Hal White: Time at MIT and Early Days of Research

Jerry Hausman

I am pleased to write a note in honor of Hal's sixtieth birthday. As memory is a tricky thing, I hope the following is accurate, but I can make no guarantees. Hal was among my first three Ph.D. students at MIT who are Roger Gordon, Hal White, and Paul Krugman. Since my first and only job has been at MIT, I found this supervision enjoyable and remarkably easy. However, using these three sample points to predict the future would have been grossly inaccurate, although mostly I have enjoyed the many MIT (and some Harvard) students I have supervised over the years.

Hal's thesis was in applied labor economics: "A Microeconomic Model of Wage Determination: Econometric Estimation and Application". His other supervisors were Bob Solow and Lester Thurow. While the thesis uses interesting econometric methods, the question may arise of why Hal did not write an econometric methodology thesis. My memory is as follows. In those days MIT students finished in four years. Then and now, MIT has much less fellowship money than Harvard and other universities which can guarantee a longer period. Nevertheless, MIT has far outdistanced other universities in the past 40 years in producing top graduate students. I told Hal that if by May of his third year he did not have an *Econometrica* level paper in process he should do an applied thesis so he could be sure to finish on time. Perhaps I was too young and inexperienced at the time to give better thesis advice. However, the initial conditions of his thesis being in an applied topic had no effect on his subsequent research which I now turn to.

Hal took three econometrics courses from me so, of course, I chose him as one of my teaching assistants (TA). He was the TA for all three courses so we talked about the topics in the courses a lot since my yellow note pages were just beginning to take shape at that time. Hal was a terrific TA, which the students appreciated given my teaching approach. He led the way for many subsequent TAs including Bernanke, Paul Ruud, Mark Machina, Whitney Newey, and Ken West over the next few years after Hal.

J. Hausman (✉)
MIT Department of Economics, Cambridge, MA, USA
e-mail: jhausman@mit.edu

X. Chen and N. R. Swanson (eds.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, DOI: 10.1007/978-1-4614-1653-1_8,

1 My Approach Before White Standard Errors

In the introductory econometrics course I discussed at length the problem of inference when the covariance matrix was not diagonal. How to estimate $(X'X)^{-1}(X'\Lambda X)(X'X)^{-1}$ where Λ is a diagonal matrix of unknown form? I pointed out that estimation of the middle matrix followed from the approach of the Berndt–Hall–Hall–Hausman (BHHH) algorithm: $D = \Sigma X_i \varepsilon_i \varepsilon_i' X_i'$ where D is an estimate of the middle matrix and (ε_i') is the least squares estimate of the residual. I missed the " White standard error" formula, but instead I took two alternative approaches. I taught that one could do FGLS (feasible GLS) using a specification for ε_i^2 using a general polynomial approach based on the X_is. I emphasized that the specification did not have to be "correct" since in large samples the estimates of the slope parameters would continue to be consistent and hopefully have reduced variance relative to the OLS estimates. However, in the absence of the correct specification for ε_i^2 the correct standard errors and t-tests were not available.

My other approach was to explore "how bad" things could be if one did OLS. Here I used " Watson's bound" (Watson 1967; Hannan 1970), to derive a formula in the single regressor setup.[1] The question at hand is how badly will OLS perform relative to GLS where the answer depends on both X and Λ. The efficiency measure is *EB* easy to compute:

$$EB(X, \Lambda) = \left(\sum_i \left(\frac{x_i^2}{\sigma_i^2}\right)\right)^{-1}.$$

Thus, the relationship of x_i and σ_i^2 is clear. Now for fixed X the approach is to maximize over Λ which is a straightforward characteristic value problem. Let us arrange the λ_is from smallest to largest, $0 \leq \lambda_1 \leq \cdots \leq \lambda_N$ where N is the sample size. Watson's bound, which is attainable, is

$$EB(X, \Lambda) \geq \frac{4\lambda_1^2\lambda_N^2}{2 + \left(\frac{\lambda_1^2}{\lambda_N^2}\right) + \left(\frac{\lambda_N^2}{\lambda_1^2}\right)}$$

where the estimates of σ_i^2 can replace the λ_i^2 in the efficiency bound formula. Calculating the bound for some illustrative values of the ratio of the largest to smallest variance yields:

Efficiency Bounds

Ratio	1.2	1.5	2	5	10	100
Bound	0.992	0.960	0.889	0.556	0.331	0.039

[1] I also did analysis with multiple right-hand side variables but I will not discuss the results here since they are more complicated.

