

# Fast and Reliable Bootstrapping of Consistent Specification Tests\*

Gábor B. Uhrin<sup>†</sup>

July 1, 2013

In this paper we investigate the applicability of reliable and fast bootstrap methods on two non-smoothing, consistent specification tests (Bierens, 1982; Escanciano, 2006). Through Monte Carlo experiments we compare the performance of these tests under the null when the null distribution is simulated by the bootstrap, the double bootstrap (Beran, 1988) and the fast double bootstrap (FDB, Davidson and MacKinnon (2007)) procedures. Based on theoretical considerations, the latter two are expected to yield refinements to the order in the error in rejection probabilities. Our main findings are: (1) The Rademacher distribution (Davidson and Flachaire, 2008) for the wild bootstrap should not be used to single bootstrap the investigated tests; (2) The two point distribution proposed by Mammen (1993) for the wild bootstrap is an adequate “omnibus” method for carrying out the bootstrap, double bootstrap and FDB; (3) The double bootstrap should not be used in small and moderate sized samples; (4) The FDB can generally be advised to be used for refining inference over the single bootstrap.

---

\*This paper is a rewritten version of my master thesis at TSE. My work was supervised by Pascal Lavergne. I thank him for suggesting this interesting topic, for his patience during the several clarifying discussions, and for his continuous support. Jean-Pierre Florens and Nour Meddahi took their time to be members of the defense committee. I thank them for this and their comments. I am also grateful for Richard Foltyn’s comments and the feedback of the participants of the Econometrics Lunch Workshop at TSE. My stay in Toulouse would not have been possible without the financial support of the Jean-Jacques Laffont foundation. The usual disclaimer applies.

<sup>†</sup>Student at Toulouse School of Economics. E-mail: gaboruh@gmail.com.

# 1. Introduction

The most basic, sometimes implicit, assumption of statistical or econometric models is that the model is correct in the sense that interdependencies in the population are indeed in accordance with the model. For example in the context of mean regressions, where the object of interest is the conditional expectation function (CEF), adequacy of the regression model means that the a priori assumptions on the (parametric) form are indeed correct. If this is not so, the scope of the interpretation of estimation results is limited—if not impossible. While it is true that the linear least squares projection of an outcome variable  $y$  on the explanatory variables  $x$  is the best linear predictor of  $y$  given  $x$  for any CEF, if we specify our regression model to be linear, but the CEF is, in fact, not linear, we cannot go beyond this result. However, if we specify the regression model correctly, the usual interpretation of the estimation results (i.e. the expectation of  $y$  if  $x$  takes on a specific value  $x_0$ ) will be correct. This understanding raises the need of testing the validity of the assumed model—the specification.

The aim of this paper is investigate the applicability of bootstrap methods that are expected to yield reliable inference for a particular type of consistent specification tests.

The literature on specification tests is rich. Our focus in this paper is on the so-called non-smoothing consistent specification tests exemplified by [Bierens \(1982\)](#). While these kind of test are consistent in the sense that their power converges to one (as  $n \rightarrow \infty$ ) for all departures from the model of the null hypothesis, they generally have unknown, or complicated asymptotic null distributions which must be simulated by bootstrap. Due to the DGP-dependent nature of their null distributions, the bootstrapped null distributions might yield undesirable levels for these tests.

Thus, we investigate by means of simulation whether we can make more reliable inferential decisions by using the double bootstrap ([Beran, 1988](#)) (otherwise known as iterated bootstrap, [Hall \(1992\)](#)) and the fast double bootstrap (FDB, [Davidson and MacKinnon \(2007\)](#)) instead of the “single” residual or wild bootstrap. While the double bootstrap is theoretically expected to yield refinements to the order in the error in rejection probabilities (ERP, the difference in the actual and nominal level of the test), its computation is somewhat costly. The fast double bootstrap overcomes this problem and, at the same time, is expected to retain the reliability of the double bootstrap.

Based on our simulations under various experimental setups considering Bierens’s test ([Bierens, 1982](#)) and the more recent Escanciano test ([Escanciano, 2006](#)) employing the residual and wild bootstrap procedures, we argue for the following theoretical and practical conclusions:

1. The Rademacher distribution ([Davidson and Flachaire, 2008](#)) for the wild bootstrap should not be used to single bootstrap the investigated tests.
2. The two point distribution proposed by [Mammen \(1993\)](#) for the wild bootstrap is an adequate “omnibus” method in the sense that it performs reasonably well for all considered bootstrap procedures (single, double, and fast double bootstrap). Its

performance is especially good under heteroskedasticity compared to the residual and Rademacher versions.

3. The Rademacher distribution for the wild bootstrap can outperform the [Mammen-wild](#) bootstrap especially with a non-symmetric error distribution.
4. The double bootstrap should not be used in small and moderate sized samples, because its performance (in terms of levels) is often worse than that of *both* the fast double bootstrap and the single bootstrap.
5. The fast double bootstrap can generally be advised to be used for refining inference over the single bootstrap. Especially if it is coupled with the appropriate wild distribution. As a thumb rule, with the [Mammen-wild](#) method, the FDB performs no worse than the single bootstrap, and it performs observably better under the practically important case of heteroskedasticity.

While the above points should be interpreted with care especially because of numerical inaccuracies, we made all attempt to give conclusions that stem from relatively stable patterns.

We believe that the last two points are of significant practical importance, especially because, to our knowledge, no paper has applied the double and fast double bootstrap for consistent specification tests. More generally, our contribution to the literature is two-fold: (1) it adds to the literature on specification testing by investigating the gains from applying the double bootstrap and fast double bootstrap methods to (theoretically) popular non-smoothing consistent specification tests; (2) it adds to the ongoing process of discovery related to the usefulness of the fast double bootstrap by providing simulation results on a testing problem that has not been considered in the existing FDB simulations.

The structure of the paper is as follows. Section 2. sets up the hypothesis testing problem, briefly surveys the literature of consistent specification tests focusing on the non-smoothing type and details the tests that we are focusing on in this paper. Section 3. introduces the bootstrap shortly and gives exposition of the double and fast double bootstrap procedures including their practical implementation. These sections help motivating the paper and fix ideas. Section 4. contains our specific contributions: it details the experimental setup and leads the reader systematically up to the conclusions stated above. Note, that most of the figures and tables are contained in Appendix A. Besides detailing the main results, we also demonstrate a power curve analysis and a robustness experiment. Finally, Section 5. concludes.

## 2. Specification tests

Let  $(Z, \mathcal{A}, \mu)$  be a probability space and let  $(y, x)$  be an  $\mathbb{R} \times \mathbb{R}^k$  random vector defined on it. Our focus in this paper is on mean regressions, where the object of interest is the conditional expectation function (CEF)  $m(x_0) = E[y \mid x = x_0]$  for any  $x_0 \in \mathbb{R}^k$ . Since we

generally have no knowledge of the explicit form of this mean regression function, we have to estimate it from a sequence  $(x_i, y_i)_{i=1}^n$  of realizations that are ideally generated from the original  $(x, y)$  (e.g. the sequence is just the  $n$ -copy of  $(x, y)$ ). While it is possible to estimate consistently a rich family of functions through nonparametric methods, it is more desirable to have a correctly specified parametric model, since such a model can produce better estimates than crude non-parametric methods (Li and Racine, 2006). Further, a well-specified parametric model is more tractable and easier to interpret.

To this end, we specify a parameter space  $B \subseteq \mathbb{R}^k$  that indexes a class of functions  $f : \mathbb{R}^k \times B \rightarrow \mathbb{R}$  that we define. That is, we impose restrictions on the functions from  $\mathbb{R}^k$  to  $\mathbb{R}$ , and each function that satisfies these restrictions are indexed with a parameter  $\beta \in B$ .

