

A CLASS OF IMPROVED HETEROSKEDASTICITY–CONSISTENT COVARIANCE MATRIX ESTIMATORS

FRANCISCO CRIBARI–NETO

*Departamento de Estatística, CCEN, Universidade Federal de Pernambuco,
Cidade Universitária, Recife/PE, 50740-540, Brazil*

NILA M.S. GALVÃO

*Departamento de Ciências Exatas e da Terra, Universidade do Estado da Bahia,
Estrada das Barreiras, s/n, Narandiba, Cabula, Salvador, BA, 41195-001, Brazil*

ABSTRACT

The heteroskedasticity-consistent covariance matrix estimator proposed by White (1980), also known as HC0, is commonly used in practical applications and is implemented into a number of statistical software. Cribari–Neto, Ferrari & Cordeiro (2000) have developed a bias-adjustment scheme that delivers bias-corrected White estimators. There are several variants of the original White estimator that also commonly used by practitioners. These include the HC1, HC2 and HC3 estimators, which have proven to have superior small-sample behavior relative to White’s estimator. This paper defines a general bias-correction mechanism that can be applied not only to White’s estimator, but to variants of this estimator as well, such as HC1, HC2 and HC3. Numerical evidence on the usefulness of the proposed corrections is also presented. Overall, the results favor the sequence of improved HC2 estimators.

Key words and phrases: Bias correction; Covariance matrix estimation; Heteroskedasticity; Linear regression; White’s estimator.

1. INTRODUCTION

Linear regression models oftentimes display heteroskedastic error structures, i.e., non-constant error variances. When that happens the usual ordinary least squares estimator (OLSE) of the linear parameters remains unbiased and consistent, albeit no longer efficient. Since the exact form of the heteroskedasticity is usually unknown, practitioners commonly use the OLSE even when heteroskedasticity is suspected. However, the usual estimator of the OLSE covariance matrix, namely $\hat{\sigma}^2(X'X)^{-1}$, $\hat{\sigma}^2$ being an estimator of the common error variance and X the matrix of covariates, is no longer unbiased nor consistent when the homoskedasticity assumption is violated. It then becomes important to develop and use alternative covariance matrix estimators that are consistent even under heteroskedasticity of unknown form. The most commonly used estimator is that proposed by Halbert White in an influential *Econometrica* paper (White, 1980). White's estimator, also known as HC0, is implemented into a number of statistical and econometric software (e.g., **SHAZAM**) and has been used in a variety of empirical analyses. A shortcoming of this estimator is that, although consistent, it can be quite biased in samples of typical size. In particular, it tends to underestimate the true variances, thus leading to associated quasi- t tests that are liberal. Cribari-Neto, Ferrari & Cordeiro (2000) have recently obtained a sequence of bias adjusted White estimators. The first corrected estimator in this sequence is obtained by bias-correcting the White estimator, the second adjusted estimator follows from applying a bias-correction to the first adjusted estimator in the sequence, and so on. The further down one advances in the sequence, the smaller the order of the bias of the corresponding estimator.

There are several alternative estimators that have been proposed in the literature, the most cited ones being the HC1, HC2 and HC3 estimators, the latter closely approximating the jackknife estimator. The finite-sample behavior of these estimators are usually superior to that of HC0 since they already incorporate small-sample corrections. Numerical evidence favors these estimators over the White estimator. For instance, the results in Cribari-Neto & Zarkos (1999) favor the HC2 estimator while the results reported by Long & Ervin (2000) and MacKinnon & White (1985) favor the HC3 estimator.

The purpose of this paper is to obtain a bias adjustment mechanism that can be used to deliver a sequence of bias-corrected estimators where the order of the bias decreases as one advances in the sequence. The adjustment mechanism can be applied to a wide class of heteroskedasticity-consistent covariance matrix estimators, including, among others, the HC0, HC1, HC2 and HC3 estimators. The results in Cribari-Neto, Ferrari & Cordeiro (2000) can thus be obtained as a special case of the results developed here. We also present numerical evidence on the small-sample performance of the proposed estimators. The numerical evidence shows that the bias-adjustment mechanism we propose indeed delivers more accurate estimates in small samples. In particular, the numerical results favor the bias-adjusted version of the HC2 estimator.

2. THE MODEL AND SOME COVARIANCE MATRIX ESTIMATORS

The model of interest is the usual linear regression model, which is defined as $y = X\beta + u$, where y and u are n -vectors of responses and random errors, respectively, β is a p -vector of unknown parameters ($p < n$), and X is an $n \times p$ matrix of fixed regressors with full column rank, i.e., $\text{rank}(X) = p$. We make the following assumptions:

- i) $E(u_t) = 0, t = 1, \dots, n$;
- ii) $E(u_t^2) = \sigma_t^2, 0 < \sigma_t^2 < \infty, t = 1, \dots, n$;
- iii) $E(u_t u_s) = 0$ for all $t \neq s$;
- iv) $\lim_{n \rightarrow \infty} (X'X/n) = Q$, where Q is a positive definite matrix.

The ordinary least squares estimator of β is $\hat{\beta} = (X'X)^{-1}X'y$. It has mean β (i.e., it is unbiased) and covariance matrix $\Psi = P\Omega P'$, where $P = (X'X)^{-1}X'$ and $\Omega = \text{cov}(u)$ is a diagonal matrix with the t th diagonal element representing the variance σ_t^2 of u_t , $t = 1, \dots, n$. For,

$$\begin{aligned} \Psi &= \text{cov}(\hat{\beta}) = \text{cov}[\beta + (X'X)^{-1}X'u] = \text{cov}[(X'X)^{-1}X'u] \\ &= (X'X)^{-1}X'\text{cov}(u)X(X'X)^{-1} = (X'X)^{-1}X'\Omega X(X'X)^{-1}. \end{aligned}$$

Under homoskedasticity, we have $\sigma_t^2 = \sigma^2$, a strictly positive finite constant, for all t , and it thus follows that $\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$, which can be estimated by $\hat{\sigma}^2(X'X)^{-1}$, where $\hat{\sigma}^2 = \hat{u}'\hat{u}/(n-p)$, $\hat{u} = (\hat{u}_1, \dots, \hat{u}_n)' = My$ being the n -vector of OLS residuals. Here, $M = I - H$, I is the $n \times n$ identity matrix, and $H = X(X'X)^{-1}X'$ (the ‘hat matrix’) is symmetric and idempotent.

The unbiasedness of $\hat{\beta}$ can be easily established:

$$E(\hat{\beta}) = E[\beta + (X'X)^{-1}X'u] = \beta + (X'X)^{-1}E(u) = \beta.$$

This property is true even under heteroskedasticity.

It is also possible to establish the consistency of the OLSE without assuming homoskedasticity. Note that $\hat{\beta} = \beta + (X'X)^{-1}X'u$ and

$$\text{plim}(\hat{\beta}) = \text{plim}[\beta + (X'X)^{-1}X'u] = \beta + \text{plim}\left(\frac{X'X}{n}\right)^{-1} \text{plim}\left(\frac{X'u}{n}\right),$$

where plim denotes limit in probability. Using Slutsky’s theorem (e.g., Rao, 1973, p. 122), we obtain

$$\text{plim}(\hat{\beta}) = \beta + \left(\lim \frac{X'X}{n}\right)^{-1} \text{plim}\left(\frac{X'u}{n}\right).$$

Recall that assumption (iv) guarantees that $(X'X)/n$ converges to a non-singular matrix Q as $n \rightarrow \infty$. Additionally, since u is a vector of uncorrelated random variables with finite variance and $X'u = \sum_{t=0}^n X'_t u_t$, where X_t is the t th row of X , such that $E(X'u) =$

$\sum_{t=0}^n X_t' E(u_t) = 0$, it thus follows from Chebyshev's weak law of large numbers (e.g., Rao, 1973, p. 112) that

$$\text{plim}\left(\frac{X'u}{n}\right) = 0.$$

Therefore, $\text{plim}(\hat{\beta}) = \beta + Q^{-1}0 = \beta$. That is, $\hat{\beta}$ is consistent for β . Again, it was not necessary to assume equal error variances to obtain the result.

Since $\hat{\beta}$ is unbiased and consistent even under heteroskedasticity of unknown form, it is commonly used by practitioners when heteroskedasticity is suspected. It is necessary, however, to obtain a consistent estimator for its covariance matrix.

White (1980) proposed a covariance matrix estimator which is consistent under both homoskedasticity and heteroskedasticity of unknown form; it is given by

$$\hat{\Psi} = P\hat{\Omega}P', \tag{1}$$

where

$$\hat{\Omega} = \text{diag}\{\hat{u}_1^2, \dots, \hat{u}_n^2\}.$$

The idea is to obtain a consistent estimator for the symmetric matrix $X'\Omega X$ (say, $X'\hat{\Omega}X$) that, unlike Ω , has at most $p(p+1)/2$ different elements, regardless of the sample size. That is, $\text{plim}(X'\hat{\Omega}X/n) = \text{plim}(X'\Omega X/n)$, as $n \rightarrow \infty$. Therefore,

$$\begin{aligned} \hat{\Psi} &= n(X'X)^{-1}(X'\hat{\Omega}X/n)(X'X)^{-1} \\ &= (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}, \end{aligned}$$

can be used to consistently estimate the covariance matrix of $\hat{\beta}$. The proposed estimator uses, as pointed out above, $\hat{\Omega} = \text{diag}\{\hat{u}_1^2, \dots, \hat{u}_n^2\}$, and is also known as HC0.

