Heteroskedasticity-robust tests of linear regression coefficients: A review and evaluation of small-sample corrections

James E. Pustejovsky and Gleb Furman
University of Texas at Austin
Educational Psychology Department

May 17, 2016

Abstract

The text of your abstract. 200 or fewer words.

Keywords: heteroskedasticity; sandwich estimator; robust covariance estimator; linear regression; Satterthwaite approximation; saddlepoint approximation; Edgeworth approximation

1 Introduction

Linear regression models, estimated by ordinary least squares (OLS), are one of the most important and ubiquitous tools in applied statistical work. Classically, hypothesis tests and confidence intervals for linear regression coefficients rely on the assumption that the model errors are homoskedastic, or have constant variance for all values of the covariates. In practice though, it can be difficult to diagnose violations of this assumption, and similarly difficult to construct and defend other assumptions about how the error variances related to the covariates. Thus, it is often desirable to use methods of inference that remain valid for models with heteroskedasticity of an unknown form.

Connect to Behrens-Fisher problem and cluster-robust variance estimation?

One well-known approach to inference in this setting is based on heteroskedasticity-consistent covariance matrix estimators (HCCMEs), which provide asymptotically consistent estimates of the sampling variance of OLS coefficient estimates under quite general conditions. HCCMEs were introduced in the statistics literature by Huber (1967) and Eicker (1967), in the survey sampling literature byand in econometrics by White (1980). They are an attractive tool because they rely on weaker assumptions than classical methods. However, they also have the drawback that it is not always clear whether a given sample is sufficently large to trust the asymptotic approximations by which they are warranted. Furthermore, when the sample size is small, it is known that some of the HCCMEs tend to be too liberal, producing variance estimates that are biased towards zero and hypothesis tests with greater than nominal size (Long & Ervin 2000).

add citations

Since White (1980) introduced the HCCME in econometrics, methods for improving the finite-sample properties of HCCMEs have been studied extensively. The most well-known strand of this work has considered modifications to the HCCME itself that lead to more accurate tests and CIs in finite samples. MacKinnon & White (1985) and Davidson & MacKinnon (1993) proposed several such modifications that are now readily available in software. Based upon an extensive set of simulations, Long & Ervin (2000) demonstrated that one of these modifications, known as HC3, performs substantially better than the others. As a result, HC3 is the default in software such as the R package `sandwich` (Zeileis 2004), although White's original HCCME remains the default in SAS `proc reg` and Stata's `regress` command with `vce(robust)`. More recently, several further variations on the

HCCMEs have been proposed (Cribari-Neto 2004, Cribari-Neto et al. 2007, Cribari-Neto & da Silva 2011), which aim to improve upon the performance of HC3 in models where the regressors exhibit high leverage. For hypothesis testing, HCCMEs are typically used to calculate t-statistics, which are compared to standard normal or $t(n-p)$ reference distributions, where n is the sample size and p is the dimension of the coefficient vector.

An alternative approach to improving the small-sample properties of hypothesis tests based on HCCMEs is to find a better approximation to the null sampling distribution of the test statistic. Several such approximations have been proposed, including Satterthwaite approximations (Lipsitz et al. 1999), Edgeworth approximations (Rothenberg 1988, Kauermann & Carroll 2001), and saddlepoint approximations (McCaffrey & Bell 2006). Although there is evidence that each of these approximations improves upon the standard, large-sample tests, their performance has been examined only under a limited range of conditions. Moreover, it appears that these approximations have been developed in isolation, without reference to previous work, and they have received little or no attention in recent reviews (e.g., none are discussed by MacKinnon 2013). In contrast to the various HC corrections, to our knowledge, none of the distributional approximations are implemented in standard software packages for data analysis.

Add more on gaps in literature.