**Definition 2.1.** A *specification* is a pair  $\mathcal{S} := (B, f)$  such that  $B \subseteq \mathbb{R}^k$  and  $f : \mathbb{R}^k \times B \rightarrow \mathbb{R}$ . A *model* is the collection

$$\mathcal{M}_{\mathcal{S}} := \{f(\cdot, \beta) : \beta \in B\}. \quad (1)$$

For example, in the standard linear regression we assume that  $f(x, \beta) = x'\beta$ ,  $B = \mathbb{R}^k$ , and the collection of these functions for all  $\beta \in B$  is the relevant model. If we have the specification  $\mathcal{S} = (B, f)$ , then the main question is: Can we find the true regression function based on this specification? Is the true regression function (at least almost surely) equal to the CEF for some  $\beta_0 \in B$ ?

It is, however, very unlikely that from the space of functions we can just specify a subset that indeed contains the right one. In fact, the likelihood of this is zero. If we, however, have some information about the problem at hand, then we might specify a model and see “how wrong” this specification is. It can be the case that the specified model is so close to the true one that the difference is statistically indistinguishable. Formally, the hypotheses of the specification testing problem are the following (e.g. Bierens (1982)):

$$H_0: \Pr [f(x, \beta_0) = m(x)] = 1 \text{ for some } \beta_0 \in B, \quad (2)$$

$$H_1: \Pr [f(x, \beta) = m(x)] \neq 1 \text{ for all } \beta \in B. \quad (3)$$

Defining  $u \equiv y - f(x, \beta_0)$  it is easy to see that the above formulation is equivalent to the following hypotheses:

$$H_0: E[u \mid x] = 0 \text{ a.s., for some } \beta_0 \in B, \quad (4)$$

$$H_1: E[u \mid x] \neq 0 \text{ on a set with positive measure, } \forall \beta \in B. \quad (5)$$

If the specification is correct in the sense of  $H_0$ , then any suitable consistent estimator that takes values in  $B$  will find the correct  $\beta_0$ .

To fix definitions, consider the probability space  $(Z, \mathcal{A}, \mu)$  and hypothetical testing problem for a parameter  $\theta_0 \in \Theta$ .  $\theta_0$  is a population parameter that we are interested in. We test  $H_0: \theta_0 \in \Theta_0$  for some restriction, against  $H_1: \theta_0 \in \Theta_1$ , where  $\Theta_0 \cup \Theta_1 \subseteq \Theta$  and

$\Theta_0 \cap \Theta_1 = \emptyset$ . We observe the sample  $z_1, \dots, z_n$ . A *test statistic* is a measurable mapping  $T_n: \mathbb{R}^n \rightarrow \mathbb{R}$ . A *test* is a decision rule that rejects or not rejects  $H_0$  for a given value of the test statistic. A test can be viewed as a random mapping from the target space of  $T$  to, say, the set  $\{0, 1\}$ . The *power* of a test is the probability of rejecting the null given that  $\theta_0$  is indeed in  $\Theta_1$ . We say that a test is *consistent* if its power converges to 1 for all  $\theta \in \Theta_1$  as the sample size converges to infinity.

The literature on specification tests is rich, but the widely applied specification tests (e.g. Hausman's) are not necessarily consistent in the sense defined above. Starting from Bierens (1982), the focus turned on consistent specification tests. These tests can be put roughly in two categories: (1) smoothing tests that employ some nonparametric (smoothing) estimator related to the true CEF in their construction (Härdle and Mammen, 1993; Zheng, 1996; Hsiao, Li, and Racine, 2007); (2) non-smoothing tests. The first type of tests usually have tractable asymptotic null distributions, but the curse of dimensionality makes their use impractical for multiple regressors and small sample sizes. These are the reasons why our focus in this paper is on consistent specification tests of the non-smoothing type.

In the following we investigate a version of the original Bierens Integrated Moment Condition Test (ICM, Bierens (1982)), and the more recent Escanciano test (Escanciano, 2006). The idea of Bierens is to use a parametrized function family  $H(x, t)$  for some  $t \in \mathcal{T} \subseteq \mathbb{R}^k$  and test whether  $E[uH(x, t)] = 0$ . If  $H(\cdot, t)$  and  $\mathcal{T}$  together form a sufficiently rich family of functions, then whenever  $E[u | x] \neq 0$ , there will be a  $t_0 \in \mathcal{T}$  such that  $E[uH(x, t_0)] \neq 0$ .<sup>1</sup> The concrete test statistics that we consider are:

1. The Bierens test (Bierens, 1982):

$$\hat{\eta}_b = \int_{\mathbb{R}^k} \left| \overbrace{\sum_{i=1}^n \frac{1}{n} (y_i - f(x_i, \hat{\beta})) \exp(\iota' x_i)}^{Z(t)} \right|^2 \phi(t) dt \quad (6)$$

$$= \frac{1}{n} \sum_i \sum_j \hat{u}_i \hat{u}_j \exp\left(-\frac{1}{2} \|x_i - x_j\|^2\right), \quad (7)$$

where  $\phi(\cdot)$  is the standard normal density,  $t \in \mathbb{R}^k$  and  $\iota = \sqrt{-1}$ . As was shown by Bierens and Ploberger (1997), the asymptotic null distribution of this test statistic is:

$$\eta_b^0 = \sum_{i=1}^{\infty} \lambda_i \gamma_i, \quad (8)$$

where  $\gamma_i \stackrel{iid}{\sim} \chi^2$ . The weight  $\lambda_i$ -s are the increasingly ordered eigenvalues of the covariance operator  $E[Z(t_1)Z(t_2)]$ , and thus they are dependent on the data

---

<sup>1</sup>A deeper reason for why Bierens-type non-smoothing tests are consistent, and why they are so for many suitable choices of  $H$  was illuminated by Stinchcombe and White (1998). In their words, the usefulness of a family of functions for consistent specification testing depends on its *comprehensive revealability*, see Definition 3.2. of Stinchcombe and White (1998, p. 302).

generating process: the distribution of  $(y_i, x_i)_{i=1}^n$  and the true model. This is the feature that necessitates the use of more reliable bootstrap inference procedures.<sup>2</sup>

2. The Escanciano test (Escanciano, 2006):

$$\hat{\eta}_e = \int_{\mathbb{R}^k} \left( \frac{1}{2} \sum_{i=1}^n \hat{u}_i \mathbb{1}[t'x_i \leq \tau] \right)^2 \hat{F}_t(d\tau) dt \quad (9)$$

$$= \frac{1}{n^2} \sum_i \sum_j \hat{u}_i \hat{u}_j \left( \sum_k \int_{\mathbb{R}^k} \mathbb{1}[t'x_i \leq t'x_k] \mathbb{1}[t'x_j \leq t'x_k] dt \right), \quad (10)$$

where  $\hat{F}_t$  is the empirical CDF of the projected  $t'x$ . The computation of the integral in the  $k$ -sum is discussed by Escanciano (2006, p. 1051). The test  $\hat{\eta}_e$  is shown to be consistent, but its asymptotic null distribution suffers from DGP-dependency—just like Bierens's test. Hence in this case, again, it is natural to employ more reliable bootstrap methods for inference.

### 3. Bootstrap methods

The method of (parametric) bootstrap is widely known among econometricians. Theoretical and practical expositions are many. For comprehensible theoretical treatments of the bootstrap we refer the reader to the books of Hall (1992) and Shao and Tu (1995). For a more eclectic and highly readable exposition, see Horowitz (2001).

In this paper we focus on wild and residual bootstrap, because the consistency of these bootstrap methods for the Bierens ICM test was shown by Dominguez (2005).