Thus, for cross-sectional data the loss in efficiency is typically not very large. Indeed, for typical *inid* cross-section Xs, a refined calculation demonstrates that the efficiency losses tend to be even smaller. However, as the number of right-hand side variables grows, the bound deteriorates. Thus, the efficiency loss in a given sample for fixed Xs could be calculated and a bound over all Xs was also available. However, unless we know the correct specification for the variances in the FGLS approach, we still did not have an estimate of the standard errors, since this period preceded the bootstrap approach.

Here, Hal solved the problem with his formula for "White standard errors", (White 1980). I first heard Hal's paper at the weekly joint Harvard–MIT Thursday econometrics seminar. I remember walking out of the seminar and saying to my close friend and co-author Zvi Griliches that Hal's paper would change the way we do econometrics. Zvi was less enthusiastic, perhaps since Hal had not been his student, but my prediction appears correct given the presence of Hal's formula in all econometric software packages and the large number of citations. Two further points. I talked with Hal when he was my TA about the "Hausman specification test" approach, and his comments were quite helpful. Second, the "Newey-West standard errors" for time series applications followed from two subsequent TAs for my econometrics courses. So I am pleased thinking that all applied econometrics computer output for regressions should have output arising from my TAs at MIT.

2 Finite Sample Approach

As Hal's thesis supervisor, I now propose a possible finite sample improvement to his approach. James MacKinnon (2011) in his paper in this volume has explored various bootstrap approaches to improve the finite sample performance of White standard error estimation. Since I am interested in inference, I will explore possible improvements in the behavior of the "t-test" using Hal's approach. However, if one is interested in the estimated standard errors, one can derive an estimate using division on the refined t-test divided by the OLS estimate of the parameter.

In large samples the asymptotic approximation assumes we know the true σ_i^2s. However, in finite samples we use estimates of these parameters. Guido Keurstiener and I explored this effect in inference from FGLS applied to difference-in-differences models in Hausman and Kuersteiner (2008). We found that taking account of the unknown variance estimates using Rothenberg (1988) second order Edgeworth expansion approach led to much more accurate sized tests. Also, we found that the second order expansion approach did better than the bootstrap in terms of power. Thus, I have applied a modification to the second order expansion to calculation of t-tests based on estimated White standard errors.[2] I have used the design framework

[2] This research is done jointly with Christopher Palmer, one of my current TAs. See Hausman and Palmer (2012).

from James MacKinnon's paper to see how his bootstrap approaches compare with the second-order refinement approach.

I consider the test statistic for linear combinations of the parameters and the null hypothesis $H_0 : c'\beta = c'\beta_0$. The associated t-statistic is

$$T = \frac{c'\hat{\beta} - c'\beta_0}{\sqrt{c'\hat{V}c}} \tag{1}$$

We first consider the size of various tests where the estimate of Σ in the middle matrix $X'\Sigma X$ takes various forms:[3]

1. HC0: White approach using the OLS residuals to estimate

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

2. HC1: this approach adjusts for degrees of freedom and is the most commonly used form:

$$\Sigma = \frac{n}{n-k}\text{diag}\left\{\hat{u}_i^2\right\} \tag{3}$$

3. HC2: this approach adjusts for the "leverage" values h_i, where h is the diagonal of the projection matrix.

$$\Sigma = \text{diag}\left\{\frac{\hat{u}_i^2}{1-h_i}\right\} \tag{4}$$

where $h = \text{diag}\,(P_X)$ and $P_X = X(X'X)^{-1}X^1$ is the projection matrix of X.

