

Bootstrapping Heteroskedasticity Consistent Covariance Matrix Estimator

by

Emmanuel Flachaire

EUREQUA, University Paris I Panthéon-Sorbonne

December 2001

Abstract

Recent results of Cribari-Neto and Zarkos (1999) show that bootstrap methods can be successfully used to estimate a heteroskedasticity robust covariance matrix estimator. In this paper, we show that the wild bootstrap estimator can be calculated directly, without simulations, as it is just a more traditional estimator. Their experimental results seem to conflict with those of MacKinnon and White (1985); we reconcile these two results.

JEL classification: C1

Keywords: wild bootstrap, heteroskedasticity

1 Introduction

In this paper, we show that the wild bootstrap covariance matrix estimator can be calculated directly, without simulation, since it is simply a traditional heteroskedasticity consistent covariance matrix estimator first proposed by Eicker (1963) and White (1980), see section 2. Simulation results in MacKinnon and White (1985) had suggested that the jackknife is a reliable heteroskedasticity-consistent variance estimator. Simulation results in Cribari-Neto and Zarkos (1999) suggest that this estimator did not perform as well as its competitors. This conclusion appears to show a conflict between different simulation results and could lead us to the wrong conclusion that the jackknife estimator is not reliable. In section 3, we reconcile these experimental results.

2 HCCME and bootstrap estimators

Consider the linear heteroskedastic model:

$$y_t = X_t\beta + u_t \quad E(u_t|X) = 0, \quad E(u_t^2|X) = \sigma_t^2 \quad (1)$$

where y_t is a dependent variable, X_t a regressor row matrix $1 \times k$, β and σ_t are respectively the unknown parameters and the unknown variances of the error terms; u_t is error term. We consider only unconditional heteroskedasticity, which means that the σ_t^2 may depend on the exogeneous regressors, but not, for instance, on lagged dependent variables. The *Heteroskedasticity Consistent Covariance Matrix Estimator*, or HCCME, is

$$(X^\top X)^{-1} X^\top \hat{\Omega} X (X^\top X)^{-1} \quad (2)$$

where the $n \times n$ diagonal matrix $\hat{\Omega}$ has typical diagonal element $a_t^2 \hat{u}_t^2$, where \hat{u}_t is the OLS residual. We refer to the basic version of the HCCME, proposed by Eicker (1963) and White (1980), as HC_0 where $a_t = 1$. Bias is reduced by multiplying the \hat{u}_t by the square root of $n/(n-k)$, thereby by multiplying the elements of $\hat{\Omega}$ by $n/(n-k)$; this procedure, analogous to the use in the homoskedastic case of the unbiased OLS estimator of the error variance, gives rise to form HC_1 of the HCCME where $a_t = (n/(n-k))^{1/2}$. In the homoskedastic case, the variance of \hat{u}_t is proportional to $1 - h_t$, where $h_t \equiv X_t(X^\top X)^{-1}X_t^\top$, the t^{th} element of the orthogonal projection matrix on to the span of the columns of X . This suggests $a_t = (1 - h_t)^{-1/2}$ in order to obtain $\hat{\Omega}$. If this is done, we obtain form HC_2 of the HCCME. Finally, arguments based on the jackknife lead MacKinnon and White to propose form HC_3 , for which $a_t = (1 - h_t)^{-1}$. MacKinnon and White (1985), and Chesher and Jewitt (1987), show that, in terms of size distortion, HC_0 is outperformed by HC_1 , which is in turn outperformed by HC_2 and HC_3 . The last two cannot be ranked in general, although HC_3 has been shown in a number of Monte carlo experiments to be superior in typical cases.

Heteroskedasticity of unknown form cannot be mimicked in the bootstrap distribution. The wild bootstrap gets round this problem by using a bootstrap data generating process, or bootstrap DGP, of the form

$$y_t^* = X_t \hat{\beta} + a_t \hat{u}_t \varepsilon_t^* \quad (3)$$

where $\hat{\beta}$ is a vector of parameter estimates, a_t takes the forms HC_0 , HC_1 , HC_2 or HC_3 and the ε_t^* are mutually independent drawings, completely independent of the original data, from some auxiliary distribution such that $E(\varepsilon_t^*) = 0$ and $E(\varepsilon_t^{*2}) = 1$. The heteroskedasticity consistent covariance matrix estimate of the OLS parameter estimate under the wild bootstrap DGP (3) is the wild bootstrap covariance matrix estimator. Cribari-Neto and Zarkos (1999) suggest to approximate the wild bootstrap estimator by the following steps:

1. Estimate the model (1) by Ordinary Least Squares (OLS): $\hat{\beta}$ and \hat{u}_t are respectively the parameter estimates and the residuals.
2. Generate a bootstrap sample, based on the bootstrap data generating process: $y_t^* = X_t \hat{\beta} + a_t \hat{u}_t \varepsilon_t^*$, where ε_t^* is drawn from a probability distribution F with zero expectation and variance equal to one. Compute the OLS bootstrap parameter estimate $\hat{\beta}^* = (X^\top X)^{-1} X^\top y^*$. Repeat this step B times.
3. The variance of the B bootstrap estimates is an approximation of the wild bootstrap covariance matrix estimate of $\hat{\beta}$.