Let a specification  $\mathcal{S} = (B, f)$  be given so that we have the regression model

$$y = f(x, \beta) + u.$$

Let  $\hat{\beta}$  be a consistent estimator of some  $\beta_0 \in B$ . For a sample  $(y_i, x_i)_{i=1}^n$  calculate

$$\hat{u}_i = y_i - f(x_i, \hat{\beta}), \quad i = 1, \dots, n,$$

the estimated residuals. Let  $(u_j^*)_{j=1}^n$  be an  $n$ -sample drawn with replacement from  $(\hat{u}_i)_{i=1}^n$ . Set

$$y_j^* = x_j \hat{\beta} + u_j^*, \quad j = 1, \dots, n, \quad (11)$$

and calculate the desired test statistic based on these data and  $\hat{\beta}^*$  that is estimated from the newly generated data  $(y_j^*, x_j)_{j=1}^n$ . Note, that in the case of the residual bootstrap we do not resample  $x$ .

<sup>2</sup>Note, that  $\lambda_i$ -s can be consistently estimated, see Theorem 3. of Carrasco and Florens (2000, p. 806). However, the reliable bootstrap methods, in principle, yield refinements over that estimate too.

Instead of creating  $u_j^*$ -s with replacement from  $\hat{u}_i$ -s, the wild bootstrap generates  $u_j^{wv} = \hat{u}_j \varepsilon_j$ ,  $j = 1, \dots, n$  for some drawing from distributions  $\varepsilon_j$  that are mutually independent and independent of the basic data. It is desirable that  $E[\varepsilon_j] = 0$ , and  $E[\varepsilon_j^2] = E[\varepsilon_j^3] = 1$ . [Mammen \(1993\)](#) recommends using a two-point distribution,

$$\varepsilon^{(1)} = \begin{cases} \frac{1-\sqrt{5}}{2} & \text{with prob. } p = (5 + \sqrt{5})/10 \\ \frac{1+\sqrt{5}}{2} & \text{with prob. } 1 - p \end{cases}. \quad (12)$$

He also suggests two continuous distributions, the better of which theoretically is

$$\varepsilon^{(2)} = (\delta_1 + \zeta_1/\sqrt{2}) \cdot (\delta_2 + \zeta_2/\sqrt{2}) - \delta_1\delta_2, \quad (13)$$

where  $\zeta_1$  and  $\zeta_2$  are random variables with independent  $\mathcal{N}(0, 1)$  distributions,  $\delta_1 = \sqrt{3/4 + \sqrt{17}/12}$  and  $\delta_2 = \sqrt{3/4 - \sqrt{17}/12}$ . [Davidson and Flachaire \(2008\)](#) recommend using the two-point distribution given by

$$\varepsilon^{(3)} = \begin{cases} -1 & \text{with prob. } p = 0.5 \\ 1 & \text{with prob. } p = 0.5 \end{cases}. \quad (14)$$

Our investigation below shows that the choice of the wild bootstrap procedure ([Mammen-wild](#), [Rademacher-wild](#)) does matter for the performance of the tests that we consider.

### 3.1. Inference refinements

Let us define a test statistic *pivotal* if its distribution does not depend on the data generating process. If only the asymptotic distribution is pivotal, then we call a statistic *asymptotically pivotal*. Define further the *error in rejection probability* (ERP) as the difference between the nominal level of a test and the actual rejection probability based on the simulated null distribution of the test statistic. From Edgeworth expansions it can be shown that if a test statistic is *not* asymptotically pivotal, we generally cannot gain in the order of the ERP over asymptotic approximations with the use of the bootstrap (see [Table 1](#), and formally [Hall \(1992\)](#)). This might mean that in finite samples, statistical inference based on bootstrap may yield severely incorrect levels.<sup>3</sup> Using analogous Edgeworth expansion arguments, [Beran \(1988\)](#) and [Beran \(1987\)](#) suggested the use of double bootstrap (also known as iterative bootstrap). The idea behind the papers of [Beran](#) is to apply a function  $h$  to the statistic of interest such that the order of ERP of  $h(T_n)$  is smaller than that of  $T_n$ . In particular, the (asymptotic) null CDF of  $T_n$  (denote it  $G$ ) is a good candidate, since  $G(T_n)$  is less heavily dependent on the data generating process, and the distribution of  $G(T_n)$ , the prepivoted statistic is  $\mathcal{U}(0, 1)$  asymptotically under general circumstances. In particular, it means that  $G(T_n)$  is asymptotically pivotal.

<sup>3</sup>Even though the order is a purely asymptotic concept, where we are not sure after what “sufficiently large”  $N$  index the result holds, there seems to be the heuristic understanding, that we can hope for better finite sample behavior if the order of the ERP is smaller.

Table 1: Error in rejection probabilities

Method	Pivot	Not pivot
Asymptotics	$O\left(n^{-\frac{1}{2}}\right)$	$O\left(n^{-\frac{1}{2}}\right)$
Bootstrap	$O\left(n^{-1}\right)$	$O\left(n^{-\frac{1}{2}}\right)$
Double Bootstrap	$O\left(n^{-\frac{3}{2}}\right)$	$O\left(n^{-1}\right)$

According to [Beran \(1988\)](#).

Prepivoting  $T_n$  with the exact finite sample  $G$  or its consistent estimator  $G_n$  yields ERP of lower order than asymptotics and simple bootstrap. This is demonstrated in Table 1 which compares the ERP-s of tests based on asymptotics, bootstrap and double bootstrap respectively for pivotal and non-pivotal test statistics.

As can be seen, double bootstrap provides a refinement regardless of the pivotal or non-pivotal nature of  $T_n$ . Along the same arguments, iterating this prepivoting procedure yields, in principle, refinements in each step, i.e. tests based on  $G_n^2(G_n^1(T_n))$  will have smaller ERP-s than those based on  $G_n^1(T_n)$ , etc. However, whether iterating this procedure until infinity yields exact rejection probabilities is not clear. While computationally infeasible, this leaves open a possibly interesting theoretical question.

The implementation of the double bootstrap for a given level  $\alpha$  consists of the following steps (see also [Beran \(1988\)](#)):

1. Calculate  $\hat{T}$ , the estimated test statistic from the original data.
2. Draw  $B_1$  bootstrap resamples according to the bootstrap method of choice. Calculate  $T_i^*$  the test statistic derived from the bootstrap sample for each  $i = 1, \dots, B_1$ .
3. For *each*  $B_1$  bootstrap sample, repeat the chosen sampling procedure and draw  $B_2$  resamples. Calculate  $T_j^{**}$  for each  $j = 1, \dots, B_2$ , and calculate

$$Z_i = \frac{1}{B_2} \sum_{j=1}^{B_2} \mathbb{1} \left[ T_j^{**} \leq T_i^* \right] \quad i = 1, \dots, B_1.$$

4. Look for the  $1 - \alpha$ -th quantile of the dataset  $(Z_i)_{i=1}^{B_1}$  and denote it  $Q^{**}$ . Look for the  $Q^{**}$ -th quantile of the set  $(T^*)_{i=1}^{B_1}$ . Denote this quantile  $Q^*$ .
5. Reject  $H_0$  if  $\hat{T} > Q^*$ .

In the case of a Monte-Carlo simulation, repeat the above procedure  $M$  times. Practically it is advisable to set  $M$ ,  $B_1$  and  $B_2$  sufficiently large.