4. HC3: this approach is an approximation to the jackknife covariance matrix HCJ, which I emit here because it is computationally more complicated and provides nearly identical results. HC3 is a slight modification of HC2:

$$\Sigma = \text{diag}\left\{\left(\frac{\hat{u}_i}{1-h_i}\right)^2\right\} \tag{5}$$

See MacKinnon (zou) for results containing HC4, which I omit because of its poor size performance in this design.

I compare these estimators to the Rothenberg second order Edgeworth approximation. This approach modifies the traditional two-sided critical values $Z_\alpha/2$ for a t-statistic of null hypothesis $\text{Ho}^2\ \text{C}'\beta = \text{C}'\beta_0$ with the equation:

$$t = \pm z_{\alpha/2}\left(1 - \frac{A}{2n}\right) \tag{6}$$

where n is the sample size and

[3] I use the same notation that MacKinnon uses in his paper in this volume.

$$A = \frac{1}{4}(1 + z_{\alpha/2}^2)V_W - a(z_{\alpha/2}^2 - 1) - b$$
$$V_W = \frac{2n}{3}\frac{\sum f_i^4 \hat{u}_i^4}{\left(\sum f_i^2 \hat{u}_i^2\right)^2}$$
$$a = \frac{\sum f_i^2 g_i^2}{\sum f_i^2 \hat{u}_i^2}$$
$$b = \frac{\sum f_i^2 Q_{ii}}{\sum f_i^2 \hat{u}_i^2}$$
$$f = nX(X'X)^{-1}\mathrm{C}$$
$$g = \frac{M\Sigma f}{\sqrt{f'\Sigma f/n}}$$
$$Q = n(M\Sigma M - \Sigma)$$
$$M = I - P_X$$

and $\hat{u}_i$ are the fitted residuals.

However, the experience of applying this formula in Hausman and Palmer (2012) was that it has significant size distortions. Thus, I apply a non-parametric bootstrap to estimate β. For B bootstrap iterations, I resample (X, y) with replacement from the original data, forming a bootstrap sample (X^*, y^*). For each iteration j, I then calculate $\beta_j^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$, and take $\hat{V}$ to be

$$\hat{V} = \frac{1}{B-1}\sum_{j=1}^{B}\left(\hat{\beta}_j^* - \bar{\hat{\beta}}^*\right)\left(\hat{\beta}_j^* - \bar{\hat{\beta}}^*\right)'. \tag{7}$$

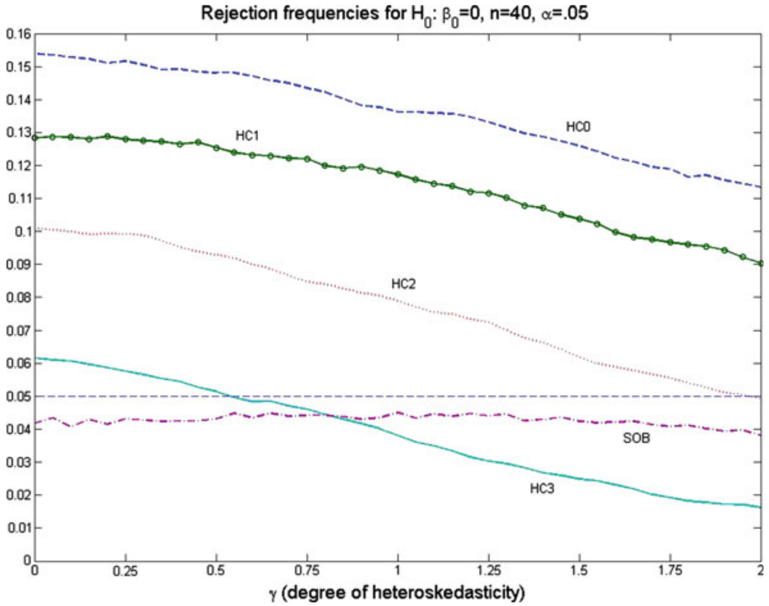
I use this estimated covariance matrix to calculate the test statistic in equation (1) and make interference by comparing it with the adjusted critical value obtained from equation (6), an approach. I refer to as the " second order bootstrap", or SOB, method.