We can now give the principal result of this section.

Theorem 1 *For a fixed number of observations and under the data generating process (3), the wild bootstrap covariance matrix estimator is equal to the Heteroskedasticity Consistent Covariance Matrix Estimator (2).*

Proof: The wild bootstrap data generating process is $y_t^* = X_t \hat{\beta} + a_t \hat{u}_t \varepsilon_t^*$. For a fixed number of observations n , the error terms are independent: $E(\varepsilon_t^* \varepsilon_s^* | X) = 0$ with $t \neq s$. The ordinary least squares parameter estimate is equal to: $\hat{\beta}^* = (X^\top X)^{-1} X^\top y^*$ and its covariance matrix is

$$(X^\top X)^{-1} X^\top \Omega^* X (X^\top X)^{-1} \quad (4)$$

where $E(u^* u^{*\top} | X) = \Omega^*$ and $u^* = a_t \hat{u}_t \varepsilon_t^*$. The off-diagonal element of Ω^* for row t and column s , where $t \neq s$, is equal to:

$$\Omega_{t,s}^* = E(u_t^* u_s^* | X) = a_t \hat{u}_t a_s \hat{u}_s E(\varepsilon_t^* \varepsilon_s^* | X) = 0 \quad (5)$$

The diagonal element t is equal to:

$$\Omega_{t,t}^* = E(u_t^{*2} | X) = a_t^2 \hat{u}_t^2 E(\varepsilon_t^{*2} | X) = a_t^2 \hat{u}_t^2 \quad (6)$$

It is clear that, for a fixed number n of observations, (4) is equal to (2). ■

This theorem implies that the wild bootstrap covariance matrix estimator is equal to the HCCME proposed by Eicker (1963), White (1980) and MacKinnon and White (1985). If we approximate the wild bootstrap covariance matrix estimator by simulations, we can observe numerical differences because a *finite* number B of simulations introduces an experimental error. Consequently, the HCCME estimator should always be preferred to the wild bootstrap estimator obtained by simulations.

3 Experimental results

Experimental results of Cribari-Neto and Zarkos (1999) show that the bias of the HC_2 form of the wild bootstrap estimator is less than that of the jackknife estimator. These results are for the bias of the *variance* estimator, but MacKinnon and White (1985) results show that *tests* based on the jackknife estimator of the variance tend to be accurate, its bias notwithstanding. The two sets of results concern fundamentally different objects. In this section we directly compare the size and the bias of different tests. The results lead us to reconcile the conflict claimed by Cribari-Neto and Zarkos. In our Monte-Carlo experiments, we consider a simple linear regression model:

$$y_t = \alpha + x_t\beta + u_t \quad u_t \sim N(0, 1) \quad (7)$$

The true parameters are: $\alpha_0 = 1$ and $\beta_0 = 0$, the error terms are independent and identically distributed following the standard Normal distribution. The sample size is small, $n = 20$, and the regressor is supposed fixed¹.

In subsection 3.1, we experimentally check the theoretical results obtained in the preceding section. In subsection 3.2, we study the bias of the estimators, as in Cribari-Neto and Zarkos (1999). In 3.3, we study the size distortion of the t -tests based on these estimators, as in MacKinnon and White (1985).

3.1 Bootstrap and HCCME estimators

We check the theoretical results of the preceding section: for a fixed number of observations n , the wild bootstrap bootstrap estimator converges to the HCCME estimator when the number of bootstrap samples B increases to infinity. In our experiment, we draw one sample (y, x) using (7), we calculate the HCCME estimator and the wild bootstrap estimator of β with $a_t = (1 - h_t)^{-1/2}$ (HC_2 form) and we increase the number of bootstrap samples: $B = 199; 499; 999; 9,999; 999,999$. We calculate two wild bootstrap estimators: one with the resampling distribution $F_1 \sim N(0, 1)$, and the other with F_2 , a bi-atomic distribution yielding the value $-(\sqrt{5} - 1)/2$ with probability $p = (\sqrt{5} + 1)/2\sqrt{5}$, and $(\sqrt{5} + 1)/2$ with probability $1 - p$, see Liu (1988).