### 3.2. Fast, reliable inference

For complex statistics the double bootstrap might be too costly to compute. If, however, it is possible to find a method that is fast and yields equally reliable inference, then we are in a win-win situation. According to the suggestion of [Davidson and MacKinnon \(2007\)](#), it is often enough to use only one double bootstrap resample per bootstrap samples. That is, set  $B_2 = 1$ . However, it leads to a slight modification of the procedure described in the previous subsection. This fast double bootstrap procedure (FDB) is best viewed from a p-value perspective and it proceeds as follows ([Davidson and MacKinnon, 2007](#)):

1. Calculate  $\hat{T}$ , the estimated test statistic from the original data.
2. Draw  $B_1$  bootstrap resamples according to the bootstrap method of choice. Calculate  $T_i^*$ , the test statistic derived from the bootstrap sample for each  $i = 1, \dots, B_1$ . Calculate

$$p^* = \frac{1}{B_1} \sum_{i=1}^{B_1} \mathbb{1}[T_i^* > \hat{T}].$$

3. For *each*  $B_1$  bootstrap sample, repeat the chosen sampling procedure and draw  $B_2 = 1$  resamples. Calculate  $T_j^{**}$  for each  $j = 1, \dots, B_1$ , and look for the  $1 - p^*$ -th quantile of the dataset  $(T_j^{**})_{j=1}^{B_1}$ . Denote it  $Q^{**}$ .
4. Calculate

$$p^{**} = \frac{1}{B_1} \sum_{i=1}^{B_1} \mathbb{1}[T_i^* > Q^{**}],$$

and reject  $H_0$  if  $p^{**} < \alpha$ .

The slight departure from the fast double bootstrap procedure is that in the case of the FDB we do not directly estimate the distribution of  $T^*$  with the numerous double bootstrap samples, each corresponding to only *one* bootstrap first-phase sample.

The theoretical background for the FDB procedure is not well-developed. The assumptions in [Giacomini, Politis, and White \(2013\)](#) are sufficiently general that we can conclude that the FDB procedure is consistent if the bootstrap (and the double bootstrap) is consistent. However, an important question is, whether the FDB yields refinements similar to the double bootstrap. What can be shown is, that whenever  $G_1(T^*)$ , the distribution of  $T^*$ , is independent of  $G_2(T^{**})$ , the distribution of  $T^{**}$ , the fast double bootstrap is equivalent to the double bootstrap.

## 4. Experiments

The goal of experiments was to investigate the small sample properties of the bootstrap, double bootstrap and fast double bootstrap applied to two types of specification tests. The first one is Bierens's ICM test as given by Equation (7), and the second, more recent

one is the Escanciano test given by Equation (10). The common property of these tests is that the asymptotic null distribution must be simulated, and is DGP-dependent. This makes the use of the more reliable inferential methods of the double and fast double bootstrap natural, perhaps even necessary.

The employed bootstrap method is that of the residual bootstrap and the wild bootstrap as it was described in Section 3. In particular, we used the discrete and continuous suggestion of Mammen (1993), and the Rademacher distribution suggested by Davidson and Flachaire (2008). Besides being a popular method, it is important that the wild bootstrap is shown to be consistent for the ICM type of tests by Dominguez (2005). Hence, based on the discussion above, we can conclude that the double bootstrap and the fast double bootstrap are consistent too. Note that for a given method all subsequent bootstrap draws (double or fast double bootstrap) were carried out using the same method.

In this section we first describe the experimental setting, then we detail our main results and perform some robustness experiments that further support our main results.

#### 4.1. Experimental design

The basic DGP that we employ is the following:

$$y_i = 0.5x_{1i} + 0.5x_{2i} + 0.5x_{3i} - 1.5x_{4i} + u_i, \quad (15)$$

where  $u_i = \varepsilon_i \sim \mathcal{N}(0, 1)$  in the homoskedastic case, or  $u_i = x_{1i}\varepsilon_i$  in the heteroskedastic case. In order to model a (left) skewed distribution, we also used  $u_i = 1/3(\varepsilon_i + \zeta_i - 4)$ , with  $\zeta_i \sim \chi^2(4)$ . The PDF of this random variable can be seen in Figure 1. It has zero mean and unit variance.

Figure 1: Skewed distribution

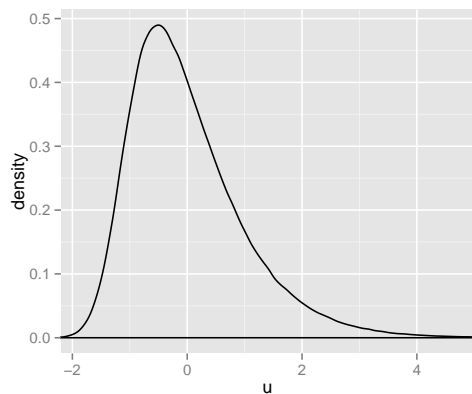
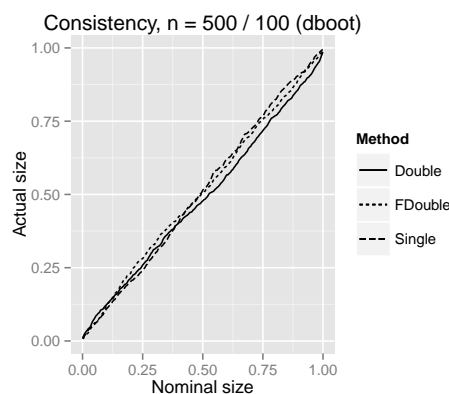


Figure 2: Consistency of bootstrap



All  $x_{ij}$ -s were drawn from  $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ , which yields unit variance. The reason why there are four explanatory variables in this DGP is that for small sample sizes the higher

number of explanatory variables give clear advantage to non-smoothing tests over smoothing tests. The number of Monte-Carlo replications ( $M$ ) was 1000, and the basic number of bootstrap samples was  $B_1 = 199$  and  $B_2 = 150$  for the first bootstrap phase and the double bootstrap phase respectively. Sample sizes ( $n$ ) that we investigated were 40, 50, 75, 100, and 500 for some simulations.

## 4.2. Main results

The first observation that we could conclude from the experiments was that for both tests, all three bootstrapping procedures are consistent, i.e. they yield correct levels for sufficiently large sample sizes. This is illustrated in Figure 2 above. The plot shows the actual versus the nominal level of the Bierens test using the wild methodology. The sample size is 500 for the single bootstrap and the fast double bootstrap, and 100 for the double bootstrap. These bootstrap procedures yield correct sizes if their plots are very close to being uniform on  $[0, 1]$ . Indeed, this is what we see in Figure 2, and the conclusion is robust to all examined specifications, including the Escanciano test. As regards tendency, the bootstrap procedures tend to yield uniformly better sizes as  $n$  grows.

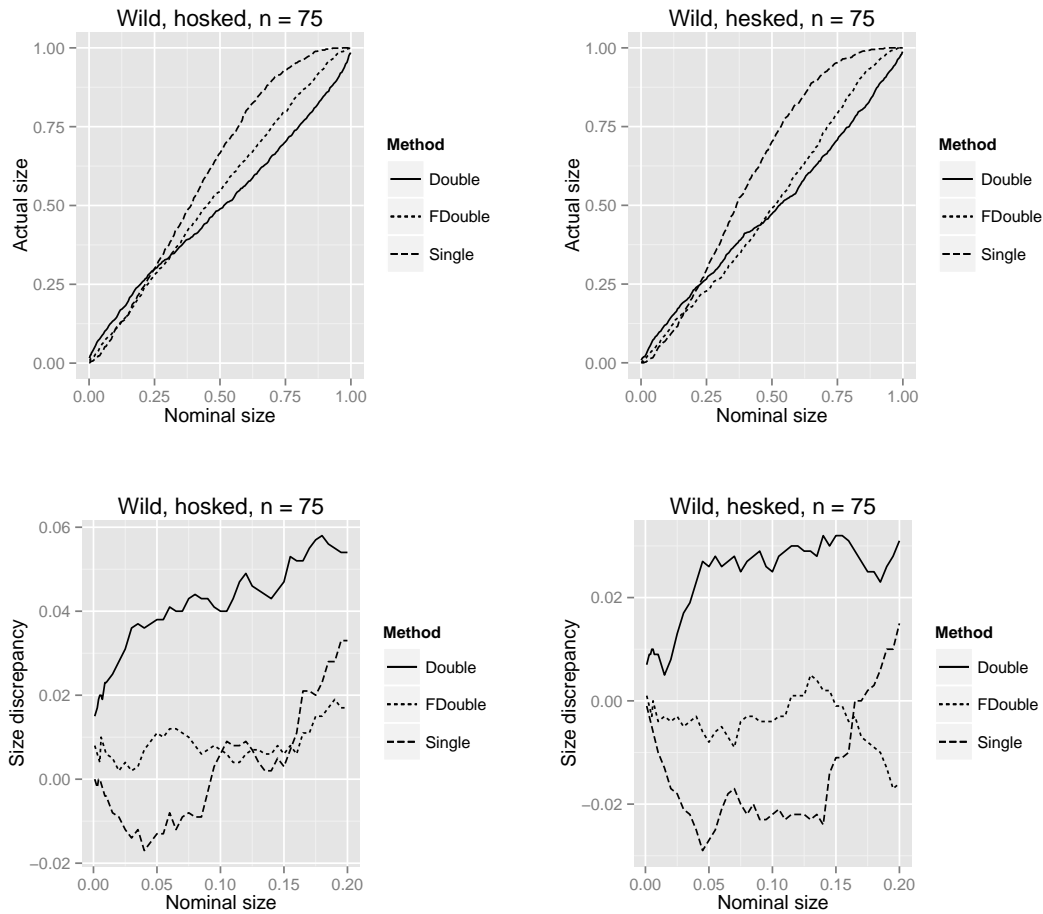
For small and moderate sample sizes, however, we can find moderate underrejection for the single bootstrap procedure. This is illustrated in Figure 3. Note, that except for Figure 3, all figures are contained in Appendix A, starting from page 20. The upper row of these plots displays the actual vs. the nominal level of the Bierens test for the wild methodology. The left column is in the case of homoskedastic data, the right is in the case of heteroskedastic data.

