Ann Inst Stat Math (2010) 62:1053–1082
DOI 10.1007/s10463-008-0201-5

Sequences of bias-adjusted covariance matrix estimators under heteroskedasticity of unknown form

Francisco Cribari-Neto · Maria da Glória A. Lima

Received: 11 October 2007 / Revised: 5 June 2008 / Published online: 18 September 2008

Abstract The linear regression model is commonly used by practitioners to model the relationship between the variable of interest and a set of explanatory variables. The assumption that all error variances are the same, known as homoskedasticity, is oftentimes violated when cross sectional data are used. Consistent standard errors for the ordinary least squares estimators of the regression parameters can be computed following the approach proposed by White (Econometrica 48:817–838, 1980). Such standard errors, however, are considerably biased in samples of typical sizes. An improved covariance matrix estimator was proposed by Qian and Wang (J Stat Comput Simul 70:161–174, 2001). In this paper, we improve upon the Qian–Wang estimator by defining a sequence of bias-adjusted estimators with increasing accuracy. The numerical results show that the Qian–Wang estimator is typically much less biased than the estimator proposed by Halbert White and that our correction to the former can be quite effective in small samples. Finally, we show that the Qian–Wang estimator can be generalized into a broad class of heteroskedasticity-consistent covariance matrix estimators, and our results can be easily extended to such a class of estimators.

Keywords Bias · Covariance matrix estimation · Heteroskedasticity · Linear regression

F. Cribari-Neto (✉)
Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária,
Recife, PE 50740-540, Brazil
e-mail: cribari@de.ufpe.br
URL: http://www.de.ufpe.br/∼cribari

M. G. A. Lima
Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco,
Dois Irmãos, Recife, PE 52171-900, Brazil
e-mail: abgloria@hotlink.com.br

1 Introduction

Homoskedasticity is a commonly violated assumption in the linear regression model. It states that the error variances are constant across all observations, regardless of the covariate values. The ordinary least squares estimator (OLSE) of the vector of regression parameters remains unbiased, consistent and asymptotically normal even when such an assumption does not hold. The OLSE is thus a valid estimator even under heteroskedasticity of unknown form. In order to perform asymptotically valid interval estimation and hypothesis testing inference, however, one needs to obtain a consistent estimator of the OLSE covariance matrix which can yield, for instance, asymptotically valid standard errors. White (1980), in an influential paper, showed that consistent standard errors can be easily obtained using a sandwich-type estimator. His estimator, which we shall call HC0, is considerably biased in finite samples; in particular, it tends to be quite optimistic, i.e., it underestimates the true variances, especially when the data contain leverage points. A more accurate estimator was proposed by Qian and Wang (2001). Their estimator usually displays much smaller biases in samples of small to moderate sizes. Our chief goal in this paper is twofold. First, we improve upon their estimator by bias-correcting it in an iterative fashion. To that end, we derive a sequence of bias-adjusted estimators such that the orders of the respective biases decrease as we move along the sequence. Our numerical results show that the proposed bias correcting scheme can be quite effective in some situations. Second, we define a class of heteroskedasticity-consistent covariance matrix estimators which includes modified versions of some well-known variants of White's estimator, and argue that the results obtained for the Qian–Wang estimator can be easily extended to this new class of estimators.

A few remarks are in order. First, bias correction may induce variance inflation, as noted by MacKinnon and Smith (1998). Indeed, our numerical results indicate that this is the case. Second, it is also possible to achieve increasing precision as far as bias is concerned by using the iterated bootstrap, which is, nonetheless, highly computer intensive. Our sequence of modified estimators achieves similar precision with almost no computational burden. For details on the relation between the two approaches (analytical and bootstrap) to iterated corrections, see Ferrari and Cribari-Neto (1998). Third, finite sample corrections to White's estimator were obtained by Cribari-Neto et al. (2000). Our results, however, apply to an estimator proposed by Qian and Wang (2001) which is more accurate than White's estimator; it is even unbiased under equal error variances. Additionally, we show that the Qian–Wang estimator can be generalized into a class that includes modified versions of well-known variants of White's estimator, and argue that the results obtained for the Qian–Wang estimator can be generalized to this broader class of heteroskedasticity-robust estimators.

The paper unfolds as follows. Section 2 introduces the linear regression model and some heteroskedasticity-consistent covariance matrix estimators. In Sect. 3 we derive a sequence of consistent estimators for the covariance matrix of the OLSE. We do so by defining a sequential bias correcting scheme which is initialized at the estimator proposed by Qian and Wang (2001). In Sect. 4 we obtain estimators for the variance of linear combinations of the elements in the vector of OLSEs. Results from a numerical evaluation are presented in Sect. 5; these are exact, not Monte Carlo results. Two empirical applications that use real data are presented and discussed in Sect. 6. In Sect. 7 we show that modified versions of variants of Halbert White's estimator can be easily obtained, and that the resulting estimators can be easily adjusted for bias; as a consequence, all of the results we derive can be extended to cover estimators other than that proposed by Qian and Wang (2001). Finally, Sect. 8 offers some concluding remarks.

2 The model and covariance matrix estimators

The model of interest is the linear regression model, which can be written as

$$y = X\beta + \varepsilon,$$

where y and ε are $n \times 1$ vectors of responses and errors, respectively, X is a full column rank fixed $n \times p$ matrix of regressors ($\text{rank}(X) = p < n$) and $\beta = (\beta_1, \ldots, \beta_p)'$ is a p-vector of unknown regression parameters. The error ε_i has mean zero, variance $0 < \sigma_i^2 < \infty$, $i = 1, \ldots, n$, and is uncorrelated to ε_j whenever $j \neq i$. Let Ω denote the covariance matrix of the errors, i.e., $\Omega = \text{cov}(\varepsilon) = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$.

The OLSE of β can be written in closed-form as $\widehat{\beta} = (X'X)^{-1}X'y$. It is unbiased, consistent and asymptotically normal even under unequal error variances. Its covariance matrix is $\Psi = \text{cov}(\widehat{\beta}) = P\Omega P'$, where $P = (X'X)^{-1}X'$. Under homoskedasticity, $\sigma_i^2 = \sigma^2$, $i = 1, \ldots, n$, where $\sigma^2 > 0$, and hence $\Psi = \sigma^2(X'X)^{-1}$. The covariance matrix Ψ can then be easily estimated as

$$\widehat{\Psi} = \widehat{\sigma}^2(X'X)^{-1},$$

where $\widehat{\sigma}^2 = (y - X\widehat{\beta})'(y - X\widehat{\beta})/(n - p)$.

Under heteroskedasticity, it is common practice to use the OLSE coupled with a consistent covariance matrix estimator. To that end, one uses an estimator $\widehat{\Omega}$ of Ω (which is $n \times n$) such that $X'\widehat{\Omega}X$ is consistent for $X'\Omega X$ (which is $p \times p$), i.e., $\text{plim}[(X'\Omega X)^{-1}(X'\widehat{\Omega}X)] = I_p$, where I_p is the p-dimensional identity matrix. In what follows, we shall omit the order subscript when denoting the identity matrix; the order must be implicitly understood.

White (1980) obtained a consistent estimator for Ψ. His estimator is consistent under both homoskedasticity and heteroskedasticity of unknown form, and can be written as

$$\text{HC0} = \widehat{\Psi} = P\widehat{\Omega}P',$$

where $\widehat{\Omega} = \text{diag}\{\widehat{\varepsilon}_1^2, \ldots, \widehat{\varepsilon}_n^2\}$. Here, $\widehat{\varepsilon}_i$ is the ith least squares residual, i.e., $\widehat{\varepsilon}_i = y_i - x_i\widehat{\beta}$, where x_i is the ith row of X, $i = 1, \ldots, n$. The vector of least squares residuals is $\widehat{\varepsilon} = (\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n)' = (I - H)y$, where $H = X(X'X)^{-1}X' = XP$ is a symmetric and idempotent matrix known as 'the hat matrix'. The diagonal elements of H ($h_1, \ldots, h_n$) assume values in the standard unit interval (0, 1) and add up to p; thus, they average $\overline{h} = p/n$. These quantities are used as measures of the leverages of the corresponding observations. A rule-of-thumb states that observations such that $h_i > 2p/n$ or $h_i > 3p/n$ are taken to be leverage points; see, e.g., Davidson and MacKinnon (1993).

The numerical evidence in Cribari-Neto and Zarkos (1999, 2001), Long and Ervin (2000) and MacKinnon and White (1985) showed that the estimator proposed by Halbert White can be quite biased in finite samples and that associated hypothesis tests can be quite liberal. Chesher and Jewitt (1987) showed that the negative HC0 bias is largely due to the presence of observations with high leverage in the data.

Several variants of the HC0 estimator were proposed in the literature, such as

(i) (Hinkley 1977) $\text{HC1} = P\widehat{\Omega}_1 P' = PD_1\widehat{\Omega}P'$, where $D_1 = (n/(n-p))I$;
(ii) (Horn et al. 1975) $\text{HC2} = P\widehat{\Omega}_2 P' = PD_2\widehat{\Omega}P'$, where $D_2 = \text{diag}\{1/(1-h_i)\}$;
(iii) (Davidson and MacKinnon 1993) $\text{HC3} = P\widehat{\Omega}_3 P' = PD_3\widehat{\Omega}P'$, where $D_3 = \text{diag}\{1/(1-h_i)^2\}$;
(iv) (Cribari-Neto 2004) $\text{HC4} = P\widehat{\Omega}_4 P' = PD_4\widehat{\Omega}P'$, where $D_4 = \text{diag}\{1/(1-h_i)^{\delta_i}\}$ and $\delta_i = \min\{4, nh_i/p\}$.

As noted earlier, the HC0 estimator is considerably biased in samples of small to moderate sizes. Cribari-Neto et al. (2000) derived bias-adjusted variants of HC0 by using an iterative bias correction mechanism. The chain of estimators was obtained by correcting HC0, then correcting the resulting adjusted estimator, and so on.