In this paper, we review the various small-sample approximations for hypothesis tests based on HCCMEs, using a common notation in order to facilitate comparisons among them. In so doing, we identify several further variations on the approximations that have not previously been considered. We then evaluate the performance of these approximations, along with the standard methods, in a large simulation study. The design of the simulation study is modeled on the earlier study of Long & Ervin (2000).

Outline paper

2 Methods

We will consider the regression model

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i, \tag{1}$$

for $i = 1, ..., n$, where y_i is the outcome, $\mathbf{x}_i$ is a $1 \times p$ row-vector of covariates (including an intercept) for observation i, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, and ϵ_i is a mean-zero error term with variance σ_i^2. We shall assume that the errors are mutually independent. For ease of notation, let $\mathbf{y} = (y_1, ..., y_n)'$ denote the $n \times 1$ vector of outcomes, $\mathbf{X} = (\mathbf{x}_1', ..., \mathbf{x}_n')'$ be the $n \times p$ design matrix, and $\boldsymbol{\epsilon}$ be the $n \times 1$ vector of errors with $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, ..., \sigma_n^2)$. Let $\mathbf{M} = (\mathbf{X}'\mathbf{X}/n)^{-1}$. Let $\hat{\boldsymbol{\beta}} = \mathbf{M}\mathbf{X}'\mathbf{y}/n$ denote the vector of OLS estimates and $e_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$, $i = 1, ..., n$ denote the residuals.

The goal is to test a hypothesis regarding a linear combination of the regression coefficients $\mathbf{c}'\boldsymbol{\beta}$, i.e., $H_0 : \mathbf{c}'\boldsymbol{\beta} = k$, with Type-I error rate α. All tests under consideration are based on the Wald statistic

$$T(V) = \frac{\mathbf{c}'\hat{\boldsymbol{\beta}} - k}{\sqrt{V}}, \tag{2}$$

where V some estimator for $\mathrm{Var}\left(\mathbf{c}'\hat{\boldsymbol{\beta}}\right)$.

If the errors are homoskedastic, so that $\sigma_i^2 = \sigma^2$ for $i = 1, ..., n$, then the hypothesis can be tested using a standard t test. The variance of $\mathbf{c}'\boldsymbol{\beta}$ is then estimated by $V^{hom} = \hat{\sigma}^2\mathbf{c}'\mathbf{M}\mathbf{c}$, where $\hat{\sigma}^2 = \left(\sum_{i=1}^n e_i^2\right)/(n-p)$. Under H_0 and assuming that the errors are normally distributed, the test statistic follows a t distribution with $n-p$ degrees of freedom. Thus, H_0 is rejected if $\left|T\left(V^{hom}\right)\right| > F_t^{-1}\left(1 - \frac{\alpha}{2}; n-p\right)$, where $F_t^{-1}(x;\eta)$ is the quantile function for a t distribution with η degrees of freedom. If the errors are instead heteroskedastic, the variance estimator V^{hom} will be inconsistent and this t test will generally have incorrect size.

2.1 HCCMEs

Under the general model that allows for heteroskedasticity, the true variance of the OLS estimator is

$$\mathrm{Var}\left(\mathbf{c}'\hat{\boldsymbol{\beta}}\right) = \frac{1}{n}\mathbf{c}'\mathbf{M}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}'\right)\mathbf{M}\mathbf{c} \tag{3}$$

The HCCCMEs estimate $\mathrm{Var}\left(\mathbf{c}'\hat{\boldsymbol{\beta}}\right)$ by replacing the σ_i^2 with estimates involving the squared residuals. They all have the same general form:

$$V^{HC} = \frac{1}{n}\mathbf{c}'\mathbf{M}\left(\frac{1}{n}\sum_{i=1}^{n}\omega_i e_i^2\mathbf{x}_i\mathbf{x}'\right)\mathbf{M}\mathbf{c} \tag{4}$$

where $\omega_1, ..., \omega_n$ are weighting terms that differ for the various HC estimators. Under weak assumptions, the weak law of large numbers ensures that the middle term in Equation (4) converges to the corresponding term in (3) as the sample size increases.