To capture the behavior of these procedures in a more relevant way, the lower row of Figure 3 plots the size discrepancy (actual minus nominal level) against the actual level truncated to the interval  $[0, 0.2]$ . While, as we have already remarked, the single bootstrap yields moderate underrejection, the fast double bootstrap performs remarkably well, especially in the heteroskedastic case (right column). The double bootstrap, however, performs unexpectedly badly seriously overrejecting on the truncated interval. Based on our theoretical discussion, this presents a puzzle.

As regards robustness, the qualitative nature of the results do not change by using the skewed error distribution. In the plots we have chosen  $n = 75$ , because this seems to display patterns in the clearest manner. Varying the sample size yields no qualitative changes besides what is implied by consistency, i.e. for larger sample size the lines gradually get closer to the uniform CDF.

The overview of the behavior of the Escanciano test is contained in Figure 6, whose structure is the same as that of Figure 3. The test performs slightly worse than the Bierens test, the undesirable behavior of both the single and the double bootstrap is reinforced both in the case of heteroskedasticity, and in the case of homoskedasticity. However, the fast double bootstrap yields the same good levels as in the case of Bierens test. For more precise comparison, see Table 2, which contains the mean absolute discrepancy of the

Figure 3: Bierens test simulation



fast double procedure and demonstrates that the FDB performs equally well in the case of both tests calculated for 48 levels on  $[0, 0.2]$ .<sup>4</sup>

As regards the behavior of the double bootstrap, it must be noted that the validity of Edgeworth expansions, on which the theory of asymptotic refinement is based, is not always guaranteed. This is especially so in the case of discrete distributions, a prime example of which in the outcome of the resampling rule of [Mammen's](#) wild bootstrap (Equation (12) on page 7). Using the continuous distribution given in Equation (12), however, provides only worse results. Besides, the Rademacher distribution (Equation (14)), which is discrete, outperforms the basic wild bootstrap (see below), hence it is not straightforward to give a reason why the double bootstrap performs unexpectedly.

<sup>4</sup>We believe that it depends on the problem at hand whether under- or overrejection is more "costly". By taking absolute differences we treat both deviations from the nominal level equally.

Table 2: Fast double bootstrap performance

Test ( $n = 75$ )	Homosked.	Heterosked.
Bierens	0.0083 (0.004)	0.0044 (0.004)
Escanciano	0.0042 (0.003)	0.0088 (0.005)
Mean absolute size discrepancy on $[0, 0.2]$ (sd in parantheses)		

Along this line of reasoning, we have tried using the plain residual bootstrap methodology instead of the wild bootstrap in the hope of utilizing more on the asymptotic considerations.<sup>5</sup> In the case of homoskedasticity the residual bootstrap yielded very precise results for small nominal levels as it is demonstrated in Figure 7 for both tests and for all procedures. In the case of heteroskedasticity, however, the residual bootstrap performs very weakly, which is expected given that the residual bootstrap assumes an iid sample. These conclusions apply to the case of the non-symmetric error distribution too.

When using the Rademacher distribution, the gains from using the double or the fast double bootstrap over the single bootstrap are remarkable for both tests. In figure 8 we report simulation results for the smallest sample size ( $n = 40$ ) with the skewed distribution to demonstrate this fact in the (a priori) worst case.<sup>6</sup> Note that in the size discrepancy plot, the line corresponding to the single bootstrap disappears, because it exits the bound we have prescribed in order to concentrate on the other two methods. The pattern extends to the Escanciano test, and for heteroskedasticity.

This strong finding raises the question whether we should prefer using the Rademacher version of the wild bootstrap for (and only for) the double and fast double bootstrap procedures, for both tests. Tables 3, 4 and 5 report the mean absolute size discrepancies (and standard deviations) for the Bierens test for all available sample sizes with homoskedastic, heteroskedastic and skewed data respectively. Even though sharp conclusions cannot be drawn from these tables (especially because of the numerical inaccuracies of our not too extensive simulations), we have mild evidence that for moderate sample sizes, under homoskedasticity or skewed data, the Rademacher-wild procedure can be the preferable choice for performing a FDB analysis. This evidence holds, even though in an even milder form, for the double bootstrap under the same conditions. A similar conclusion could not be established for the Escanciano test.

Based on the same tables, we can observe that the popular wild bootstrap (Mammen, 1993) yields balanced and relatively reliable results for the Bierens test (especially

<sup>5</sup>In the residual bootstrap, the estimated  $\hat{u}$ -s get resampled instead of being “perturbed” by a different, independent distribution.

<sup>6</sup>While Davidson and Flachaire (2008) argue for the superiority of the Rademacher distribution over the more popular distribution of Mammen (1993) for the wild bootstrap, it is not clear how well their method work for non-symmetric distributions.

in the case of heteroskedasticity), and similar conclusions hold for the Escanciano test. Moreover, its performance in the single bootstrap is adequate compared to the Rademacher-wild, which does not behave in a desired way, which fact undermines on-the-spot size comparisons in applied work. Therefore, based on our evidences, it might be more advisable to employ the **Mammen**-wild method for performing specification tests regardless of the procedure (single, double or fast double bootstrap) being employed.

We have, thus, argued that **Mammen**-wild is an appropriate “omnibus” method, even though for solely the FDB procedure, using the Rademacher-wild might be advisable. From the casual observation of Figures 3 and 6, we have concluded that the fast double bootstrap procedure has its place in a specification tester’s toolbox. From the same graphs we can also argue *against* the usage of double bootstrap, because (1) its performance is often worse than that of the other two procedures and (2) it is computationally costly on top of that.

However, we have also seen that the single bootstrap does not perform particularly poorly on the size interval of interest ( $[0, 0.2]$ ). Hence the question remains, whether we should actually *prefer* the fast double bootstrap over the single bootstrap (the two having nearly similar computation costs). Tables 6, 7, and 8 compare the mean absolute discrepancy of the FDB (wild) and single bootstrap (wild) for both tests. For homo- and heteroskedastic data, the fast double bootstrap yields clearly better results for both tests, but especially for the Bierens test. For skewed data the pattern is not so clear, but the FDB is not strictly worse than the single bootstrap. However, given that the Rademacher-wild (FDB) performs better under skewness than the **Mammen**-wild, we can also see in the said tables that the Rademacher-wild FDB provides clear advantage over the single bootstrap under skewed data, for both tests.