Let $(A)_d$ denote the diagonal matrix obtained by setting the nondiagonal elements of the square matrix A equal to zero. Note that $\widehat{\Omega} = (\widehat{\varepsilon}\widehat{\varepsilon}')_d$. Thus,

$$\begin{aligned}\mathbb{E}(\widehat{\varepsilon}\widehat{\varepsilon}') &= \text{cov}(\widehat{\varepsilon}) + \mathbb{E}(\widehat{\varepsilon})\mathbb{E}(\widehat{\varepsilon}') \\ &= (I-H)\Omega(I-H)\end{aligned}$$

since $(I-H)X = 0$. It thus follows that $\mathbb{E}(\widehat{\Omega}) = \{(I-H)\Omega(I-H)\}_d$ and $\mathbb{E}(\widehat{\Psi}) = P\mathbb{E}(\widehat{\Omega})P'$. Hence, the biases of $\widehat{\Omega}$ and $\widehat{\Psi}$ as estimators of Ω and Ψ are

$$B_{\widehat{\Omega}}(\Omega) = \mathbb{E}(\widehat{\Omega}) - \Omega = \{H\Omega(H-2I)\}_d$$

and

$$B_{\widehat{\Psi}}(\Omega) = \mathbb{E}(\widehat{\Psi}) - \Psi = PB_{\widehat{\Omega}}(\Omega)P',$$

respectively.

Cribari-Neto et al. (2000) define the bias corrected estimator

$$\widehat{\Omega}^{(1)} = \widehat{\Omega} - B_{\widehat{\Omega}}(\widehat{\Omega}).$$

This estimator can be in turn bias corrected:

$$\widehat{\Omega}^{(2)} = \widehat{\Omega}^{(1)} - B_{\widehat{\Omega}^{(1)}}(\widehat{\Omega}),$$

and so on. After k iterations of the bias correcting scheme one obtains

$$\widehat{\Omega}^{(k)} = \widehat{\Omega}^{(k-1)} - B_{\widehat{\Omega}^{(k-1)}}(\widehat{\Omega}).$$

Consider the following recursive function of an $n \times n$ diagonal matrix A:

$$M^{(k+1)}(A) = M^{(1)}(M^{(k)}(A)), \quad k = 0, 1, \ldots,$$

where $M^{(0)}(A) = A$, $M^{(1)}(A) = \{HA(H - 2I)\}_d$, and H is as before. Given two $n \times n$ matrices A and B, it is not difficult to show that, for $k = 0, 1, \ldots$,

P1: $M^{(k)}(A) + M^{(k)}(B) = M^{(k)}(A + B)$;
P2: $M^{(k)}(M^{(1)}(A)) = M^{(k+1)}(A)$;
P3: $\mathbb{E}[M^{(k)}(A)] = M^{(k)}(\mathbb{E}(A))$.

Note that it follows from [P2] that $M^{(2)}(A) = M^{(1)}(M^{(1)}(A))$, $M^{(3)}(A) = M^{(2)}(M^{(1)}(A))$, and so on. We can then write $B_{\widehat{\Omega}}(\Omega) = M^{(1)}(\Omega)$. By induction, it can be shown that the kth-order bias-corrected estimator and its respective bias can be written as

$$\widehat{\Omega}^{(k)} = \sum_{j=0}^{k} (-1)^j M^{(j)}(\widehat{\Omega})$$

and

$$B_{\widehat{\Omega}^{(k)}}(\Omega) = (-1)^k M^{(k+1)}(\Omega), \tag{1}$$

for $k = 1, 2, \ldots$.

It is now possible to define a sequence of bias-corrected covariance matrix estimators as $\{\widehat{\Psi}^{(k)}, k = 1, 2, \ldots\}$, where

$$\widehat{\Psi}^{(k)} = P\widehat{\Omega}^{(k)}P'. \tag{2}$$

The bias of $\widehat{\Psi}^{(k)}$ is

$$B_{\widehat{\Psi}^{(k)}}(\Omega) = (-1)^k PM^{(k+1)}(\Omega)P',$$

$k = 1, 2, \ldots$.

Now assume that the design matrix X is such that P and H are $O(n^{-1})$ and assume that Ω is $O(1)$. In particular, note that the leverage measures $h_1, \ldots, h_n$ converge to zero as $n \to \infty$. Let A be a diagonal matrix such that $A = O(n^{-r})$ for some $r \geq 0$. Thus,

C1: $PAP' = O(n^{-(r+1)})$;
C2: $M^{(1)}(A) = \{HA(H - 2I)\}_d = O(n^{-(r+1)})$.

Since $\Omega = O(n^0)$, it follows from [C1] and [C2] that

$$M^{(1)}(\Omega) = \{H\Omega(H - 2I)\}_d = O(n^{-1});$$

hence, $B_{\widehat{\Omega}}(\Omega) = M^{(1)}(\Omega) = O(n^{-1})$ and the bias of HC0 is

$$B_{\widehat{\Psi}}(\Omega) = PB_{\widehat{\Omega}}(\Omega)P' = O(n^{-2}).$$

Note that

$$M^{(2)}(\Omega) = M^{(1)}(M^{(1)}(\Omega)) = \{H\{H\Omega(H-2I)\}_d(H-2I)\}_d = O(n^{-2}).$$

Since $M^{(k+1)}(\Omega) = M^{(1)}(M^{(k)}(\Omega))$, then $M^{(k+1)}(\Omega) = O(n^{-(k+1)})$ and, thus, $B_{\widehat{\Omega}^{(k)}}(\Omega) = O(n^{-(k+1)})$. Using [C1] one can show that $B_{\widehat{\Psi}^{(k)}}(\Omega) = O(n^{-(k+2)})$. That is, the bias of the k-times corrected estimator is of order $O(n^{-(k+2)})$, whereas the bias of Halbert White's estimator is $O(n^{-2})$.

3 A new class of bias-adjusted estimators

An alternative estimator was proposed by Qian and Wang (2001). It is, as we shall see, a bias-adjusted variant of HC0. Let $K = (H)_d = \text{diag}\{h_1, \ldots, h_n\}$, i.e., K is the diagonal matrix containing the leverage measures, and let $C_i = X(X'X)^{-1}x_i'$ denote the ith column of the hat matrix H.

Following Qian and Wang (2001), define

$$D^{(1)} = \text{diag}\{d_i\} = \text{diag}\left\{(\widehat{\varepsilon}_i^2 - \widehat{b}_i)g_{ii}\right\},$$

where

$$g_{ii} = \left(1 + C_i'KC_i - 2h_i^2\right)^{-1}$$

and

$$\widehat{b}_i = C_i'\left(\widehat{\Omega} - 2\widehat{\varepsilon}_i^2 I\right)C_i.$$

The Qian–Wang estimator can be written as

$$\widehat{V}^{(1)} = PD^{(1)}P'. \tag{3}$$

At the outset, we shall show that the estimator in (3) is a bias-corrected version of the estimator proposed by Halbert White except for an additional correction factor. Note that

$$\begin{aligned} d_i &= \left(\widehat{\varepsilon}_i^2 - \widehat{b}_i\right)g_{ii} \\ &= \left(\widehat{\varepsilon}_i^2 - C_i'\widehat{\Omega}C_i + 2\widehat{\varepsilon}_i^2 C_i'C_i\right)g_{ii}. \end{aligned} \tag{4}$$

The bias corrected estimator in (2) obtained using $k = 1$ (one-step correction) can be written as $\widehat{\Psi}^{(1)} = P\widehat{\Omega}^{(1)}P'$, where

$$\begin{aligned}\widehat{\Omega}^{(1)} &= \widehat{\Omega} - M^{(1)}(\widehat{\Omega}) \\ &= \widehat{\Omega} - \{H\widehat{\Omega}(H-2I)\}_d \\ &= \operatorname{diag}\left\{\widehat{\varepsilon}_i^2 - C_i'\widehat{\Omega}C_i + 2\widehat{\varepsilon}_i^2 h_i\right\}. \end{aligned} \tag{5}$$

Since $h_i = C_i'C_i$, it is easy to see that (4) equals the ith diagonal element of $\widehat{\Omega}^{(1)}$ in (5), apart from multiplication by g_{ii}. Thus,

$$D^{(1)} = [\widehat{\Omega} - \{H\widehat{\Omega}(H-2I)\}_d]G,$$

where $G = \{I + HKH - 2KK\}_d^{-1}$.

Qian and Wang (2001) have shown that $\widehat{V}^{(1)}$ is unbiased for Ψ under homoskedasticity; under heteroskedasticity, the bias of $D^{(1)}$ is $O(n^{-2})$, as we shall show.

We shall now improve upon the Qian–Wang estimator by obtaining a sequence of bias-adjusted estimators with biases of smaller order than that of the estimator in (3) under unequal error variances. At the outset, note that

$$\begin{aligned}D^{(1)} &= (\widehat{\Omega} - M^{(1)}(\widehat{\Omega}))G \\ &= M^{(0)}(\widehat{\Omega})G - M^{(1)}(\widehat{\Omega})G.\end{aligned}$$

Therefore,

$$\begin{aligned}B_{D^{(1)}}(\Omega) &= \mathbb{E}(D^{(1)}) - \Omega \\ &= \mathbb{E}[\widehat{\Omega}G - M^{(1)}(\widehat{\Omega})G] - \Omega \\ &= \mathbb{E}(\widehat{\Omega}G - \Omega) - \mathbb{E}[M^{(1)}(\widehat{\Omega}) - M^{(1)}(\Omega)]G - M^{(1)}(\Omega)G.\end{aligned}$$

Since $\mathbb{E}(\widehat{\Omega} - \Omega) = B_{\widehat{\Omega}}(\Omega) = \{H\Omega(H-2I)\}_d = M^{(1)}(\Omega)$, it then follows that

$$\begin{aligned}\mathbb{E}[M^{(1)}(\widehat{\Omega}) - M^{(1)}(\Omega)] &= \mathbb{E}[M^{(1)}(\widehat{\Omega} - \Omega)] = M^{(1)}(\mathbb{E}(\widehat{\Omega} - \Omega)) \\ &= M^{(1)}(M^{(1)}(\Omega)) = M^{(2)}(\Omega).\end{aligned}$$

