

The Exact Wild Bootstrap

Bruce E. Hansen*

University of Wisconsin[†]

August 2026

Abstract

This paper proposes new methods for exact computation of wild bootstrap p-values and confidence intervals without simulation. The methods are applicable to regression coefficients in linear models. The proposed methods are considerably more accurate, and computationally much faster, than conventional simulation methods. The methods are developed for calculation of the bootstrap distribution, p-values for tests on linear coefficients, and confidence intervals for linear coefficients.

*Research support from the Phipps Chair is gratefully acknowledged.

[†]Department of Economics, 1180 Observatory Drive, University of Wisconsin, Madison WI 53706.

1 Introduction

This paper proposes new methods for exact computation of wild bootstrap p-values and confidence intervals without simulation. The methods are applicable to regression coefficients in linear models. The proposed methods are considerably more accurate, and computationally much faster, than conventional simulation methods. The methods apply exclusively to the wild bootstrap with normal auxiliary random variables, and can be applied for inference on any t-ratio constructed with a variance estimator which is a quadratic form in the dependent variable, which includes all standard heteroskedasticity-consistent variance estimators. The methods extend to the cluster-robust wild bootstrap, though this extension is not detailed in this paper.

The wild bootstrap is appropriate for linear regression models with conditional heteroskedasticity. The family of heteroskedasticity-consistent variance estimators are often written with the labels HC_0 , HC_1 , HC_2 , and HC_3 . The HC_0 version was introduced by Eicker (1963), Huber (1967), and White (1980). The degrees of freedom correction known as HC_1 was suggested by Hinkley (1977), and became ubiquitous in applied econometric practice by its designation as the default “r” robust option in Stata. Together, HC_0 and HC_1 are known as Eicker-Huber-White (EHW) variance estimators. HC_2 and HC_3 were introduced by MacKinnon and White (1985) as unbiased estimators under homoskedasticity and the jackknife principle, respectively.

The wild bootstrap was developed by Liu (1988) and Mammen (1993), building on suggestions of Wu (1986) and Beran (1986). Papers in the econometrics literature which develop aspects of the wild bootstrap for linear regression include Davidson and Flachaire (2008), MacKinnon (2015, 2023), MacKinnon and Webb (2017, 2018, 2020), Djogbenou, MacKinnon, and Nielsen (2019), Roodman, MacKinnon, Nielsen, and Webb (2019), Canay, Santos, and Shaikh (2021), and MacKinnon, Nielsen, and Webb (2021, 2023a, 2023b, 2023c).

This paper proposes two algorithms for computation of the wild bootstrap distribution of a regression t-ratio. The algorithms are based on the theory of Hansen (2026b), which develops an integral representation for the exact distribution of regression t-ratios in the normal regression model. This theory applies to the wild bootstrap when the bootstrap auxiliary variables are normally distributed. The first algorithm is an exact calculation, and is most appropriately applied in small samples (e.g., $n \leq 1000$). The second algorithm is based on an eigenvalue approximation, and can be applied for any sample size.

The methods in this paper are developed for the calculation of the bootstrap distribution of regression t-statistics for linear coefficients, for p-values of tests of linear hypotheses, and for confidence intervals for such coefficients.

The paper is organized as follows. Section 2 illustrates the problem in a simple numerical example. Section 3 is a general description of the wild bootstrap for regression models. Section 4 presents the theory for the wild bootstrap distribution. Section 5 presents the theory for the wild bootstrap distribution under a simplified eigenvalue structure. Section 6 shows how to use the method to plot a wild bootstrap distribution function. Section 7 discusses calculation of wild bootstrap critical values. Section 8 extends the results to the restricted wild bootstrap. Section 9 is a conclusion. Appendix A contains a description of the randomized low rank eigenvalue decomposition method used for the eigenvalue approximation.

Appendix B describes the proofs of the main theory.

The numerical work reported in this paper was written in R. The code is posted on the author’s website.

2 Illustration

We illustrate the problem using a simple example taken from Davidson and Flachaire (2008). These authors generated a sample of $n = 10$ regressors X_{1i} with strong leverage. They did so by drawing 10 values from the $N(0, 1)$ distribution, and replacing one value with the number 10. These regressors were fixed and are printed in Table 1 of their paper. Indeed this sample has strong leverage, as the largest leverage value is 0.93 (while the sum of the remaining leverage values is 0.07).

For our experiment we take this regressor as given, and using it, generate a sample of dependent variables from the normal model $Y_i \sim N(0, \sigma_i^2)$ with the standard deviation either homoskedastic $\sigma_i = 1$ or heteroskedastic $\sigma_i = |X_{1i}|$ (again following Davidson-Flachaire).

We then we calculate the least squares regression coefficient and its HC₁ standard error. Which critical value should we use to construct a 95% confidence interval? The conventional t_9 critical value of 2.26 is likely to be too small given the strong leverage present in the regressors, a problem expositied by Chesher and Jewitt (1987), Chesher (1989), and Chesher and Austin (1991). However, as the error distribution is normal, the exact critical value can be calculated by the method of Hansen (2026b). These 95% critical values are displayed in the first column of Table 1, and equal 7.69 and 28.78 for the two variance specifications. What we can see is: (1) the exact critical values vastly exceed the conventional t_9 critical value; (2) the exact critical values vary considerably depending on the true (unknown) error variances.

Another way to illustrate the problem is to calculate the rejection probability of a two-sided test using the conventional 2.26 critical value. These probabilities can also be calculated using the exact method of Hansen (2026b), are displayed in the second column of Table 1, and equal 0.46 and 0.90 for the two variance specifications. Again, we see that that the rejection probabilities are considerably greater than the nominal target of 0.05, and these probabilities vary considerably depending on the unknown error variances.

What can be done? One approach is to use the critical value associated with a reference model. In the regression context the reference model is typically taken as the homoskedastic case, which in this example means the critical value of 7.69. A downside with this suggestion is that it the critical value does not adapt to the possible presence of heteroskedasticity. In this example, the critical value of 7.69 is too small when the the error variances are heteroskedastic.

A second approach, advocated by Pötscher and Preinerstorfer (2023), is to seek the worst-case critical value across all possible values of the error variances σ_i . There are two difficulties with this suggestion. First, this optimization is computationally challenging (if we allow for arbitrary variances) and may be infeasible when n is moderately large. Second, as we can see from Table 1, the worst-case 95% critical value is at least 28.78. The Pötscher-Preinerstorfer suggestion is to use this value, regardless of the context. This will produce unreasonably excessively large confidence intervals in most contexts.

The bootstrap enacts a compromise: use the data to inform us about the unknown error variances. The wild bootstrap, in particular, uses a residual as an estimate of the unknown error standard deviation. While this is a noisy estimate, it is still informative. To illustrate, the rejection probabilities of a two-sided test using our recommended¹ wild bootstrap p-value are displayed² in the third column of Table 1, and in the final column we report the rejection probability when we make the further adjustment to use HC₃ standard errors instead of HC₁ standard errors. What we see is that these rejection probabilities are much closer to the nominal 5%. In particular, the rejection probabilities of our wild bootstrap using HC₃ standard errors are 8-9% in the two variance specifications. This is remarkably excellent performance given the small sample size, high regressor leverage, and dismal performance of the conventional inference methods.

Table 1: Illustration

	95% Exact Critical Value	Rejection Probability		
		t_9	Wild Bootstrap	Using HC ₃
$\sigma_i = 1$	7.69	0.455	0.072	0.090
$\sigma_i = X_{1i} $	28.78	0.902	0.165	0.078

3 Wild Bootstrap

The context is the linear regression

$$Y_i = X_{1i}\theta + X_{2i}'\beta + e_i$$

where X_{1i} and θ are scalars, and X_{2i} and β are $(k-1) \times 1$. The regressors have been constructed so that the hypothesis of interest is $H_0 : \theta = 0$. Written in matrix notation the regression is

$$\mathbf{Y} = \mathbf{X}_1\theta + \mathbf{X}_2\beta + \mathbf{e} \quad (1)$$

where $\mathbf{Y}$, $\mathbf{X}_1$, $\mathbf{e}$ are $n \times 1$, and $\mathbf{X}_2$ is $n \times (k-1)$. Write $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, which we assume has full column rank. Let $(\hat{\theta}, \hat{\beta})$ denote the least squares estimator of (θ, β) , and let $\hat{e}_i = Y_i - X_{1i}\hat{\theta} - X_{2i}'\hat{\beta}$ denote the least squares residual.

Let $\hat{v}$ be a standard error for $\hat{\theta}$. (We discuss specific choices in Section 3.1.) Define the absolute value of the t -ratio

$$T = \left| \frac{\hat{\theta}}{\hat{v}} \right|. \quad (2)$$

Let $\hat{\sigma}_i$ denote an estimate of the error standard deviation $\sigma_i = (\text{var}[e_i])^{1/2}$. We discuss choices in Section 3.2.

¹Our recommended method generates wild bootstrap observations using the SD₃ estimator for the error standard deviation (see Section 3.2) and normal auxiliary random variables.

²These are calculated by simulation using 50,000 simulation replications.

Let ξ_i denote an i.i.d. auxillary random variable, independent of the data, satisfying $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[\xi_i^2] = 1$. The choice of ξ_i is discussed in Section 3.3.