White (1980) originally described the HCCME without any correction factor, which is equivalent to taking $\omega_i = 1$ for $i = 1, ..., n$. This form has come to be known as HC0. Subsequently, various correction factors have been proposed that aim to improve on the finite-sample behavior of HC0. Following common convention, we refer to these correction factors by number. Letting $h_i = n\mathbf{x}_i\mathbf{M}\mathbf{x}_i'$ denote the hat value for unit i, the correction factors are as follows:

$$
\begin{aligned}
\text{HC1:}\quad & \omega_i = n/(n-p) & & \\
\text{HC2:}\quad & \omega_i = (1-h_i)^{-1} & & \\
\text{HC3:}\quad & \omega_i = (1-h_i)^{-2} & & \\
\text{HC4:}\quad & \omega_i = (1-h_i)^{-\delta}, & & \delta_i = \min\{h_i n/p, 4\} \\
\text{HC4m:}\quad & \omega_i = (1-h_i)^{-\delta}, & & \delta_i = \min\left\{h_i n/p, 1\right\} + \min\left\{h_i n/p, 1.5\right\} \\
\text{HC5:}\quad & \omega_i = (1-h_i)^{-\delta}, & & \delta_i = \frac{1}{2}\min\left\{h_i n/p, \max\left\{4, 0.7 h_{(n)} n/p\right\}\right\}
\end{aligned}
$$

MacKinnon & White (1985) suggested HC1, which uses an ad hoc correction similar to the correction used for $\hat{\sigma}^2$, and HC2, which has the property that V^{HC2} is exactly unbiased when the errors are homoskedastic. Davidson & MacKinnon (1993) proposed HC3 as an approximation to the leave-on-out jackknife variance estimator.

Cribari-Neto and colleagues subsequently proposed three further variations, HC4 (Cribari-Neto 2004), HC4m (Cribari-Neto & da Silva 2011), and HC5 (Cribari-Neto et al. 2007), all of which aim to improve upon HC3 for design matrices where some observations are very influential. All of these correction factors inflate the squared residual terms to a greater extent when the observation has a higher degree of leverage. HC4 truncates the degree of inflation at 4 times the average leverage. Compared to HC4, HC4m inflates observations with lower leverage more strongly, but it also truncates the maximum degree of inflation at 2.5 times the average. In HC5, the truncation depends on the maximum leverage value but the degree of inflation will tend to be smaller than HC4.

For any of the HCCMEs, the robust Wald statistic $T\left(V^{HC}\right)$ converges in distribution to $N(0,1)$ as n increases to infinity. Thus, any asymptotically correct test can be constructed by rejecting H_0 when $\left|T\left(V^{HC}\right)\right|$ is greater than the $1-\alpha/2$ critical value from a standard normal distribution. In practice, it is common to instead use the critical value from a t distribution with $n-p$ degrees of freedom. However, use of the t_{n-p} reference distribution is only an ad hoc approximation.

The following subsections review several distinct, better-grounded approximations to the null sampling distribution of $T\left(V^{HC}\right)$. As will be seen, all of the approximations depend on the structure of the unknown error variances. A key consideration in developing these approximations is how to estimate the error variances. Past proposals have each considered different strategies, including estimating the errors empirically (as in the HCCME itself) or by assuming that they follow a known structure.

2.2 Satterthwaite approximation

Lipsitz et al. (1999) proposed a hypothesis testing procedure that is based on a Satterthwaite approximation for the distribution of $T\left(V^{HC}\right)$, using the HC2 form of the variance estimator. The Satterthwaite approximation involves approximating the distribution of V^{HC} by a multiple of a χ^2 distribution with degrees of freedom η chosen by match the first two moments of V^{HC}. In the abstract, the degrees of freedom are given by

$$\eta = 2\left[\mathrm{E}\left(V^{HC}\right)\right]^2 / \mathrm{Var}\left(V^{HC}\right).$$

The null hypothesis is then tested by comparing $T\left(V^{HC}\right)$ to a t-distribution with η degrees of freedom.