### 4.3. Robustness experiments

Two questions remain to support our results. First, do the double bootstrap and FDB procedures alter the power curves compared to the single bootstrap in such a way that we have to take into account the tradeoff between level and power when choosing a bootstrap procedure? Second, do our results vary in a significant way by altering  $B_1$  and  $B_2$ ?

The answer to the first question is negative: the double and fast double bootstrap procedures simply shift upwards the power curve yielding higher level and also higher power under (local) alternatives, and they do not alter the slope of power curves. To demonstrate this, we conducted the power analysis for the following local alternatives:

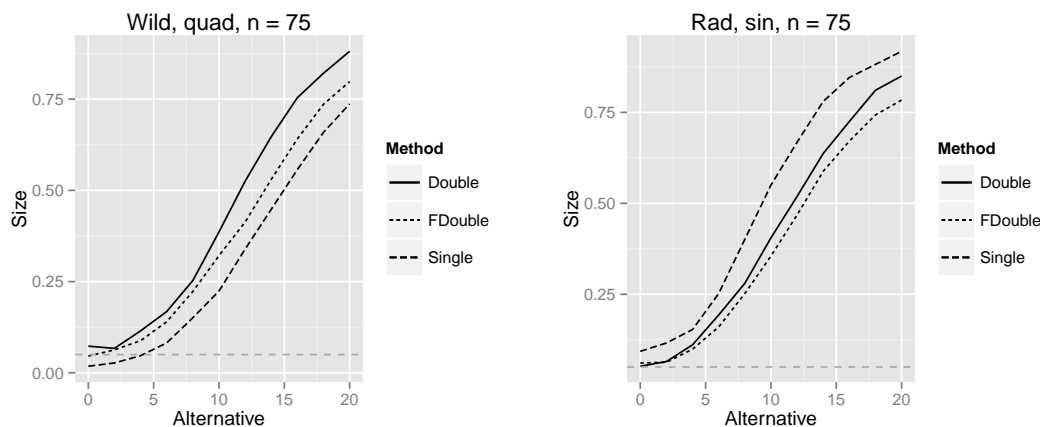
$$y_{\text{alt}} = y + d \cdot A(x'\gamma),$$

with

$$A(z) = \begin{cases} 0.1 \times (z)^2 & \text{(quadratic)} \\ 0.1 \times \sin(\pi z) & \text{(sine)} \end{cases}$$

The “direction”  $\gamma$  was chosen to be  $\gamma = (1, 2, 3, -2) / \sqrt{18}$ , and the “distance”  $d$  was set  $d \in \{0, 2, 4, \dots, 20\}$ .<sup>7</sup> Note, that the sine alternative is not easy to distinguish from the linear null. As of now, only the results for Bierens’s test are available, but we do not expect the conclusions to differ for the Escanciano test. The power curves are contained in Figure 4. The left figure shows the power curve under the quadratic alternative for  $n = 75$  and for the *Mammen-wild*. The right figure depicts the power curve for the *Rademacher-wild* and for the sine alternative. The nominal level is  $\alpha = 0.5$ . Note, that in

Figure 4: Power curve



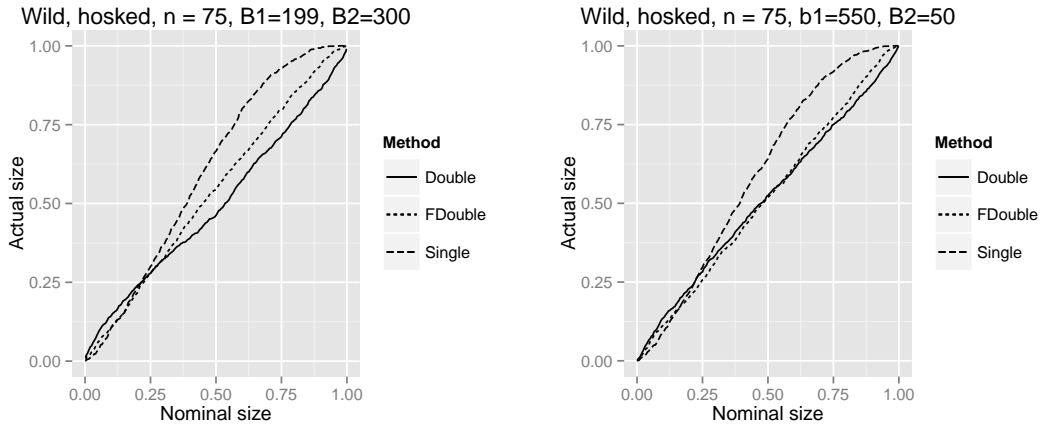
the right figure the power curve of the *single* bootstrap is above the other two procedures. This is in accord with our conclusions in the previous section: the actual level of the Rademacher-wild single bootstrap is significantly higher, which is undesirable. Besides this observation, the pattern holds for all alternative–bootstrap method combinations and supports our conclusions stated above.

Next we investigated the robustness of our results concerning the double bootstrap with respect to varying  $B_1$  and  $B_2$  in relation to each other. [Booth and Hall \(1994\)](#) derive optimal values establishing that  $B_2$  should be a constant multiple of  $\sqrt{B_1}$ . In practice this means that  $B_2$  is advised to be significantly lower than  $B_1$ , which is not the case in our experiments above. Figure 5 demonstrates two scenarios for the Bierens test, *Mammen-wild* bootstrap and  $n = 75$ . In the left figure we let  $B_1$  remain equal to 199 as in the original experiments and we increase  $B_2$  to 300 instead of the original 150. In the right figure, instead of decreasing only  $B_2$  to 50, we employ a  $B_1$  that is close to being optimal in relation with  $B_2 = 50$  according to [Booth and Hall \(1994\)](#).

The left figure supports the claims of [Booth and Hall \(1994\)](#) in that it shows that increasing  $B_2$  yields even worse results for the double bootstrap. This finding also supports our results with the setup  $150 = B_2 < B_1 = 199$ . The right figure indeed

<sup>7</sup>Note that  $\sqrt{18}$  is the norm of the vector  $(1, 2, 3, -2)$ .

Figure 5: Bootstrap variations



improves upon the picture in Figure 3, however, the double bootstrap curve is still above the fast double bootstrap curve. This finding, therefore, does not alter our conclusions regarding the double bootstrap. Overall, it can be concluded that varying  $B_1$  and  $B_2$  in relation to each other only marginally matters.

## 5. Conclusion

In the preceding pages we investigated the applicability of the double and fast double bootstrap procedures on the Bierens and Escanciano test. Section 2. and 3. provided exposition and theoretical motivation for the question. In Section 4. we used numerical experiments to investigate this question under various data generating processes. Based on the experimental evidence we have concluded the following:

1. The Rademacher distribution (Davidson and Flachaire, 2008) for the wild bootstrap should not be used to single bootstrap the investigated tests.
2. The two point distribution proposed by Mammen (1993) for the wild bootstrap is an adequate “omnibus” method in the sense that it performs reasonably well for all considered bootstrap procedures (single, double, and fast double bootstrap). Its performance is especially good under heteroskedasticity compared to the residual and Rademacher versions.
3. The Rademacher distribution for the wild bootstrap can outperform the Mammen-wild bootstrap especially with a non-symmetric error distribution.
4. The double bootstrap should not be used in small and moderate sized samples, because its performance (in terms of levels) is often worse than that of *both* the fast



double bootstrap and the single bootstrap.

5. The fast double bootstrap can generally be advised to be used for refining inference over the single bootstrap. Especially if it is coupled with the appropriate wild distribution. As a thumb rule, with the [Mammen-wild](#) method, the FDB performs no worse than the single bootstrap, and it performs observably better under the practically important case of heteroskedasticity.