White's estimator is useful since it allows practitioners to perform inference which is asymptotically correct regardless of the presence of heteroskedasticity in the data or even of the form of such heteroskedasticity if it indeed exists. However, this estimator can be substantially biased in small samples: it tends to underestimate the true variances, thus being a bit too optimistic. In order to overcome this shortcoming, other estimators have been proposed and are also widely used. They can also be written as (1), but with different structures for $\hat{\Omega}$. The HC1 estimator (Hinkley, 1977) uses

$$\hat{\Omega} = \frac{n}{n-p} \text{diag}\{\hat{u}_1^2, \dots, \hat{u}_n^2\}.$$

It thus includes a finite-sample correction, namely $n/(n-p)$, which accounts for the fact that the OLS residuals tend to fluctuate less than the unknown errors. The HC2 estimator uses

$$\hat{\Omega} = \text{diag}\left\{\frac{\hat{u}_1^2}{1-h_1}, \dots, \frac{\hat{u}_n^2}{1-h_n}\right\},$$

where h_1, \dots, h_n are the diagonal elements of H . These quantities are usually viewed as measures of the leverage of the corresponding observations. The HC3 estimator discounts the squared residuals more heavily; it uses

$$\widehat{\Omega} = \text{diag} \left\{ \frac{\widehat{u}_1^2}{(1-h_1)^2}, \dots, \frac{\widehat{u}_n^2}{(1-h_n)^2} \right\}.$$

The finite-sample corrections included in the definition of the HC2 and HC3 estimators are therefore based on the degrees of leverage of the different observations. Since $0 < h_t < 1$ for all t , the greater h_t , the more we inflate the t th squared residual. This is because, as noted by Chesher & Jewitt (1987, p. 1219), the possibility of severe downward bias in the White estimator arises when there are large h_t , because the associated least squares residuals have small magnitude on average and the White estimator takes small residuals as evidence of small error variances.

The HC2 estimator, initially proposed by Horn, Horn & Duncan (1975), can be shown to be unbiased under homoskedasticity. Note that $E(\widehat{u}) = E(My) = E[M(X\beta - u)]$. Since $MX = (I - H)X = 0$ it follows that $E(\widehat{u}) = E(Mu) = 0$, and hence

$$E(\widehat{u}\widehat{u}') = \text{cov}(\widehat{u}) + E(\widehat{u})E(\widehat{u}') = \text{cov}(Mu) = M\Omega M'.$$

When the model is homoskedastic, we have that $\Omega = \sigma^2 I$. Since M is symmetric and idempotent, we obtain

$$E(\widehat{u}\widehat{u}') = M\sigma^2 IM = \sigma^2 M = \sigma^2(I - H),$$

where each diagonal element is given by $E(\widehat{u}_t^2) = \sigma^2(1 - h_t)$.

The four covariance matrix estimators presented above are commonly used by practitioners and have implementations in several econometric and statistical software, such as Shazam (<http://www.shazam.econ.ubc.ca>), Stata (<http://www.stata.com>) and gretl (<http://gretl.sourceforge.net>). The next section develops bias adjusted versions of these estimators. The biases of the corrected estimators are shown to converge faster to zero than those of the unmodified estimators.

3. A CLASS OF IMPROVED COVARIANCE MATRIX ESTIMATORS

We consider the four covariance matrix estimators described in the previous section, namely:

$$\begin{aligned} \widehat{\Psi}_i &= (X'X)^{-1} X' \widehat{\Omega}_i X (X'X)^{-1} \\ &= P \widehat{\Omega}_i P', \quad i = 0, 1, 2, 3. \end{aligned}$$

We write $\widehat{\Omega}_i = D_i \widehat{\Omega}$, where $\widehat{\Omega} = \text{diag}(\widehat{u}_1^2, \dots, \widehat{u}_n^2)$ and \widehat{u}_t^2 denotes the t th OLS squared residual. The diagonal matrices D_i 's are defined, respectively, as

$$\begin{aligned}
\text{HC0 : } & D_0 = I, \\
\text{HC1 : } & D_1 = \left(\frac{n}{n-p} \right) I, \\
\text{HC2 : } & D_2 = \text{diag}\{1/(1-h_t)\}, \\
\text{HC3 : } & D_3 = \text{diag}\{1/(1-h_t)^2\},
\end{aligned}$$

where again h_t is the t th diagonal element of $H = X(X'X)^{-1}X'$ and I is the $n \times n$ identity matrix. This notation will allow us to treat the four estimators in a unified manner.

At the outset, note that $\widehat{\Omega}$ can be written as $\widehat{\Omega} = (\widehat{u}\widehat{u}')_d$, where \widehat{u} is the vector of OLS residuals and the subscript d denotes that a diagonal matrix was formed out of the original matrix by setting all non-diagonal elements equal to zero. We then have that

$$\begin{aligned}
E(\widehat{\Omega}) &= E\{(\widehat{u}\widehat{u}')_d\} = \{E(\widehat{u}\widehat{u}')\}_d = \{M\Omega M\}_d \\
&= \{(I-H)\Omega(I-H)\}_d = \{H\Omega(H-2I) + \Omega\}_d.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E(\widehat{\Omega}_i) &= E(D_i\widehat{\Omega}) = D_iE(\widehat{\Omega}) = D_i\{H\Omega(H-2I) + \Omega\}_d \\
&= \{D_iH\Omega(H-2I) + D_i\Omega\}_d.
\end{aligned}$$

The mean of the estimator $\widehat{\Psi}_i$ is thus given by

$$E(\widehat{\Psi}_i) = E\{P\widehat{\Omega}_iP'\} = PE(\widehat{\Omega}_i)P' = P\{D_iH\Omega(H-2I) + D_i\Omega\}_dP'.$$

Using the equations above we obtain the bias function of $\widehat{\Omega}_i$:

$$B_{\widehat{\Omega}_i}(\Omega) = E(\widehat{\Omega}_i) - \Omega = \{D_iH\Omega(H-2I) + D_i\Omega\}_d - \Omega.$$

Since the matrix $D_i\Omega$ is diagonal, for all i , we can write

$$\begin{aligned}
B_{\widehat{\Omega}_i}(\Omega) &= \{D_iH\Omega(H-2I)\}_d + D_i\Omega - \Omega \\
&= D_i\{H\Omega H - 2H\Omega\}_d + (D_i - I)\Omega
\end{aligned} \tag{2}$$

and

$$\begin{aligned}
B_{\widehat{\Psi}_i}(\Omega) &= E(\widehat{\Psi}_i) - \Psi = PE(\widehat{\Omega}_i)P' - P\Omega P' = P\{B_{\widehat{\Omega}_i}(\Omega)\}_dP' \\
&= P[D_i\{H\Omega H - 2H\Omega\}_d + (D_i - I)\Omega]P'.
\end{aligned} \tag{3}$$

Equations (2) and (3) give closed-form expressions for the biases of $\widehat{\Omega}_i$ and $\widehat{\Psi}_i$, for $i = 0, \dots, 3$, as estimators of Ω and Ψ , respectively. It is noteworthy that one can obtain

the exact bias of any of the four estimators described in the previous section by using the corresponding expression for the diagonal matrix D_i .

The next step is to define sequences of bias adjusted covariance matrix estimators, denoted $\{\widehat{\Psi}_i^{(k)}, k = 1, 2, \dots\}$, $i = 0, 1, 2, 3$, for the covariance matrix of $\widehat{\beta}$. At the outset, we define sequences of adjusted estimators for Ω , say $\{\widehat{\Omega}_i^{(k)}, k = 1, 2, \dots\}$, $i = 0, 1, 2, 3$, which will be used to deliver sequences of improved estimators for Ψ .

Consider the recursive function $M^{(k+1)}(A) = M^{(1)}(M^{(k)}(A))$, $k = 0, 1, 2, \dots$, with $M^{(1)}(A) = \{HA(H - 2I)\}_d$ e $M^{(0)}(A) = A$, where H is defined as before and A is an $n \times n$ diagonal matrix. It is possible to show that the following properties hold. Let A and B be $n \times n$ diagonal matrices, and let $k = 0, 1, 2, \dots$. Then:

$$[\text{P1}] \quad M^{(k)}(A) + M^{(k)}(B) = M^{(k)}(A + B);$$

$$[\text{P2}] \quad M^{(k)}(\{HA(H - 2I)\}_d) = M^{(k)}(M^{(1)}(A)) = M^{(k+1)}(A);$$

$$[\text{P3}] \quad E\{M^{(k)}(A)\} = M^{(k)}(E(A)).$$

Using the notation and properties outlined above, it is possible to write (2) and (3), respectively, as

$$\begin{aligned} B_{\widehat{\Omega}_i}(\Omega) &= D_i M^{(1)}(\Omega) + (D_i - I)M^{(0)}(\Omega) \\ &= D_i[M^{(0)}(\Omega) + M^{(1)}(\Omega)] - M^{(0)}(\Omega) \end{aligned} \quad (4)$$

and

$$B_{\widehat{\Psi}_i}(\Omega) = P\{D_i[M^{(0)}(\Omega) + M^{(1)}(\Omega)] - M^{(0)}(\Omega)\}P', \quad (5)$$

for $i = 0, 1, 2, 3$. An estimator for Ω can be defined as

$$\widehat{\Omega}_i^{(1)} = \widehat{\Omega}_i - B_{\widehat{\Omega}_i}(\widehat{\Omega}) = \widehat{\Omega}_i - \{D_i[M^{(0)}(\widehat{\Omega}) + M^{(1)}(\widehat{\Omega})] - M^{(0)}(\widehat{\Omega})\},$$

where the last equality follows from (4). Since $\widehat{\Omega}_i = D_i\widehat{\Omega}$ and $M^{(0)}(\widehat{\Omega}) = \widehat{\Omega}$, we have that

$$\begin{aligned} \widehat{\Omega}_i^{(1)} &= D_i M^{(0)}(\widehat{\Omega}) - D_i M^{(0)}(\widehat{\Omega}) - D_i M^{(1)}(\widehat{\Omega}) + M^{(0)}(\widehat{\Omega}) = M^{(0)}(\widehat{\Omega}) - D_i M^{(1)}(\widehat{\Omega}) \\ &= \widehat{\Omega} - D_i M^{(1)}(\widehat{\Omega}). \end{aligned} \quad (6)$$

Note that it would not be feasible to compute the bias of $\widehat{\Omega}_i$, since it is a function of Ω . Replacing Ω by $\widehat{\Omega}$, we obtain the adjusted estimator defined in (6), which is nearly unbiased. Any bias in this estimator is due solely to the fact that $M^{(1)}(\widehat{\Omega})$ is used in place of the unknown matrix $M^{(1)}(\Omega)$. The next step is, thus, to correct for any bias induced by the estimation of $B_{\widehat{\Omega}_i}(\Omega)$.