The moments of V^{HC} can be obtained by observing that the variance estimator is a quadratic form in the errors, which can be written as

$$V^{HC} = \sum_{i=1}^{n} \omega_i \left(g_i e_i / n\right)^2 = \mathbf{e}'\mathbf{A}\mathbf{e},$$

where $g_i = \mathbf{x}_i\mathbf{M}\mathbf{c}$ and $\mathbf{A} = \mathrm{diag}\left(\omega_1 g_1^2/n^2, ..., \omega_n g_n^2/n^2\right)$. It follows from the properties of quadratic forms that $\mathrm{E}\left(V^{HC}\right) = \mathrm{tr}\left[\left(\mathbf{I}-\mathbf{H}\right)\mathbf{A}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{\Sigma}\right]$, where $\mathbf{I}$ is an $n\times n$ identity matrix and $\mathbf{H} = n\mathbf{X}\mathbf{M}\mathbf{X}'$ is the hat matrix. Furthermore, assuming that the errors are normally distributed, the variance of the quadratic form is

$$\begin{aligned}
\mathrm{Var}\left(V^{HC}\right) &= 2\mathrm{tr}\left[\left(\mathbf{I}-\mathbf{H}\right)\mathbf{A}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{\Sigma}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{A}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{\Sigma}\right] \\
&= 2\mathrm{tr}\left[\left(\mathbf{I}-\mathbf{H}\right)\mathbf{A}\left(\mathbf{I}-\mathbf{H}\right)\left[\left(\left(\mathbf{I}-\mathbf{H}\right)\mathbf{A}\left(\mathbf{I}-\mathbf{H}\right)\right)\circ\mathbf{S}\right]\right],
\end{aligned}$$

where $\circ$ denotes the element-wise (Hadamard) product and $\mathbf{S}$ has entries $S_{ij} = \sigma_i^2 \sigma_j^2$.

The mean and variance involve the unknown quantity $\mathbf{\Sigma}$ and thus must be estimated. Lipsitz and colleagues propose to treat V^{HC} as unbiased and to estimate $\mathbf{S}$ using the matrix with entries

$$\hat{S}_{ii} = \frac{e_i^4}{3(1-h_{ii})^2} \quad \text{for } i = 1, ..., n \quad \text{and} \quad \hat{S}_{ij} = \frac{e_i^2 e_j^2}{2h_{ij}^2 + (1-h_{ii})(1-h_{jj})} \quad \text{for } i \neq j.$$

They then construct estimated degrees of freedom by substituting $\mathbf{c}'\mathbf{V}^{HC}\mathbf{c}$ in place of its expectation and taking

$$\eta_E = \frac{\left(\mathbf{c}'\mathbf{V}^{HC}\mathbf{c}\right)^2}{\text{tr}\left[\left(\mathbf{I}-\mathbf{H}\right)\mathbf{A}_{2q}\left(\mathbf{I}-\mathbf{H}\right)\left[\left(\left(\mathbf{I}-\mathbf{H}\right)\left(\mathbf{I}-\mathbf{H}\right)\right)\circ\hat{\mathbf{S}}\right]\right]}. \tag{5}$$