The bias of $D^{(1)}$ can be written in closed-form as

$$\begin{aligned}B_{D^{(1)}}(\Omega) &= \mathbb{E}(\widehat{\Omega}G - \Omega G + \Omega G - \Omega) - M^{(2)}(\Omega)G - M^{(1)}(\Omega)G \\ &= M^{(1)}(\Omega)G - M^{(2)}(\Omega)G - M^{(1)}(\Omega)G + \Omega(G-I) \\ &= M^{(0)}(\Omega)(G-I) - M^{(2)}(\Omega)G.\end{aligned}$$

We can now define a bias corrected estimator by subtracting from $D^{(1)}$ its estimated bias:

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

The bias of $D^{(2)}$ is

$$\begin{aligned} B_{D^{(2)}}(\Omega) &= \mathbb{E}(D^{(2)}) - \Omega \\ &= \mathbb{E}[\widehat{\Omega} - M^{(1)}(\widehat{\Omega})G + M^{(2)}(\widehat{\Omega})G] - \Omega \\ &= \mathbb{E}(\widehat{\Omega} - \Omega) - \mathbb{E}[M^{(1)}(\widehat{\Omega}) - M^{(1)}(\Omega)]G - M^{(1)}(\Omega)G \\ &\quad + \mathbb{E}[M^{(2)}(\widehat{\Omega}) - M^{(2)}(\Omega)]G + M^{(2)}(\Omega)G. \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{E}[M^{(2)}(\widehat{\Omega}) - M^{(2)}(\Omega)] &= \mathbb{E}[M^{(2)}(\widehat{\Omega} - \Omega)] = M^{(2)}(\mathbb{E}(\widehat{\Omega} - \Omega)) \\ &= M^{(2)}(M^{(1)}(\Omega)) = M^{(3)}(\Omega). \end{aligned}$$

It then follows that

$$B_{D^{(2)}}(\Omega) = -M^{(1)}(\Omega)(G - I) + M^{(3)}(\Omega)G.$$

In similar fashion,

$$D^{(3)} = \widehat{\Omega} - M^{(1)}(\widehat{\Omega}) + M^{(2)}(\widehat{\Omega})G - M^{(3)}(\widehat{\Omega})G$$

is a bias-corrected version of $D^{(2)}$. Its bias can be expressed as

$$B_{D^{(3)}}(\Omega) = M^{(2)}(\Omega)(G - I) - M^{(4)}(\Omega)G.$$

It is possible to bias correct $D^{(3)}$. To that end, we obtain the following corrected estimator:

$$D^{(4)} = \widehat{\Omega} - M^{(1)}(\widehat{\Omega}) + M^{(2)}(\widehat{\Omega}) - M^{(3)}(\widehat{\Omega})G + M^{(4)}(\widehat{\Omega})G$$

whose bias is

$$B_{D^{(4)}}(\Omega) = -M^{(3)}(\Omega)(G - I) + M^{(5)}(\Omega)G.$$

Note that this estimator can be in turn corrected for bias.

More generally, after k iterations of the bias correcting scheme we obtain

$$\begin{aligned} D^{(k)} &= 1_{(k>1)} \times M^{(0)}(\widehat{\Omega}) + 1_{(k>2)} \times \sum_{j=1}^{k-2} (-1)^j M^{(j)}(\widehat{\Omega}) \\ &\quad + \sum_{j=k-1}^{k} (-1)^j M^{(j)}(\widehat{\Omega})G, \end{aligned}$$

$k = 1, 2, \ldots$, where $1_{(\cdot)}$ is the indicator function. Its bias is

$$B_{D^{(k)}}(\Omega) = (-1)^{k-1} M^{(k-1)}(\Omega)(G - I) + (-1)^k M^{(k+1)}(\Omega) G, \tag{6}$$

$k = 1, 2, \ldots$.

We can now define a sequence $\{\widehat{V}^{(k)}, k = 1, 2, \ldots\}$ of bias-adjusted estimators for Ψ, where

$$\widehat{V}^{(k)} = P D^{(k)} P' \tag{7}$$

is the kth-order bias-corrected estimator of Ψ. The bias of $\widehat{V}^{(k)}$ follows from (6) and (7):

$$B_{\widehat{V}^{(k)}}(\Omega) = P[B_{D^{(k)}}(\Omega)]P'. \tag{8}$$

We shall now obtain the order of the bias in (8). To that end, we make the same assumptions on the matrices X, P, H and Ω as in Sect. 2. We saw in (6) that

$$B_{D^{(k)}}(\Omega) = (-1)^{k-1} M^{(k-1)}(\Omega)(G - I) + (-1)^k M^{(k+1)}(\Omega) G.$$

Note that, if $G = I$, the Qian–Wang estimator reduces to the one-step corrected HC0 estimator of Cribari-Neto et al. (2000) and

$$B_{D^{(k)}}(\Omega) = (-1)^k M^{(k+1)}(\Omega),$$

as in (1). Note also that $M^{(k-1)}(\Omega) = O(n^{-(k-1)})$ and $M^{(k+1)}(\Omega) = O(n^{-(k+1)})$, as we have seen in Sect. 2.

To obtain the order of $G = \{I + HKH - 2KK\}_d^{-1}$, we write $G = \{I + A\}_d^{-1}$, where $A = HKH - 2KK$. Let a_{ii} and g_{ii} denote the ith diagonal elements of A_d and G, respectively, $i = 1, \ldots, n$. Thus,

$$g_{ii} = 1/(1 + a_{ii}), \quad i = 1, \ldots, n.$$

The matrix $G - I$ is also diagonal, its ith diagonal element being

$$t_{ii} = 1/(1 + a_{ii}) - 1 = -a_{ii}/(1 + a_{ii}), \quad i = 1, \ldots, n.$$

Since $H = O(n^{-1})$, then $K = O(n^{-1})$. Thus, $HKH = O(n^{-2})$, $KK = O(n^{-2})$, $A = HKH - 2KK = O(n^{-2})$, $G^{-1} = I + A_d = O(n^0)$ and $G = O(n^0)$. The order of t_{ii} can now be established:

$$t_{ii} = -a_{ii}/(1 + a_{ii}) = -(a_{ii})(1 + a_{ii})^{-1} = O(n^{-2}),$$

$i = 1, \ldots, n$, since $1 + a_{ii} = O(n^0) + O(n^{-2}) = O(n^0)$. That is, $G - I = O(n^{-2})$. Thus,

$$B_{D^{(k)}}(\Omega) = O(n^{-(k+1)}),$$

which leads to

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

Therefore, the order of the bias of the kth-order corrected Qian–Wang estimator is the same as that of the kth-order White estimator of Cribari-Neto et al. (2000); see Sect. 2. (Recall, however, that $k = 1$ here yields the unmodified Qian–Wang estimator, which is itself a correction to White's estimator.)

4 Variance estimation of linear combinations of the elements of $\widehat{\beta}$

Let c be a p-vector of constants such that $c'\widehat{\beta}$ is a linear combination of the elements of $\widehat{\beta}$. Define

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

The kth-order corrected estimator of our sequence of bias corrected estimators, given in (7), is

$$\widehat{V}^{(k)} = \widehat{\Psi}^{(k)}_{QW} = PD^{(k)}P',$$

and hence

$$\widehat{\Phi}^{(k)}_{QW} = c'\widehat{\Psi}^{(k)}_{QW}c = c'PD^{(k)}P'c$$

is the kth-order element of a sequence of bias-adjusted estimators for Φ, where, as before,

$$\begin{aligned} D^{(k)} &= 1_{(k>1)} \times M^{(0)}(\widehat{\Omega}) + 1_{(k>2)} \times \sum_{j=1}^{k-2} (-1)^j M^{(j)}(\widehat{\Omega}) \\ &\quad + \sum_{j=k-1}^{k} (-1)^j M^{(j)}(\widehat{\Omega})G, \end{aligned}$$

$k = 1, 2, \ldots$.

Recall that when $k = 1$ we obtain the Qian–Wang estimator. Using this estimator, we obtain

$$\widehat{\Phi}^{(1)}_{QW} = c'\widehat{\Psi}^{(1)}_{QW}c = c'PD^{(1)}P'c,$$

where

$$D^{(1)} = \widehat{\Omega}G - M^{(1)}(\widehat{\Omega})G = G^{1/2}\widehat{\Omega}G^{1/2} - G^{1/2}M^{(1)}(\widehat{\Omega})G^{1/2}.$$

Let $W = (ww')_d$, where $w = G^{1/2}P'c$. We can now write

$$\begin{aligned}\widehat{\Phi}^{(1)}_{QW} &= c'P[G^{1/2}\widehat{\Omega}G^{1/2} - G^{1/2}M^{(1)}(\widehat{\Omega})G^{1/2}]P'c \\ &= w'\widehat{\Omega}w - w'M^{(1)}(\widehat{\Omega})w.\end{aligned}$$

Note that $w'\widehat{\Omega}w = w'[(\widehat{\varepsilon}\widehat{\varepsilon}')_d]w = \widehat{\varepsilon}'[(ww')_d]\widehat{\varepsilon} = \widehat{\varepsilon}'W\widehat{\varepsilon}$ and that

$$w'M^{(1)}(\widehat{\Omega})w = \sum_{s=1}^{n}\widehat{\alpha}_s w_s^2,$$

where $\widehat{\alpha}_s$ is the sth diagonal element of $M^{(1)}(\widehat{\Omega}) = \{H\widehat{\Omega}(H-2I)\}_d$ and w_s is the sth element of the vector w. Thus,

$$\widehat{\Phi}^{(1)}_{QW} = \widehat{\varepsilon}'W\widehat{\varepsilon} - \sum_{s=1}^{n}\widehat{\alpha}_s w_s^2. \tag{9}$$

Given that

$$\widehat{\alpha}_s = \sum_{t=1}^{n} h_{st}^2\widehat{\varepsilon}_t^2 - 2h_{ss}\widehat{\varepsilon}_s^2, \tag{10}$$

where h_{st} denotes the (s, t) element of H, the summation in (9) can be expanded as

$$\begin{aligned}\sum_{s=1}^{n}\widehat{\alpha}_s w_s^2 &= \sum_{s=1}^{n} w_s^2\widehat{\alpha}_s \\ &= \sum_{s=1}^{n} w_s^2\left(\sum_{t=1}^{n} h_{st}^2\widehat{\varepsilon}_t^2 - 2h_{ss}\widehat{\varepsilon}_s^2\right) \\ &= \sum_{t=1}^{n}\widehat{\varepsilon}_t^2\sum_{s=1}^{n} h_{st}^2 w_s^2 - 2\sum_{t=1}^{n}\widehat{\varepsilon}_t^2 h_{tt} w_t^2 \\ &= \sum_{t=1}^{n}\widehat{\varepsilon}_t^2\widehat{\delta}_t,\end{aligned}$$

where $\widehat{\delta}_t = \sum_{s=1}^{n} h_{st}^2 w_s^2 - 2h_{tt}w_t^2$.