The bias of $\widehat{\Omega}_i^{(1)}$ can be written as

$$\begin{aligned} B_{\widehat{\Omega}_i^{(1)}}(\Omega) &= \mathbb{E}\{\widehat{\Omega} - D_i M^{(1)}(\widehat{\Omega})\} - \Omega \\ &= B_{\widehat{\Omega}}(\Omega) - D_i \mathbb{E}\{M^{(1)}(\widehat{\Omega}) - M^{(1)}(\Omega)\} - D_i M^{(1)}(\Omega). \end{aligned}$$

Note that $\widehat{\Omega}_0 = D_0 \widehat{\Omega} = \widehat{\Omega}$, since $D_0 = I$. Thus, using (4) we arrive at $B_{\widehat{\Omega}}(\Omega) = B_{\widehat{\Omega}_0}(\Omega) = M^{(1)}(\Omega)$. Finally, using properties [P1] and [P3], it follows that

$$\begin{aligned} B_{\widehat{\Omega}_i^{(1)}}(\Omega) &= M^{(1)}(\Omega) - D_i M^{(1)}(M^{(1)}(\Omega)) - D_i M^{(1)}(\Omega) \\ &= M^{(1)}(\Omega) - D_i M^{(2)}(\Omega) - D_i M^{(1)}(\Omega), \end{aligned}$$

that is,

$$B_{\widehat{\Omega}_i^{(1)}}(\Omega) = M^{(1)}(\Omega) - D_i M^{(1)}(\Omega) - D_i M^{(2)}(\Omega). \quad (7)$$

Applying again the procedure described above and results (6) and (7), we define a second corrected estimator for Ω , namely:

$$\begin{aligned} \widehat{\Omega}_i^{(2)} &= \widehat{\Omega}_i^{(1)} - B_{\widehat{\Omega}_i^{(1)}}(\widehat{\Omega}) \\ &= \widehat{\Omega} - D_i M^{(1)}(\widehat{\Omega}) - \{M^{(1)}(\widehat{\Omega}) - D_i M^{(1)}(\widehat{\Omega}) - D_i M^{(2)}(\widehat{\Omega})\} \\ &= \widehat{\Omega} - M^{(1)}(\widehat{\Omega}) + D_i M^{(2)}(\widehat{\Omega}). \end{aligned}$$

The procedure can be repeated recursively, and after k iterations we arrive at

$$\widehat{\Omega}_i^{(k)} = \sum_{j=0}^{k-1} (-1)^j M^{(j)}(\widehat{\Omega}) + (-1)^k D_i M^{(k)}(\widehat{\Omega}), \quad k = 1, 2, \dots, \quad (8)$$

for $i = 0, 1, 2, 3$. The bias of $\widehat{\Omega}_i^{(k)}$ can be derived using properties [P1]—[P3] and equation (8) as

$$\begin{aligned} B_{\widehat{\Omega}_i^{(k)}}(\Omega) &= \mathbb{E}(\widehat{\Omega}_i^{(k)}) - \Omega \\ &= \mathbb{E}\left\{\sum_{j=0}^{k-1} (-1)^j M^{(j)}(\widehat{\Omega}) + (-1)^k D_i M^{(k)}(\widehat{\Omega})\right\} - \Omega \\ &= \sum_{j=0}^{k-1} (-1)^j \mathbb{E}\{M^{(j)}(\widehat{\Omega}) + (-1)^k D_i M^{(k)}(\widehat{\Omega})\} - \Omega \\ &= \sum_{j=0}^{k-1} (-1)^j \{\mathbb{E}[M^{(j)}(\widehat{\Omega}) - M^{(j)}(\Omega)] + M^{(j)}(\Omega)\} \\ &\quad + (-1)^k D_i \mathbb{E}[M^{(k)}(\widehat{\Omega}) - M^{(k)}(\Omega)] + (-1)^k D_i M^{(k)}(\Omega) - \Omega \\ &= (-1)^k \{D_i M^{(k+1)}(\Omega) + D_i M^{(k)}(\Omega) - M^{(k)}(\Omega)\}. \end{aligned} \quad (9)$$

Therefore, we define sequences $\{\widehat{\Omega}_i^{(k)}, k = 1, 2, \dots\}$, $i = 0, 1, 2, 3$, of estimators for Ω with $\widehat{\Omega}_i^{(k)}$ and their biases given in (8) and (9), respectively. We shall now define sequences of estimators for Ψ , say $\{\widehat{\Psi}_i^{(k)}, k = 1, 2, \dots\}$, $i = 0, 1, 2, 3$, where

$$\widehat{\Psi}_i^{(k)} = P\widehat{\Omega}_i^{(k)}P', \quad k = 1, 2, \dots, \quad (10)$$

for $i = 0, \dots, 3$, so that $\widehat{\Psi}_i^{(k)}$ denotes the i th estimator, corrected up to the k th iteration. The bias of $\widehat{\Psi}_i^{(k)}$ follows from (9) and (10) as

$$\begin{aligned} B_{\widehat{\Psi}_i^{(k)}}(\Omega) &= E(\widehat{\Psi}_i^{(k)}) - \Psi = PE(\widehat{\Omega}_i^{(k)})P' - P\Omega P' = P\{B_{\widehat{\Omega}_i^{(k)}}(\Omega)\}P' \\ &= (-1)^k P\{D_i M^{(k+1)}(\Omega) + D_i M^{(k)}(\Omega) - M^{(k)}(\Omega)\}P'. \end{aligned} \quad (11)$$

Equation (11) yields a closed-form expression for the biases of the different estimators for Ψ corrected up to the k th iteration.

We can now obtain the orders of the biases of the sequences of corrected estimators. From assumption (iv) we have that $\lim(X'X/n) = Q$, as $n \rightarrow \infty$, where Q is a positive definite matrix. Equivalently, we have that $X'X = O(n)$, and therefore $(X'X)^{-1} = O(n^{-1})$. It then follows that:

1. The matrices $P = (X'X)^{-1}X'$ and $H = X(X'X)^{-1}X'$ are $O(n^{-1})$. In particular, the diagonal elements of H , h_t for $t = 1, \dots, n$, converge to zero as the sample size increases, i.e., $h_t = o(1)$, for all t , which means that the regression design must be balanced in large samples.
2. The elements of the matrices D_i 's, defined earlier, converge to finite constants as $n \rightarrow \infty$, i.e., $D_i = O(1)$ for all i .

Using the above results it can be shown that for a given diagonal matrix A such that $A = O(n^{-r})$, for some $r \geq 0$, then

1. $PAP' = O(n^{-(r+1)})$;
2. $M^{(k)}(A) = O(n^{-(r+k)}) = D_i M^{(k)}(A)$, $i = 0, 1, 2, 3$.

Also, from assumption (ii) we have that $\Omega = O(1)$. It is now easy to obtain the order of the biases of the sequences of estimators in (8) and (10).

As defined in (9), $B_{\widehat{\Omega}_i^{(k)}}(\Omega) = (-1)^k \{D_i M^{(k+1)}(\Omega) + D_i M^{(k)}(\Omega) - M^{(k)}(\Omega)\}$, and thus

$$B_{\widehat{\Omega}_i^{(k)}}(\Omega) = O(n^{-(k+1)}).$$

We also establish that

$$B_{\widehat{\Psi}_i^{(k)}}(\Omega) = O(n^{-(k+2)}).$$

That is, at each iteration of the bias correction scheme the order of the bias is reduced. The result holds for all different estimators of Ψ considered.

In order to be general, we shall next consider the variance estimation of linear combinations of the elements of $\widehat{\beta}$. Let c denote a $p \times 1$ vector of constants so that $c'\widehat{\beta}$ defines a linear combination of the elements of $\widehat{\beta}$. Note that

$$\phi = \text{var}(c'\widehat{\beta}) = c'[\text{cov}(\widehat{\beta})]c = c'\Psi c.$$

Thus, a sequence of corrected estimators for ϕ can be defined as

$$\widehat{\phi}_i^{(k)} = c'\widehat{\Psi}_i^{(k)}c, \quad i = 0, 1, 2, 3, \quad (12)$$

with $k = 0$ corresponding to the original HCs, as functions of the different D_i 's matrices. The quantity $c'\widehat{\Psi}_i^{(k)}c$ then represents the i th estimator for the variance of $c'\widehat{\beta}$, corrected up to the k th iteration.

