

Heteroskedasticity-Robust Inference in Finite Samples

Jerry Hausman* and Christopher Palmer†

Massachusetts Institute of Technology

December 2011

Abstract

Since the advent of heteroskedasticity-robust standard errors, several papers have proposed adjustments to the original White formulation. We replicate earlier findings that each of these adjusted estimators performs quite poorly in finite samples. We propose a class of alternative heteroskedasticity-robust tests of linear hypotheses based on an Edgeworth expansion of the test statistic distribution. Our preferred test outperforms existing methods in both size and power for low, moderate, and severe levels of heteroskedasticity.

Keywords: Heteroskedasticity; finite samples; Edgeworth expansion; bootstrap

JEL Codes: C1, C12

1 Introduction

The use of “White standard errors” (White, 1980) is now prevalent in economics. However, it has long been known that t-tests based on White standard errors over-reject when the null hypothesis is true and the sample is not large. Indeed, it is not uncommon for the actual size of the test to be 0.15 when the nominal size is the usual 0.05. Various fixes to estimating the middle matrix ($X'\Sigma X$) in equation (2) below have been introduced. We consider the performance of some of these methods in this paper; see MacKinnon (2011) for a more comprehensive discussion. The major finding seems to be that these attempted fixes do not solve the problem, as we demonstrate subsequently.

The other major approach has been to bootstrap the t-test, which will get the correct size (on average). We benchmark the performance of techniques in this paper with the Wild bootstrap (WB), which MacKinnon (2011) finds to perform best in terms of power. Hall (1992) has demonstrated that with a pivotal test statistic, as occurs here, the bootstrapped test will be accurate to the second order in n rather than to the first order, which underlies the asymptotic expansion used for the White approach.

In this paper, we directly apply the second-order Edgeworth approximation approach to the test statistic distribution using the results of Rothenberg (1988). Hausman and Kursteiner (2008) used this approach to estimate the covariance of the feasible generalized least squares estimator (FGLS) and found a marked improvement. However, we find that the second-order Edgeworth approach has significant size distortions in this setting. Instead, we nonparametrically bootstrap the covariance matrix of the parameter vector β and then use the second-order Edgeworth expansion to modify the t-statistic critical value. Using MacKinnon’s (2011) Monte Carlo design, we find this approach has excellent size properties and has power that is generally superior to the Wild bootstrap approach. We call this technique the “second-order bootstrap” (SOB) approach and recommend it for use in applied research, particularly when there are sample size concerns.

*Corresponding Author. John and Jennie S. MacDonald Professor of Economics, Economics Department, Massachusetts Institute of Technology, E52-271D, 50 Memorial Drive, Cambridge, MA, 02142, USA and NBER; E-mail: jhausman@mit.edu; Telephone: (617) 253-3644; Fax: (617) 253-1330.

†PhD Candidate, MIT Economics Department; E-mail: cjpalmer@mit.edu. Palmer acknowledges support from the National Science Foundation Graduate Research Fellowship under Grant No. 0645960.

2 Traditional Robust Standard Error Estimators

For the model

$$y = X\beta + u \quad (1)$$

with $Var(u) = \Sigma$, the variance of the parameter vector $\hat{\beta}$ estimated by OLS is

$$Var(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}. \quad (2)$$

Let n denote the sample size and k denote the dimension of β . All of the heteroskedasticity-consistent covariance matrix estimators in the literature (commonly denoted HC_j for $j = 0, 1, 2$, etc.) have the same sandwich estimator form with variations in the estimated sample matrix that is used for Σ .

We are interested in test statistics of the form

$$T = \frac{c'\hat{\beta} - c'\beta_0}{\sqrt{c'\hat{V}c}} \quad (3)$$

corresponding to a null hypothesis about a linear combination of the estimated parameters $H_0 : c'\beta = c'\beta_0$, where \hat{V} is an asymptotically valid estimate of $Var(\hat{\beta})$. The following are approaches to estimating $(X'\Sigma X)$ that have appeared in the literature.¹

1. HC_0 (White, 1980) is the original formulation used in White standard errors. White's (1980) contribution was to recognize that $X'\widehat{\Sigma}X$ is a consistent estimator of $X'\Sigma X$ when using the sample matrix

$$\Sigma = \text{diag} \{ \hat{u}_i^2 \}$$

where \hat{u}_i^2 are the fitted residuals from estimating (1) via OLS.