Using (10) and the symmetry of H, it is easy to see that $\widehat{\delta}_t$ is the tth diagonal element of $\{HW(H-2I)\}_d = M^{(1)}(W)$, and thus

$$w'M^{(1)}(\widehat{\Omega})w = \sum_{t=1}^{n}\widehat{\varepsilon}_t{}^2\widehat{\delta}_t = \widehat{\varepsilon}'[M^{(1)}(W)]\widehat{\varepsilon}.$$

Equation (9) can now be written in matrix form as

$$\begin{aligned}\widehat{\Phi}_{QW}^{(1)} &= \widehat{\varepsilon}' W \widehat{\varepsilon} - \widehat{\varepsilon}'[M^{(1)}(W)]\widehat{\varepsilon} \\ &= \widehat{\varepsilon}'[W - M^{(1)}(W)]\widehat{\varepsilon}.\end{aligned}$$

We shall now obtain $\widehat{\Phi}_{QW}^{(2)}$. We have seen that

$$D^{(2)} = \widehat{\Omega} - M^{(1)}(\widehat{\Omega})G + M^{(2)}(\widehat{\Omega})G.$$

Therefore,

$$\begin{aligned}\widehat{\Phi}_{QW}^{(2)} &= c'P[\widehat{\Omega} - M^{(1)}(\widehat{\Omega})G + M^{(2)}(\widehat{\Omega})G]P'c \\ &= c'P\widehat{\Omega}P'c - c'PG^{1/2}M^{(1)}(\widehat{\Omega})G^{1/2}P'c \\ &\quad + c'PG^{1/2}M^{(2)}(\widehat{\Omega})G^{1/2}P'c.\end{aligned}$$

Let $b = P'c$ and $B = (bb')_d$. It then follows that

$$\widehat{\Phi}_{QW}^{(2)} = b'\widehat{\Omega}b - w'M^{(1)}(\widehat{\Omega})w + w'M^{(2)}(\widehat{\Omega})w.$$

Note that

$$b'\widehat{\Omega}b = b'[(\widehat{\varepsilon}\widehat{\varepsilon}')_d]b = \widehat{\varepsilon}'[(bb')_d]\widehat{\varepsilon} = \widehat{\varepsilon}'B\widehat{\varepsilon}.$$

Similarly to the case where $k = 1$, it can be shown that

$$w'M^{(k)}(\widehat{\Omega})w = \widehat{\varepsilon}'M^{(k)}(W)\widehat{\varepsilon}, \quad k = 2, 3, \ldots.$$

Thus,

$$\widehat{\Phi}_{QW}^{(2)} = \widehat{\varepsilon}'[B - M^{(1)}(W) + M^{(2)}(W)]\widehat{\varepsilon}.$$

It can also be shown that

$$\widehat{\Phi}_{QW}^{(3)} = \widehat{\varepsilon}'[B - M^{(1)}(B) + M^{(2)}(W) - M^{(3)}(W)]\widehat{\varepsilon}.$$

More generally,

$$\begin{aligned}\widehat{\Phi}_{QW}^{(k)} &= c'\widehat{\Psi}_{QW}^{(k)}c \\ &= \widehat{\varepsilon}'Q^{(k)}\widehat{\varepsilon}, \quad k = 1, 2, \ldots,\end{aligned} \qquad (11)$$

where $Q^{(k)} = 1_{(k>1)} \times \sum_{j=0}^{k-2}(-1)^j M^{(j)}(B) + \sum_{j=k-1}^{k}(-1)^j M^{(j)}(W)$.

Cribari-Neto et al. (2000) have shown that the HC0 variance estimator of $c'\widehat{\beta}$ is given by

$$\begin{aligned}\widehat{\Phi}_W^{(k)} &= c'\widehat{\Psi}^{(k)}c \\ &= \widehat{\varepsilon}' Q_W^{(k)}\widehat{\varepsilon}, \quad k = 0, 1, 2, \ldots, \end{aligned} \tag{12}$$

where $Q_W^{(k)} = \sum_{j=0}^{k}(-1)^j M^{(j)}(B)$. It is noteworthy that when $G = I$, the Qian–Wang estimator reduces to the one-step bias-adjusted HC0 estimator and, as a consequence, $W = B$ and (11) reduces to (12) for $k \geq 1$.

We shall now write the quadratic form in (11) as a quadratic form in a vector of uncorrelated, mean zero and unit variance random variates.

We have seen in Sect. 2 that $\widehat{\varepsilon} = (I - H)y$. We can then write

$$\begin{aligned}\widehat{\Phi}_{QW}^{(k)} &= \widehat{\varepsilon}' Q^{(k)}\widehat{\varepsilon} \\ &= y'(I - H)Q^{(k)}(I - H)y \\ &= y'\Omega^{-1/2}\Omega^{1/2}(I - H)Q^{(k)}(I - H)\Omega^{1/2}\Omega^{-1/2}y \\ &= z'C_{QW}^{(k)}z, \end{aligned} \tag{13}$$

where $C_{QW}^{(k)} = \Omega^{1/2}(I - H)Q^{(k)}(I - H)\Omega^{1/2}$ is an $n \times n$ symmetric matrix and $z = \Omega^{-1/2}y$ is an n-vector whose mean is $\theta = \Omega^{-1/2}X\beta$ and whose covariance matrix is $\mathrm{cov}(z) = \mathrm{cov}(\Omega^{-1/2}y) = I$.

Note that $\theta' C_{QW}^{(k)} = \beta' X' \Omega^{-1/2}\Omega^{1/2}(I - H)Q^{(k)}(I - H)\Omega^{1/2} = \beta' X'(I - H) Q^{(k)}(I - H)\Omega^{1/2}$. Since $X'(I - H) = 0$, then $\theta' C_{QW}^{(k)} = 0$. Hence, equation (13) can be written as

$$z'C_{QW}^{(k)}z = (z - \theta)'C_{QW}^{(k)}(z - \theta),$$

i.e.,

$$\widehat{\Phi}_{QW}^{(k)} = z'C_{QW}^{(k)}z = a'C_{QW}^{(k)}a,$$

where $a = (z - \theta) = \Omega^{-1/2}(y - X\beta) = \Omega^{-1/2}\varepsilon$, such that $\mathbb{E}(a) = 0$ and $\mathrm{cov}(a) = I$. It then follows that

$$\begin{aligned}\mathrm{var}\left(\widehat{\Phi}_{QW}^{(k)}\right) &= \mathrm{var}\left(a'C_{QW}^{(k)}a\right) \\ &= \mathbb{E}\left[\left(a'C_{QW}^{(k)}a\right)^2\right] - \left[\mathbb{E}(a'C_{QW}^{(k)}a)\right]^2. \end{aligned}$$

(In what follows, we shall write $C_{QW}^{(k)}$ simply as C_{QW} to simplify the notation.)

When the errors are independent, it follows that

$$\mathrm{var}\left(\widehat{\Phi}_{QW}^{(k)}\right) = d'\Lambda d + 2\mathrm{tr}\left(C_{QW}^2\right), \tag{14}$$

where d is a column vector formed out of the diagonal elements of C_{QW}, $\Lambda = \text{diag}(\gamma_1, \ldots, \gamma_n)$, where $\gamma_i = (\mu_{4i} - 3\sigma_i^4)/\sigma_i^4$ is the excess kurtosis of the ith error and $\text{tr}(C_{QW})$ is the trace of C_{QW}. When the errors are independent and normally distributed, $\gamma_i = 0$. Thus, $\Lambda = 0$ and (14) simplifies to

$$\text{var}\left(\widehat{\Phi}_{QW}^{(k)}\right) = \text{var}\left(c'\widehat{\Psi}_{QW}^{(k)}c\right) = 2\text{tr}\left(C_{QW}^2\right).$$

For the sequence of corrected HC0 estimators, one obtains (Cribari-Neto et al. 2000)

$$\text{var}\left(\widehat{\Phi}_{W}^{(k)}\right) = 2\text{tr}\left(C_{W}^2\right),$$

where $C_W = \Omega^{1/2}(I - H)Q_W^{(k)}(I - H)\Omega^{1/2}$.

5 Numerical results

In this section we shall numerically evaluate the effectiveness of the finite-sample corrections to the White (HC0) and Qian–Wang estimators. To that end, we shall use the exact expressions obtained for the biases and for the variances of linear combinations of the elements of $\widehat{\beta}$. We shall also report results on the root mean squared errors and maximal biases of the different estimators.

The model used in the numerical evaluation is

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are uncorrelated with $\mathbb{E}(\varepsilon_i) = 0$ and $\text{var}(\varepsilon_i) = \exp(a x_{i2})$, $i = 1, \ldots, n$. We have used different values of a in order to vary the strength of heteroskedasticity, which we measure as $\lambda = \max\{\sigma_i^2\}/\min\{\sigma_i^2\}$, $i = 1, \ldots, n$. The sample sizes considered were $n = 20, 40, 60$. For $n = 20$, the covariates values x_{i2} and x_{i3} were obtained as random draws from the following distributions: standard uniform $\mathcal{U}(0, 1)$ and standard lognormal $\text{LN}(0, 1)$; under the latter design the data contain leverage points. These twenty covariates values were replicated two and three times when the sample sizes were 40 and 60, respectively. This was done so that the degree of heteroskedasticity (λ) would not change with n.

