)	6.2978 (1.0000)	9.9961 (1.0000)
Implied	50	1.9373 (0.9423)	2.2273 (0.9093)	2.3083 (0.8301)	2.3659 (0.7011)
	100	1.9089 (0.9431)	2.1906 (0.9313)	2.2733 (0.9108)	2.3340 (0.8039)

Table 3: **Quantiles of the pairs and implied probability bootstrap for the just-identified case.** The rows labelled "True" report the simulated quantiles of the distribution of $|t_n|$ based on 20,000 realizations. The rows labelled "Pairs" report the pairs bootstrap quantiles. The rows labelled "Implied" report the implied probability bootstrap quantiles. The sample sizes are $n = 50$ and 100. In brackets, we report the empirical coverages of 95% bootstrap confidence intervals.

		No Outlier	Outlier		
$\lvert T_{n,over}\rvert$	n		$Y_{(n)} = 5Y_{max}$	$Y_{(n)} = 10Y_{max}$	$Y_{(n)} = 20Y_{max}$
True	50	1.5831	8.7738	18.0977	36.9841
	100	1.5488	6.8020	13.8711	28.2121
Pairs	50	1.5597 (0.9492)	7.3790 (0.9974)	14.9219 (0.9992)	30.1704 (0.9990)
	100	1.5411 (0.9494)	5.4235 (0.9915)	10.8790 (0.9986)	21.7734 (0.9998)
Implied	50	1.5365 (0.9454)	2.2006 (0.5007)	2.3037 (0.3268)	2.3477 (0.1884)
	100	1.5207 (0.9406)	1.9708 (0.5101)	1.9935 (0.3314)	1.9692 (0.1978)
		No Outlier	Outlier		
$\lvert T^d_{n,over}\rvert$	n		$Y_{(n)} = 5Y_{max}$	$Y_{(n)} = 10Y_{max}$	$Y_{(n)} = 20Y_{max}$
True	50	1.5112	1.6134	1.6134	1.6134
	100	1.5031	1.5524	1.5524	1.5524
Pairs	50	1.5432 (0.9597)	6.8614 (0.9976)	13.8689 (0.9990)	28.0197 (0.9998)
	100	1.5292 (0.9581)	5.2663 (0.9927)	10.4683 (0.9964)	21.2601 (0.9999)
Implied	50	1.5012 (0.9474)	1.5698 (0.9571)	1.5363 (0.9545)	1.5228 (0.9470)
	100	1.4921 (0.9454)	1.5617 (0.9565)	1.5311 (0.9540)	1.5201 (0.9497)

Table 4: **Quantiles of the pairs and implied probability bootstrap for the over-identified case.** The rows labelled "True" report the simulated quantiles of the distribution of $|T_{n,over}|$, and $\left|T^d_{n,over}\right|$ based on 20,000 realizations. The rows labelled "Pairs" report the pairs bootstrap quantiles. The rows labelled "Implied" report the implied probability bootstrap quantiles. The sample sizes are $n = 50$ and 100. In brackets, we report the empirical coverages of 95% bootstrap confidence intervals.

		No Outlier	Outlier		
$\lvert t_{n,over}\rvert$	n		$X_{(n)} = 5X_{max}$	$X_{(n)} = 10X_{max}$	$X_{(n)} = 20X_{max}$
True	50	1.9766	3.2956	5.5371	11.0175
	100	1.9641	2.5796	3.5563	6.2190
Pairs	50	1.9541 (0.9441)	5.7298 (0.9981)	9.6094 (0.9989)	14.2579 (0.9992)
	100	1.9530 (0.9472)	4.4665 (0.9990)	7.9467 (0.9997)	13.3555 (0.9998)
Implied	50	1.9454 (0.9427)	2.2110 (0.8528)	2.3243 (0.6876)	2.4643 (0.5021)
	100	1.9431 (0.9424)	2.1326 (0.9206)	2.1774 (0.8365)	2.2308 (0.6269)

Table 5: **Quantiles of the pairs and implied probability bootstrap for the over-identified case.** The rows labelled "True" report the simulated quantiles of the distribution of $|t_{n,over}|$ based on 20,000 realizations. The rows labelled "Pairs" report the pairs bootstrap quantiles. The rows labelled "Implied" report the implied probability bootstrap quantiles. The sample sizes are $n = 50$ and 100. In brackets, we report the empirical coverages of 95% bootstrap confidence intervals.