2. HC_1 (MacKinnon and White, 1985) adjusts for degrees of freedom and is the most commonly used robust standard error estimator and is employed by Stata's `robust` option.

$$\Sigma = \frac{n}{n-k} \text{diag} \{ \hat{u}_i^2 \}$$

3. HC_2 (MacKinnon and White, 1985) adjusts for the leverage values h_i where h is the diagonal of the projection matrix $P_X = X(X'X)^{-1}X'$.

$$\Sigma = \text{diag} \left\{ \frac{\hat{u}_i^2}{1-h_i} \right\}$$

4. HC_J (MacKinnon and White, 1985) is the jackknife covariance matrix estimator.

$$\Sigma = \frac{n-1}{n} \left(\text{diag} \{ \tilde{u}_i^2 \} - \frac{1}{n} \tilde{u} \tilde{u}' \right)$$

where $\tilde{u}_i = \frac{\hat{u}_i}{1-h_i}$.

5. HC_3 (Davidson and MacKinnon, 1993) is an approximation to HC_J and is a slight modification of HC_2

$$\Sigma = \text{diag} \left\{ \left(\frac{\hat{u}_i}{1-h_i} \right)^2 \right\}$$

6. HC_4 (Cribari-Neto, 2004) adjusts the residuals by a leverage factor that increases with the leverage.

$$\Sigma = \text{diag} \left\{ \frac{\hat{u}_i^2}{(1-h_i)^{\delta_i}} \right\}$$

where $\delta_i = \min \{ 4, nh_i/k \}$

We consider these approaches in terms of their size in Section 6 below. We find using MacKinnon's (2011) research design that each of the HC_j estimators continues to have significant size distortions when n is of moderate size.

¹We omit HC_5 (Cribari-Neto et al., 2007) from our analysis as it is nearly identical to HC_4 .

3 Bootstrap Estimators

Another class of heteroskedasticity robust estimators uses the Wild bootstrap to estimate the distribution of a given test statistic, forming a rejection region based on the realized bootstrap distribution. The Wild bootstrap involves forming B bootstrap samples using the data generating process

$$y_i^* = X_i \tilde{\beta} + f(\tilde{u}_i) v_i^*$$

where \tilde{u}_i are residuals from an estimate $\tilde{\beta}$ of β , $f(\cdot)$ is one of several candidate transformations of the estimated residuals, and v_i^* is an independent random variable with mean 0 and variance 1. For each bootstrap sample $\{X_i, y_i^*\}$, we estimate $\hat{\beta}_j^*$ where j indexes the bootstrap sample, $j = 1, \dots, B$, and calculate the test statistic of interest T_j^* , as in (3), using a particular heteroskedasticity-robust estimator of the variance of $\hat{\beta}$. Inference is then based on comparing the original test statistic to the $\alpha/2$ and $1 - \alpha/2$ percentiles of $\{T_j^*\}$.

MacKinnon (2011) shows that using the Wild bootstrap to estimate the distribution of test statistics based on HC_1 , using $v_i^* \in \{-1, 1\}$ with equal probability, restricted residuals (i.e. $\tilde{\beta}$ is estimated imposing the null hypothesis), and a transformation of the residuals corresponding to HC_3 , $f(\tilde{u}_i) = \frac{\tilde{u}_i}{1 - \tilde{h}_i}$ (where \tilde{h}_i an element of the diagonal of the restricted projection matrix $P_{\tilde{X}}$) performs best in terms of size and power. The bootstrap will have correct size on average by construction, so its power characteristics determine the usefulness of the approach. We will benchmark our results with this particular variant of the Wild bootstrap and show that our preferred estimator performs comparably in size and much better in power.

4 Second-Order Correction to Test Statistic Distribution

Rothenberg (1988) derives n^{-1} Edgeworth approximations for the distribution functions of test statistics that are linear functions combinations of β such as (3), assuming that the errors are normally distributed. Hall (1992) demonstrates that the second-order Edgeworth expansion approach and the bootstrap approach have the same order of approximation in the case of pivotal test statistics. If the traditional, first-order critical values are $\pm z_{\alpha/2}$, then the second-order approximation critical values t for the test of the null hypothesis $H_0 : c'\beta = c'\beta_0$ are a multiplicative adjustment to $z_{\alpha/2}$:

$$t = \pm z_{\alpha/2} \left(1 - \frac{1}{12}(1 + z_{\alpha/2}^2)V + \frac{a(z_{\alpha/2}^2 - 1) + b}{2n} \right) = \pm z_{\alpha/2} \cdot h \quad (4)$$

where n is the sample size and

$$\begin{aligned} V &= \frac{\sum f_i^4 \hat{u}_i^4}{(\sum f_i^2 \hat{u}_i^2)^2} \\ a &= \frac{\sum f_i^2 g_i^2}{\sum f_i^2 \hat{u}_i^2} \\ b &= \frac{\sum f_i^2 Q_{ii}}{\sum f_i^2 \hat{u}_i^2} \\ f &= nX(X'X)^{-1}c \\ g &= \frac{(I - P_X)\Sigma f}{\sqrt{f'\Sigma f/n}} \\ Q &= nP_X\Sigma(P_X - 2I) \end{aligned}$$

and \hat{u}_i are the fitted residuals and Σ is estimated with HC_0 .