Let $V_i = \text{diag}\{v_{i1}^2, \dots, v_{in}^2\} = (v_i v_i')_d$ be a matrix with $v_i = D_i^{1/2}P'c$ for $i = 0, 1, 2, 3$, and let the matrices D_i 's be as defined earlier, such that

$$\widehat{\phi}_i = c'\widehat{\Psi}_i c = c'P\widehat{\Omega}_i P'c = c'PD_i\widehat{\Omega}P'c = c'PD_i^{1/2}\widehat{\Omega}D_i^{1/2}P'c = v_i'\widehat{\Omega}v_i.$$

Since $\widehat{\Omega} = (\widehat{u}\widehat{u}')_d$, it follows that

$$\widehat{\phi}_i = v_i'[(\widehat{u}\widehat{u}')_d]v_i = \widehat{u}'[(v_i v_i')_d]\widehat{u} = \widehat{u}'V_i\widehat{u}, \quad i = 0, 1, 2, 3.$$

Additionally, since $\widehat{\Omega} = (\widehat{u}\widehat{u}')_d$, we have that

$$\widehat{\phi}_i = v_i'[(\widehat{u}\widehat{u}')_d]v_i = \widehat{u}'[(v_i v_i')_d]\widehat{u} = \widehat{u}'V_i\widehat{u}, \quad i = 0, 1, 2, 3.$$

Therefore, $\widehat{\phi}_i = c'\widehat{\Psi}_i c$ can be defined as a quadratic form in the OLS residuals. Similarly,

$$\begin{aligned} \widehat{\phi}_i^{(1)} &= c'P\widehat{\Omega}_i^{(1)}P'c = c'P\{\widehat{\Omega} - D_i M^{(1)}(\widehat{\Omega})\}P'c = c'P\widehat{\Omega}P'c - c'PD_i^{1/2}M^{(1)}(\widehat{\Omega})D_i^{1/2}P'c \\ &= c'\widehat{\Psi}_0 c - v_i'\{H\widehat{\Omega}(H - 2I)\}_d v_i = \widehat{u}'V_0\widehat{u} - \sum_{s=1}^n \widehat{\alpha}_s v_{is}^2, \end{aligned} \quad (13)$$

where $\widehat{\alpha}_s$ is the s th diagonal element of the matrix $\{H\widehat{\Omega}(H - 2I)\}_d$ and v_{is} is the s th element of the vector v_i , $i = 0, 1, 2, 3$. Since $\widehat{\alpha}_s = \sum_{t=1}^n h_{st}^2 \widehat{u}_t^2 - 2h_{ss} \widehat{u}_s^2$, where h_{st} is the (s, t) element of the matrix H , the summation in (13) can be computed as follows:

$$\begin{aligned} \sum_{s=1}^n \widehat{\alpha}_s v_{is}^2 &= \sum_{s=1}^n v_{is}^2 \widehat{\alpha}_s = \sum_{s=1}^n v_{is}^2 \left(\sum_{t=1}^n h_{st}^2 \widehat{u}_t^2 - 2h_{ss} \widehat{u}_s^2 \right) \\ &= \sum_{s=1}^n \sum_{t=1}^n v_{is}^2 h_{st}^2 \widehat{u}_t^2 - 2 \sum_{s=1}^n v_{is}^2 h_{ss} \widehat{u}_s^2 = \sum_{t=1}^n \widehat{u}_t^2 \left(\sum_{s=1}^n h_{st}^2 v_{is}^2 - 2h_{tt} \widehat{u}_t^2 \right). \end{aligned}$$

We thus obtain

$$\widehat{\phi}_i^{(1)} = \widehat{u}'V_0\widehat{u} - \sum_{t=1}^n \widehat{u}_t^2 \widehat{\delta}_{it}, \quad (14)$$

where $\widehat{\delta}_{it} = \sum_{s=1}^n h_{st}^2 v_{is}^2 - 2h_{it}v_{it}^2$ is the t th diagonal element of $\{HV_i(H - 2I)\}_d$. Equation (14) can be written in matrix form as

$$\begin{aligned} \widehat{\phi}_i^{(1)} &= \widehat{u}'V_0\widehat{u} - \widehat{u}'\{HV_i(H - 2I)\}_d\widehat{u} \\ &= \widehat{u}'\{V_0 - M^{(1)}(V_i)\}\widehat{u}. \end{aligned}$$

Generalizing this result, we obtain

$$\widehat{\phi}_i^{(k)} = c'\widehat{\Psi}_i^{(k)}c = \widehat{u}'Q_i^{(k)}\widehat{u}, \quad i = 0, 1, 2, 3, \quad (15)$$

where $Q_i^{(k)} = \sum_{j=0}^{k-1} (-1)^j M^{(j)}(V_0) + (-1)^k M^{(k)}(V_i)$ for $k = 1, 2, \dots$, and $Q_i^{(k)} = V_i$ for $k = 0$.

A standard random quadratic form is defined with respect to a vector of uncorrelated random variables that have mean zero and unit variance. The vector of OLS residuals, however, has covariance structure that is not typically equal to the identity matrix. In order to simplify the variance computation of the sequence of estimators defined by the quadratic form in (15), we can standardize it, as described below. Using the fact that $\widehat{u} = My$ we can write

$$\begin{aligned} \widehat{\phi}_i^{(k)} &= \widehat{u}'Q_i^{(k)}\widehat{u} = y'MQ_i^{(k)}My = y'\Omega^{-1/2}\Omega^{1/2}MQ_i^{(k)}M\Omega^{1/2}\Omega^{-1/2}y \\ &= z'G_i^{(k)}z, \end{aligned} \quad (16)$$

where $G_i^{(k)} = \Omega^{1/2}MQ_i^{(k)}M\Omega^{1/2}$ is a symmetric matrix of dimension $n \forall i$ and $z = \Omega^{-1/2}y$ is an n -vector with mean $\theta = \Omega^{-1/2}X\beta$ and covariance matrix given by

$$\text{cov}(z) = \text{cov}(\Omega^{-1/2}y) = \Omega^{-1/2}\text{cov}(y)\Omega^{-1/2} = \Omega^{-1/2}\text{cov}(u)\Omega^{-1/2} = I.$$

Note that $\theta'G_i^{(k)} = \beta'X'\Omega^{-1/2}\Omega^{1/2}MQ_i^{(k)}M\Omega^{1/2} = \beta'X'MQ_i^{(k)}M\Omega^{1/2}$. Since $X'(I - H) = X'\{I - X(X'X)^{-1}X'\} = 0$, then $\theta'G_i^{(k)} = 0 \forall i$. In order to simplify the notation, we shall denote $G_i^{(k)}$ as G_i , the index k being left implicit in the expressions that follow.

Expression (16) can be written as

$$z'G_iz = (z - \theta)'G_i(z - \theta) + 2\theta'G_i(z - \theta) + \theta'G_i\theta.$$

Since $\theta'G_i = 0$, it follows that $z'G_iz = (z - \theta)'G_i(z - \theta)$, i.e., $\widehat{\phi}_i^{(k)} = c'\widehat{\Psi}_i^{(k)}c = w'G_iw$, where $w = (z - \theta) = \Omega^{-1/2}(y - X\beta) = \Omega^{-1/2}u$, so that $E(w) = 0$ and $\text{cov}(w) = I$. Therefore,

$$\text{var}(\widehat{\phi}_i^{(k)}) = \text{var}(w'G_iw) = E\{(w'G_iw)^2\} - \{E(w'G_iw)\}^2.$$

The computation of the variance of $\widehat{\phi}_i^{(k)}$ then boils down to computing $E(w'G_iw)$ and $E\{(w'G_iw)^2\}$. Using standard results for expected values of quadratic forms in random vectors (Seber, 1977, p.14), we have that

$$\begin{aligned} E(w'G_iw) &= \text{tr}[G_i\text{cov}(w)] + E(w)'G_iE(w) \\ &= \text{tr}(G_i). \end{aligned}$$

We now need to evaluate $E\{(w'G_iw)^2\}$. To that end, we use

$$(w'G_iw)^2 = \sum_{q,r,s,t=1}^n g_{iqr}g_{ist}w_qw_rw_sw_t,$$

where g_{ist} is the (s, t) element of the matrix G_i , $i = 0, 1, 2, 3$. Since $w = \Omega^{-1/2}u$, then $w_j = u_j/\sigma_j$, $j = 1, 2, \dots, n$. We can therefore write

$$(w'G_iw)^2 = \sum_{q,r,s,t=1}^n g_{iqr}g_{ist} \frac{u_qu_ru_su_t}{\sigma_q\sigma_r\sigma_s\sigma_t},$$

and hence

$$\text{var}(\widehat{\phi}_i^{(k)}) = \text{var}(c'\widehat{\Psi}_i^{(k)}c) = \sum_{q,r,s,t=1}^n \frac{g_{iqr}g_{ist}E(u_qu_ru_su_t)}{\sigma_q\sigma_r\sigma_s\sigma_t} - \{\text{tr}(G_i)\}^2, \quad (17)$$

for $i = 0, 1, 2, 3$, assuming that $E(u_qu_ru_su_t)$ exists for all q, r, s, t . Equation (17) gives a general closed-form expression for the variance of estimators of the variance of linear combinations of ordinary least squares estimators in terms of fourth order joint moments of the errors and of the elements of G_i .