Given $\hat{\sigma}_i$ and ξ_i the bootstrap observations are defined as $Y_i^* = \hat{\sigma}_i \xi_i$. Repeating for all observations we obtain the bootstrap sample $\mathbf{Y}^*$. The wild bootstrap statistics are constructed by applying the estimators to the bootstrap sample. Least squares regression of $\mathbf{Y}^*$ on $\mathbf{X}$ yields the coefficients $(\hat{\theta}^*, \hat{\beta}^*)$, with a standard error $\hat{v}^*$ for $\hat{\theta}^*$. The bootstrap t-ratio is

$$T^* = \left| \frac{\hat{\theta}^*}{\hat{v}^*} \right|.$$

The wild bootstrap distribution is the conditional distribution of T^* given the original observations:

$$F^*(x) = \Pr[T^* \leq x \mid \mathbf{X}, \hat{\sigma}].$$

The wild bootstrap p-value for the test of $H_0: \theta = 0$ is

$$p^* = 1 - F^*(T).$$

The wild bootstrap critical value c^* is the $1 - \alpha$ quantile of the distribution F^* . The symmetric wild bootstrap confidence interval for θ is

$$CI^* = \hat{\theta} \pm c^* \hat{v}.$$

The bootstrap p-value and confidence interval are determined by the bootstrap distribution $F^*(x)$. In practice this distribution is unknown. The standard implementation is to estimate $F^*(x)$ by simulation. This is done as follows. The user simulates n i.i.d. draws ξ_i and constructs the statistic T^* . This is one draw from the bootstrap distribution. The user repeats this procedure B times. The estimate of $F^*(x)$ is the empirical distribution $F_B^*(x)$ of the simulated T^* . The estimates of the bootstrap p-value and confidence interval are $p_B^* = 1 - F_B^*(T)$ and $CI_B^* = \hat{\theta} \pm c_B^* \hat{v}$ where c_B^* is a $1 - \alpha$ quantile estimate from the empirical distribution F_B^* .

For this calculation, the number of bootstrap replications B induces a trade-off between computation cost and accuracy. To assess accuracy, observe that to obtain simulation estimates within δ of the bootstrap p-value with 95% probability, a user needs to set³ $B \geq 3.84/4\delta^2$. This requires $B \geq 9600$ if $\delta = 0.01$, $B \geq 38,400$ if $\delta = 0.005$, or $B \geq 960,000$ if $\delta = 0.001$. Thus, for accurate bootstrap p-value estimation, a large number of simulation draws are required, which can be computationally costly.

To illustrate the computation cost for the wild bootstrap, Table 2 displays the computation time (in seconds) for the calculation⁴ of the wild bootstrap⁵ distribution at a single point⁶. The first panel displays the time required for the simulation method using $B = 20,000$ bootstrap replications. The calculation is

³As $B \rightarrow \infty$ the simulation estimate p_B^* is asymptotically normal with variance $p^*(1 - p^*)/B \leq 1/4B$. Thus $\Pr\left[\left|p_B^* - p^*\right| \leq \delta\right] \simeq \Pr\left[\chi_1^2 \leq \delta^2 B / p(1 - p)\right] \geq \Pr\left[\chi_1^2 \leq \delta^2 4B\right] \geq 0.95$ if $\delta^2 4B \geq 3.84$ as claimed.

⁴The calculations were performed in R under Windows 11 on a standard office computer.

⁵Our recommended procedure, using HC₃ standard errors, SD₃ standard deviation estimates, and normal auxillary variables.

⁶The bootstrap distribution $F^*(x)$ where x equals the 97.5% quantile of the t_{n-k} distribution.

Table 2: Computation Time (Seconds) for Bootstrap Distribution Calculation

	$k = 20$	$k = 50$	$k = 100$	$k = 200$
Simulation Method				
$n = 100$	0.19	0.26	NA	NA
$n = 500$	0.64	1.15	1.93	3.33
$n = 1000$	1.34	2.25	3.69	6.92
$n = 5000$	6.42	11.05	18.45	34.67
$n = 10,000$	12.69	22.56	37.91	73.89
$n = 100,000$	137.98	283.21	513.55	978.05
Algorithm 1				
$n = 100$	0.001	0.001	NA	NA
$n = 500$	0.027	0.034	0.046	0.076
$n = 1000$	0.180	0.203	0.234	0.328
$n = 5000$	22.212	22.759	23.400	25.141
$n = 10,000$	NA	NA	NA	NA
$n = 100,000$	NA	NA	NA	NA
Algorithm 2				
$n = 100$	0.001	0.001	NA	NA
$n = 500$	0.002	0.004	0.011	0.035
$n = 1000$	0.003	0.007	0.021	0.071
$n = 5000$	0.014	0.041	0.118	0.412
$n = 10,000$	0.026	0.084	0.245	0.847
$n = 100,000$	0.310	0.950	3.040	10.340

performed for a variety of sample sizes n and regressors k . The calculation takes less than one second for small n and k , but is increasing in n and k .

3.1 Standard Errors

The t-ratio (2) is constructed with a standard error $\hat{v}$ which is the square root of an estimator $\hat{v}^2$ of $\text{var}[\hat{\theta}]$. Our exact method applies to any estimator $\hat{v}^2$ which can be written as a non-negative quadratic in the dependent variable.

Assumption 1 $\hat{v}^2 = \mathbf{Y}' \mathbf{A}' \mathbf{A} \mathbf{Y}$ for some matrix $\mathbf{A}$ which is a function only of the regressors $\mathbf{X}$.

Assumption 1 includes the common HC_0 , HC_1 , HC_2 , and HC_3 variance estimators, which we write as $\hat{v}_0^2$, $\hat{v}_1^2$, $\hat{v}_2^2$, and $\hat{v}_3^2$.

In current econometric practice, HC_1 is the dominant variance estimator for cross-section regression, perhaps due to its designation as the default “r” robust option in Stata. However, numerous theoretical and simulation investigations, including MacKinnon and White (1985), Chesher and Austin (1991), Davidson and Flachaire (2008), MacKinnon, Nielsen, and Webb (2023a, 2023b, 2023c), and Hansen (2026a), have demonstrated advantages of HC_3 over the other estimators. Consequently, we repeat this endorsement here, and recommend HC_3 (the jackknife) for empirical practice. However, our results apply to any HC variance estimator as described.

For HC_0 , HC_1 , HC_2 , and HC_3 , the variance estimator $\hat{v}^2$ takes the form in Assumption 1, with $\mathbf{A} = \mathbf{DM}$, where $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the standard regression projection matrix and $\mathbf{D}$ is a diagonal matrix with diagonal elements d_i which vary across estimators. These elements equal, respectively,

$$\begin{aligned} d_{0i} &= |Z_i| \\ d_{1i} &= \sqrt{\frac{n}{n-k}} |Z_i| \\ d_{2i} &= \frac{1}{\sqrt{1-m_i}} |Z_i| \\ d_{3i} &= \frac{1}{1-m_i} |Z_i| \end{aligned}$$

where $m_i = \mathbf{X}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i$ are the leverage values, and

$$Z_i = X_{1i} - X_{2i}'(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1 \quad (3)$$

are the residuals from the regression of X_{1i} on X_{2i} .

The standard definitions for HC_2 and HC_3 , and the expressions above, require $m_i < 1$ for all i , or equivalently, that $\mathbf{X}$ maintains full column rank after deleting any individual observation. This excludes models with regressors which are single-observation dummy variables. As discussed in Hansen (2026a), in this context a modified HC_3 estimator exists based on the jackknife principle. This estimator also takes the quadratic form $\hat{v}^2 = \mathbf{Y}'\mathbf{A}'\mathbf{A}\mathbf{Y}$ so Assumption 1 holds, but $\mathbf{A}$ does not take the simple form $\mathbf{A} = \mathbf{DM}$. The exact theory of Section 4 applies, but the approximation method of Section 5 does not apply without modification. We do not explore this extension in this paper.

3.2 Standard Deviation Estimates

The wild bootstrap distribution $Y_i^* = \hat{\sigma}_i \xi_i$ requires a standard deviation estimator $\hat{\sigma}_i$ to capture the unknown heteroskedasticity of the errors e_i . Liu (1988) and Mammen (1993) set

$$\hat{\sigma}_{0i} = |\hat{e}_i|$$

where $\hat{e}_i$ is the least squares residual. An interesting implication of this choice is $\text{var}[\hat{\theta}^*] = \hat{v}_0^2$. Thus, when $\hat{\sigma}_{0i}$ is used, the variance of the bootstrap estimator exactly equals the HC_0 variance estimator. In analogy, we call $\hat{\sigma}_{0i}$ the SD_0 standard deviation estimator.

As we have discussed, it is well known that HC_0 is biased downwards, and in highly leveraged contexts can be a poor estimator of the sampling variance. This means that when SD_0 is used, the variance of the bootstrap estimator can be considerably biased downwards.

Alternatively, we can construct standard deviation estimators SD_1 , SD_2 , and SD_3 which are analogs

of HC₁, HC₂, and HC₃, respectively. These are

$$\begin{aligned}\hat{\sigma}_{1i} &= \sqrt{\frac{n}{n-k}} |\hat{e}_i| \\ \hat{\sigma}_{2i} &= \frac{1}{\sqrt{1-m_i}} |\hat{e}_i|\end{aligned}$$

and

$$\hat{\sigma}_{3i} = \frac{1}{1-m_i} |\hat{e}_i|.$$

We can also write $\hat{\sigma}_{3i} = |\tilde{e}_i|$ where $\tilde{e}_i = Y_i - X_{1i}\hat{\theta}_{-i} + X'_{2i}\hat{\beta}_{-i}$ and $(\hat{\theta}_{-i}, \hat{\beta}_{-i})$ is the leave-one-out least squares estimator. $\tilde{e}_i$ is known as the leave-one-out prediction error.

It is straightforward to calculate that when $\hat{\sigma}_{1i}$, $\hat{\sigma}_{2i}$, or $\hat{\sigma}_{3i}$ is used, then $\text{var}[\hat{\theta}^*] = \hat{v}_1^2$, $\text{var}[\hat{\theta}^*] = \hat{v}_2^2$, and $\text{var}[\hat{\theta}^*] = \hat{v}_3^2$, respectively. This means that when by using one of these choices, the variance of the bootstrap estimator will exactly equal an HC variance estimator. One implication is that the wild bootstrap should never be used as a method to calculate a standard error for a linear regression coefficient. The simulation estimate is simply the HC estimator plus noise. It is also worth pointing out that bootstrap p-values and confidence intervals calculated with SD₀ and SD₁ will be identical, as the scaling appears in both the numerator and denominator, and hence cancels. Since HC₃ generally has the best performance as a variance estimator, it seems reasonable to expect that SD₃ will have the best performance as a standard deviation estimator, and is hence our recommended choice.