We then calculate the test statistic in equation (3) and make inference by comparing it with the adjusted critical value obtained from equation (4). In other words, we reject the null hypothesis if the test statistic exceeds the adjusted critical value in magnitude $|\hat{T}| > |t|$ and fail to reject otherwise. We refer to this test as the second-order (SO) approach. Applied researchers implementing a SO adjustment may find it convenient to calculate “virtual standard errors” by multiplying given standard errors by the adjustment factor h in (4) and comparing the resulting t-statistics to the traditional asymptotic critical value.

4.1 Bootstrapped \hat{V}

While any estimate \hat{V} of $Var(\hat{\beta})$ can be used in (3), simulation results show that for small samples, the empirical covariance matrix of a vector of nonparametrically bootstrapped $\hat{\beta}$ estimates performs best. To compute this estimated covariance matrix, for $B = 400$ bootstrap iterations we resample (X, y) with replacement from the original data, forming a pairs bootstrap sample (X^*, y^*) . For each iteration j , we then calculate $\hat{\beta}_j^* = (X^{*'}X^*)^{-1}X^{*'}y^*$, and take \hat{V} to be

$$\hat{V} = \frac{1}{B-1} \sum_{j=1}^B \left(\hat{\beta}_j^* - \bar{\beta}^* \right) \left(\hat{\beta}_j^* - \bar{\beta}^* \right)' \quad (5)$$

We refer to inference based on using \hat{V} from (5) in equation (3) compared to the regular asymptotic critical values $\pm z_{\alpha/2}$ as the variance bootstrap (VB). When comparing the VB test statistic to the adjusted critical values in (4), we call this the “second-order bootstrap” (SOB) approach.

5 Simulation Design

The data generating process for the simulations follows MacKinnon (2011) with a sample size of $n = 40$

$$\begin{aligned} y_i &= \beta_1 + \sum_{k=2}^5 \beta_k X_{ik} + u_i \\ u_i &= \sigma_i \varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, 1) \\ \sigma_i &= z(\gamma) (X_i \beta)^\gamma \\ X_{ik} &\sim LN(0, 1), k > 1 \\ \beta_k &= 1, k < 5 \\ \beta_5 &= 0 \end{aligned} \quad (6)$$

where $z(\gamma)$ is a scaling factor that ensures that the average variance of u_i is equal to 1. $\gamma = 0$ corresponds to homoskedasticity, and the degree of heteroskedasticity increases with γ . For context, in this simulation design when $\gamma = 1$, HC_1 robust standard errors are 44% larger than their homoskedastic counterparts, and $\gamma = 2$ corresponds to standard errors that are 70% larger than the corresponding homoskedastic standard errors.

6 Size Results

We compare the performance of the various variance estimators in the test $H_0 : \beta_5 = 0$ with significance level $\alpha = 0.05$ for 10,000 Monte Carlo simulations with varying degrees of heteroskedasticity using the research design in (6). Since the data was generated with $\beta_5 = 0$, this test should reject in approximately 5% of simulations. Table 1 below shows rejection frequencies for three levels of heteroskedasticity, where none, moderate and severe correspond to $\gamma = 0, 1, 2$, respectively. The rejection frequencies of each of the HC_j are decreasing in the degree of the heteroskedasticity. Of the HC_j estimators, HC_3 and HC_J perform the best, although they both over-reject for homoskedasticity and drastically under-reject for severe heteroskedasticity.

The rejection frequencies when we use the Rothenberg second-order critical values with test statistics based on the HC_0 variance estimates, denoted SO, show that the exact adjustment proposed by Rothenberg (1988) performs quite poorly. Indeed, the SO test size is approximately the same as the original White estimator.