2.3 Kauermann and Carroll's Edgeworth approximation

Kauermann & Carroll (2001) proposed a method of constructing confidence intervals based on HC variance estimators that is based on a somewhat simpler Edgeworth approximation. The hypothesis testing procedure corresponding to their proposed confidence intervals rejects the null if $|t_{HC}| > z_{\tilde{\alpha}}$, where $\tilde{\alpha}$ is implicitly defined as the solution to

$$\alpha = \tilde{\alpha} + \frac{\phi\left(z_{\tilde{\alpha}}\right)}{2\nu_q}\left(z_{\tilde{\alpha}}^3 + z_{\tilde{\alpha}}\right), \tag{6}$$

where $\phi(\cdot)$ is the density of the standard normal distribution and

$$\nu_q = \frac{2\left[\text{Var}(\hat{\beta}_q)\right]^2}{\text{Var}\left(V_{qq}^{HC}\right)}$$

is a degrees of freedom measure. Equivalently, the p-value for the test is given by

$$p = 2\left[1 - \Phi\left(\left|t_q^{HC}\right|\right)\right] + \frac{\phi\left(t_q^{HC}\right)}{2\nu_q}\left(\left|t_q^{HC}\right|^3 + \left|t_q^{HC}\right|\right),$$

These authors also offer a further approximation for the critical value $z_{\tilde{\alpha}}$, which saves the trouble of solving Equation (6):

$$z_{\tilde{\alpha}} = F_t^{-1}\left(1 - \frac{\alpha}{2}; n-p\right) + \frac{\left(z_\alpha^3 + z_\alpha\right)}{4}\left(\frac{1}{\nu_q} - \frac{\left(\sum_{i=1}^n g_{qi}^2\right)^2}{n}\right).$$

In contrast to the degrees of freedom estimator used by Lipsitz and colleagues (as given in Equation **??**), Kauermann and Carroll calculate the degrees of freedom under the working assumption that the errors are actually homoskedastic. Under this working assumption, the HC2 variance estimator is unbiased, with degrees of freedom are given by

$$\nu_q = \left(\sum_{i=1}^n g_{qi}^2\right)^2 \left(\sum_{i=1}^n g_{qi}^4 + \sum_{i=1}^n \sum_{j \neq i} \frac{g_{qi}^2 g_{qj}^2 h_{ij}^2}{(1-h_{ii})(1-h_{jj})}\right)^{-1}$$

.

2.4 Rothenberg's Edgeworth approximation

Rothenberg (1988) developed an Edgeworth approximation for the distribution of Wald-type t-statistics under the assumption that the errors are normally distributed. The original approximation was developed for tests based on the HC0 variance estimator, but extending it other HCCMEs is straightforward; here, we state the more general form. Let

$$\begin{aligned}
g_i &= \mathbf{x}_i \mathbf{M} \mathbf{c} \\
z_i &= \sigma_i^2 g_i - \mathbf{x}_i \mathbf{M} \mathbf{X}' \mathbf{\Sigma} \mathbf{X} \mathbf{M} \mathbf{c} / n \\
q_i &= \frac{1}{n^2} \mathbf{x}_i \mathbf{M} \mathbf{X}' \mathbf{\Sigma} \mathbf{X} \mathbf{M} \mathbf{x}_i' - 2h_i \\
a &= \frac{\sum_{i=1}^n \omega_i g_i^2 z_i^2}{\left(\sum_{i=1}^n g_i^2 \sigma_i^2\right)^2} \\
b &= \frac{\sum_{i=1}^n \omega_i g_i^2 q_i}{\sum_{i=1}^n g_i^2 \sigma_i^2} \\
\nu &= \frac{2\left(\sum_{i=1}^n g_i^2 \sigma_i^2\right)^2}{\sum_{i=1}^n \omega_i^2 g_i^4 \sigma_i^4}
\end{aligned}$$

Rothenberg's Edgeworth approximation is then given by

$$\Pr\left(T(\mathbf{V}^{HC}) \leq t\right) \approx \Phi\left[t\left(1 - \frac{1+t^2}{2\nu} + \frac{a\,(t^2-1)+b}{2}\right)\right],$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Based on the Edgeworth approximation, Rothenberg (1988) proposed a test in which the null hypothesis is rejected if the observed test statistic is greater than the critical value defined by

$$t_{crit} = z_\alpha \left[1 + \frac{z_\alpha^2 + 1}{2\nu} - \frac{a\,(z_\alpha^2 - 1) + b}{2}\right], \tag{7}$$

where z_α is the $1 - \alpha/2$ standard normal critical value.