This conjecture is consistent with the simulations and recommendations reported in Davidson and Flachaire (2008) and MacKinnon, Nielsen, and Webb (2023a), which compare the performance of wild bootstrap methods using SD₀ and SD₃. These authors found that SD₃ generally produces p-values with better size control. In Section 3.4 we extend this analysis and provide further strong evidence for the recommendation of SD₃ over SD₀.

The standard deviation estimates $\hat{\sigma}_{0i}$ through $\hat{\sigma}_{3i}$ as described are all unrestricted estimators; that is, they do not impose the null hypothesis to be tested. It is also possible to use restricted estimators which are estimated under the null hypothesis. Such estimators have been recommended by several authors. They have distinct properties and will be discussed separately in Section 8.

3.3 Auxillary Variables

The wild bootstrap distribution $Y_i^* = \hat{\sigma}_i \xi_i$ requires an auxillary random variable ξ_i satisfying $\mathbb{E}[\xi_i] = 0$ and $\mathbb{E}[\xi_i^2] = 1$. This induces the bootstrap randomness. Liu (1988) and Mammen (1993) recommended restricting attention to distributions satisfying $\mathbb{E}[\xi_i^3] = 1$ in order to establish asymptotic refinements. Liu (1988) proposed two continuous distributions, Mammen (1993) one discrete distribution⁷, and Webb (2023) a six-point distribution, which satisfy this property.

Davidson and Flachaire (2008) provided a counter-argument, pointing out that other useful properties hold when ξ_i is symmetrically distributed and satisfies $\mathbb{E}[\xi_i^4] = 1$. The latter is uniquely satisfied by the

⁷Mammen (1993) also proposed two continuous distributions. However, one is identical to that proposed by Liu (1988) and the second does not satisfy the property $\mathbb{E}[\xi_i^3] = 1$ as claimed.

Rademacher distribution. Combined with simulation results comparing the Mammen and Rademacher distributions, Davidson-Flachaire recommended the Rademacher distribution for applications. Based on their recommendation, this has become a common choice in econometric applications.

In contrast to the above recommendations, in this paper we advocate for the standard normal distribution, $\xi_i \sim N(0, 1)$. Our primary motivation is practical, as this choice permits the application of the exact distribution theory of Hansen (2026b) and hence exact and computationally efficient calculation of the wild bootstrap distribution. However, there are additional advantages of the normal distribution which can be articulated.

First, $\xi_i \sim N(0, 1)$ implies the bootstrap distributions of the coefficient estimator $\hat{\theta}^*$ and t-ratio T^* are continuous and differentiable. One implication is that the bootstrap critical value c^* is uniquely defined. It also implies that higher-order expansions of the bootstrap distributions are likely to exist. In contrast, when ξ_i is discrete then the bootstrap distributions are discrete, critical values are not uniquely defined, and expansions do not exist. Some of these problems are explored in MacKinnon and Webb (2018). These distinctions may be important in small samples or when the regressors have high leverage. In such contexts, the discreteness of the bootstrap distribution can result in poor coverage properties.

Second, the choice $\xi_i \sim N(0, 1)$ works well in simulations, especially in contexts of high leverage. This observation emerged in the simulation work reported in Hansen (2026a). We report additional numerical findings in in the next section.

3.4 Illustration Revisited

We return to the illustration of Section 2, which is based on the example of Davidson and Flachaire (2008). As we described, these authors generated a sample of $n = 10$ regressors X_{1i} with high leverage for a simple ($k = 1$) linear regression. Furthermore, they generated five additional regressors. Their second regressor X_{2i} was set at 1, and their subsequent regressors were drawn from the $N(0, 1)$ distribution. These were all fixed and reported in their Table 1. We use these values, and report results for models with $k = 1$, $k = 2$, and $k = 3$ regressors.

Following Davidson and Flachaire (2008) we generate the regression errors as $e_i \sim \sigma_i \zeta_i$ with $\sigma_i = |X_{1i}|$ and $\zeta_i \sim N(0, 1)$ (normal errors) or $\zeta_i \sim \chi_2^2/2 - 1$ (skewed errors). We generate 50,000 samples from each model, and on each sample calculate a wild bootstrap p-value for a test of significance for the coefficient on X_{1i} . We explore 8 implementations of the wild bootstrap, by combination of HC_1 and HC_3 standard errors, SD_1 and SD_3 standard deviation estimates, and Rademacher (F_1) and Normal (F_2) auxillary variables. For the case of normal auxillary variables we use the exact methods described in Section 4 (and labeled as “Algorithm 1”). For the case of Rademacher auxillary variables we use simulation with 10,000 simulation replications.

For each case we calculate the p-value discrepancy, as proposed by Davidson and MacKinnon (1998), which is the difference between the empirical distribution of the bootstrap p-values and their nominal size (the 45° line). Ideally, the empirical distribution should be uniform and the p-value discrepancy should equal zero. Deviations from zero equal the distortions of reported p-values from their correct values. For example, if the p-value discrepancy at $x = 0.08$ is 0.05, then the wild bootstrap will incorrectly

report a p-value of 0.08 as 0.13. The p-value discrepancy plot is a graph of the p-value discrepancy (on the y axis) against the nominal size (on the x axis). These are displayed in Figure 1 (for normal errors) and Figure 2 (for skewed errors).

In each figure, six plots are displayed. Those on the left correspond to the wild bootstrap using HC_1 standard errors; those on the right use HC_3 standard errors. The three rows correspond to $k = 1$, $k = 2$, and $k = 3$. Each plot displays four p-value discrepancy lines, for different combinations of the standard deviation estimator and auxillary distribution. In each plot, the top two lines correspond to SD_1 standard deviation estimators and the bottom two lines correspond to SD_3 standard deviation estimators. The lines labeled “F₁” use Rademacher auxillary variables, the lines labeled “F₂” use normal auxillary variables.

Better methods are those with the smallest p-value discrepancies, especially for smaller values of the nominal size. No single method dominates the others (in terms of having uniformly smaller distortions). However, several strong patterns emerge. First, in every context, SD_3 strongly dominates SD_1 . This strongly reinforces the findings of Davidson and Flachaire (2008) and MacKinnon, Nielsen, and Webb (2023a). Second, in general, HC_3 dominates HC_1 . This in particular holds for the case of normal auxillary variables, where the p-value discrepancy plots are uniformly reduced by the use of HC_3 . Third, neither Rademacher nor normal auxillary variables dominate the other. Rademacher auxillaries work relatively poorly for $k = 1$ (the case of strongest leverage) but relatively better for $k = 3$. The differences are small in most cases. Finally, we observe that the p-value discrepancy plots for normal auxillary variables are uniformly positive, implying that wild bootstrap p-values will be too large, and hence conservative.

Our deduction is that the Davidson-Flachaire endorsement of the Rademacher distribution should be tempered. Normal auxillary variables produce excellent results, and may be preferred under extreme leverage. Combined with the computational advantages which arise in the next sections, we endorse the use of normal auxillary random variables for wild bootstrap application, and proceed for the remainder of the paper under this assumption.

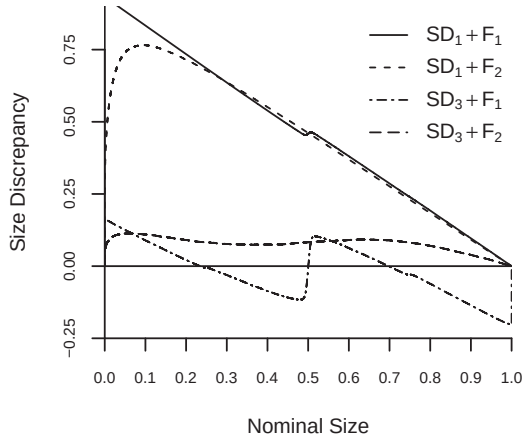
In summary, the simulations confirm our theoretical recommendations. For the wild bootstrap, we recommend using HC_3 standard errors, SD_3 standard deviation estimates, and normal auxillary random variables.

4 Bootstrap Distribution

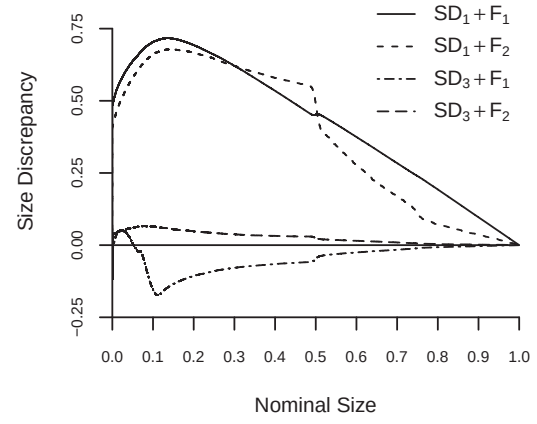
We now restrict attention to normal auxillary variables.