The three bootstrap methods perform more consistently across the heteroskedasticity spectrum. It is worth noting that whereas the rejection frequencies of the HC_j estimators decline significantly with the degree of heteroskedasticity, the bootstrap tests perform quite well even under homoskedasticity. The variance

Table 1: Rejection Frequencies for Nominal Size 0.05 Test

| Test statistic | Level of Heteroskedasticity | | |
|----------------|-----------------------------|--------------------------|------------------------|
| | None $\gamma = 0$ | Moderate $\gamma = 1$ | Severe $\gamma = 2$ |
| HC_0 | 0.159 | 0.144 | 0.110 |
| HC_1 | 0.135 | 0.121 | 0.090 |
| HC_2 | 0.106 | 0.085 | 0.049 |
| HC_J | 0.069 | 0.043 | 0.018 |
| HC_3 | 0.067 | 0.041 | 0.017 |
| HC_4 | 0.034 | 0.015 | 0.004 |
| SO | 0.156 | 0.149 | 0.134 |
| VB | 0.042 | 0.033 | 0.021 |
| WB | 0.046 | 0.050 | 0.040 |
| SOB | 0.045 | 0.045 | 0.039 |

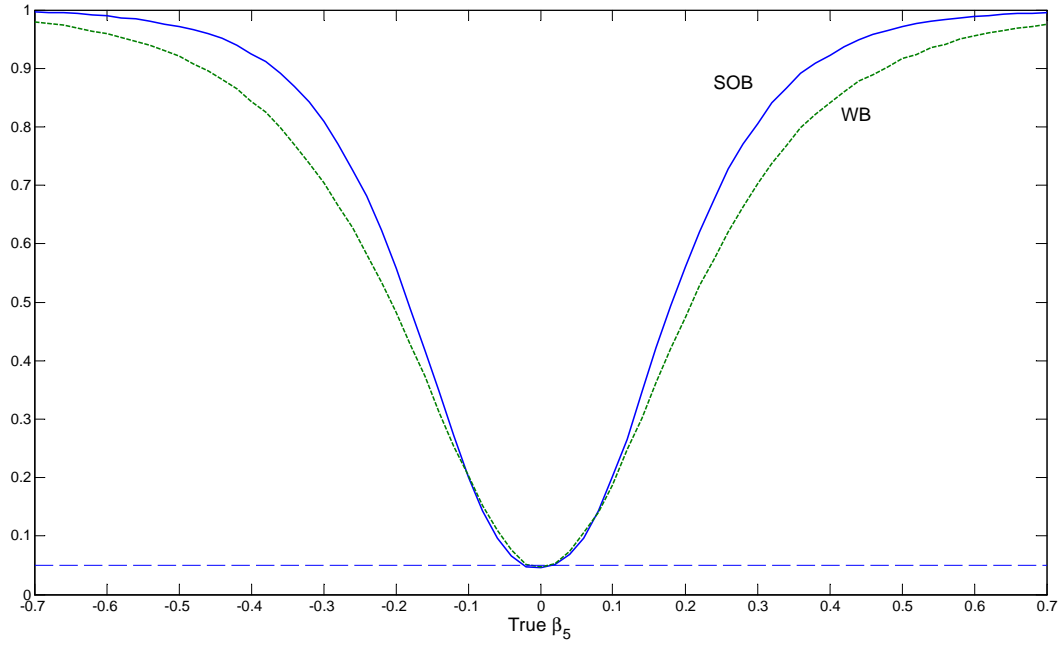
bootstrap (VB) uses the bootstrapped covariance matrix of $\hat{\beta}$ from (5) with $B = 400$ to compute the test statistic (3) and compares it to the regular critical value, i.e. 1.96. Applying the Edgeworth expansion to the distribution of the test statistic to adjust the critical value for the VB test statistic as in (4) (the SOB approach) improves the rejection frequencies considerably for high degrees of heteroskedasticity. Rejection frequencies from the Wild bootstrap (WB) approach are consistently close to their nominal value. Identifying the Wild bootstrap and second-order bootstrap tests as having the best size properties, we now compare the power of these two approaches.

7 Power Results

Figures 1 and 2 examine the power of the Wild bootstrap and second-order bootstrap tests. In each graph, we vary the true value of β_5 and report rejection frequencies of the null hypothesis that $\beta_5 = 0$, with $\alpha = 0.05$. Note that when the true $\beta_5 = 0$, the rejection frequency is the size of the test statistic. Accordingly, a test statistic has good size the closer its rejection frequency is to 0.05 when $\beta_5 = 0$ and has greater power the higher its rejection frequency is for $\beta_5 \neq 0$.