Table 1 presents the maximal leverages ($h_{\max}$) for the two regression designs used in the simulations (i.e., values of the covariates selected as random draws from the uniform and lognormal distributions). The threshold values commonly used to identify leverage points ($2p/n$ and $3p/n$) are also presented. It is noteworthy that the data contain observations with very high leverage when the values of the regressor are selected as random lognormal draws.

Table 2 presents the total relative bias of the OLS variance estimator. [The bias of the OLS covariance matrix estimator is given by $(n-p)^{-1}\text{tr}\{\Omega(I-H)\}(X'X)^{-1} - P\Omega P'$; here, $p = 3$.] Total relative bias is defined as the sum of the absolute values of the individual relative biases; relative bias is the difference between the estimated

Table 1 Maximal leverages for the two regression designs

n	$\mathcal{U}(0,1)$	LN(0, 1)	Threshold	
	$h_{\max}$	$h_{\max}$	$2p/n$	$3p/n$
20	0.288	0.625	0.300	0.450
40	0.144	0.312	0.150	0.225
60	0.096	0.208	0.100	0.150

Table 2 Total relative bias of the OLS variance estimator

n	$\mathcal{U}(0,1)$			LN(0, 1)		
	$\lambda = 1$	$\lambda \approx 9$	$\lambda \approx 49$	$\lambda = 1$	$\lambda \approx 9$	$\lambda \approx 49$
20	0.000	1.143	1.810	0.000	1.528	3.129
40	0.000	1.139	1.811	0.000	1.581	3.332
60	0.000	1.138	1.811	0.000	1.597	3.393

variance of $\widehat{\beta}_j$ and the corresponding true variance divided by the latter, $j = 1, 2, 3$. As expected, the OLS variance estimator is unbiased under homoskedasticity ($\lambda = 1$), and becomes more biased as heteroskedasticity becomes stronger; also, this estimator is more biased when the regression design includes leverage points. Note that the biases do not vanish as the sample size increases; indeed, they remain approximately constant across different sample sizes.

Table 3 contains the total relative biases of HC0, its first four bias-corrected counterparts (HC01, HC02, HC03 and HC04), the Qian–Wang estimator ($\widehat{V}^{(1)}$) and the first four corresponding bias-adjusted estimators ($\widehat{V}1^{(1)}$, $\widehat{V}2^{(1)}$, $\widehat{V}3^{(1)}$ and $\widehat{V}4^{(1)}$). Note that, following the notation used in Sects. 2 and 3, HC04 and $\widehat{V}4^{(1)}$, for example, correspond to $\widehat{\Psi}^{(4)}$ and $\widehat{V}^{(5)}$, respectively. First, note that the Qian–Wang estimator is unbiased when all errors share the same variance ($\lambda = 1$). Additionally, we note that our corrections to this estimator can be effective under heteroskedastic errors, even though it did not behave well under homoskedasticity with unbalanced regression design (leverage points in the data, values of the covariates obtained as random lognormal draws) and small sample size ($n = 20$). Consider, for instance, the situation where the regression design is unbalanced, $n = 20$, $\lambda \approx 9$ ($\lambda \approx 49$). The total relative bias of the Qian–Wang estimator exceeds 22% (exceeds 44%), whereas the fourth-order bias-corrected estimator has total relative bias of 5.1% (less than 2%). In particular, when heteroskedasticity is strong ($\lambda \approx 49$), the bias adjustment achieves a reduction in the total relative bias of over 23 times. This is certainly a sizeable improvement. It is also noteworthy that the bias corrected Qian–Wang estimators outperform the corresponding HC0 bias-corrected estimators.

Table 4 contains the square roots of the total relative mean squared errors, which are defined as the sums of the individual mean squared errors standardized by the corresponding true variances. First, note that the figures for the Qian–Wang estimator are slightly larger than those for the HC0 estimator. Second, it is noteworthy that the total relative root mean squared errors of the corrected Qian–Wang estimators

Table 3 Total relative biases

Covariates	n	λ	HC0	HC01	HC02	HC03	HC04	$\widehat{V}^{(1)}$	$\widehat{V}1^{(1)}$	$\widehat{V}2^{(1)}$	$\widehat{V}3^{(1)}$	$\widehat{V}4^{(1)}$
$\mathcal{U}(0, 1)$	20	1	0.551	0.124	0.035	0.012	0.005	0.000	0.007	0.004	0.002	0.001
		≈ 9	0.478	0.082	0.013	0.001	0.001	0.033	0.006	0.002	0.001	0.000
		≈ 49	0.464	0.073	0.009	0.002	0.002	0.044	0.009	0.003	0.002	0.001
	40	1	0.276	0.031	0.004	0.001	0.000	0.000	0.001	0.000	0.000	0.000
		≈ 9	0.239	0.020	0.002	0.000	0.000	0.007	0.001	0.000	0.000	0.000
		≈ 49	0.232	0.018	0.001	0.000	0.000	0.010	0.001	0.000	0.000	0.000
	60	1	0.184	0.014	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.159	0.009	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000
		≈ 49	0.155	0.008	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
LN(0, 1)	20	1	0.801	0.415	0.305	0.252	0.215	0.000	0.155	0.156	0.139	0.122
		≈ 9	0.733	0.289	0.166	0.118	0.094	0.222	0.049	0.066	0.059	0.051
		≈ 49	0.601	0.260	0.132	0.071	0.043	0.443	0.100	0.038	0.024	0.019
	40	1	0.401	0.104	0.038	0.016	0.007	0.000	0.010	0.005	0.002	0.001
		≈ 9	0.366	0.072	0.021	0.007	0.003	0.034	0.004	0.002	0.001	0.000
		≈ 49	0.301	0.065	0.016	0.004	0.001	0.069	0.009	0.002	0.000	0.000
	60	1	0.267	0.046	0.011	0.003	0.001	0.000	0.003	0.001	0.000	0.000
		≈ 9	0.244	0.032	0.006	0.001	0.000	0.014	0.001	0.000	0.000	0.000
		≈ 49	0.200	0.029	0.005	0.001	0.000	0.029	0.002	0.000	0.000	0.000

The values of the covariates were selected as random uniform and lognormal draws

are approximately equal to those of the corresponding corrected HC0 estimators, especially when $n = 40, 60$. Third, the total relative root mean squared errors are larger when the values of the covariates were selected as random uniform draws, since the variances are considerably larger when the data contain no influential point. Fourth, it is noteworthy that bias correction leads to variance inflation and even to slight increase in the mean squared error, which is true for the corrected estimators we propose and also for those proposed by Cribari-Neto et al. (2000).

We shall now determine the linear combination of the regression parameter estimators that yields the maximal estimated variance bias, i.e., we shall find the p-vector c (normalized such that $c'c = 1$) that maximizes $\mathbb{E}[\widehat{\text{var}}(c'\widehat{\beta})] - \text{var}(c'\widehat{\beta})$. In order for negative biases not to offset positive ones, we shall work with matrices of absolute biases. Since such matrices are symmetric, the maximum value of the bias of the estimated variances of linear combinations of the $\widehat{\beta}$'s is given by the maximal eigenvalues of the corresponding (absolute) bias matrices. Recall that if A is a symmetric matrix, then $\max_c c'Ac/c'c$ equals the largest eigenvalue of A; see, e.g., Rao (1973, p. 62). The results are presented in Table 5. The figures in this table reveal that the sequence of corrections we propose to improve the finite-sample performance of the Qian–Wang estimator can be quite effective in some cases. For instance, when $n = 20$, $\lambda \approx 49$ and the covariate values were selected as random uniform draws, the maximal bias of the Qian–Wang estimator is reduced from 0.285 to 0.012 after four iterations of our bias adjusting scheme; i.e., there is a reduction in bias of nearly 24 times (the reduction

Table 4 Square roots of the total relative mean squared errors

Covariates	n	λ	HC0	HC01	HC02	HC03	HC04	$\widehat{V}^{(1)}$	$\widehat{V}1^{(1)}$	$\widehat{V}2^{(1)}$	$\widehat{V}3^{(1)}$	$\widehat{V}4^{(1)}$
$\mathcal{U}(0, 1)$	20	1	0.540	0.612	0.649	0.664	0.670	0.647	0.662	0.669	0.672	0.674
		≈ 9	1.173	1.348	1.413	1.433	1.438	1.408	1.429	1.437	1.440	1.440
		≈ 49	2.621	3.065	3.212	3.253	3.264	3.200	3.244	3.260	3.265	3.266
	40	1	0.277	0.299	0.305	0.306	0.306	0.303	0.306	0.306	0.306	0.306
		≈ 9	0.612	0.663	0.672	0.673	0.673	0.670	0.673	0.673	0.673	0.673
		≈ 49	1.406	1.537	1.558	1.561	1.561	1.554	1.560	1.561	1.561	1.561
	60	1	0.186	0.197	0.198	0.199	0.199	0.198	0.199	0.199	0.199	0.199
		≈ 9	0.413	0.437	0.440	0.440	0.440	0.439	0.440	0.440	0.440	0.440
		≈ 49	0.957	1.019	1.025	1.026	1.026	1.023	1.025	1.026	1.026	1.026
LN(0, 1)	20	1	0.269	0.297	0.321	0.341	0.360	0.361	0.377	0.394	0.409	0.422
		≈ 9	0.496	0.585	0.642	0.676	0.701	0.690	0.712	0.734	0.752	0.767
		≈ 49	1.177	1.435	1.573	1.643	1.682	1.638	1.676	1.711	1.736	1.753
	40	1	0.142	0.157	0.164	0.167	0.168	0.165	0.167	0.168	0.169	0.169
		≈ 9	0.263	0.293	0.302	0.305	0.306	0.302	0.305	0.306	0.306	0.306
		≈ 49	0.641	0.723	0.744	0.749	0.750	0.743	0.749	0.750	0.751	0.751
	60	1	0.096	0.104	0.106	0.107	0.107	0.106	0.107	0.107	0.107	0.107
		≈ 9	0.178	0.193	0.196	0.196	0.196	0.195	0.196	0.196	0.196	0.196
		≈ 49	0.438	0.477	0.484	0.485	0.485	0.483	0.485	0.485	0.485	0.485

The values of the covariates were selected as random uniform and lognormal draws

is of almost 22 times when the covariate values are selected as random lognormal draws). Note that these figures are *not* relative. The corrections to the HC0 estimator proposed by Cribari-Neto et al. (2000) also prove effective.