Assumption 2 *The auxillary variables ξ_i are independent of the sample, i.i.d., and satisfy $\xi_i \sim N(0, 1)$.*

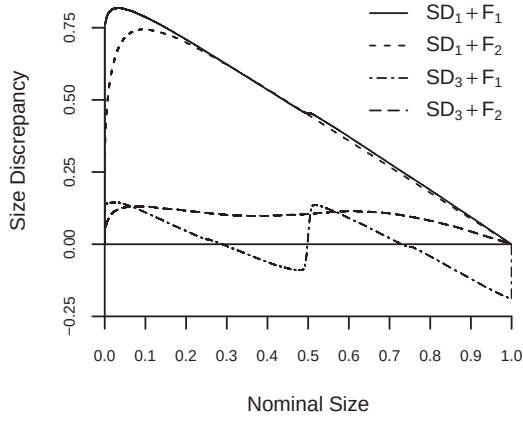
Recall that in Section 3.1 we assumed that the standard error $\hat{v}$ is the square root of an estimator $\hat{v}^2 = \mathbf{Y}'\mathbf{A}'\mathbf{A}\mathbf{Y}$ for some matrix $\mathbf{A}$. Let $\mathbf{Z}$ be the $n \times 1$ vector of stacked Z_i from (3). Let $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_n^2)$ where $\hat{\sigma}_i$ are the standard deviation estimators. Let s equal the number of observations with $\hat{\sigma}_i > 0$. In most applications, $s = n$, but in atypical cases we can have $\hat{\sigma}_i = 0$ for some observations. Factor $\hat{\Sigma} = \mathbf{S}\mathbf{S}'$



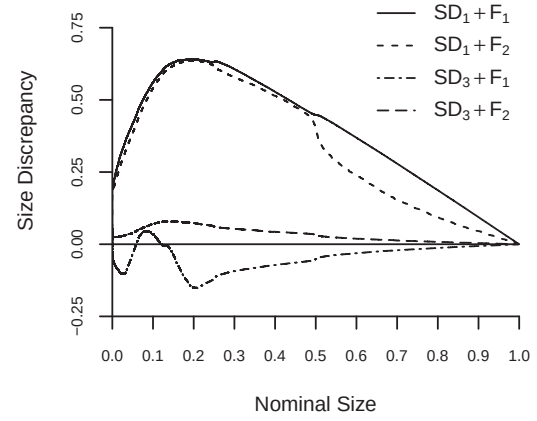
(a) $k = 1, HC_1$



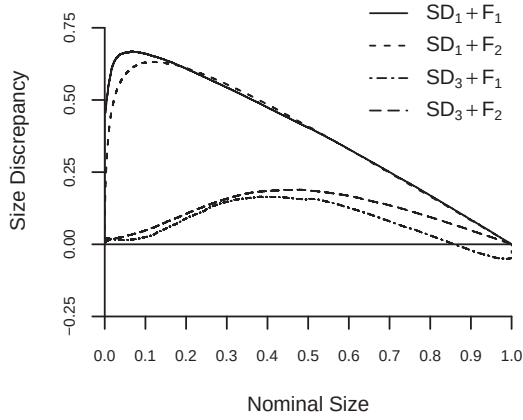
(b) $k = 1, HC_3$



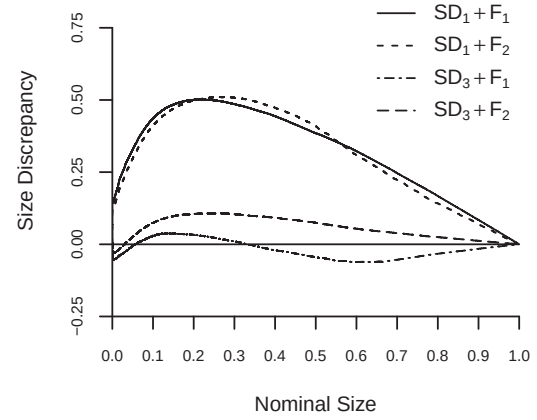
(c) $k = 2, HC_1$



(d) $k = 2, HC_3$

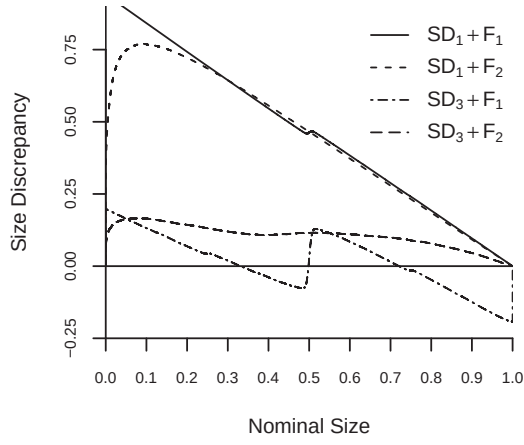


(e) $k = 3, HC_1$

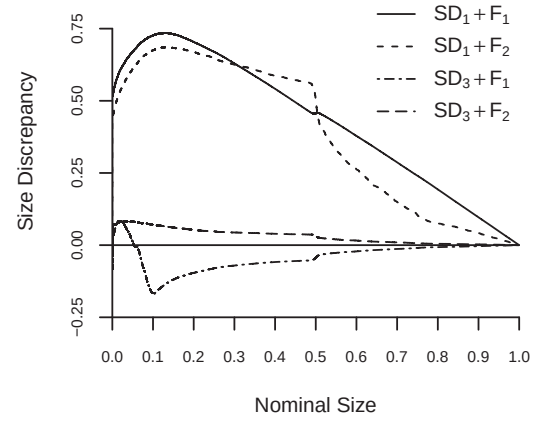


(f) $k = 3, HC_3$

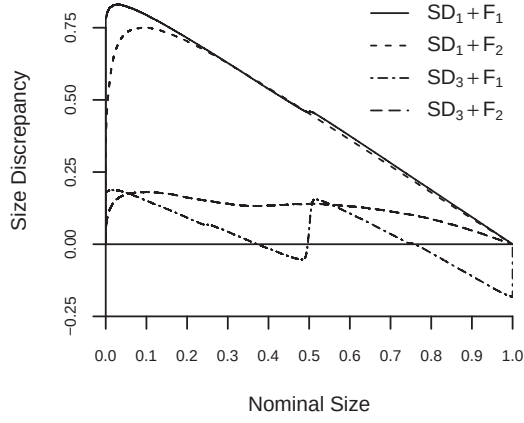
Figure 1: p-value Discrepancy Plots, Normal Errors



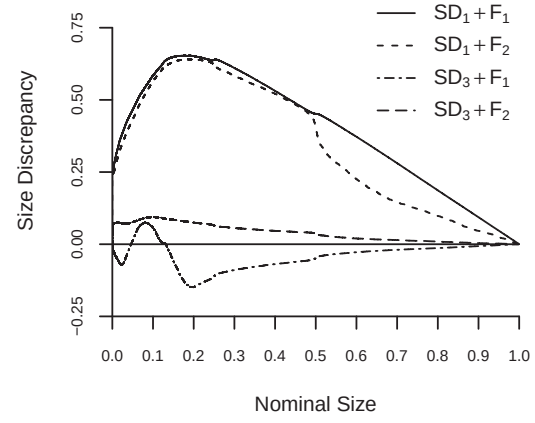
(a) $k = 1, HC_1$



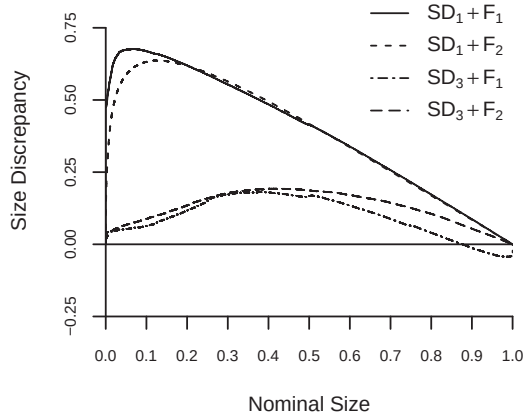
(b) $k = 1, HC_3$



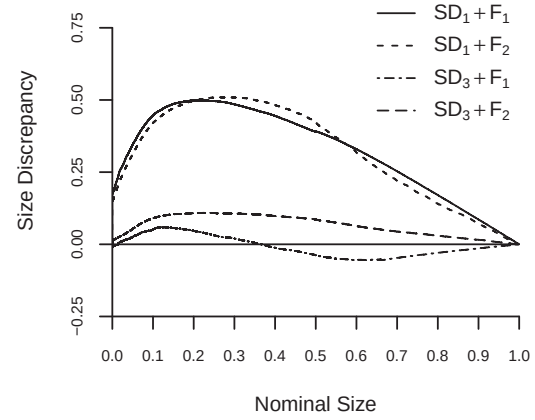
(c) $k = 2, HC_1$



(d) $k = 2, HC_3$



(e) $k = 3, HC_1$



(f) $k = 3, HC_3$

Figure 2: p-value Discrepancy Plots, Skewed Errors

where $\mathbf{S}$ is $n \times s$ is full rank. If $s = n$, $\mathbf{S} = \text{diag}(\hat{\sigma}_1, \dots, \hat{\sigma}_n)$. When $s < n$, $\mathbf{S}$ is this matrix, deleting columns corresponding to $\hat{\sigma}_i = 0$.

Define the $s \times s$ symmetric matrix $\mathbf{B} = \mathbf{S}'\mathbf{A}'\mathbf{A}\mathbf{S}$ and the $s \times 1$ vector $\mathbf{a} = \mathbf{S}'\mathbf{Z}$. Define $r = \text{rank}(\mathbf{B})$. We use the following technical condition.

Assumption 3 *The vector $\mathbf{a}$ and matrix $\mathbf{B}$ satisfy: (1) $\mathbf{a}'\mathbf{a} > 0$; (2) $r > 0$; (3) $\mathbf{a} \notin \text{range}(\mathbf{B})$.*

The conditions of Assumption 3 are necessary to exclude the possibility of degenerate $\hat{\theta}^*$, degenerate $\hat{v}^*$, or perfect collinearity between these variables.

We now present an expression for the wild bootstrap distribution function.

Theorem 1 *Under Assumptions 1-3, the wild bootstrap distribution equals*

$$F^*(x) = 1 - \int_0^1 \left(\prod_{\ell=1}^r \left(1 - \frac{\lambda_\ell}{\lambda_s} \sec^2\left(\frac{\pi}{2}t\right) \right) \right)^{-1/2} dt \quad (4)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s$ are the eigenvalues of the $s \times s$ symmetric matrix

$$\mathbf{C} = x^2 \mathbf{B} - \mathbf{a}\mathbf{a}'. \quad (5)$$

The eigenvalues satisfy $\lambda_\ell > 0$ for $1 \leq \ell \leq r$, $\lambda_\ell = 0$ for $r+1 \leq \ell \leq s-1$, and $\lambda_s < 0$. The function $F^*(x)$ is continuous, differentiable, and strictly increasing for $x > 0$.