In the special case where the errors are independent, we have that

$$E(u_qu_ru_su_t) = \begin{cases} E(u_j^4) = \mu_{4j}, & \text{for } q = r = s = t; \\ \sigma_j^2\sigma_l^2, & \text{for } q = r, s = t; q = s, r = t; q = t, r = s; \\ 0, & \text{otherwise.} \end{cases}$$

Thus,

$$\begin{aligned} \text{var}(\widehat{\Psi}_i^{(k)}) &= \sum_{s=1}^n g_{iss}^2 \frac{\mu_{4s}}{\sigma_s^4} + \sum_{s \neq t} g_{iss}g_{itt} \frac{E(u_s^2)E(u_t^2)}{\sigma_s^2\sigma_t^2} + \\ &+ \sum_{s \neq t} g_{ist}^2 \frac{E(u_s^2)E(u_t^2)}{\sigma_s^2\sigma_t^2} + \sum_{s \neq t} g_{ist}g_{its} \frac{E(u_s^2)E(u_t^2)}{\sigma_s^2\sigma_t^2} - \{\text{tr}(G_i)\}^2. \end{aligned}$$

Since G_i is a symmetric matrix such that $g_{ist} = g_{its} \forall i, t, s$ and $E(u_t^2) = \sigma_t^2 \forall t$, then

$$\text{var}(\widehat{\phi}_i^{(k)}) = \text{var}(c' \widehat{\Psi}_i^{(k)} c) = \sum_{s=1}^n g_{iss}^2 \frac{\mu_{4s}}{\sigma_s^4} + \sum_{s \neq t}^n g_{iss} g_{itt} + 2 \sum_{s \neq t}^n g_{ist}^2 - \{\text{tr}(G_i)\}^2. \quad (18)$$

Note that $\{\text{tr}(G_i)\}^2 = \sum_{s=1}^n g_{iss}^2 + \sum_{s \neq t}^n g_{iss} g_{itt}$ and $\text{tr}(G_i^2) = 2 \left\{ \sum_{s=1}^n g_{iss}^2 + \sum_{s \neq t}^n g_{ist}^2 \right\} \forall i$. Using these results together with (18) we obtain a matrix expression for $\text{var}(\widehat{\phi}_i^{(k)})$ in the case where the errors are independent:

$$\begin{aligned} \text{var}(\widehat{\phi}_i^{(k)}) &= \sum_{s=1}^n g_{iss}^2 \frac{\mu_{4s}}{\sigma_s^4} - \sum_{s=1}^n g_{iss}^2 - 3 \sum_{s=1}^n g_{iss}^2 + 2 \left\{ \sum_{s=1}^n g_{iss}^2 + \sum_{s \neq t}^n g_{ist}^2 \right\} \\ &= \sum_{s=1}^n g_{iss}^2 \frac{\mu_{4s} - 3\sigma_s^4}{\sigma_s^4} + 2 \left\{ \sum_{s=1}^n g_{iss}^2 + \sum_{s \neq t}^n g_{ist}^2 \right\} = \sum_{s=1}^n g_{iss}^2 \gamma_s + 2 \text{tr}(G_i^2) \\ &= g_i' \Lambda g_i + 2 \text{tr}(G_i^2), \end{aligned} \quad (19)$$

where g_i is a column vector containing the diagonal elements of G_i , for $i = 0, 1, 2, 3$, and $\Lambda = \text{diag}(\gamma_1, \dots, \gamma_n)$, where $\gamma_j = \frac{\mu_{4j} - 3\sigma_j^4}{\sigma_j^4}$ is the coefficient of excess kurtosis of the j th error term.

If we assume that the errors are independent and normally distributed, then $\gamma_j = 0$ for $j = 1, \dots, n$. Thus, $\Lambda = 0$ and expression (19) simplifies to

$$\text{var}(\widehat{\phi}_i^{(k)}) = \text{var}(c' \widehat{\Psi}_i^{(k)} c) = 2 \text{tr}(G_i^2), \quad i = 0, 1, 2, 3. \quad (20)$$

It is important to note that the corrected estimators developed in this section can be easily computed. Their computation only requires standard matrix operations, and can be performed using most statistical software and matrix programming languages. It should also be noted that the results above generalize those of Cribari–Neto, Ferrari & Cordeiro (2000), who have obtained corrections that can only be applied to the White estimator. Our results, on the other hand, can be applied to any estimator of Ψ that can be written as $\widehat{\Psi} = (X'X)^{-1} X' D \widehat{\Omega} X (X'X)^{-1}$, provided that mild restrictions are placed on X and D .

4. NUMERICAL RESULTS

The numerical results presented below were obtained using the regression model $y_t = \beta_1 + \beta_2 x_t + u_t$, $t = 1, \dots, n$, where u_1, \dots, u_n are uncorrelated, each u_t having mean 0 and variance $h(\alpha_1 x_t + \alpha_2 x_t^2)$, h being a skedastic function. Note that

we need not specify the error distribution. We consider two regression designs. In the first design, the values of x were obtained as independent random draws from a uniform $\mathcal{U}(0, 1)$ distribution, and we used $\alpha_1 = \alpha_2 = 0.0, 1.0, 2.0, 2.5$. The second design aims at introducing high leverage points in the design matrix; for that, the values of x were obtained as independent random draws from a t_3 distribution, and $\alpha_1 = \alpha_2 = 0.000, 0.025, 0.050, 0.075, 0.100, 0.125, 0.150$. In both cases, $h(\cdot) = \exp(\cdot)$ (multiplicative heteroskedasticity) and $\lambda = \max(\sigma_t^2) / \min(\sigma_t^2)$, a measure of the degree of heteroskedasticity, ranges from 1 (which corresponds to homoskedasticity) to over 100. The sample sizes considered were $n = 50, 100, 150, 200$. The covariate values for $n = 100, 150, 200$ were obtained by replicating twice, three times and four times, respectively, the x values for $n = 50$. This was done in order for the degree of heteroskedasticity to remain constant as the sample size increased. All computations were performed using the matrix programming language `Ox` (Doornik, 2001). Finally, note that the results presented below were not obtained from Monte Carlo simulations, but from the exact expressions derived in Section 3.

At the outset, Tables 1 and 2 (each divided into two parts, a and b) include the results of the bias evaluation for the OLS, HC0, HC1, HC2 and HC3 variance estimators, and the first four corrected versions of HC0–HC3. Table 1 corresponds to the case where the covariate values were obtained from a uniform distribution, whereas Table 2 relates to the situation where x follows a t_3 distribution. In the notation used, HC03, e.g., corresponds to $\widehat{\Psi}_0^{(3)}$, i.e., the HC0 (White) estimator corrected up to the third iteration of the bias correcting mechanism. The figures in these tables are the total relative bias of each estimator, which is the sum of absolute relative biases of the variance estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_2$, i.e., it is defined as

$$\frac{|\mathbb{E}\{\widehat{\text{var}}(\widehat{\beta}_1)\} - \text{var}(\widehat{\beta}_1)|}{\text{var}(\widehat{\beta}_1)} + \frac{|\mathbb{E}\{\widehat{\text{var}}(\widehat{\beta}_2)\} - \text{var}(\widehat{\beta}_2)|}{\text{var}(\widehat{\beta}_2)},$$

where ‘ $\widehat{\text{var}}$ ’ denotes the variance estimator of interest. The total relative bias is a measure of global bias, and the comparisons that follow will be based on this measure. The biases of the unmodified and corrected consistent variance estimators were computed using expression (11). The bias of the usual OLS covariance matrix estimator is given by

$$\frac{\text{tr}\{\Omega(I - H)\}}{n - p} (X'X)^{-1} - P\Omega P',$$

where here $p = 2$.

The results contained in Table 1 yield several important conclusions. First, as expected, the OLS estimator is unbiased under homoskedasticity, but is substantially biased when the error variances are unequal. For instance, its total relative bias for $n = 100$ and $\alpha_1 = \alpha_2 = 1.50$ exceeds 100%. Second, the HC0 (White) estimator is considerably biased in some cases. For $n = 50$ its total relative bias is approximately

equal to 12% (being equal to 9.9% under homoskedasticity). Third, the HC1 and HC2 estimators outperform the White estimator both under homoskedasticity (in which case the HC2 estimator is unbiased) and heteroskedasticity. The finite-sample performance of the HC2 estimator is slightly better than that of the HC1 estimator. For example, when $n = 150$ and $\lambda > 1$, the maximum total relative bias for HC2 (0.6% for $\lambda = 104.47$) is smaller than the minimum total relative bias for HC1 (2.1% for $\lambda = 2.53$). Fourth, the HC3 estimator is the unmodified consistent estimator with poorest performance, except under strong heteroskedasticity in which case it outperforms the HC0 estimator.

Fifth, although the individual bias results are not presented (only the total relative biases are presented), we note that the estimator proposed by Halbert White (HC0) always underestimated the true variances. This behavior was attenuated by HC1 and HC2 (more effectively by the latter). The HC3 estimator, on the hand, displayed a tendency to overestimate the true variances.

Sixth, it is noteworthy that HC3 is the estimator least sensitive to the degree of heteroskedasticity present in the data. Considering $n = 50$ and the different values of α_1 and α_2 used, the difference between the maximum and minimum total relative biases equals 0.005, whereas for the other unmodified consistent estimators this quantity reaches 0.054 (HC1).

Seventh, the biases of the corrected estimators are all very small, even for samples of small size. Note, for instance, that while the total relative bias for HC3 when $n = 50$ reaches 11%, the total relative bias for the corrected HC3 estimator based on only one iteration does not exceed 0.6%. When $n = 200$ the corrected estimators display bias values nearly equal to zero, whereas the total relative biases of the HC0 and HC2 estimators are close to 3.5% and 0.5%, respectively.

Tables 2a and 2b present the bias evaluation for the case where x is obtained from a t_3 distribution. We see that the unmodified consistent estimators become substantially more biased when there are high leverage points in the design matrix. For example, when $n = 100$ and $\alpha_1 = \alpha_2 = 0.125$ ($\lambda = 56.50$), the total relative biases of HC0 and HC2 reach 22.7% and 5.6%, respectively. The corrected estimators are also sensitive to leverage points; however, their total relative biases are much smaller than those of the respective unmodified estimators and converge faster to zero as the sample size increases. Consider, e.g., the case where $n = 150$ and $\alpha_1 = \alpha_2 = 0.150$, where total relative biases of the HC1 and HC3 estimators equal 14.1% and 10.0%, respectively. The total relative biases of the respective corrected estimators up to three iterations are 2.5% and 1.1%, respectively. It should also be noted that in the second design (where there are leverage points) HC3 outperforms HC0 and HC1 under moderate to strong heteroskedasticity. HC2 is the best performer among the unmodified estimators. A comparison of the improved estimators favor the bias adjusted HC2 estimators. For instance, when $n = 50$ and $\alpha_1 = \alpha_2 = 0.15$, the total relative biases of HC02, HC12, HC22 and HC32 (corrected estimators based on two iterations of the bias adjusting mechanism) are, respectively, 6.5%, 6.1%, 2.5% and 2.8%. Therefore, there are gains from starting the bias adjustment sequence at the HC2 estimator.