While the above points should be interpreted with care especially because of numerical inaccuracies, we made all attempt to give conclusions that stem from relatively stable patterns.

## References

- BERAN, R. (1987): "Prepivoting to Reduce Level Error of Confidence Sets," *Biometrika*, 74(3), 457–458.
- (1988): "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83(403), 687–697.
- BIERENS, H. J. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20(1), 105–134.
- BIERENS, H. J., AND W. PLOBERGER (1997): "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65(5), 1129–1152.
- BOOTH, J. G., AND P. HALL (1994): "Monte Carlo Approximation and the Iterated Bootstrap," *Biometrika*, 81(2), 331–340.
- CARRASCO, M., AND J.-P. FLORENS (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 794–834.
- DAVIDSON, R., AND E. FLACHAIRE (2008): "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146(1), 162–169.
- DAVIDSON, R., AND J. G. MACKINNON (2007): "Improving the Reliability of Bootstrap Tests With the Fast Double Bootstrap," *Computational Statistics & Data Analysis*, 51(7), 3259–3281.
- DOMINGUEZ, M. A. (2005): "On the Power of Bootstrapped Specification Tests," *Econometric Reviews*, 23(3), 215–228.
- ESCANCIANO, J. C. (2006): "A Consistent Diagnostic Test For Regression Models Using Projections," *Econometric Theory*, 22(06), 1030–1051.
- GIACOMINI, R., D. N. POLITIS, AND H. WHITE (2013): "A Warp-Speed Method For Conducting Monte Carlo Experiments Involving Bootstrap Estimators," *Econometric Theory*, 29(3), 567–589.
- HALL, P. (1992): *The Bootstrap and Edgeworth Expansion*, Springer Series in Statistics. Springer-Verlag, New York, NY.
- HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21(4), 1926–1947.
- HOROWITZ, J. L. (2001): "The bootstrap," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, chap. 52, pp. 3159–3228. North-Holland.
- HSIAO, C., Q. LI, AND J. S. RACINE (2007): "A consistent model specification test with mixed discrete and continuous data," *Journal of Econometrics*, 140(2), 802–826.

- LI, Q., AND J. S. RACINE (2006): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, Princeton, NJ.
- MAMMEN, E. (1993): "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *The Annals of Statistics*, 21(1), 255–285.
- SHAO, J., AND D. TU (1995): *The Jackknife and Bootstrap*, Springer Series in Statistics. Springer-Verlag, New York, NY.
- STINCHCOMBE, M. B., AND H. WHITE (1998): "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric Theory*, 14(3), 295–325.
- ZHENG, J. X. (1996): "A consistent test of functional form via nonparametric estimation technique," *Journal of Econometrics*, 75(2), 263–289.

## A. Tables and figures

Figure 6: Escanciano test simulation

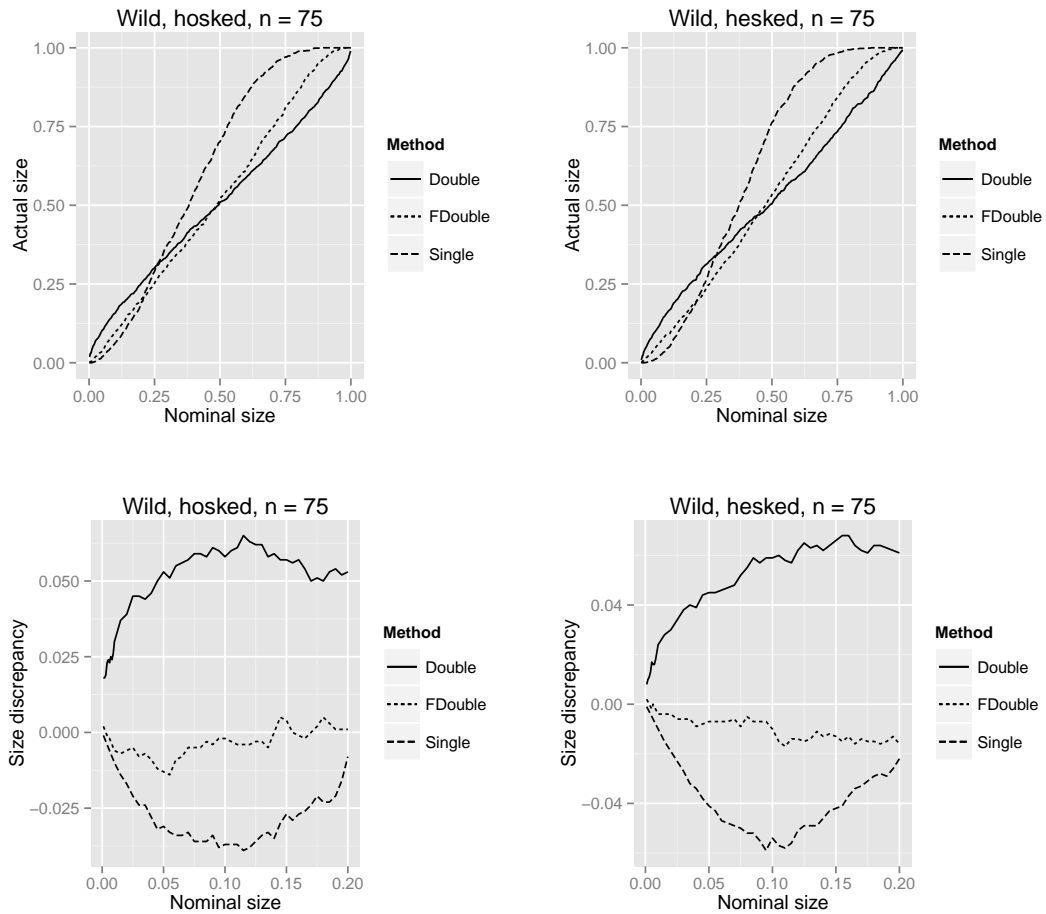


Figure 7: Residual bootstrap discrepancy



Figure 8: Rademacher (F)DB improvement for the Bierens test

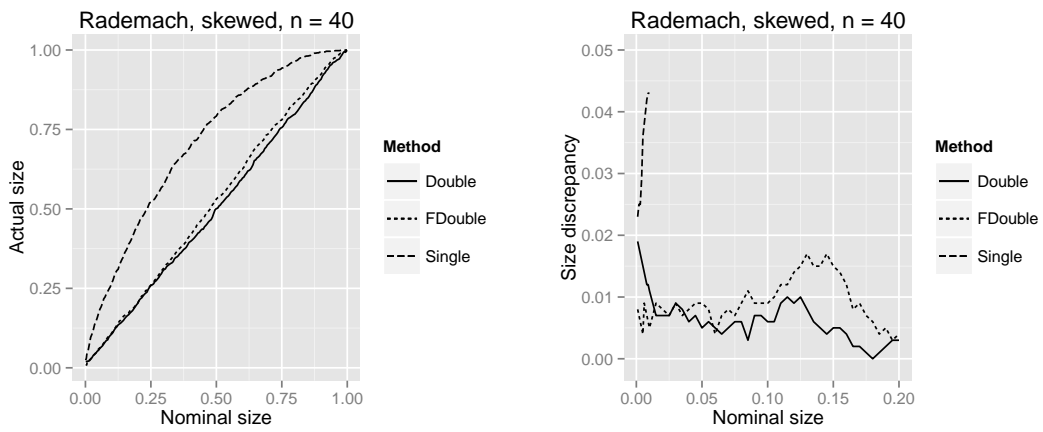


Table 3: Wild mean absolute size discrepancies for Bierens, homoskedastic data

Procedure		Sample size ( $n$ )			
		40	50	75	100
FDB	Residual	0.0080 <sub>(0.005)</sub>	0.0074 <sub>(0.005)</sub>	0.0031 <sub>(0.003)</sub>	0.0092 <sub>(0.006)</sub>
	Wild	0.0063 <sub>(0.003)</sub>	0.0095 <sub>(0.007)</sub>	0.0083 <sub>(0.004)</sub>	0.0088 <sub>(0.005)</sub>
	Rademacher-wild	0.0149 <sub>(0.008)</sub>	0.0060 <sub>(0.002)</sub>	0.0061 <sub>(0.003)</sub>	0.0042 <sub>(0.003)</sub>
Double	Residual	0.0050 <sub>(0.002)</sub>	0.0052 <sub>(0.003)</sub>	0.0105 <sub>(0.008)</sub>	0.0058 <sub>(0.003)</sub>
	Wild	0.0291 <sub>(0.013)</sub>	0.0278 <sub>(0.014)</sub>	0.0389 <sub>(0.012)</sub>	0.0196 <sub>(0.006)</sub>
	Rademacher-wild	0.0203 <sub>(0.003)</sub>	0.0140 <sub>(0.007)</sub>	0.0158 <sub>(0.003)</sub>	0.0194 <sub>(0.004)</sub>

Standard deviations in parentheses. Calculated from 48 observations on  $[0, 0.2]$ .