Equation (4) is the exact wild bootstrap distribution function, and equals a one-dimensional definite integral of a monotonic bounded integrand. It is an application of Theorem 1 of Hansen (2026b), is derived from a fundamental result of Makshanov and Shalaevskii (1977), and is closely related to a result in Müller and Watson (2022). Theorem 1 follows from the assumption that the auxillary variables are normal, for then the bootstrap regression is exactly a normal regression, under which Theorem 1 of Hansen (2026b) applies. Theorem 1 requires the auxillary variables (which are under our control) to be normal, but makes absolutely no restrictions on the true regression model (1). This is because Theorem 1 is describing the exact distribution of the bootstrap distribution; it is not describing the sampling distribution of the sample statistics.

We recommend numerical calculation of (4) by Gaussian quadrature. Given the symmetry of the integrand about 0, this estimate is

$$\tilde{F}^*(x) = 1 - \sum_{j=1}^{N/2} w_j \left(\prod_{\ell=1}^r \left(1 - \frac{\lambda_\ell}{\lambda_s} \sec^2\left(\frac{\pi}{2}t_j\right) \right) \right)^{-1/2} \quad (6)$$

where t_j are the positive roots of the N th-order Legendre polynomial and w_j are the associated quadrature weights⁸. For our numerical implementation we set $N = 64$. If r is unknown we can replace it with an upper bound, including $s - 1$.

⁸The roots and weights are challenging to calculate, but tables are widely available. One excellent set of tables is available at <https://pomax.github.io/bezierinfo/legendre-gauss>.

Given the eigenvalues λ_ℓ , the calculation (6) is numerically fast and accurate. Calculations reported in Hansen (2026b) indicate that (6) is equally accurate as the method of Imhof (1961), yet considerably faster computationally⁹.

This discussion leads to our first algorithm to calculate the bootstrap distribution, which we now summarize.

Algorithm 1

1. Given $\mathbf{A}$, $\mathbf{Z}$, and $\mathbf{S}$, calculate $\mathbf{a}$ and $\mathbf{B}$.
2. Given x , calculate $\mathbf{C}$.
3. Calculate the s eigenvalues of $\mathbf{C}$.
4. Calculate $F^*(x)$ using (6) and $N = 64$.

To illustrate the computation cost, Table 2 (second panel) displays the computation time (in seconds) for the calculation of the bootstrap distribution at a single point using Algorithm 1. To obtain accurate estimates, the calculations were repeated 100 times and the average computation time reported. What we can see is that for small samples (e.g., $n = 100$), Algorithm 1 is very fast, with computation time of 0.001 seconds. For sample sizes up to $n = 1000$, Algorithm 1 is considerably faster than the simulation method. However, computation time increases quickly with n , and Algorithm 1 is computationally slower than simulation for $n \geq 5000$.

The primary computational burden in Algorithm 1 is the calculation of the eigenvalues of $\mathbf{C}$. Eigenvalue calculation is fast when s is small (roughly, when $s \leq 200$), but increases at the rate $O(s^3)$. In the next section we propose a computationally attractive alternative.

5 Approximately Exact Wild Bootstrap

The expression (4) requires calculation of the eigenvalues of the $s \times s$ matrix $\mathbf{C}$. Eigenvalue calculation is fast when s is small. However, calculation time increases rapidly with s , and is effectively infeasible for $s \geq 5000$. In this section we describe a computationally fast approximation with nearly identical accuracy.

As $\mathbf{C}$ is a rank-one perturbation of $\mathbf{B}$, the eigenvalues of $\mathbf{C}$ and $\mathbf{B}$ are closely related. By a combination of targeted computation, algebra, and simplification, we can approximate the eigenstructure of $\mathbf{B}$ with minimal computation, and with this knowledge can approximate the eigenvalues of $\mathbf{C}$.

The matrix $\mathbf{B}$ is positive semi-definite, so is largely determined by its largest eigenpairs; the structure of the remaining eigenpairs is of reduced importance. It is well known that numerical computation of a modest number $M \ll s$ of eigenpairs $(b_\ell, \mathbf{h}_\ell)$ of $\mathbf{B}$ can be achieved with much greater speed than the

⁹The method of Imhof (1961) is computationally slow, and may not converge, when r is large. It can also have convergence problems when the largest eigenvalue λ_1 is much larger than the others.

full set of eigenvalues. This calculation can be implemented by well-studied methods including subset, Arnoldi, and Lanczos iteration. An excellent R library which implements these methods is RSpectra. An alternative method (which we adopt) is the randomized eigenvalue decomposition of Halko, Martinsson, and Tropp (2011), which we describe in Appendix A. This is computationally faster (in our R implementation) than RSpectra yet is simple to program and does not rely on specialized programs, tuning parameters, or convergence criteria. Any of these algorithms can be implemented without explicit construction of $\mathbf{B}$, but only require a program which calculates the mapping $\mathbf{g} = \mathbf{B}\mathbf{h}$.

When $\mathbf{B}$ has repeating eigenvalues, deflation methods can reduce the dimensionality of the eigenvalue problem for $\mathbf{C}$. A known set of repeating eigenvalues of $\mathbf{B}$ are its zeros, which are due to rank deficiency in $\mathbf{B}$. Furthermore, when the remaining positive eigenvalues of $\mathbf{B}$ (other than the first M) are identical, the computation is greatly simplified. The key is the following condition.

Assumption 4 *The positive eigenvalues of $\mathbf{B}$ are $b_1 > b_2 > \dots > b_M > \bar{b} > 0$ where $\bar{b}$ repeats with multiplicity $r - M > 0$.*

Assumption 4 specifies that the first M eigenvalues are unique, and the remaining take the common value $\bar{b}$. This restricts the smaller eigenvalues to be identical. This restriction is of minor importance when M is large.

Given that the eigenvalues b_ℓ for $\ell \leq M$ have been calculated, the repeated eigenvalue $\bar{b}$ can be calculated from the relationship

$$\bar{b} = \frac{\text{tr} \mathbf{B} - \sum_{\ell=1}^M b_\ell}{r - M}. \quad (7)$$

To calculate $\text{tr} \mathbf{B}$, when n is large it is best to avoid explicit construction of $\mathbf{B}$. For HC₀, HC₁, HC₂, and HC₃ we can use the following formula

$$\text{tr} \mathbf{B} = \sum_{i=1}^n d_i^2 \hat{\sigma}_i^2 (1 - 2m_i) + \text{tr} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{D}^2 \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma} \mathbf{X} \quad (8)$$

where $m_i = \mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i$ are the leverage values.

We now introduce some notation. For $\ell = 1, \dots, M$, let $\mathbf{h}_\ell$ denote the eigenvector of $\mathbf{B}$ associated with the eigenvalues b_ℓ . Let $\mathbf{H}_0$ denote the $s \times (s - r)$ matrix of eigenvectors of $\mathbf{B}$ associated with the 0 eigenvalues. (We discuss calculation of $\mathbf{H}_0$ below.) Set

$$\bar{\mathbf{a}}_0 = (\mathbf{a}' \mathbf{H}_0 \mathbf{H}_0' \mathbf{a})^{1/2} \quad (9)$$

$$\bar{\mathbf{a}}_\ell = \mathbf{h}'_\ell \mathbf{a}, \text{ for } \ell = 1, \dots, M \quad (10)$$

$$\bar{\mathbf{a}}_{M+1} = \left(\mathbf{a}' \mathbf{a} - \sum_{\ell=0}^M \bar{\mathbf{a}}_\ell^2 \right)^{1/2} \quad (11)$$

$$\bar{\mathbf{a}} = (\bar{\mathbf{a}}_0, \bar{\mathbf{a}}_1, \dots, \bar{\mathbf{a}}_{M+1})' \quad (12)$$

$$\bar{\mathbf{B}} = \text{diag}(0, b_1, \dots, b_M, \bar{b}). \quad (13)$$

Let $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \dots \geq \bar{\lambda}_{M+2}$ be the eigenvalues of the $(M+2) \times (M+2)$ symmetric matrix

$$\bar{\mathbf{C}} = x^2 \bar{\mathbf{B}} - \bar{\mathbf{a}} \bar{\mathbf{a}}'. \quad (14)$$

Given these definitions we have an expression for the wild bootstrap distribution function.

Theorem 2 *Under Assumptions 1-4,*

$$F^*(x) = 1 - \int_0^1 \left(\prod_{\ell=1}^{M+1} \left(1 - \frac{\bar{\lambda}_\ell}{\bar{\lambda}_{M+2}} \sec^2 \left(\frac{\pi}{2} t \right) \right) \right)^{-1/2} \left(1 - \frac{x^2 \bar{b}}{\bar{\lambda}_{M+2}} \sec^2 \left(\frac{\pi}{2} t \right) \right)^{-(r-M-1)/2} dt. \quad (15)$$

Theorem 3 provides an alternative expression for the exact bootstrap distribution which does not require explicit construction of $\mathbf{C}$ nor direct calculation of its eigenvalues. The additional requirement is Assumption 4. When this assumption fails then (15) can be viewed as an approximation to the exact distribution. It is an approximation as (15) approaches (4) as M increases. Consequently, we call (15) the approximately exact wild bootstrap.

To compute (15) we need the matrix $\mathbf{H}_0$, which can be constructed from the null space of $\mathbf{B}$. In practice, this calculation depends on the specific structure of $\mathbf{B}$. We now restrict ourselves to the standard HC variance estimators.

Assumption 5 $\hat{v}^2$ is calculated by HC_0 , HC_1 , HC_2 , or HC_3 .