Table 1a. Bias evaluation I. The model is $y_t = \beta_1 + \beta_2 x_t + u_t$, $u_t \sim (0, \exp\{\alpha_1 x_t + \alpha_2 x_t^2\})$, $t = 1, \dots, n$. The covariate x_t is obtained as random draws from a $\mathcal{U}(0, 1)$ distribution. The quantities in the table below are the total relative biases of the HC0 and HC1 estimators and their respective corrected versions, and also the total relative bias of the OLS estimator.

n	α_1, α_2	λ	OLS	HC0	HC01	HC02	HC03	HC04	HC1	HC11	HC12	HC13	HC14
50	0.00	1.00	0.000	0.099	0.006	0.000	0.000	0.000	0.010	0.001	0.000	0.000	0.000
	0.50	2.53	0.346	0.102	0.007	0.001	0.000	0.000	0.023	0.003	0.000	0.000	0.000
	1.00	6.42	0.818	0.109	0.008	0.001	0.000	0.000	0.030	0.004	0.000	0.000	0.000
	1.50	16.27	1.321	0.118	0.009	0.001	0.000	0.000	0.040	0.005	0.001	0.000	0.000
	2.00	41.23	1.716	0.130	0.011	0.001	0.000	0.000	0.052	0.006	0.001	0.000	0.000
100	2.50	104.47	1.914	0.142	0.013	0.001	0.000	0.000	0.064	0.008	0.001	0.000	0.000
	0.00	1.00	0.000	0.049	0.001	0.000	0.000	0.000	0.010	0.001	0.000	0.000	0.000
	0.50	2.53	0.347	0.051	0.002	0.000	0.000	0.000	0.011	0.001	0.000	0.000	0.000
	1.00	6.42	0.819	0.054	0.002	0.000	0.000	0.000	0.015	0.001	0.000	0.000	0.000
	1.50	16.27	1.326	0.059	0.002	0.000	0.000	0.000	0.019	0.001	0.000	0.000	0.000
150	2.00	41.23	1.724	0.065	0.003	0.000	0.000	0.000	0.025	0.002	0.000	0.000	0.000
	2.50	104.47	1.927	0.071	0.003	0.000	0.000	0.000	0.031	0.002	0.000	0.000	0.000
	0.00	1.00	0.000	0.033	0.001	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000
	0.50	2.53	0.348	0.034	0.001	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000
	1.00	6.42	0.820	0.036	0.001	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000
200	1.50	16.27	1.327	0.039	0.001	0.000	0.000	0.000	0.013	0.001	0.000	0.000	0.000
	2.00	41.23	1.727	0.043	0.001	0.000	0.000	0.000	0.017	0.001	0.000	0.000	0.000
	2.50	104.47	1.931	0.047	0.001	0.000	0.000	0.000	0.021	0.001	0.000	0.000	0.000
	0.00	1.00	0.000	0.025	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000
	0.50	2.53	0.348	0.026	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000
200	1.00	6.42	0.820	0.027	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000
	1.50	16.27	1.328	0.030	0.001	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000
	2.00	41.23	1.729	0.032	0.001	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000
	2.50	104.47	1.933	0.035	0.001	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000

Table 1b. Bias evaluation I. The model is $y_t = \beta_1 + \beta_2 x_t + u_t$, $u_t \sim (0, \exp\{\alpha_1 x_t + \alpha_2 x_t^2\})$, $t = 1, \dots, n$. The covariate x_t is obtained as random draws from a $\mathcal{U}(0, 1)$ distribution. The quantities in the table below are the total relative biases of the HC2 and HC3 estimators and their respective corrected versions.

n	α_1, α_2	λ	HC2	HC21	HC22	HC23	HC24	HC3	HC31	HC32	HC33	HC34
50	0.00	1.00	0.000	0.001	0.000	0.000	0.000	0.105	0.005	0.000	0.000	0.000
	0.50	2.53	0.005	0.001	0.000	0.000	0.000	0.106	0.005	0.000	0.000	0.000
	1.00	6.42	0.011	0.001	0.000	0.000	0.000	0.108	0.005	0.000	0.000	0.000
	1.50	16.27	0.016	0.002	0.000	0.000	0.000	0.110	0.006	0.000	0.000	0.000
	2.00	41.23	0.018	0.003	0.000	0.000	0.000	0.110	0.006	0.001	0.000	0.000
	2.50	104.47	0.020	0.004	0.000	0.000	0.000	0.111	0.006	0.001	0.000	0.000
100	0.00	1.00	0.000	0.000	0.000	0.000	0.000	0.051	0.001	0.000	0.000	0.000
	0.50	2.53	0.002	0.000	0.000	0.000	0.000	0.052	0.001	0.000	0.000	0.000
	1.00	6.42	0.005	0.000	0.000	0.000	0.000	0.053	0.001	0.000	0.000	0.000
	1.50	16.27	0.008	0.000	0.000	0.000	0.000	0.054	0.001	0.000	0.000	0.000
	2.00	41.23	0.009	0.001	0.000	0.000	0.000	0.054	0.002	0.000	0.000	0.000
	2.50	104.47	0.009	0.001	0.000	0.000	0.000	0.054	0.002	0.000	0.000	0.000
150	0.00	1.00	0.000	0.000	0.000	0.000	0.000	0.034	0.001	0.000	0.000	0.000
	0.50	2.53	0.002	0.000	0.000	0.000	0.000	0.034	0.001	0.000	0.000	0.000
	1.00	6.42	0.003	0.000	0.000	0.000	0.000	0.035	0.001	0.000	0.000	0.000
	1.50	16.27	0.005	0.000	0.000	0.000	0.000	0.035	0.001	0.000	0.000	0.000
	2.00	41.23	0.006	0.000	0.000	0.000	0.000	0.036	0.001	0.000	0.000	0.000
	2.50	104.47	0.006	0.000	0.000	0.000	0.000	0.036	0.001	0.000	0.000	0.000
200	0.00	1.00	0.000	0.000	0.000	0.000	0.000	0.025	0.000	0.000	0.000	0.000
	0.50	2.53	0.001	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000
	1.00	6.42	0.003	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000
	1.50	16.27	0.004	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000
	2.00	41.23	0.004	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.000	0.000
	2.50	104.47	0.005	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.000	0.000

Table 2a. Bias evaluation II. The model is $y_t = \beta_1 + \beta_2 x_t + u_t$, $u_t \sim (0, \exp\{\alpha_1 x_t + \alpha_2 x_t^2\})$, $t = 1, \dots, n$. The covariate x_t is obtained as random draws from a t_3 distribution. The quantities in the table below are the total relative biases of the HC0 and HC1 estimators and their respective corrected versions, and also the total relative bias of the OLS estimator.

n	α_1, α_2	λ	OLS	HC0	HC01	HC02	HC03	HC04	HC1	HC11	HC12	HC13	HC14
50	0.000	1.00	0.000	0.176	0.038	0.012	0.004	0.002	0.100	0.032	0.011	0.004	0.001
	0.025	2.24	0.333	0.236	0.061	0.021	0.008	0.003	0.163	0.054	0.019	0.007	0.003
	0.050	5.02	0.583	0.294	0.085	0.030	0.011	0.004	0.223	0.076	0.027	0.010	0.004
	0.075	11.25	0.761	0.351	0.108	0.039	0.015	0.006	0.282	0.098	0.036	0.014	0.006
	0.100	25.21	0.887	0.405	0.131	0.048	0.019	0.008	0.338	0.119	0.045	0.018	0.007
100	0.125	56.50	0.976	0.455	0.153	0.057	0.023	0.009	0.391	0.140	0.053	0.021	0.009
	0.150	126.60	1.040	0.497	0.172	0.065	0.026	0.011	0.435	0.158	0.061	0.025	0.010
	0.00	1.00	0.000	0.088	0.009	0.001	0.000	0.000	0.049	0.008	0.001	0.000	0.000
	0.025	2.24	0.325	0.118	0.015	0.003	0.000	0.000	0.080	0.013	0.002	0.000	0.000
	0.050	5.02	0.565	0.147	0.021	0.004	0.001	0.000	0.109	0.019	0.003	0.001	0.000
150	0.075	11.25	0.730	0.175	0.027	0.005	0.001	0.000	0.138	0.024	0.004	0.001	0.000
	0.100	25.21	0.840	0.202	0.033	0.006	0.001	0.000	0.166	0.029	0.005	0.001	0.000
	0.125	56.50	0.913	0.227	0.038	0.007	0.001	0.000	0.191	0.034	0.007	0.001	0.000
	0.150	126.60	0.962	0.249	0.043	0.008	0.002	0.000	0.213	0.039	0.007	0.002	0.000
	0.00	1.00	0.000	0.059	0.004	0.000	0.000	0.000	0.033	0.003	0.000	0.000	0.000
200	0.025	2.24	0.322	0.079	0.007	0.001	0.000	0.000	0.053	0.006	0.001	0.000	0.000
	0.050	5.02	0.559	0.098	0.009	0.001	0.000	0.000	0.072	0.008	0.001	0.000	0.000
	0.075	11.25	0.720	0.117	0.012	0.001	0.000	0.000	0.091	0.011	0.001	0.000	0.000
	0.100	25.21	0.824	0.135	0.015	0.002	0.000	0.000	0.110	0.013	0.002	0.000	0.000
	0.125	56.50	0.892	0.152	0.017	0.002	0.000	0.000	0.127	0.015	0.002	0.000	0.000
200	0.150	126.60	0.936	0.166	0.019	0.002	0.000	0.000	0.141	0.017	0.002	0.000	0.000
	0.00	1.00	0.000	0.044	0.002	0.000	0.000	0.000	0.024	0.002	0.000	0.000	0.000
	0.025	2.24	0.321	0.059	0.004	0.000	0.000	0.000	0.039	0.003	0.000	0.000	0.000
	0.050	5.02	0.556	0.074	0.005	0.000	0.000	0.000	0.054	0.005	0.000	0.000	0.000
	0.075	11.25	0.714	0.088	0.007	0.001	0.000	0.000	0.068	0.006	0.001	0.000	0.000
200	0.100	25.21	0.817	0.101	0.008	0.001	0.000	0.000	0.082	0.007	0.001	0.000	0.000
	0.125	56.50	0.882	0.114	0.010	0.001	0.000	0.000	0.095	0.008	0.001	0.000	0.000
	0.150	126.60	0.924	0.124	0.011	0.001	0.000	0.000	0.105	0.010	0.001	0.000	0.000