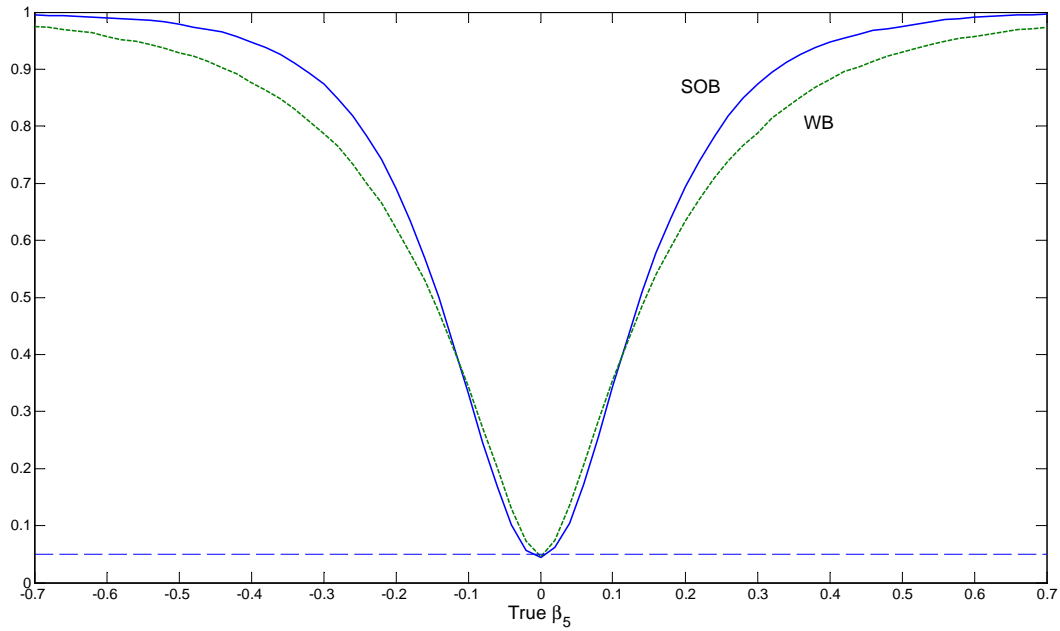
For both $\gamma = 1$ and $\gamma = 2$, the Wild bootstrap and second-order bootstrap tests have quite good size. However, the SOB approach has better power performance than the Wild bootstrap (WB), which is the best of the bootstrap approaches for this design. For any magnitude of the true β_5 greater than approximately 0.1 (where the power performance of the two tests is quite similar), the SOB rejection frequencies are often much higher than the WB rejection frequencies.

Figure 1: Power Results: Moderate Heteroskedasticity



Graph shows rejection frequencies for $H_0 : \beta_5 = 0$ given varying values of the true value of β_5 . $\gamma = 1$, $n = 40$, $\alpha = 0.05$.

Figure 2: Power Results: Severe Heteroskedasticity



Graph shows rejection frequencies for $H_0 : \beta_5 = 0$ given varying values of the true value of β_5 . $\gamma = 2$, $n = 40$, $\alpha = 0.05$.

8 Conclusion

White robust standard errors are universally used in econometrics. Their finite sample properties lead to over-rejection under the null hypothesis, sometimes by a large amount. Over the past 25 years numerous approaches have been suggested to fix the problem. In this paper, we suggest a second-order bootstrap (SOB) approach that has approximately the correct size and superior power properties to the best of the bootstrap approaches.

References

- Cribari-Neto, F. (2004). "Asymptotic inference under heteroskedasticity of unknown form," *Computational Statistics and Data Analysis*, 45, 215-233.
- Cribari-Neto, F., T. C. Souza, and K. L. P. Vasconcellos (2007). "Inference under heteroskedasticity and leveraged data," *Communications in Statistics: Theory and Methods*, 36, 1977-1988.
- Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*, New York, Oxford University Press.
- Davidson, R. and J. G. MacKinnon (2010). "Wild bootstrap tests for IV regression," *Journal of Business and Economic Statistics*, 28, 128-144.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.
- Hannan, E.J. (1970). *Multiple Time Series*. New York: Wiley.
- Hausman, J. and G. Kursteiner (2008). "Difference in difference meets generalized least squares: higher order properties of hypotheses tests," *Journal of Econometrics*, 144, 371-391.
- MacKinnon, J. G., and H. White (1985). "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305-325.
- MacKinnon, J.G. (2011). "Thirty years of heteroskedasticity-robust inference," Queen's Economics Department Working Paper No. 1268.
- Rothenberg, T. J. (1988). "Approximate power functions for some robust tests of regression coefficients," *Econometrica*, 56, 997-1019.
- White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817-838.