Recall that $\mathbf{B} = \mathbf{S}' \mathbf{A}' \mathbf{A} \mathbf{S}$. Under Assumption 5, the matrix $\mathbf{A}$ takes the special form $\mathbf{A} = \mathbf{D} \mathbf{M}$ where $\mathbf{D}$ is diagonal (with diagonal elements d_i) and $\mathbf{M}$ is the residual projection matrix. In typical applications, $\hat{\sigma}_i > 0$ and $d_i > 0$ for all i . In this case it is straightforward to see that

$$\begin{aligned} r &= \text{rank}(\mathbf{B}) = n - k \\ \text{null}(\mathbf{B}) &= \hat{\mathbf{\Sigma}}^{-1/2} \mathbf{X} \\ \mathbf{H}_0 &= \hat{\mathbf{\Sigma}}^{-1/2} \mathbf{X} \left(\mathbf{X}' \hat{\mathbf{\Sigma}}^{-1} \mathbf{X} \right)^{-1/2} \\ \bar{\mathbf{a}}_0 &= \left(\mathbf{R}' \left(\mathbf{X}' \hat{\mathbf{\Sigma}}^{-1} \mathbf{X} \right)^{-1} \mathbf{R} \right)^{1/2} \end{aligned} \quad (16)$$

where $\mathbf{R}$ is the first $k \times 1$ unit vector. (The final expression can be derived from (9) using $\mathbf{a} = \hat{\mathbf{\Sigma}}^{1/2} \mathbf{Z}$ and $\mathbf{Z} = \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{R}$.)

The calculations are a bit more subtle in the atypical case where $\hat{\sigma}_i = 0$ or $d_i = 0$ for some i . Let n_0 be the number of observations for which either $\hat{\sigma}_i = 0$ or $d_i = 0$, and let $\mathbf{X}_0$ denote the $n_0 \times k$ matrix consisting of the rows of $\mathbf{X}$ corresponding to these observations. Let n_d be the number of observations for which $\hat{\sigma}_i > 0$ and $d_i = 0$, and let $\mathbf{X}_d$ denote the $n_d \times k$ matrix consisting of the corresponding rows of $\mathbf{X}$. Let $\mathbf{X}_0^\perp$ be a full rank $k \times \text{nullity}(\mathbf{X}_0)$ matrix spanning the orthogonal complement of the row space of $\mathbf{X}_0$, so that $\mathbf{X}_0 \mathbf{X}_0^\perp = \mathbf{0}$. The matrix $\mathbf{X}_0^\perp$ can be calculated from $\mathbf{X}_0$ using the QR decomposition or by row operations. When $\hat{\sigma}_i > 0$ and $d_i > 0$ for all i , then $\mathbf{X}_0$ is null and $\mathbf{X}_0^\perp = \mathbf{I}_k$. When $\mathbf{X}_0$ has full column rank then $\mathbf{X}_0^\perp$ is null. When there are no observations with $\hat{\sigma}_i > 0$ and $d_i = 0$, then $\mathbf{X}_d$ is null.

We use the following technical condition.

Assumption 6 $\text{rank}(\mathbf{D}\mathbf{X}) = k$.

Assumption 6 is equivalent to $\mathbf{X}$ maintaining full column rank after omitting observations with $d_i = 0$. Assumption 6 is not essential, but simplifies our characterization of the null space of $\mathbf{B}$.

Theorem 3 *Under Assumptions 1-6,*

$$r = n - k - n_0 + \text{rank}(\mathbf{X}_0) + \text{rank}(\mathbf{X}_d) \quad (17)$$

$$\bar{a}_0 = \left(R' \mathbf{X}_0^\perp \left(\mathbf{X}_0^{\perp'} \mathbf{X}' \hat{\Sigma}^+ \mathbf{X} \mathbf{X}_0^\perp \right)^{-1} \mathbf{X}_0^{\perp'} R \right)^{1/2} \quad (18)$$

where $\mathbf{A}^+$ denotes the Moore-Penrose inverse of $\mathbf{A}$.

In practice, we can approximate (17) as

$$r = n - k - n_0 + \min(n_0, k) + \min(n_d, k).$$

The computational savings of (15) relative to (4) is that rather than calculating the full set of n eigenvalues of $\mathbf{C}$, (15) requires calculating the first M eigenpairs of $\mathbf{B}$ plus the $M + 2$ eigenvalues of $\bar{\mathbf{C}}$. While the latter requires solving two eigenvalue problems, they are considerably computationally faster when M is small relative to n .

Calculation of (15) by Gaussian quadrature is

$$F^*(x) \simeq 1 - \sum_{j=1}^{N/2} w_j \left(\prod_{\ell=1}^{M+1} \left(1 - \frac{\bar{\lambda}_\ell}{\bar{\lambda}_{M+2}} \sec^2 \left(\frac{\pi}{2} t_j \right) \right) \right)^{-1/2} \left(1 - \frac{x^2 \bar{b}}{\bar{\lambda}_{M+2}} \sec^2 \left(\frac{\pi}{2} t_j \right) \right)^{-(r-M-1)/2}. \quad (19)$$

A practical issue is the choice of M . Larger values yield more accurate calculations, but with increased computation cost (roughly linear in M). Based on our experiments, we set $M = 15$, as with this choice p-values calculated by (4) and (15) are within 10^{-5} of each other.

We summarize the calculations described in this section as the following algorithm.

Algorithm 2

1. Given $\mathbf{A}$, $\mathbf{Z}$, and $\hat{\Sigma}$, calculate $\mathbf{a}$ and $\mathbf{B}$.
2. Calculate the approximate largest $M = 15$ eigenvalue/eigenvector pairs $(b_\ell, \mathbf{h}_\ell)$ of $\mathbf{B}$ using the randomized eigenvalue decomposition.
3. Calculate r using either $r = n - k$ or (17).
4. Calculate $\bar{b}$ from (7) using (8).
5. Calculate $\bar{a}_0$ using (16) or (18).

6. Calculate $\bar{\mathbf{a}}$ and $\bar{\mathbf{B}}$ as defined in (12)-(13) using (10)-(11).
7. Given x , calculate $\bar{\mathbf{C}}$ from (14).
8. Calculate the eigenvalues $\bar{\lambda}_\ell$ of $\bar{\mathbf{C}}$.
9. Calculate $F^*(x)$ using (19) and $N = 64$.

Algorithm 2 has more steps than Algorithm 1, but is computationally much faster. The step with the greatest computational burden is step 2, which calculates the M largest eigenpairs of $\mathbf{B}$. However, if M is moderately small, this calculation is considerably faster than the s -dimensional eigenvalue calculation required for Algorithm 1.

To illustrate the computation cost, Table 2 (third panel) displays the computation time (in seconds) for the calculation of the bootstrap distribution at a single point using Algorithm 2. To obtain accurate estimates, the calculations (for $n \leq 10,000$) were repeated 100 times and the average computation time reported. What we can see is that the computation time of Algorithm 2 is uniformly smaller than both the simulation method and Algorithm 1. For $n = 100$ the computation time of Algorithms 1 and 2 are similar. For larger n , the computation time of Algorithm 2 is consider less. The computation time is less than one second for sample sizes up to $n = 10,000$. In all cases, computation time is considerably less than simulation. An important feature is that the computation time is scalable as n increases, approximately linear in n .

6 Plotting a Bootstrap Distribution Function

In some cases it is desired to calculate the bootstrap distribution function $F^*(x)$ at multiple values of x , for example, when plotting the bootstrap distribution. For Algorithm 1, the matrix $\mathbf{C}$ depends on x , and therefore the eigenvalues need to be re-calculated for each x , which is costly. For Algorithm 2, the matrices $\bar{\mathbf{a}}$ and $\bar{\mathbf{B}}$ can be calculated once, and then for each x calculate $\bar{\mathbf{C}}$ and its eigenvalues. This is computationally efficient, as most of the computation time is expended in calculation of $\bar{\mathbf{a}}$ and $\bar{\mathbf{B}}$. Hence, we generally recommend the use of Algorithm 2.

We illustrate with a simple example. We generated a single sample¹⁰ with $n = 10,000$ observations and $k = 20$ regressors. We calculated the wild bootstrap distribution for the HC₃ t-statistic for the first regression coefficient, evaluated at 250 points. We calculated the distribution using the simulation method with $B = 1000$ and $B = 20,000$ simulation replications, and with Algorithm 2. The estimated distributions are displayed in Figure 3.

What you can see in the figure is that there is a visual discrepancy between the simulation estimate with $B = 1000$ and the exact distribution (Algorithm 2). There is essentially no visual discrepancy between the latter and the simulation estimate with $B = 20,000$. The exact distribution was calculated in 0.05 seconds, and the simulation estimate with $B = 20,000$ in 12 seconds.

¹⁰The regressors were generated from the LogNormal distribution (plus one intercept), and the equation errors were generated as normal with standard deviation equal to the first regressor.

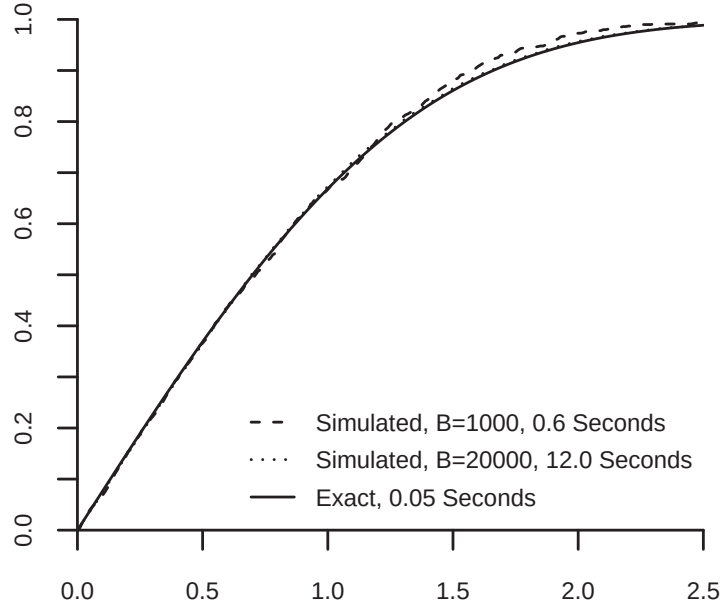


Figure 3: Wild Bootstrap Distribution Function, $n = 10,000$

7 Critical Values and Confidence Intervals

The wild bootstrap critical value c^* is the $1 - \alpha$ quantile of the distribution F^* . These are used to construct the symmetric wild bootstrap confidence interval $\hat{\theta} \pm c^* \hat{v}$. Algorithms 1 and 2 provide methods to compute the distribution function $F^*(x)$ but not a direct method to calculate its quantiles. Consequently, we calculate the critical values numerically by solving the equation

$$F^*(x) = 1 - \alpha.$$

As the function $F^*(x)$ is strictly increasing (Theorem 1), the solution is unique, and can be found by many root-finding algorithms.