$\lvert T_n \rvert$		2.0	2.1	2.2	2.3
Pairs	No Outlier	0.0519	0.2233	0.6245	0.8845
	Outlier	0.0033	0.0208	0.0622	0.1389
Implied	No Outlier	0.0539	0.1818	0.5543	0.8257
	Outlier	0.6235	0.7804	0.8446	0.8681
$\lvert T_n^d \rvert$		2.0	2.1	2.2	2.3
Pairs	No Outlier	0.0409	0.2045	0.6169	0.9132
	Outlier	0.0024	0.0050	0.0249	0.0846
Implied	No Outlier	0.0469	0.1794	0.5373	0.8122
	Outlier	0.0413	0.1691	0.5123	0.8010

Table 6: **Empirical rejection frequencies just-identified case.** We report the empirical rejection frequencies of the null hypothesis $H_0 : \theta_0 = 2$ under different parameter values $\theta_0 = 2.0, 2.1, 2.2, 2.3$. We consider the pairs bootstrap ("Pairs") and the implied probability bootstrap ("Implied") applied to the statistics $|T_n|$ and $|T_n^d|$. The sample sizes is $n = 100$, the significance level is 5%.

$\lvert T_n \rvert$		2.0	2.1	2.2	2.3
R. Residual	No Outlier	0.0556	0.1978	0.5803	0.8447
	Outlier	0.0273	0.0491	0.0658	0.0959
U. Residual	No Outlier	0.0546	0.2297	0.6312	0.8895
	Outlier	0.0267	0.0545	0.0682	0.0961

Table 7: **Empirical rejection frequencies just-identified case.** We report the empirical rejection frequencies of the null hypothesis $H_0 : \theta_0 = 2$ under different parameter values $\theta_0 = 2.0, 2.1, 2.2, 2.3$. We consider the restricted residual bootstrap ("R. Residual") and the unrestricted residual bootstrap ("U. Residual") applied to the statistic $|T_n|$. The sample sizes is $n = 100$, the significance level is 5%.

$\lvert T_{n,over} \rvert$		2.0	2.1	2.2	2.3
Pairs	No Outlier	0.0506	0.2808	0.6618	0.9303
	Outlier	0.0014	0.0142	0.0680	0.1443
Implied	No Outlier	0.0594	0.2512	0.6131	0.8707
	Outlier	0.6686	0.7773	0.7864	0.7363
$\lvert T^d_{n,over} \rvert$		2.0	2.1	2.2	2.3
Pairs	No Outlier	0.0419	0.2415	0.6517	0.9118
	Outlier	0.0036	0.0042	0.0398	0.0905
Implied	No Outlier	0.0546	0.2222	0.6021	0.8575
	Outlier	0.0460	0.2158	0.5920	0.8497

Table 8: **Empirical rejection frequencies over-identified case.** We report the empirical rejection frequencies of the null hypothesis $H_0 : \theta_0 = 2$ under different parameter values $\theta_0 = 2.0, 2.1, 2.2, 2.3$. We consider the pairs bootstrap ("Pairs") and the implied probability bootstrap ("Implied") applied to statistics $|T_{n,over}|$ and $|T^d_{n,over}|$. The sample sizes is $n = 100$, the significance level is 5%.

$\hat{\beta}$	Quantile	Confidence Interval	Length
Pairs Bootstrap	90%	$[-0.5890; -0.0768]$	0.5122
	95%	$[-0.6437; -0.0221]$	0.6216
	99%	$[-0.7972; \quad 0.1314]$	0.9286
Implied Bootstrap	90%	$[-0.5079; -0.1579]$	0.3500
	95%	$[-0.5301; -0.1357]$	0.3944
	99%	$[-0.6004; -0.0654]$	0.5350
$\hat{\beta}_{GMTM}$	Quantile	Confidence Interval	Length
Pairs Bootstrap	90%	$[-0.3689; \quad 0.0257]$	0.3946
	95%	$[-0.4341; \quad 0.0910]$	0.5251
	99%	$[-0.5717; \quad 0.2285]$	0.8002
Implied Bootstrap	90%	$[-0.2950; -0.0482]$	0.2468
	95%	$[-0.3240; -0.0192]$	0.3048
	99%	$[-0.3831; \quad 0.0399]$	0.4230

Table 9: **Confidence Intervals by the pairs and implied probability bootstrap.** The rows labelled "Pairs Bootstrap" report the pairs bootstrap confidence intervals. The rows labelled "Implied Bootstrap" report the implied probability bootstrap confidence intervals.