In practice, the quantities a, b, and ν must be estimated because they depend on the unknown error variances. Rothenberg proposed to do so by replacing values of σ_i^2 with $\omega_i e_i^2$ and values of σ_i^4 with $\omega_i^2 e_i^4/3$. Alternately, one could assume that $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$, in which case $a = 0$,

$$b = -\frac{\sum_{i=1}^n h_i \omega_i g_i^2}{\sum_{i=1}^n g_i^2}, \qquad \text{and} \qquad \nu = \frac{2\left(\sum_{i=1}^n g_i^2\right)^2}{\sum_{i=1}^n \omega_i^2 g_i^4}.$$

2.5 Saddlepoint approximation

McCaffrey & Bell (2006) developed small-sample adjustments to test statistics based on cluster-robust variance estimators, of which HC variance estimators are a special case. They consider both a Satterthwaite approximation (similar to Lipsitz et al.) and a saddlepoint approximation for the distribution of the test statistic, finding that the latter produced tests with more accurate size. The saddlepoint approximation is obtained as follows. Let Observe that the cumulative distribution of t_q^{HC} can be expressed as

$$\Pr\left(t_q^{HC} \leq t\right) = \Pr\left(\frac{\left(\hat\beta_q - c\right)^2}{\text{Var}(\hat\beta_q)} - t^2 \frac{V_{qq}^{HC}}{\text{Var}(\hat\beta_q)} \leq 0\right).$$

Note that $\left(\hat\beta_q - c\right)^2 / \text{Var}(\hat\beta_q)\ \chi_1^2$ and that V_{qq}^{HC} is distributed as a weighted sum of χ_1^2 random variables, with weights given by the eigen-values $\lambda_1, ..., \lambda_n$ of the matrix $\mathbf{A}_{xq}\left(\mathbf{I} - \mathbf{H}\right)\mathbf{\Sigma}\left(\mathbf{I} - \mathbf{H}\right)$. Assuming that V_{qq}^{HC} is unbiased, so that

$$\text{E}\left(V_{qq}^{HC}\right) = \text{tr}\left[\mathbf{A}_{xq}\left(\mathbf{I} - \mathbf{H}\right)\mathbf{\Sigma}\left(\mathbf{I} - \mathbf{H}\right)\right] = \sum_{j=1}^n \lambda_j,$$

and that $\hat\beta_q$ is independent of V_{qq}^{HC}, it follows that the $\Pr\left(t_q^{HC} \leq t\right)$ can be expressed as $\Pr(Z \leq 0)$, where $Z = \sum_{i=0}^n \gamma_i z_i$, $\gamma_0 = 1$, $\gamma_i = -t^2 \lambda_i / \sum_{j=1}^n \lambda_j$, and $z_0, ..., z_n \stackrel{\text{iid}}{\sim} \chi_1^2$.

The saddlepoint technique is a means to approximate the distribution of Z. Let s be the saddlepoint, defined implicitly as the solution to

$$\sum_{i=0}^n \frac{\gamma_i}{1 - 2\gamma_i s} = 0.$$

Note that the solution to the saddlepoint equation will be in the range $\left(\left(2\min\left\{\gamma_0, ..., \gamma_n\right\}\right)^{-1}, 0\right)$ if $\sum_{i=0}^n \gamma_i > 0$ and in the range $\left(0, \left(2\max\left\{\gamma_0, ..., \gamma_n\right\}\right)^{-1}\right)$ if $\sum_{i=0}^n \gamma_i \leq 0$. Define the quantities r and q as

$$r = \text{sign}(s)\sqrt{2sz + \sum_{i=0}^{n} \log\left(1 - 2\gamma_i s\right)}, \qquad q = s\sqrt{2\sum_{i=0}^{n} \frac{\gamma_i^2}{\left(1 - 2\gamma_i s\right)^2}}$$