Table 2b. Bias evaluation II. The model is $y_t = \beta_1 + \beta_2 x_t + u_t$, $u_t \sim (0, \exp\{\alpha_1 x_t + \alpha_2 x_t^2\})$, $t = 1, \dots, n$. The covariate x_t is obtained as random draws from a t_3 distribution. The quantities in the table below are the total relative biases of the HC2 and HC3 estimators and their respective corrected versions.

n	α_1, α_2	λ	HC2	HC21	HC22	HC23	HC24	HC3	HC31	HC32	HC33	HC34
50	0.000	1.00	0.000	0.011	0.005	0.002	0.001	0.214	0.022	0.005	0.001	0.000
	0.025	2.24	0.038	0.021	0.008	0.003	0.001	0.206	0.029	0.008	0.002	0.001
	0.050	5.02	0.071	0.031	0.012	0.005	0.002	0.207	0.038	0.011	0.004	0.001
	0.075	11.25	0.099	0.040	0.016	0.006	0.003	0.218	0.047	0.015	0.005	0.002
	0.100	25.21	0.122	0.049	0.019	0.008	0.003	0.238	0.059	0.019	0.007	0.003
	0.125	56.498	0.141	0.056	0.023	0.009	0.004	0.261	0.071	0.024	0.009	0.003
	0.150	126.60	0.157	0.063	0.025	0.011	0.004	0.283	0.081	0.028	0.011	0.004
100	0.00	1.00	0.000	0.002	0.000	0.000	0.000	0.096	0.005	0.001	0.000	0.000
	0.025	2.24	0.016	0.005	0.001	0.000	0.000	0.096	0.008	0.001	0.000	0.000
	0.050	5.02	0.030	0.006	0.001	0.000	0.000	0.100	0.010	0.002	0.000	0.000
	0.075	11.25	0.041	0.008	0.002	0.000	0.000	0.109	0.013	0.002	0.000	0.000
	0.100	25.21	0.050	0.010	0.002	0.000	0.000	0.122	0.016	0.003	0.001	0.000
	0.125	56.498	0.056	0.011	0.002	0.000	0.000	0.136	0.020	0.003	0.001	0.000
	0.150	126.60	0.062	0.012	0.002	0.001	0.000	0.149	0.023	0.004	0.001	0.000
150	0.00	1.00	0.000	0.001	0.000	0.000	0.000	0.062	0.002	0.000	0.000	0.000
	0.025	2.24	0.010	0.002	0.000	0.000	0.000	0.063	0.003	0.000	0.000	0.000
	0.050	5.02	0.019	0.003	0.000	0.000	0.000	0.066	0.004	0.000	0.000	0.000
	0.075	11.25	0.026	0.003	0.000	0.000	0.000	0.072	0.006	0.001	0.000	0.000
	0.100	25.21	0.031	0.004	0.001	0.000	0.000	0.081	0.007	0.001	0.000	0.000
	0.125	56.498	0.035	0.005	0.001	0.000	0.000	0.091	0.009	0.001	0.000	0.000
	0.150	126.60	0.038	0.005	0.001	0.000	0.100	0.010	0.001	0.001	0.000	0.000
200	0.00	1.00	0.000	0.001	0.000	0.000	0.000	0.046	0.001	0.000	0.000	0.000
	0.025	2.24	0.007	0.001	0.000	0.000	0.000	0.047	0.002	0.000	0.000	0.000
	0.050	5.02	0.014	0.001	0.000	0.000	0.000	0.049	0.003	0.000	0.000	0.000
	0.075	11.25	0.019	0.002	0.000	0.000	0.000	0.054	0.003	0.000	0.000	0.000
	0.100	25.21	0.023	0.002	0.000	0.000	0.000	0.061	0.004	0.000	0.000	0.000
	0.125	56.498	0.025	0.003	0.000	0.000	0.000	0.068	0.005	0.000	0.000	0.000
	0.150	126.60	0.028	0.003	0.000	0.000	0.075	0.006	0.001	0.001	0.000	0.000

A quick comparison between Tables 1 and 2 indicate that the presence of high leverage points in the data is more decisive to the reliability of the different covariance matrix estimators than the degree of heteroskedasticity itself. For example, when $n = 50$ and the design matrix does not contain high leverage points the total relative biases of the consistent estimators under *strong heteroskedasticity* are smaller than the respective biases in the situation where there are leverage points and the data are *homoskedastic*.

It is possible to identify the presence of high leverage points in the second design by looking at h_t , $t = 1, \dots, n$, the diagonal elements of H . It can be shown that $0 < h_t < 1$ for all t . Note also that $\sum_{t=1}^n h_t = \text{tr}(H) = \text{tr}[X(X'X)^{-1}X'] = p$. The h_t 's therefore have an average value of p/n . A general rule-of-thumb is that values of h_t in excess of two or three times the average are regarded as influential and worthy of further investigation (Judge et al., 1988, p. 893). As noted by Chesher & Jewitt (1987, p. 1219), the possibility of severe downward bias in the White estimator arises when there are large h_t , because the associated least squares residuals have small magnitude on average and the White estimator takes small residuals as evidence of small error variances.

Table 3 displays the maximum and minimum values of h_t for the two regression designs under consideration; it also includes the benchmark values of $2p/n$ and $3p/n$. We note that when the covariate values are obtained from a t_3 distribution the maximum value of h_t exceeds $3p/n$, thus suggesting that the data contain potentially influential observations. The same does not happen when the covariate values are obtained from a uniform distribution, in which case the design matrix contains no leverage point.

Table 3. Identification of high leverage points.

n	$x_t \sim U(0, 1)$		$x_t \sim t_3$		$2p/n$	$3p/n$
	h_{min}	h_{max}	h_{min}	h_{max}		
50	0.020	0.078	0.020	0.250	0.080	0.120
100	0.010	0.034	0.010	0.125	0.040	0.060
150	0.007	0.026	0.007	0.083	0.027	0.040
200	0.005	0.019	0.005	0.062	0.020	0.030

Variance estimators of different linear combinations of the ordinary least squares estimator are affected differently by the heteroskedasticity present in the data. We shall now determine the linear combination of the regression parameter estimators that yields the maximal estimated variance bias, that is the p -vector c such that $c'c = 1$ and $E\{\widehat{\text{var}}(c'\widehat{\beta})\} - \text{var}(c'\widehat{\beta})$ is maximized. Since the bias matrices given in (11) are symmetric, the maximum absolute value of the bias of the estimated variances of linear combinations of the $\widehat{\beta}$'s is the maximum of the absolute values of the eigenvalues of the corresponding bias matrices. The results are presented in Table 4 (which is divided into 4a and 4b) for the case where the covariate values are obtained from a uniform distribution.

We note that HC2 is the unmodified consistent estimator with best performance, and that HC1 displays maximal biases considerably smaller than those of HC0 and

Table 4a. Bias evaluation III. The entries in the table below are the absolute value of the maximal bias of $\widehat{\text{var}}(c'\beta)$ for different variance estimators.

n	α_1, α_2	HC0	HC01	HC02	HC03	HC04	HC1	HC11	HC12	HC13	HC14
50	0.00	0.014	0.001	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000
	0.50	0.023	0.002	0.000	0.000	0.000	0.007	0.001	0.000	0.000	0.000
	1.00	0.044	0.003	0.000	0.000	0.000	0.016	0.002	0.000	0.000	0.000
	1.50	0.094	0.008	0.001	0.000	0.000	0.041	0.005	0.001	0.000	0.000
	2.00	0.216	0.020	0.002	0.000	0.000	0.105	0.012	0.001	0.000	0.000
100	2.50	0.509	0.050	0.006	0.001	0.000	0.265	0.031	0.004	0.000	0.000
	0.00	0.003	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	0.50	0.006	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
	1.00	0.011	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
	1.50	0.024	0.001	0.000	0.000	0.000	0.010	0.001	0.000	0.000	0.000
150	2.00	0.054	0.003	0.000	0.000	0.000	0.026	0.001	0.000	0.000	0.000
	2.50	0.127	0.006	0.000	0.000	0.000	0.065	0.004	0.000	0.000	0.000
	0.00	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.50	0.003	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	1.00	0.005	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
200	1.50	0.010	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
	2.00	0.024	0.001	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000
	2.50	0.057	0.002	0.000	0.000	0.000	0.029	0.001	0.000	0.000	0.000
	0.00	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.50	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
200	1.00	0.003	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	1.50	0.006	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
	2.00	0.014	0.000	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000
	2.50	0.032	0.001	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000
	0.00	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 4b. Bias evaluation III. The entries in the table below are the absolute value of the maximal bias of $\widehat{\text{var}}(c'\hat{\beta})$ for different variance estimators.

n	α_1, α_2	HC2	HC21	HC22	HC23	HC24	HC3	HC31	HC32	HC33	HC34
50	0.00	0.000	0.000	0.000	0.000	0.000	0.015	0.001	0.000	0.000	0.000
	0.50	0.001	0.000	0.000	0.000	0.000	0.022	0.001	0.000	0.000	0.000
	1.00	0.005	0.001	0.000	0.000	0.000	0.037	0.002	0.000	0.000	0.000
	1.50	0.015	0.002	0.000	0.000	0.000	0.069	0.004	0.000	0.000	0.000
	2.00	0.045	0.006	0.001	0.000	0.000	0.139	0.009	0.001	0.000	0.000
100	2.50	0.121	0.016	0.002	0.001	0.000	0.295	0.020	0.002	0.000	0.000
	0.00	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
	0.50	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000
	1.00	0.001	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.000
	1.50	0.004	0.000	0.000	0.000	0.000	0.017	0.001	0.000	0.000	0.000
150	2.00	0.011	0.001	0.000	0.000	0.000	0.034	0.001	0.000	0.000	0.000
	2.50	0.029	0.002	0.000	0.000	0.000	0.073	0.003	0.000	0.000	0.000
	0.00	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
	0.50	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
	1.00	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
200	1.50	0.002	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000
	2.00	0.005	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.000
	2.50	0.013	0.001	0.000	0.000	0.000	0.032	0.001	0.000	0.000	0.000
	0.00	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	0.50	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
200	1.00	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
	1.50	0.001	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
	2.00	0.003	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000
	2.50	0.007	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000
	0.00	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000

HC3. Under homoskedasticity, HC0 and HC3 have approximately the same finite-sample behavior, and HC3 outperforms HC0 under unequal error variances. It is noteworthy that the maximal biases are always much larger for the unmodified estimators than for their respective corrected versions. For $n = 50$ and $\alpha_1 = \alpha_2 = 2.50$, for instance, we obtain maximal biases for the one-iteration corrected estimators that are 10 times smaller than those of the respective original estimators.