We now consider the simple regression model $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, \ldots, n$, where the errors have mean zero, are uncorrelated, and each ε_i has variance $\sigma_i^2 = \exp\{a x_i\}$. The covariate values are n equally spaced points between zero and one. The sample size is set at $n = 40$. We gradually increase the last covariate value (x_{40}) so as to get increased maximal leverages. The maximal biases were computed as in the previous table, and the results are presented in Table 6. First, note that the maximal biases of HC0 are considerably more pronounced than those of the Qian–Wang estimator under heteroskedasticity and increased maximal leverages. For instance, under the strongest level of heteroskedasticity ($\lambda \approx 49$) and $h_{\max} = 0.289$, the maximal biases of these two estimators are 1.587 and 0.356, respectively. Second, note that corrections proposed in this paper can be quite effective under unequal error variances. As an illustration, consider again the setting under strongest heteroskedasticity and maximal leverage of almost twice the threshold value $3p/n = 0.150$. The bias of the Qian–Wang estimator shrinks from 0.356 to 0.024 after four iterations of our bias correcting scheme, which amounts to a bias reduction of nearly 15 times.

Our focus lies in obtaining accurate (nearly unbiased) point estimates of variances and covariances of OLSEs. As noted by a referee, however, such estimates are oftentimes used for performing inferences on the regression parameters. We have run a

Table 5 Maximal biases

Covariates	n	λ	HC0	HC01	HC02	HC03	HC04	$\widehat{V}^{(1)}$	$\widehat{V}1^{(1)}$	$\widehat{V}2^{(1)}$	$\widehat{V}3^{(1)}$	$\widehat{V}4^{(1)}$
$\mathcal{U}(0, 1)$	20	1	0.208	0.052	0.016	0.006	0.003	0.000	0.004	0.002	0.001	0.001
		≈ 9	0.775	0.142	0.025	0.007	0.004	0.057	0.014	0.006	0.003	0.002
		≈ 49	2.848	0.443	0.100	0.048	0.034	0.285	0.093	0.045	0.024	0.012
	40	1	0.052	0.006	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.194	0.018	0.002	0.000	0.000	0.006	0.001	0.000	0.000	0.000
		≈ 49	0.712	0.055	0.006	0.001	0.001	0.032	0.005	0.001	0.000	0.000
	60	1	0.023	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.086	0.005	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000
		≈ 49	0.316	0.016	0.001	0.000	0.000	0.009	0.001	0.000	0.000	0.000
LN(0,1)	20	1	0.074	0.034	0.025	0.020	0.018	0.000	0.013	0.013	0.011	0.010
		≈ 9	0.196	0.067	0.030	0.018	0.014	0.028	0.009	0.009	0.008	0.007
		≈ 49	0.892	0.323	0.128	0.053	0.022	0.131	0.047	0.021	0.010	0.006
	40	1	0.018	0.004	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.049	0.008	0.002	0.001	0.000	0.002	0.000	0.000	0.000	0.000
		≈ 49	0.223	0.040	0.008	0.002	0.000	0.012	0.003	0.001	0.000	0.000
	60	1	0.008	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.022	0.002	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
		≈ 49	0.099	0.012	0.002	0.000	0.000	0.003	0.000	0.000	0.000	0.000

The values of the covariates were selected as random uniform and lognormal draws

small Monte Carlo experiment in order to evaluate the finite sample performance of quasi-t tests based on the HC0 and Qian–Wang estimators and also on their corrected versions up to four iterations of the bias correcting schemes. The regression model is $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$, $i = 1, \ldots, n$. The interest lies in testing $\mathcal{H}_0 : \beta_3 = 0$ against $\mathcal{H}_1 : \beta_3 \neq 0$. The errors are independent and normally distributed with mean zero and variance $\sigma_i^2 = \exp(a x_{i2})$. The covariate values were obtained as random draws from the t_3 distribution; there are leverage points, $n = 20$, $\lambda \approx 49$ and the number of Monte Carlo replications was 10,000. Here, $h_{\max} = 5.66p/n$, so there is an observation with very high leverage. The null rejection rates at the 5% nominal level of the HC0 test and of the tests based on standard errors obtained from the corrected HC0 estimators (one, two, there and four iterations of the bias correcting scheme) were, respectively, 17.46, 16.20, 18.31, 18.71 and 15.97%; the corresponding figures for the Qian–Wang test and the four tests based on the corrected Qian–Wang estimators were 11.66, 7.07, 6.44, 5.87 and 5.71%. The tests based on the corrected Qian–Wang estimators were also less size-distorted than the Qian–Wang test when $\lambda = 1$ (15.28% for the Qian–Wang test and 8.35, 7.59, 7.04 and 6.58% for the tests based on the corrected standard errors) and $\lambda \approx 9$ (Qian–Wang: 12.50%; corrected: 6.86, 6.25, 5.93 and 5.60%). We thus notice that the finite sample corrections we propose may yield more accurate hypothesis testing inference in addition to more accurate point estimates. Even though we do not present all Monte Carlo results to save space,

Table 6 Maximal biases, $n = 40$

λ	h_{max}	HC0	HC01	HC02	HC03	HC04	$\widehat{V}^{(1)}$	$\widehat{V}1^{(1)}$	$\widehat{V}2^{(1)}$	$\widehat{V}3^{(1)}$	$\widehat{V}4^{(1)}$
1	0.096	0.025	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.154	0.026	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.220	0.028	0.005	0.002	0.001	0.000	0.000	0.001	0.000	0.000	0.000
	0.289	0.033	0.010	0.004	0.002	0.001	0.000	0.002	0.001	0.000	0.000
	0.357	0.039	0.016	0.009	0.005	0.003	0.000	0.004	0.003	0.002	0.001
	0.422	0.044	0.023	0.015	0.010	0.007	0.000	0.007	0.005	0.004	0.002
	0.482	0.049	0.030	0.022	0.016	0.012	0.000	0.011	0.009	0.006	0.005
≈ 9	0.096	0.109	0.011	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	0.154	0.126	0.023	0.005	0.001	0.000	0.007	0.002	0.001	0.000	0.000
	0.220	0.197	0.063	0.024	0.009	0.004	0.024	0.012	0.005	0.002	0.001
	0.289	0.296	0.133	0.065	0.033	0.016	0.055	0.033	0.017	0.008	0.004
	0.357	0.405	0.226	0.133	0.079	0.047	0.098	0.068	0.041	0.024	0.014
	0.422	0.495	0.320	0.214	0.143	0.096	0.144	0.113	0.077	0.052	0.035
	0.482	0.571	0.410	0.301	0.221	0.163	0.192	0.165	0.123	0.090	0.066
≈ 49	0.096	0.482	0.052	0.006	0.001	0.000	0.012	0.002	0.000	0.000	0.000
	0.154	0.588	0.126	0.033	0.009	0.003	0.049	0.015	0.004	0.001	0.000
	0.220	1.006	0.358	0.139	0.055	0.022	0.160	0.069	0.028	0.011	0.004
	0.289	1.587	0.757	0.376	0.187	0.094	0.356	0.191	0.097	0.048	0.024
	0.357	2.170	1.255	0.741	0.439	0.260	0.611	0.386	0.230	0.136	0.081
	0.422	2.700	1.787	1.197	0.803	0.539	0.901	0.640	0.432	0.290	0.194
	0.482	3.117	2.274	1.671	1.230	0.905	1.189	0.922	0.681	0.501	0.369

Single regression model with covariate values chosen as a sequence of equally spaced points in the standard unit interval. The last point is gradually increased in order for the maximal leverage to increase; here, $3p/n = 0.150$

we note that the tests based on the Qian–Wang estimator and its corrected versions displayed similar behavior for larger sample sizes (40 observations or more).

We have also performed simulations in which the wild bootstrap was used to obtain a critical value for the HC3-based quasi-t test statistic. As suggested by Flachaire (2005), resampling in the wild bootstrap scheme was performing using the Rademacher population. The number of Monte Carlo replications was 5,000 and there were 500 bootstrap replicates for each Monte Carlo sample. The null rejection rates at the 5% nominal level for $n = 20$, covariate values obtained as t_3 random draws and $\lambda = 1$, $\lambda \approx 9$ and $\lambda \approx 49$ were 17.62, 14.76 and 11.34%, respectively. We noticed that the wild bootstrap worked well in the balanced case (no leverage point in the data) for all sample sizes. In the unbalanced case (leveraged data), it only yielded satisfactory results for $n \geq 60$. The wild bootstrap performed considerably better under less extreme leverage, e.g., when $h_{max} < 4p/n$.

It is interesting to note that in the design used in the simulations, all tests performed better under stronger heteroskedasticity. We note that there is a reduction in the variance and in the kurtosis of the test statistics as λ increases.