To test our approach I use the same simulation design as MacKinnon (2011)

$$y_i = \beta_1 + \sum_{k=2}^{5}\beta_k X_{ik} + u_i$$
$$u_i = \sigma_i \varepsilon_i$$
$$\varepsilon_i \sim \mathcal{N}(0, 1)$$
$$\sigma_i = z(\gamma)\,(X_i\beta)^{\gamma}$$
$$X_{ik} \sim LN(0, 1) \qquad \text{for} \quad k \geq 2$$
$$\beta_k = 1 \qquad \text{for} \quad k < 5$$
$$\beta_5 = 0$$

where $\gamma = 0$ corresponds to homoskedasticity, and the degree of heteroskedasticity increases with γ, and $z(\gamma)$ is a scaling factor which ensures that the average variance of u_i is equal to 1.

I first consider the sizes of the various approaches. I test $H_0 : \beta_5 = 0$ with a size of $\alpha = 0.05$. In Graph 1 we see that HC0 rejects much too often as is well-recognized.



The alternatives HC1, HC2, and HC3 offer improvements, but all have significant size distortions. The "second order bootstrap" (SOB) approach has acceptable size performance, being the best of the alternatives considered.[4]

I now consider power performance and compare the SOB approach to a bootstrap approach to the White test. It is well recognized, e.g., (Hall 1992), the bootstrapped test statistic for a pivotal situation has the same order of approximation as the second-order approach. MacKinnon finds the wild bootstrap to perform the best using the following specification. The wild bootstrap involves forming B bootstrap samples using the data generating process

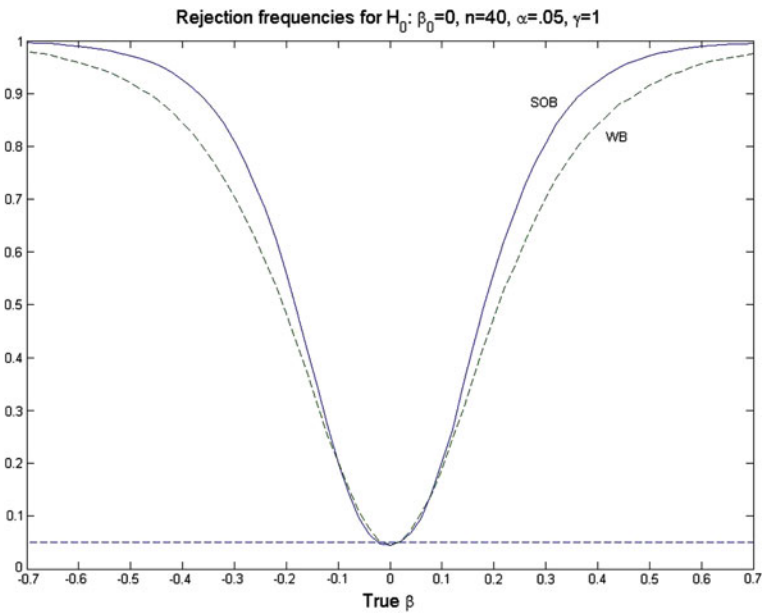
$$y_i^* = X_i \tilde{\beta} + f(\tilde{u}_i) v_i^*,$$

[4] I do not consider the bootstrap form of the White test since Hausman and Palmer (2012) find it has significant size distortions and will be inferior in terms of the higher order expansions to the bootstrap version of the test I consider below.