In order to evaluate the variances of the different variance estimators of linear combinations of the elements of $\widehat{\beta}$ we use the exact expression given in (20). The results for uniform covariate values and $c = (0, 1)'$ are given in Table 5 (again divided into 5a and 5b). The entries correspond to the standard errors of the estimators of $\text{var}(\widehat{\beta}_2)$. It is well known that bias correction oftentimes induces variance inflation, and we note that this occurs here. It is important to note, however, that the variance inflation induced by the bias correction mechanism is not large, especially when heteroskedasticity is not strong. For example, when $n = 50$ and $\alpha_1 = \alpha_2 = 1.00$ ($\lambda = 6.42$), the standard errors of the HC0 estimator and its first four corrected versions are, respectively, 0.209, 0.228, 0.230, 0.230, 0.230. Two remarks are in order here: 1) The bias correction mechanism reduces the variance of the estimator when applied to HC3, conversely to what happens to their estimators. 2) From the third iteration of the bias correcting mechanism on, the variances of the improved estimators become stable.

5. CONCLUDING REMARKS

Cross sectional data oftentimes display some form of heteroskedasticity. It is common practice to still use the OLSE of the vector of regression parameters, since it remains unbiased and consistent. Its covariance matrix, however, has to be consistently estimated in order for inference to be performed. In this paper, we propose a sequence of bias adjusted heteroskedasticity-consistent covariance matrix estimators. The sequence of improved estimators is defined by sequentially transforming a selected consistent estimator. The proposed sequence is general enough to be applicable to a number of well known estimators. Overall, the numerical results favor improved estimators obtaining by modifying the HC2 heteroskedasticity-consistent covariance matrix estimator.

ACKNOWLEDGEMENTS

We thank Spyros Zarkos for comments on a previous draft. The first author gratefully acknowledges financial support from CNPq.

Table 5a. Variance evaluation. The entries in the table below are the standard error of the different variance estimators for the OLS slope estimator in a simple linear regression model.

n	α_1, α_2	HC0	HC01	HC02	HC03	HC04	HC1	HC11	HC12	HC13	HC14
50	0.00	0.054	0.058	0.058	0.058	0.058	0.056	0.058	0.058	0.058	0.058
	0.50	0.097	0.105	0.105	0.105	0.105	0.101	0.105	0.105	0.105	0.105
	1.00	0.209	0.228	0.230	0.230	0.230	0.218	0.228	0.230	0.230	0.230
	1.50	0.484	0.529	0.534	0.534	0.534	0.504	0.531	0.534	0.534	0.534
	2.00	1.148	1.258	1.270	1.272	1.272	1.195	1.262	1.271	1.272	1.272
100	2.50	2.754	3.026	3.057	3.060	3.061	2.869	3.037	3.058	3.061	3.061
	0.00	0.019	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020
	0.50	0.035	0.036	0.037	0.037	0.037	0.036	0.037	0.037	0.037	0.037
	1.00	0.076	0.080	0.080	0.080	0.080	0.078	0.080	0.080	0.080	0.080
	1.50	0.177	0.185	0.186	0.186	0.186	0.181	0.185	0.186	0.186	0.186
150	2.00	0.421	0.441	0.442	0.443	0.443	0.430	0.442	0.442	0.443	0.443
	2.50	1.013	1.063	1.066	1.066	1.066	1.034	1.064	1.066	1.066	1.066
	0.00	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
	0.50	0.019	0.020	0.020	0.020	0.020	0.019	0.020	0.020	0.020	0.020
	1.00	0.042	0.043	0.043	0.043	0.043	0.042	0.043	0.043	0.043	0.043
200	1.50	0.097	0.100	0.101	0.101	0.101	0.099	0.100	0.101	0.101	0.101
	2.00	0.232	0.239	0.240	0.240	0.240	0.235	0.239	0.240	0.240	0.240
	2.50	0.558	0.577	0.578	0.578	0.578	0.566	0.577	0.578	0.578	0.578
	0.00	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
	0.50	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
200	1.00	0.027	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
	1.50	0.064	0.065	0.065	0.065	0.065	0.064	0.065	0.065	0.065	0.065
	2.00	0.151	0.155	0.155	0.155	0.155	0.153	0.155	0.155	0.155	0.155
	2.50	0.365	0.374	0.374	0.374	0.374	0.369	0.374	0.374	0.374	0.374

Table 5b. Variance evaluation. The entries in the table below are the standard error of the different variance estimators for the OLS slope estimator in a simple linear regression model.

n	α_1, α_2	HC2	HC21	HC22	HC23	HC24	HC3	HC31	HC32	HC33	HC34
50	0.00	0.057	0.058	0.058	0.058	0.058	0.061	0.058	0.058	0.058	0.058
	0.50	0.103	0.105	0.105	0.105	0.105	0.110	0.106	0.105	0.105	0.105
	1.00	0.224	0.229	0.230	0.230	0.230	0.240	0.231	0.230	0.230	0.230
	1.50	0.519	0.532	0.534	0.534	0.534	0.557	0.536	0.535	0.534	0.534
	2.00	1.232	1.267	1.271	1.272	1.272	1.323	1.276	1.272	1.272	1.272
100	2.50	2.961	3.048	3.059	3.061	3.061	3.184	3.184	3.072	3.062	3.061
	0.00	0.020	0.020	0.020	0.020	0.020	0.021	0.020	0.020	0.020	0.020
	0.50	0.036	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.037
	1.00	0.079	0.080	0.080	0.080	0.080	0.081	0.080	0.080	0.080	0.080
	1.50	0.183	0.186	0.186	0.186	0.186	0.190	0.186	0.186	0.186	0.186
150	2.00	0.436	0.442	0.442	0.443	0.443	0.452	0.443	0.443	0.443	0.443
	2.50	1.050	1.065	1.066	1.066	1.066	1.088	1.067	1.066	1.066	1.066
	0.00	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.011
	0.50	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020
	1.00	0.043	0.043	0.043	0.043	0.043	0.044	0.043	0.043	0.043	0.043
200	1.50	0.100	0.100	0.101	0.101	0.101	0.102	0.101	0.101	0.101	0.101
	2.00	0.237	0.240	0.240	0.240	0.240	0.243	0.240	0.240	0.240	0.240
	2.50	0.572	0.577	0.578	0.578	0.578	0.585	0.578	0.578	0.578	0.578
	0.00	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007
	0.50	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
200	1.00	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028
	1.50	0.065	0.065	0.065	0.065	0.065	0.066	0.065	0.065	0.065	0.065
	2.00	0.154	0.155	0.155	0.155	0.155	0.157	0.155	0.155	0.155	0.155
	2.50	0.371	0.374	0.374	0.374	0.374	0.378	0.374	0.374	0.374	0.374
	0.00	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007

REFERENCES

- [1] Chesher, A. & Jewitt, I. (1987). The bias of a heteroskedasticity consistent covariance matrix estimator. *Econometrica*, 55, 1217–1222.
- [2] Cribari–Neto, F., Ferrari, S.L.P. & Cordeiro, G.M. (2000). Improved heteroscedasticity–consistent covariance matrix estimators. *Biometrika*, 87, 907–918.
- [3] Cribari–Neto, F. & Zarkos, S.G. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews*, 18, 211–228.
- [4] Doornik, J.A. (2001). *Ox: an Object-oriented Matrix Programming Language*, 4th ed. London: Timberlake Consultants & Oxford: <http://www.nuff.ox.ac.uk/Users/Doornik>.
- [5] Hinkley, D.V. (1977). Jackknifing in unbalanced situations. *Technometrics*, 19, 285–292.
- [6] Horn, S.D., Horn, R.A. & Duncan, D.B. (1975). Estimating heteroskedastic variances in linear models. *Journal of the American Statistical Association*, 70, 380–385.
- [7] Judge, G.C., Hill, R.C., Griffiths, W.E., Lutkepohl, H. & Lee, T.C. (1988). *Introduction to the Theory and Practice of Econometrics*, 2nd ed. New York: Wiley.
- [8] Long, J.S. & Ervin, L.H. (2000). Using heteroskedasticity-consistent standard errors in the linear regression model. *The American Statistician*, 54, 217–224.
- [9] MacKinnon, J.G. & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite-sample properties. *Journal of Econometrics*, 29, 305–25.
- [10] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.
- [11] Seber, G.A.F. (1977). *Linear Regression Analysis*. New York: Wiley.
- [12] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817–38.