Design	B=199	B=499	B=999	B=9,999	B=999,999
F_1, HC_2	0.06088	0.05557	0.05465	0.05545	0.05637
F_2, HC_2	0.04592	0.05515	0.05539	0.05516	0.05632

Table 1: Wild bootstrap estimator

With our data, the HC_2 form of the HCCME estimator is equal to 0.0562873, the two wild bootstrap estimators are presented in Table 1. These results make clear that, for a fixed number of observations $n = 20$, the wild bootstrap estimators converge to the HCCME estimator as $B \rightarrow \infty$.

¹ $x = -2.2824, -0.435864, 2.27108, -1.05705, -1.10142, 0.648927, 0.143281, -0.25922, 1.87924, -1.32969, 0.013618, -0.303695, 1.24507, 0.670023, 0.658823, 0.521237, -0.0656568, -0.370603, -0.0734635, -0.169986$

3.2 Bias

In this subsection, we report the bias of the wild bootstrap and HCCME estimators of $\hat{\beta}$. In our experiment, we consider the following estimators: the two wild bootstrap estimators described in the preceding subsection, the HC_2 form of the HCCME, and the HC_3 form of the HCCME. MacKinnon and White (1985) investigated the performance of the jackknife estimator, but subsequent computer simulations suggested that the HC_3 form of the HCCME behaves very similarly to the jackknife, see Davidson and MacKinnon (1993, section 16.3).

We calculate the bias as the (mean) estimated variance minus the true variance of $\hat{\beta}$. In addition, we calculate the mean squared error (MSE). The number of replications is $N = 9,999$ and the number of bootstrap samples is $B = 499$. The results are presented in Table 2.

Design	Bias	MSE
wboot, F_1 , HC_2	0.000206	0.000828
wboot, F_2 , HC_2	0.000180	0.000823
HCCME, HC_2	0.000078	0.000809
HCCME, HC_3	0.013018	0.001645

Table 2: Bias and MSE

We find the same results as Cribari-Neto and Zarkos (1999): the bias of the HC_2 form of the wild bootstrap estimator is smaller than that of the HC_3 form of the HCCME. Furthermore, these results make clear that the same forms of the wild bootstrap and of the HCCME estimators have similar finite sample behavior, and that the wild bootstrap estimators (which include experimental error) are substantially worse, as theory suggests.

3.3 Size distortion

In this subsection we report the size distortion of tests based on the HC_2 and HC_3 forms of the HCCME of $\hat{\beta}$. We do not study the wild bootstrap estimator because we have shown that its finite sample behavior is quite similar to that of the HCCME. We test the null hypothesis $H_0 : \beta = 0$ with the squared t -statistic, which is asymptotically $\chi^2(1)$: $\tau = t^2 = [\hat{\beta}/\hat{S}(\hat{\beta})]^2$, where $\hat{\beta}$ is the ordinary least square parameter estimate and $\hat{S}(\hat{\beta})^2$ the second diagonal element of the covariance matrix estimator. The realizations of the statistics $\hat{\tau}$ are used to calculate a P -value at the nominal level $\alpha = 0.05$. The number of replications is $N = 9,999$ and the number of bootstrap samples is $B = 499$.

estimator	<i>Empirical level</i>
HCCME, HC_2	0.0774
HCCME, HC_3	0.0547

Table 3: Empirical level ($\alpha = 0.05$)

Table 3 presents the empirical level of the t -statistics based on the HC_2 and HC_3 forms of the HCCME estimators. Tests based on the HC_2 estimator reject the null hypothesis

more than 5% of the time when it is true (7.74%), whereas tests based on the HC_3 reject near 5%. We find here the same results as those of MacKinnon and White (1985): tests based on the jackknife estimator, or HC_3 , are more reliable than others.

4 Conclusion

In this paper, we show that the wild bootstrap is not useful to estimate a heteroskedasticity consistent covariance matrix. However, it is well known that the bootstrap yields better results when applied to asymptotic pivots, and it is therefore a better idea to use HCCME in order to construct such asymptotic pivots before using the bootstrap, see Horowitz (2000) for an account of the relevant issues.

References

- Chesher, A. and I. Jewitt (1987). “The bias of a heteroskedasticity consistent covariance matrix estimator”. *Econometrica* 55, 1217–1222.
- Cribari-Neto, F. and S. G. Zarkos (1999). “Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing”. *Econometric Reviews* 18, 211–228.
- Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Eicker, B. (1963). “Limit theorems for regression with unequal and dependant errors”. *Annals of Mathematical statistics* 34, 447–456.
- Horowitz, J. L. (2000). “The bootstrap”. In *Handbook of Econometrics*, Volume 5. J. J. Heckman and E. E. Leamer (eds), Elsevier Science.
- Liu, R. Y. (1988). “Bootstrap procedure under some non-i.i.d. models”. *Annals of Statistics* 16, 1696–1708.
- MacKinnon, J. G. and H. L. White (1985). “Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties”. *Journal of Econometrics* 21, 53–70.
- White, H. (1980). “A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity”. *Econometrica* 48, 817–838.