For our numerical implementation, we use the root-finding method known as the Illinois algorithm, which is an improved version of *regula falsi*. This algorithm can typically find the critical value with a small number of function evaluations. The method requires a bracket $[L, U]$ which contains the critical value. We initially set $L = 0$ and $U = 2$, and make the shifts $\{(L, U) \rightarrow (U, 2U)\}$ until $F^*(U) > 1 - \alpha$ so that the interval brackets c^* .

Algorithm 2 is well-suited for calculation of critical values. The matrices $\bar{\mathbf{a}}$ and $\bar{\mathbf{B}}$ are calculated once, and taken as given for the root-finding algorithm. As most of the computation time is expended in

calculation of $\bar{\mathbf{a}}$ and $\bar{\mathbf{B}}$, this results in a computation time which is similar to a single calculation of the distribution function.

This is verified in a simulation experiment. The design for Table 2 is repeated, with the simulation method and Algorithm 2 compared for calculation of critical values. We find that the computation times are nearly identical to those reported in Table 2. As these results are nearly identical to Table 2, they are not reported here.

8 Restricted Wild Bootstrap

The methods described so far in this paper are known as the “unrestricted” wild bootstrap, as the standard deviation estimates are calculated from unrestricted least-squares residuals. In contrast, an alternative set of methods are known as the “restricted” wild bootstrap, where the standard deviation estimates are calculated from restricted least-squares residuals, where the restriction is the null hypothesis being tested. In most of the econometric literature exploring the properties of the wild bootstrap there has been a general preference for the restricted wild bootstrap, in the sense that it tends to have better finite sample size. The natural explanation is that by calculating residuals imposing the null hypothesis, the standard deviation estimates are more accurate estimates of the true standard deviations, and hence should produce bootstrap distributions which are closer to the true sampling distribution.

However, the extensive simulation evidence in MacKinnon, Nielsen, and Webb (2023a) provides a caveat. These authors found that when t-statistics are calculated with the HC₃ standard error and the standard deviation estimate using SD₃ (our recommended procedure), that the finite sample performance of the wild bootstrap was generally excellent, and there was not a clear preference for restricted over unrestricted residual estimates. Essentially, using improved standard error and standard deviation estimators may eliminate the need for the extra complication of restricted residuals.

Regardless, we describe in this section how to extend the exact methods to include the restricted wild bootstrap.

We amend the description of the wild bootstrap from Section 3. The primary distinction is that the standard deviation estimates $\hat{\sigma}_i(\theta)$ are indexed by the parameter θ . We label these as RSD₀ through RSD₃ for “restricted standard deviation” estimates. The RSD₀ estimate is $\tilde{\sigma}_{0i}(\theta) = |\tilde{e}_i(\theta)|$ where $\tilde{e}_i(\theta)$ is the restricted least squares residual imposing the value θ , which can be obtained as the least squares residual from the regression of $Y_i - X_{1i}\theta$ on X_{2i} . The RSD₃ estimate is $\tilde{\sigma}_{3i}(\theta) = |\tilde{e}_i(\theta)| / (1 - X_{2i}'(\mathbf{X}_2'\mathbf{X}_2)^{-1}X_{2i})$. The others are defined analogously.

Other statistics are also redefined to be indexed by θ . For example, the t-ratio T is redefined as $T(\theta) = |\hat{\theta} - \theta| / \hat{v}$. The bootstrap sample is $Y_i^*(\theta) = X_{1i}\theta + \hat{\sigma}_i(\theta)\xi_i$, the bootstrap coefficient estimate is $\hat{\theta}^*(\theta)$, its standard error is $\hat{v}^*(\theta)$, and its t-ratio is $T^*(\theta) = |\hat{\theta}^*(\theta) - \theta| / \hat{v}^*(\theta)$. Given these definitions, the wild bootstrap distribution is

$$F^*(x, \theta) = \Pr[T^*(\theta) \leq x \mid \mathbf{X}, \hat{\sigma}(\theta)].$$

Observe that we index the distribution by both x and θ .

The wild bootstrap p-value for the test of $H_0 : \theta = \theta_0$ is

$$p^* = 1 - F^*(T(\theta_0), \theta_0).$$

Computationally this is not a major adjustment over the case with unrestricted standard deviation estimates. The only substantive difference is that the restricted least squares residuals (under the null hypothesis) are calculated.

A substantive difference is in the construction of confidence intervals. They do not take a simple form, but rather must be solved numerically using test inversion. One way to do this is to find the set of parameter values for which the t-ratio is less than its critical value. The restricted wild bootstrap critical value $c^*(\theta)$ is the solution to

$$F^*(c^*(\theta), \theta) = 1 - \alpha$$

and (in general) depends on θ . For any θ this critical value can be found numerically using a root finding method as described in Section 7. The $1 - \alpha$ wild bootstrap confidence interval for θ is

$$CI^* = \{\theta : T(\theta) \leq c^*(\theta)\}.$$

As the critical value is a function of θ , this set can take a complicated form. In extreme cases the confidence set may not be convex.

We illustrate with a simple yet extreme example. We generate¹¹ one sample with $n = 10,000$ observations and $k = 20$ regressors. The latter are drawn from the LogNormal distribution (with one intercept), and the dependent variable from $N(0, X_{1i}^2)$. We calculate the HC₁ t-ratio, and the wild bootstrap using the RSD₀ standard deviation estimator. In this sample the least squares estimator is $\hat{\theta} = -0.93$ with standard error $\hat{\nu} = 0.25$. In Figure 4 we plot two lines as a function of the parameter θ . One is the t-ratio function $T(\theta) = |\hat{\theta} - \theta| / \hat{\nu}$, which is the simple “V” shape. This shows the value of the t-statistic evaluated at specific null hypotheses. The function is minimized at $\theta = \hat{\theta}$ by construction. The second line is the restricted wild bootstrap 95% critical value function $c^*(\theta)$, which is found numerically for each θ . This shows us the 0.95 quantile of the bootstrap distribution of the t-ratio, using the restricted estimates given θ . In this example, we see that the critical value function is highly non-linear, reaching a low of 1.17 and a high of 5.88. (In contrast, the conventional critical value is 1.96.) The function is so non-linear that $c^*(\theta)$ and $T(\theta)$ cross four times, at $\theta = -2.29, -1.36, -1.25$, and 0.35 .

It is not standard to display figures such as Figure 4, rather a user will most likely only calculate a confidence interval. Based on the no-rejection principle, the test inversion confidence interval is $[-2.29, -1.36] \cup [-1.25, 0.35]$. In this context it would be natural to prefer the convexified set $[-2.29, 0.35]$. In principle the endpoints of the confidence set can be found by root-finding algorithms (finding the roots of $T(\theta) - c^*(\theta)$) but these are not guaranteed to find the outer endpoints when there are multiple crossings. This raises an additional computational challenge for generic implementation of confidence interval construction for the restricted wild bootstrap. Furthermore, the vast fluctuations in $c^*(\theta)$ across

¹¹The random number seed was selected to generate an “extreme” numerical example. In that sense this should not be viewed as a “random” sample.

θ indicate a high degree of uncertainty concerning the critical value. The bootstrap critical value is an estimate, based on a noisy estimate of the unknown standard deviations. If this estimate is noisy, this adds additional uncertainty to confidence interval construction, and reduces its accuracy.

It should be acknowledged that the situation in Figure 4 is not typical; it is extreme. However, the point raised here is that this situation *can* arise. I see no easy fixes, other than to caution that the restricted wild bootstrap raises these challenges. Based on all of the evidence, a simple solution is to use the simpler conventional (unrestricted) wild bootstrap, for which such problems do not arise.

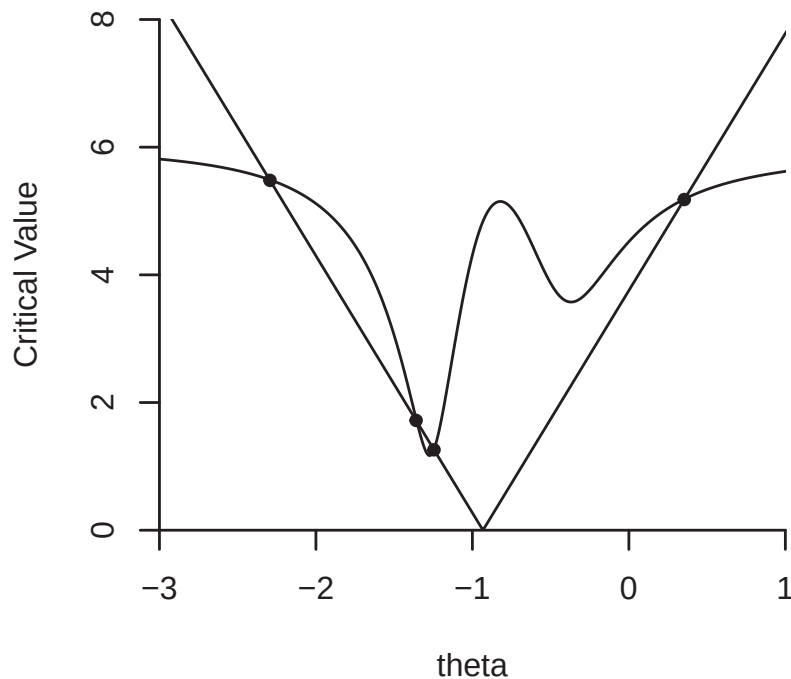


Figure 4: Restricted Wild Bootstrap Confidence Interval Construction

9 Conclusion

The wild bootstrap is a powerful method to improve the accuracy of inference in finite samples over conventional asymptotic approximations. The methods explored in this paper are appropriate for inference on linear coefficients of linear regression models. While this is a narrow context, it is the most common setting in applied econometrics. The existing method to calculate wild bootstrap p-values and confidence intervals is based on simulation, which requires a high number of bootstrap replications for accuracy, and thus requires some computation cost. The method proposed in this paper is an exact calculation, which does not require simulation, but rather one-dimensional numerical integration. The

proposed method is considerably faster than simulation methods, and considerably more accurate. The computer code to implement the methods is simple and does not use specialized packages.