Table 7 Standard errors, first application

Case	OLS	HC0	HC01	HC02	HC03	HC04	$\widehat{V}^{(1)}$	$\widehat{V}1^{(1)}$	$\widehat{V}2^{(1)}$	$\widehat{V}3^{(1)}$	$\widehat{V}4^{(1)}$
1											
$\widehat{\beta}_1$	327.29	460.89	551.94	603.90	641.57	672.03	741.35	722.21	730.28	745.04	760.64
$\widehat{\beta}_2$	828.99	1,243.04	1,495.05	1, 638.07	1,741.22	1,824.42	2,011.74	1,960.72	1,983.10	2,023.45	2,066.01
$\widehat{\beta}_3$	519.08	829.99	1,001.78	1,098.54	1,167.94	1,223.77	1,348.36	1,314.92	1,330.15	1,357.25	1,385.77
2											
$\widehat{\beta}_1$	405.22	345.73	381.36	404.39	422.51	436.99	454.51	445.82	453.91	461.93	468.58
$\widehat{\beta}_2$	1,064.0	936.92	1,039.39	1,104.93	1,156.01	1,196.63	1,243.19	1,220.43	1,243.39	1,265.96	1,284.65
$\widehat{\beta}_3$	691.32	626.68	699.16	745.03	780.48	808.55	839.28	824.47	840.49	856.12	869.04
3											
$\widehat{\beta}_1$	529.15	505.34	529.71	532.04	531.57	530.95	535.68	531.74	530.96	530.55	530.31
$\widehat{\beta}_2$	1,419.9	1,394.09	1,465.84	1,473.92	1,473.28	1,471.89	1,482.49	1,473.60	1,471.90	1,470.92	1,470.34
$\widehat{\beta}_3$	942.71	949.41	1,001.46	1,008.06	1,008.04	1,007.28	1,013.03	1,008.16	1,007.27	1,006.71	1,006.36
4											
$\widehat{\beta}_1$	619.28	625.87	660.52	666.34	667.47	667.66	667.20	667.45	667.65	667.67	667.65
$\widehat{\beta}_2$	1,647.6	1,699.02	1,797.21	1,814.12	1,817.45	1,818.01	1,816.07	1,817.34	1,817.98	1,818.05	1,818.00
$\widehat{\beta}_3$	1,085.1	1,140.63	1,209.57	1,221.72	1,224.14	1,224.56	1,222.82	1,224.02	1,224.53	1,224.59	1,224.56

6 Empirical illustrations

In what follows we shall present two empirical applications that use real data. In the first application, the dependent variable (y) is per capita spending on public schools and the independent variables, x and x^2, are per capita income by state in 1979 in the United States and its square; income is scaled by 10^{-4}. Wisconsin was not considered since it had missing data, and Washington D.C. was included. The data are presented in Greene (1997, Table 12.1, p. 541) and their original source is the US Department of Commerce. The regression model is

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i, \quad i = 1, \ldots, 50.$$

The ordinary least squares estimates for the linear parameters are $\widehat{\beta}_1 = 832.91$, $\widehat{\beta}_2 = -1{,}834.20$ and $\widehat{\beta}_3 = -1{,}587.04$. The Breusch–Pagan–Godfrey test of homoskedasticity rejects this hypothesis at the 1% nominal level, thus indicating that there is heteroskedasticity in the data. It should be noted that the data contain three leverage points, namely: Alaska, Mississippi and Washington, D.C. (their leverage measures are 0.651, 0.200 and 0.208, respectively; note that $3p/n = 0.180$).

In Table 7 we present the standard errors for the regression parameter estimates. We consider four designs: (i) *case 1*: all 50 observations were used; (ii) *case 2*: Alaska (the strongest leverage point) was removed from the data ($n = 49$); (iii) *case 3*: Alaska and Washington D.C. were removed from the data ($n = 48$); (iv) *case 4*: Alaska, Washington D.C. and Mississippi were removed from data ($n = 47$). Table 8 contains information on the detection of leverage points in these four situations.

Table 8 Leverage measures, thresholds for detecting leverage points and ratio between $h_{\max}$ and $3p/n$; first application

Case	n	h_{ii}	$2p/n$	$3p/n$	$h_{\max}/(3p/n)$
1	50	0.651 ($h_{\max}$)	0.12	0.18	3.62
		0.208			
		0.200			
2	49	0.562 ($h_{\max}$)	0.122	0.184	3.05
		0.250			
3	48	0.312 ($h_{\max}$)	0.125	0.187	1.67
		0.197			
4	47	0.209 ($h_{\max}$)	0.128	0.191	1.09

The figures in Table 7 reveal that when all 50 observations are used (case 1, three leverage points in the data), the HC0 standard errors are considerably smaller than the Qian–Wang standard errors; the same pattern holds for their corresponding bias-adjusted versions. For instance, the standard errors of $\widehat{\beta}_3$ are 829.99 (HC0) and 1,348.36 (Qian–Wang). The same discrepancy holds for case 2, i.e., when Alaska is removed from the data. The HC0 and Qian–Wang standard errors, however, are somewhat similar in cases 3 (Alaska and Washington D.C. are not in the data) and 4 (Alaska, Mississippi and Washington D.C. are not in the data). In these cases, the ratios between $h_{\max}$ and $3p/n$ are smaller than 2. It is also noteworthy that in case 4 the fourth-order corrected HC0 and Qian–Wang standard errors are nearly equal.

It is particularly interesting to note that a scatterplot shows a satisfactorily linear scatter except for a single high leverage point: Alaska. The HC0 standard error of $\widehat{\beta}_3$ equals 829.99 when the sample contains all 50 observations ($\widehat{\beta}_3 = 1{,}587.04$); this standard error is highly biased in favor of the quadratic model specification. The Qian–Wang standard error equals 1,348.36, thus indicating greater uncertainty relative to the possible nonlinear effect of x_t on $\mathbb{E}(y_t)$. Our fourth-order corrected estimate is even greater: 1,385.77.

The data for the second application were obtained from Cagan (1974, Table 1, p. 4). The dependent variable (y) is the percent rate of change in stock prices (% per year) and the independent variable (x) is the percent rate of change in consumer prices (% per year). There are observations for 20 countries ($n = 20$) in the period that goes from the post-World War II through 1969. The regression model is

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, \ldots, n.$$

Table 9 contains information on leverage points. Note that the data contain a strong leverage point, namely: Chile ($h_{\text{Chile}} = 0.931$). When such an observation is removed from the data ($n = 19$), a new leverage point emerges (Israel). The data become well balanced when these two observations are removed from the sample ($n = 18$).

Table 10 presents the standard errors of the two regression parameter estimates. Case 1 corresponds to the complete dataset ($n = 20$), case 2 relates to the situation where Chile (the first leverage point) is not in the data ($n = 19$), and case 3 corresponds to the well balanced design ($n = 18$). When all 20 observations are used, the HC0

Table 9 Leverage measures, thresholds for detecting leverage points and ratio between $h_{\max}$ and $3p/n$; second application

Case	n	$h_{\max}$	$2p/n$	$3p/n$	$h_{\max}/(3p/n)$
1	20	0.931	0.200	0.300	3.10
2	19	0.559	0.210	0.316	1.77
3	18	0.225	0.220	0.330	0.68

Table 10 Standard errors; second application

Case	OLS	HC0	HC01	HC02	HC03	HC04	$\widehat{V}^{(1)}$	$\widehat{V}1^{(1)}$	$\widehat{V}2^{(1)}$	$\widehat{V}3^{(1)}$	$\widehat{V}4^{(1)}$
1											
$\widehat{\beta}_1$	1.09	0.95	0.99	0.99	0.99	0.99	1.14	1.04	1.03	1.04	1.04
$\widehat{\beta}_2$	0.15	0.07	0.07	0.07	0.07	0.07	0.16	0.11	0.10	0.10	0.10
2											
$\widehat{\beta}_1$	2.38	2.00	2.03	1.94	1.83	1.74	1.94	1.72	1.63	1.56	1.50
$\widehat{\beta}_2$	0.55	0.42	0.40	0.36	0.31	0.26	0.37	0.26	0.20	0.16	0.10
3											
$\widehat{\beta}_1$	3.31	3.41	3.74	3.81	3.83	3.83	3.82	3.83	3.83	3.83	3.83
$\widehat{\beta}_2$	0.84	0.87	0.95	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97

standard errors are again considerably smaller than the Qian–Wang ones. For instance, the HC0 standard error of $\widehat{\beta}_2$ is 0.07 whereas the Qian–Wang counterpart equals 0.16 (note that the latter is more than twice the former); the discrepancy is smaller when their fourth-order corrected versions are used (0.07 and 0.10, respectively). The discrepancies between the HC0 and Qian–Wang standard errors are reduced in cases 2 ($n = 19$) and 3 ($n = 18$). Finally, note that all four corrected Qian–Wang standard errors are equal when the data are well balanced and that they agree with the HC0 bias corrected standard errors.

7 A generalization of the Qian–Wang estimator

In this section, we shall show that the Qian–Wang estimator can be obtained by bias-correcting HC0 and then modifying the adjusted estimator so that it becomes unbiased under equal error variances. We shall also show that this approach can be applied to the variants of HC0 introduced in Sect. 2. It then follows that all of the results we have derived can be easily extended to cover modified versions of variants of HC0.

At the outset, note that Halbert White's HC0 estimator can be written as HC0 $= \widehat{\Psi}_0 = P\widehat{\Omega}_0 P' = PD_0\widehat{\Omega}P'$, where $D_0 = I$. In Sect. 2 we have presented some variants of HC0, namely:

(i) HC1 $= \widehat{\Psi}_1 = P\widehat{\Omega}_1 P' = PD_1\widehat{\Omega}P'$, $D_1 = (n/(n-p))I$;
(ii) HC2 $= \widehat{\Psi}_2 = P\widehat{\Omega}_2 P' = PD_2\widehat{\Omega}P'$, $D_2 = \mathrm{diag}\{1/(1-h_i)\}$;
(iii) HC3 $= \widehat{\Psi}_3 = P\widehat{\Omega}_3 P' = PD_3\widehat{\Omega}P'$, $D_3 = \mathrm{diag}\{1/(1-h_i)^2\}$;

(iv) HC4 $= \widehat{\Psi}_4 = P\widehat{\Omega}_4 P' = PD_4\widehat{\Omega}P'$, $D_4 = \text{diag}\{1/(1-h_i)^{\delta_i}\}$ and $\delta_i = \min\{4, nh_i/p\}$.

In what follows, we shall denote these estimators as HCi, $i = 0, 1, 2, 3, 4$.