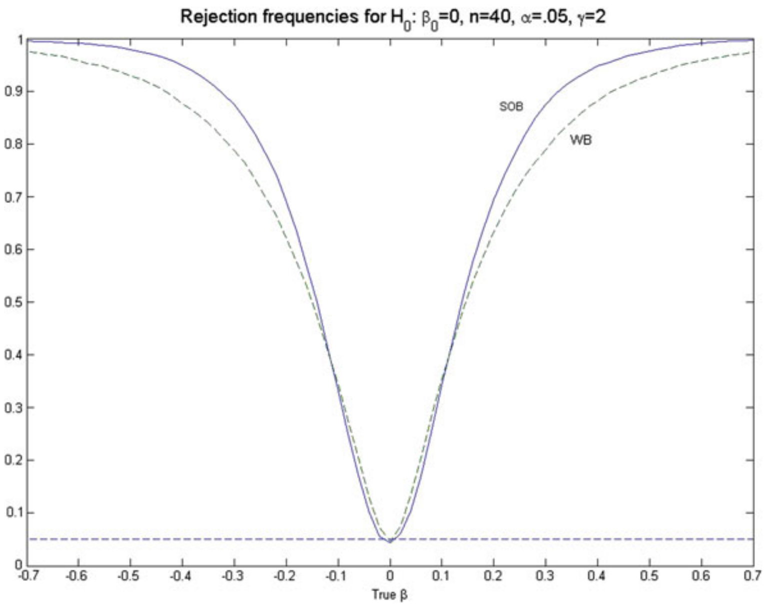
where $\tilde{u}_i$ are residuals from an estimate $\tilde{\beta}$ of β, $f(\cdot)$ is one of several candidate transformations of the estimated residuals, and v_i^* is a random variable with mean 0 and variance 1, such that $E\left(f(\tilde{u}_i)v_i^*\right) = 0$. For each bootstrap sample $\left\{X_i, y_i^*\right\}$, I estimate $\hat{\beta}_j^*$ where j indexes the bootstrap sample, $j = 1, \ldots, B$, and calculates the test statistic of interest, as in (1), using a particular heteroskedasticity-robust estimator of the variance of $\hat{\beta}$.

MacKinnon (2011) shows that using the wild bootstrap to estimate the distribution of test statistics based on $HC1$, using $v_i^* \in \{-1, 1\}$ with equal probability, restricted residuals (i.e. $\tilde{\beta}$ is estimated imposing the null hyptothesis), and a transformation of the residuals corresponding to $HC3$, $f(\tilde{u}_i) = \frac{\tilde{u}_i}{1-\tilde{h}_i}$ (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.

I now compare the second order-bootstrap (SOB) approach to the best bootstrap approach found by MacKinnon. In Graph 2, we see that the SOB approach has good size properties as does the wild bootstrap (WB), by construction.



However, the SOB statistic has considerably greater power than the WB. In Graph 3 for the case of severe heteroskedasticity, we find a similar result.



The size of both tests is quite accurate, but the power of the SOB approach exceeds the power of the WB by a considerable margin. Thus, I conclude that the second order bootstrap (SOB) approach appears to be better than alternative approaches to calculating the White test in finite samples.

The refinement to the t-tests arising from the second-order bootstrap approach is straightforward to program for econometric software. Thus, I recommend that econometric software providers include the refined SOB formula since it is typically (weakly) more accurate than the standard White formula.[5]

Hal has written many other important papers since his heteroscedasticity paper. I recommend to the reader the papers in this volume to see the breadth of Hal's research interests and contributions. I take great pride in Hal's accomplishments over the years and congratulate him and the conference organizers for celebrating Hal's sixtieth birthday.

Acknowledgments I thank Christopher Palmer for assisting in the preparation of this note.

[5] Similar refinements could be quite useful in the case of Newey-West and GMM estimated covariance matrices, where the number of unknown parameters estimates is significantly larger than in the heteroscedasticity situation.

References

Cribari-Neto, F. (2004). " Asymptotic inference under heteroskedasticity of unknown form", Computational Statistics and Data Analysis, 45, 215–233.

Davidson, R. and J. G. MacKinnon (1993). Estimation and Inference in Econometrics, New York, Oxford University Press.

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. Kuersteiner (2008). " Difference in difference meets generalized least squares: higher order properties of hypotheses tests," Journal of Econometrics, 144, 371–391.

Hausman J. and C. Palmer (2012). " Heteroskedasticity Robust Inference in Finite Samples, Economics Letters, 116(2), 232–235.

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.

Watson, G.S. (1967). " Linear least squares regression", Annals of Mathematical Statistics, 38, 1679–99.

White, H. (1980). " A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", Econometrica, 48, 817–838.