Table 4: Wild mean absolute size discrepancies for Bierens, heteroskedastic data

Procedure		Sample size ( $n$ )			
		40	50	75	100
FDB	Residual	0.0231 <sub>(0.017)</sub>	0.0202 <sub>(0.020)</sub>	0.0454 <sub>(0.026)</sub>	0.0498 <sub>(0.034)</sub>
	Wild	0.0196 <sub>(0.010)</sub>	0.0046 <sub>(0.003)</sub>	0.0044 <sub>(0.004)</sub>	0.0041 <sub>(0.003)</sub>
	Rademacher-wild	0.0233 <sub>(0.011)</sub>	0.0127 <sub>(0.007)</sub>	0.0174 <sub>(0.007)</sub>	0.0064 <sub>(0.003)</sub>
Double	Residual	0.0506 <sub>(0.030)</sub>	0.0660 <sub>(0.035)</sub>	0.0591 <sub>(0.035)</sub>	0.0716 <sub>(0.034)</sub>
	Wild	0.0293 <sub>(0.016)</sub>	0.0236 <sub>(0.013)</sub>	0.0223 <sub>(0.009)</sub>	0.0269 <sub>(0.012)</sub>
	Rademacher-wild	0.0320 <sub>(0.004)</sub>	0.0287 <sub>(0.007)</sub>	0.0203 <sub>(0.006)</sub>	0.0154 <sub>(0.004)</sub>

Standard deviations in parentheses. Calculated from 48 observations on  $[0, 0.2]$ .

Table 5: Wild mean absolute size discrepancies for Bierens, skewed data

Procedure		Sample size ( $n$ )			
		40	50	75	100
FDB	Residual	0.0163 <sub>(0.008)</sub>	0.0153 <sub>(0.009)</sub>	0.0163 <sub>(0.012)</sub>	0.0072 <sub>(0.005)</sub>
	Wild	0.0084 <sub>(0.004)</sub>	0.0215 <sub>(0.014)</sub>	0.0138 <sub>(0.007)</sub>	0.0100 <sub>(0.006)</sub>
	Rademacher-wild	0.0089 <sub>(0.004)</sub>	0.0056 <sub>(0.003)</sub>	0.0075 <sub>(0.004)</sub>	0.0043 <sub>(0.002)</sub>
Double	Residual	0.0098 <sub>(0.005)</sub>	0.0039 <sub>(0.003)</sub>	0.0056 <sub>(0.005)</sub>	0.0058 <sub>(0.003)</sub>
	Wild	0.0144 <sub>(0.008)</sub>	0.0156 <sub>(0.010)</sub>	0.0129 <sub>(0.007)</sub>	0.0306 <sub>(0.015)</sub>
	Rademacher-wild	0.0074 <sub>(0.005)</sub>	0.0088 <sub>(0.004)</sub>	0.0166 <sub>(0.009)</sub>	0.0068 <sub>(0.004)</sub>

Standard deviations in parentheses. Calculated from 48 observations on  $[0, 0.2]$ .

Table 6: FDB–Single mean absolute size discrepancies, homoskedastic data

Procedure		Sample size ( $n$ )			
		40	50	75	100
Bierens	FDB (wild)	0.0063 <sub>(0.003)</sub>	0.0095 <sub>(0.007)</sub>	0.0083 <sub>(0.004)</sub>	0.0088 <sub>(0.005)</sub>
	Single (wild)	0.0199 <sub>(0.019)</sub>	0.0158 <sub>(0.015)</sub>	0.0100 <sub>(0.009)</sub>	0.0118 <sub>(0.010)</sub>
Escanciano	FDB (wild)	0.0307 <sub>(0.016)</sub>	0.0217 <sub>(0.012)</sub>	0.0042 <sub>(0.003)</sub>	0.0033 <sub>(0.002)</sub>
	Single (wild)	0.0287 <sub>(0.017)</sub>	0.0223 <sub>(0.015)</sub>	0.0242 <sub>(0.012)</sub>	0.0149 <sub>(0.008)</sub>

Standard deviations in parentheses. Calculated from 48 observations on  $[0, 0.2]$ .

Table 7: FDB–Single mean absolute size discrepancies, heteroskedastic data

Procedure		Sample size ( $n$ )			
		40	50	75	100
Bierens	FDB (wild)	0.0196 <sub>(0.010)</sub>	0.0046 <sub>(0.003)</sub>	0.0044 <sub>(0.004)</sub>	0.0041 <sub>(0.003)</sub>
	Single (wild)	0.0209 <sub>(0.020)</sub>	0.0176 <sub>(0.013)</sub>	0.0148 <sub>(0.008)</sub>	0.0120 <sub>(0.009)</sub>
Escanciano	FDB (wild)	0.0219 <sub>(0.010)</sub>	0.0321 <sub>(0.017)</sub>	0.0076 <sub>(0.004)</sub>	0.0093 <sub>(0.004)</sub>
	Single (wild)	0.0291 <sub>(0.019)</sub>	0.0271 <sub>(0.017)</sub>	0.0235 <sub>(0.014)</sub>	0.0231 <sub>(0.014)</sub>

Standard deviations in parentheses. Calculated from 48 observations on  $[0, 0.2]$ .

Table 8: FDB–Single mean absolute size discrepancies, skewed data

Procedure		Sample size ( $n$ )			
		40	50	75	100
Bierens	FDB (Rademach)	0.0089 <sub>(0.004)</sub>	0.0056 <sub>(0.003)</sub>	0.0075 <sub>(0.004)</sub>	0.0043 <sub>(0.002)</sub>
	FDB (wild)	0.0084 <sub>(0.004)</sub>	0.0215 <sub>(0.014)</sub>	0.0138 <sub>(0.007)</sub>	0.0100 <sub>(0.006)</sub>
	Single (wild)	0.0218 <sub>(0.018)</sub>	0.0167 <sub>(0.011)</sub>	0.0134 <sub>(0.009)</sub>	0.0208 <sub>(0.012)</sub>
Escanciano	FDB (Rademach)	0.0126 <sub>(0.010)</sub>	0.0098 <sub>(0.008)</sub>	0.0025 <sub>(0.002)</sub>	0.0061 <sub>(0.005)</sub>
	FDB (wild)	0.0439 <sub>(0.028)</sub>	0.0317 <sub>(0.020)</sub>	0.0326 <sub>(0.021)</sub>	0.0270 <sub>(0.018)</sub>
	Single (wild)	0.0289 <sub>(0.019)</sub>	0.0337 <sub>(0.018)</sub>	0.0374 <sub>(0.021)</sub>	0.0332 <sub>(0.019)</sub>

Standard deviations in parentheses. Calculated from 48 observations on  $[0, 0.2]$ .