A Appendix: Randomized Low Rank Eigenvalue Decomposition

In this section we describe an application of the randomized low rank eigenvalue decomposition of Halko, Martinsson, and Tropp (2011), which combines their Algorithms 4.1 and 5.3.

The algorithm takes an $n \times n$ symmetric matrix $\mathbf{B} = \mathbf{G}'\mathbf{G}$ (or functions which compute the linear maps $\mathbf{G}\mathbf{h}$ and $\mathbf{G}'\mathbf{h}$) and outputs an estimate of the M largest eigenvalues of $\mathbf{B}$ and their associated eigenvectors.

The algorithm is:

1. Draw an $n \times M$ Gaussian random matrix $\mathbf{U}$ (whose elements are i.i.d. $N(0, 1)$)
2. Form the $n \times M$ matrix $\mathbf{V} = \mathbf{G}'\mathbf{G}\mathbf{U}$
3. Construct an $n \times M$ matrix $\mathbf{F}$ whose columns form an orthonormal basis for the range of $\mathbf{V}$, e.g. using the QR factorization
4. Form the $M \times M$ matrix $\mathbf{W} = (\mathbf{G}\mathbf{F})'(\mathbf{G}\mathbf{F})$
5. Compute an eigenvalue decomposition $\mathbf{W} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$ where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$
6. Form the $n \times M$ matrix $\mathbf{H} = \mathbf{F}\mathbf{P}$
7. The M eigenvalue and eigenvector estimates are $\lambda_1, \lambda_2, \dots, \lambda_M$ and the columns of $\mathbf{H}$.

The properties of this and related randomized decompositions are discussed in Halko, Martinsson, and Tropp (2011) and Martinsson and Tropp (2020).

For our applications, $\mathbf{G} = \mathbf{D}\mathbf{M}\hat{\mathbf{\Sigma}}^{1/2} = \mathbf{S} - \mathbf{Q}_1\mathbf{Q}_2'$, where $\mathbf{S} = \mathbf{D}\hat{\mathbf{\Sigma}}^{1/2} = \text{diag}(d_1\hat{\sigma}_1, d_2\hat{\sigma}_2, \dots, d_n\hat{\sigma}_n)$, $\mathbf{Q}_1 = \mathbf{D}\mathbf{Q}$, $\mathbf{Q}_2 = \hat{\mathbf{\Sigma}}^{1/2}\mathbf{Q}$, and $\mathbf{Q}$ is the $n \times k$ orthonormal basis for $\mathbf{X}$ obtained by the QR factorization.

B Appendix: Mathematical Proofs

Proof of Theorem 1: Conditional on $\mathbf{X}$ and $\hat{\sigma}$, the bootstrap errors $\hat{\sigma}_i\xi_i$ are mutually independent and distributed $N(0, \hat{\sigma}_i^2)$. Therefore $\mathbf{Y}^* \sim N(0, \hat{\mathbf{\Sigma}})$, which implies that the bootstrap sample is an exact normal regression model. Applying Theorem 1 of Hansen (2026b), we obtain the statements in the theorem.

■

Proof of Theorems 2 and 3: Follows Theorems 2 and 3 of Hansen (2026), replacing $\mathbf{\Sigma}$ with $\hat{\mathbf{\Sigma}}$.

■

References

- [1] Beran, Rudolf (1986): "Discussion of 'Jackknife bootstrap and other resampling methods in regression analysis' by C. F. J. Wu," *Annals of Statistics*, 14, 1295-1298.
- [2] Canay, Ivan A., Andres Santos, and Azeem M. Shaikh (2021): "The wild bootstrap with a small number of large clusters," *Review of Economics and Statistics*, 103, 346-363.
- [3] Chesher, Andrew D. (1989): "Hajek inequalities, measures of leverage, and the size of heteroskedasticity robust Wald tests," *Econometrica*, 57, 971-977.
- [4] Chesher, Andrew D. and Gerard Austin (1991): "The finite-sample distributions of heteroskedasticity robust Wald statistics," *Journal of Econometrics*, 47, 153-173.
- [5] Chesher, Andrew D. and Ian D. Jewitt (1987): "The bias of the heteroskedasticity consistent covariance matrix estimator," *Econometrica*, 55, 1217-1272.
- [6] Davidson, Russel and Emmanuel Flachaire (2008): "The wild bootstrap, tamed at last," *Journal of Econometrics*, 146, 162-169.
- [7] Davidson, Russell and James G. MacKinnon (1998): "Graphical methods for investigating the size and power of hypothesis tests," *The Manchester School*, 66, 1-26.
- [8] Djogbenou, Antoine A., James G. MacKinnon, and Morten Ø. Nielsen (2019): "Asymptotic theory and wild bootstrap inference with clustered errors," *Journal of Econometrics*, 212, 393-412.
- [9] Eicker, Friedhelm (1963): "Asymptotic normality and consistency of the least squares estimators for families of linear regressions," *Annals of Mathematical Statistics*, 34, 447-456.
- [10] Halko, Nathan, Per Gunner Martinsson, and Joel A. Tropp (2011): "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions", *SIAM Review*, 53, 217-288.
- [11] Hansen, Bruce E. (2026a): "Jackknife standard errors for clustered regression," *Review of Economic Studies*, forthcoming.
- [12] Hansen, Bruce E. (2026b): "Exact inference in linear regression," in preparation.
- [13] Hinkley, David V. (1977): "Jackknifing in unbalanced situations," *Technometrics*, 19, 285-292.
- [14] Huber, Peter J. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, Lucien M. Le Cam and Jerzy Neyman, editors, 1, 221-223.
- [15] Imhof, J. P. (1961): "Computing the distribution of quadratic forms in normal variables," *Biometrika*, 48, 419-426.

- [16] Liu, Regina Y. (1988): "Bootstrap procedures under some non-I.I.D. models," *Annals of Statistics*, 16, 1696-1708.
- [17] MacKinnon, James G. (2015): "Wild cluster bootstrap confidence intervals," *L'Actualité économique*, 91, 11-33.
- [18] MacKinnon, James G. (2023): "Fast cluster bootstrap methods for linear regression models," *Econometrics and Statistics*, 26, 52-71.
- [19] MacKinnon, James G., Morten Ø. Nielsen, and Matthew D. Webb (2021): "Wild bootstrap and asymptotic inference with multiway clustering," *Journal of Business and Economic Statistics*, 39, 505-519.
- [20] MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb (2023a): "Fast and reliable jackknife and bootstrap methods for cluster-robust inference," *Journal of Applied Econometrics*, 38, 671-694.
- [21] MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb (2023b): "Cluster-robust inference: A guide to empirical practice," *Journal of Econometrics*, 232, 272-299.
- [22] MacKinnon, James G., Morten Ørregaard Nielsen, and Matthew D. Webb (2023c): "Leverage, influence, and the jackknife in clustered regression models: Reliable inference using summclust," *Stata Journal*, 4, 942-982.
- [23] MacKinnon, James G., and Matthew D. Webb (2017): "Wild bootstrap inference for wildly different cluster sizes," *Journal of Applied Econometrics*, 32, 233-254.
- [24] MacKinnon, James G., and Matthew D. Webb (2018): "The wild bootstrap for few (treated) clusters," *The Econometrics Journal*, 21, 114-135.
- [25] MacKinnon, James G., and Matthew D. Webb (2020): "Randomization inference for difference-in-differences with few treated clusters," *Journal of Econometrics*, 218, 435-450.
- [26] MacKinnon, James G. and Halbert White (1985): "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305-325.
- [27] Makshanov, A. V. and O. V. Shalaeviskii (1977): "Some problems of asymptotic approximation of distributions," *Zapiski Nauchnykh Seminarov LOMI*, 72, 118-138 (in Russian). Reprinted in English in (1986) *Journal of Soviet Mathematics*, 34, 1433-1445.
- [28] Mammen, Enno (1993): "Bootstrap and wild bootstrap for high dimensional linear models," *Annals of Statistics*, 21, 255-285.
- [29] Martinsson, Per Gunner and Joel A. Tropp (2020): "Randomized numerical linear algebra: Foundations and algorithms," *Acta Numerica*, 403-572.

- [30] Müller, Ulrich K., and Mark W. Watson (2022): “Spatial correlation robust inference”, *Econometrica*, 90, 2901-2935.
- [31] Pötscher, Benedikt M. and David Preinerstorfer (2023): “Valid heteroskedasticity robust testing,” *Econometric Theory*, 41, 249-301.
- [32] Roodman, David, James G. MacKinnon, Morten Ø. Nielsen, and Matthew D. Webb (2019): “Fast and wild: Bootstrap inference in Stata using boottest,” *Stata Journal*, 19, 4-60.
- [33] Webb, Matthew D. (2023): “Reworking wild bootstrap based inference for clustered errors,” *Canadian Journal of Economics*, 56, 839-858.
- [34] White, Halbert (1980): “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity,” *Econometrica*, 48, 817-838.
- [35] Wu, C. F. J. (1986): “Jackknife, bootstrap and other resampling methods in regression analysis,” *Annals of Statistics*, 14, 1261-1295.