for a constant z. The saddlepoint approximation is then

$$\Pr(Z \leq z) \approx \begin{cases} \Phi(r) + \phi(r)\left[\frac{1}{r} - \frac{1}{q}\right] & s \neq 0 \\ \frac{1}{2} + \frac{\sum_{i=0}^{n} \gamma_i^3}{3\sqrt{\pi}\left(\sum_{i=0}^{n} \gamma_i^2\right)^{3/2}} & s = 0. \end{cases} \tag{8}$$

Given an observed value for the t-statistic t_q^{HC}, a p-value for H_0 can be calculated by taking $\gamma_i = -\left(t_q^{HC}\right)^2 \lambda_i / \sum_{j=1}^{n} \lambda_j$ for $i = 1, ..., n$, finding s, r, and q, and evaluating $1 - \Pr(Z \leq 0)$ using Equation (8).

3 Simulation study

4 Conclusion

SUPPLEMENTARY MATERIAL

Title: Brief description. (file type)

R-package for MYNEW routine: R-package ?MYNEW? containing code to perform the diagnostic methods described in the article. The package also contains all datasets used as examples in the article. (GNU zipped tar file)

HIV data set: Data set used in the illustration of MYNEW method in Section 3.2. (.txt file)

References

Cribari-Neto, F. (2004), 'Asymptotic inference under heteroskedasticity of unknown form', *Computational Statistics and Data Analysis* **45**(2), 215–233.

Cribari-Neto, F. & da Silva, W. B. (2011), 'A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model', *Advances in Statistical Analysis* **95**(2), 129–146.

Cribari-Neto, F., Souza, T. C. & Vasconcellos, K. L. P. (2007), 'Inference under heteroskedasticity and leveraged data', *Communications in Statistics - Theory and Methods* **36**(10), 1877–1888.

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

Eicker, F. (1967), Limit theorems for regressions with unequal and dependent errors, *in* 'Proceedings of the Fifth Berkeley symposium on Mathematical Statistics and Probability', University of California Press, Berkeley, CA, pp. 59–82.

Huber, P. J. (1967), The behavior of maximum likelihood estimates under nonstandard conditions, *in* 'Proceedings of the Fifth Berkeley symposium on Mathematical Statistics and Probability', University of California Press, Berkeley, CA, pp. 221–233.

Kauermann, G. & Carroll, R. J. (2001), 'A note on the efficiency of sandwich covariance matrix estimation', *Journal of the American Statistical Association* **96**(456), 1387–1396.

Lipsitz, S. R., Ibrahim, J. G. & Parzen, M. (1999), 'A degrees-of-freedom approximation for a t-statistic with heterogeneous variance', *Journal of the Royal Statistical Society: Series D (The Statistician)* **48**(4), 495–506.

Long, J. S. & Ervin, L. H. (2000), 'Using heteroscedasticity consistent standard errors in the linear regression model', *The American Statistician* **54**(3), 217–224.

MacKinnon, J. G. (2013), Thirty years of heteroskedasticity-robust inference, *in* X. Chen & N. R. Swanson, eds, 'Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis', Springer New York, New York, NY.

MacKinnon, J. G. & White, H. (1985), 'Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties', *Journal of Econometrics* **29**, 305–325.

McCaffrey, D. F. & Bell, R. M. (2006), 'Improved hypothesis testing for coefficients in generalized estimating equations with small samples of clusters.', *Statistics in medicine* **25**(23), 4081–98.

Rothenberg, T. (1988), 'Approximate power functions for some robust tests of regression coefficients', *Econometrica* **56**(5), 997–1019.

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**(4), 817–838.

Zeileis, A. (2004), 'Econometric computing with HC and HAC covariance matrix estimators', *Journal of Statistical Software* **11**(10), 1–17.