We have shown that

$$\mathbb{E}(\widehat{\Omega}) = \{(I-H)\Omega(I-H)\}_d = M^{(1)}(\Omega) + \Omega.$$

Note that

$$\mathbb{E}(\widehat{\Omega}_i) = \mathbb{E}(D_i\widehat{\Omega}) = D_i\mathbb{E}(\widehat{\Omega}) = D_i M^{(1)}(\Omega) + D_i\Omega$$

and

$$B_{\widehat{\Omega}_i}(\Omega) = \mathbb{E}(\widehat{\Omega}_i) - \Omega = D_i M^{(1)}(\Omega) + (D_i - I)\Omega.$$

As we have done in Sect. 2, we can write $\widehat{\Psi}_i^{(1)} = P\widehat{\Omega}_i^{(1)}P'$, where

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

Recall that $\mathbb{E}(\widehat{\Omega} - \Omega) = M^{(1)}(\Omega)$ and that $M^{(1)}(M^{(1)}(\Omega)) = M^{(2)}(\Omega)$. Thus,

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

When $\Omega = \sigma^2 I$ (homoskedasticity), it follows that

$$\begin{aligned}M^{(1)}(\sigma^2 I) &= \{H\sigma^2 I(H-2I)\}_d\\ &= \sigma^2\{-H\}_d\\ &= -\sigma^2 K.\end{aligned}$$

Additionally,

$$\begin{aligned}M^{(2)}(\sigma^2 I) &= M^{(1)}(M^{(1)}(\sigma^2 I))\\ &= \{H(-\sigma^2 K)(H-2I)\}_d\\ &= \sigma^2\{-HKH + 2KK\}_d.\end{aligned}$$

(Note that we have used the fact that H is idempotent, that $K = (H)_d$ and that $(HK)_d = (KK)_d$.) Therefore, under homoskedasticity,

$$\begin{aligned}\mathbb{E}(\widehat{\Omega}_i^{(1)}) &= -\sigma^2 K + D_i \sigma^2 K + \sigma^2 I - \sigma^2 D_i\{-HKH + 2KK\}_d \\ &= \sigma^2[(I - K) + D_i\{K + HKH - 2KK\}_d] \\ &= \sigma^2 A_i,\end{aligned}$$

where $A_i = (I - K) + D_i\{K + HKH - 2KK\}_d$. We shall now obtain the expected value of $\widehat{\Psi}_i^{(1)}$ when $\Omega = \sigma^2 I$ (homoskedastic errors):

$$\begin{aligned}\mathbb{E}\left(\widehat{\Psi}_i^{(1)}\right) &= \mathbb{E}\left(P\widehat{\Omega}_i^{(1)}P'\right) \\ &= \sigma^2 P A_i P'.\end{aligned}$$

Hence, the estimator

$$\widehat{\Psi}_{iA}^{(1)} = P\widehat{\Omega}_{iA}^{(1)}P' = P\widehat{\Omega}_i^{(1)}A_i^{-1}P'$$

is unbiased:

$$\begin{aligned}\mathbb{E}(\widehat{\Psi}_{iA}^{(1)}) &= \mathbb{E}\left(P\widehat{\Omega}_i^{(1)}A_i^{-1}P'\right) \\ &= P\sigma^2 A_i A_i^{-1} P' \\ &= P\sigma^2 I P' \\ &= P\Omega P' \\ &= \Psi.\end{aligned}$$

It is noteworthy that the Qian–Wang estimator given in Sect. 3 is a particular case of $\widehat{\Psi}_{iA}^{(1)}$ when $i = 0$, i.e., when $D_0 = I$. Indeed, note that

$$\widehat{\Psi}_{0A}^{(1)} = P\widehat{\Omega}_0^{(1)}A_0^{-1}P' = PD^{(1)}P' = \widehat{V}^{(1)},$$

where $\widehat{\Omega}_0^{(1)} = \widehat{\Omega} - M^{(1)}(\widehat{\Omega})$ and $A_0 = \{I + HKH - 2KK\}_d$. [In Sect. 2, $\widehat{\Omega}_0^{(1)}$ was denoted as $\widehat{\Omega}^{(1)}$.]

We shall now derive the bias of $\widehat{\Psi}_{iA}^{(1)}$ under heteroskedasticity. Note that

$$\begin{aligned}B_{\widehat{\Omega}_{iA}^{(1)}}(\Omega) &= \mathbb{E}\left(\widehat{\Omega}_{iA}^{(1)}\right) - \Omega \\ &= \left[M^{(1)}(\Omega) - D_i M^{(1)}(\Omega) + \Omega - D_i M^{(2)}(\Omega)\right] A_i^{-1} - \Omega \\ &= \Omega(A_i^{-1} - I) + (I - D_i)M^{(1)}(\Omega)A_i^{-1} - D_i M^{(2)}(\Omega)A_i^{-1}.\end{aligned}$$

Hence,

$$\begin{aligned} B_{\widehat{\Psi}^{(1)}_{iA}}(\Omega) &= \mathbb{E}(\widehat{\Psi}^{(1)}_{iA}) - \Psi \\ &= P\left[B_{\widehat{\Omega}^{(1)}_{iA}}(\Omega)\right]P'. \end{aligned}$$

This is a closed-form expression for the bias of the class of estimators we have considered in this section. In particular, it can be used to further bias correct the estimators. Indeed, it is important to note that all of the results in Sects. 3 and 4 can be easily extended to the more general class of estimators considered here.

We shall obtain a sequence of estimators adjusted for bias starting from the modified estimator

$$\widehat{\Psi}^{(1)}_{iA} = P\widehat{\Omega}^{(1)}_{iA}P' = P\widehat{\Omega}^{(1)}_{i}A_i^{-1}P',$$

for $i = 1, \ldots, 4$. (The case $i = 0$ was already addressed when we bias corrected the Qian–Wang estimator. Note that the results presented below agree with the ones obtained for $\widehat{V}^{(1)}$ when we let $D_0 = I$.) Let $G_i = A_i^{-1}$.

The one iteration bias-adjusted estimator is

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

Its bias can be expressed as

$$\begin{aligned} B_{\widehat{\Omega}^{(2)}_{iA}}(\widehat{\Omega}) &= \mathbb{E}(\widehat{\Omega}^{(2)}_{iA}) - \Omega \\ &= \mathbb{E}(\widehat{\Omega} - M^{(1)}(\widehat{\Omega})G_i + D_i M^{(2)}(\widehat{\Omega})G_i) - \Omega \\ &= \mathbb{E}(\widehat{\Omega} - \Omega) - \mathbb{E}(M^{(1)}(\widehat{\Omega}) - M^{(1)}(\Omega))G_i - M^{(1)}(\Omega)G_i \\ &\quad + D_i\mathbb{E}(M^{(2)}(\widehat{\Omega}) - M^{(2)}(\Omega))G_i + D_i M^{(2)}(\Omega)G_i \\ &= -M^{(1)}(\Omega)(G_i - I) - (I - D_i)M^{(2)}(\Omega)G_i + D_i M^{(3)}(\Omega)G_i. \end{aligned}$$

After k iterations of the bias correcting scheme we obtain

$$\begin{aligned} \widehat{\Omega}^{(k)}_{iA} &= 1_{(k>1)} \times M^{(0)}(\widehat{\Omega}) + 1_{(k>2)} \times \sum_{j=1}^{k-2}(-1)^j M^{(j)}(\widehat{\Omega}) \\ &\quad + (-1)^{k-1}M^{(k-1)}(\widehat{\Omega})G_i + (-1)^k D_i M^{(k)}(\widehat{\Omega})G_i, \end{aligned}$$

$k = 1, 2, \ldots$ The bias of this estimator is given by

$$B_{\widehat{\Omega}_{iA}^{(k)}}(\Omega) = (-1)^{k-1} M^{(k-1)}(\Omega)(G_i - I) + (-1)^{k-1}(I - D_i) M^{(k)}(\Omega) G_i + (-1)^k D_i M^{(k+1)}(\Omega) G_i,$$

$k = 1, 2, \ldots$

We can now define a sequence $\{\widehat{\Psi}_{iA}^{(k)}, k = 1, 2, \ldots\}$ of bias-adjusted estimators for Ψ, where

$$\widehat{\Psi}_{iA}^{(k)} = P \widehat{\Omega}_{iA}^{(k)} P'$$

is the kth-order bias-corrected estimator of Ψ and its bias is

$$B_{\widehat{\Psi}_{iA}^{(k)}}(\Omega) = P \left[B_{\widehat{\Omega}_{iA}^{(k)}}(\Omega) \right] P'.$$

Next, we shall investigate the order of the bias of $\widehat{\Omega}_{iA}^{(k)}$ given above for $k = 1, 2, \ldots$. Recall that $G_i = A_i^{-1}$ with $A_i = (I - K) + D_i\{K + HKH - 2KK\}_d$. Recall also that $\Omega = O(n^0)$, $P = O(n^{-1})$, $H = O(n^{-1})$. Let us obtain the order of G_i, $i = 1, \ldots, 4$. Note that $I - K = O(n^0)$. Additionally, for $i = 1, \ldots, 4$, it is easy to show that $D_i = O(n^0)$ and $I - D_i = O(n^{-1})$. Also, $K + HKH - 2KK = O(n^{-1})$. Thus, $G_i^{-1} = O(n^0)$ and, as a consequence, $G_i = O(n^0)$. Let us now move to the order of $G_i - I$. Let d_j, b_j and g_j denote the jth diagonal elements of D_i, $\{K + HKH - 2KK\}_d$ and G_i, respectively. Hence,

$$g_j - 1 = \frac{1}{(1 - h_j) + d_j b_j} - 1 = \frac{h_j - d_j b_j}{1 - h_j + d_j b_j}.$$

Note that $h_j - d_j b_j = O(n^{-1})$ and that the order of the denominator is $O(n^0)$ since it is a diagonal element of A_i. Therefore, $G_i - I = O(n^{-1})$. Now recall that $M^{(k-1)}(\Omega) = O(n^{-(k-1)})$, $M^{(k)}(\Omega) = O(n^{-k})$ and $M^{(k+1)}(\Omega) = O(n^{-(k+1)})$. We then obtain that $B_{\widehat{\Omega}_{iA}^{(k)}}(\Omega)$ is of order $O(n^{-k})$ which implies that $B_{\widehat{\Psi}_{iA}^{(k)}}(\Omega) = O(n^{-(k+1)})$, $i = 1, \ldots, 4$.

By letting $k = 1$, we see that the order of the biases of the estimators we proposed in this section is larger than that of $\widehat{\Psi}_{0A}^{(1)}$ (the Qian–Wang estimator), which we have shown to be $O(n^{-3})$. It also follows that in order to obtain the same precision with the bias-corrected estimators given here relative to those given in Sect. 3 one needs to go one step further in the sequence of bias adjustment iterations. In that sense, even though the Qian–Wang estimator is a particular case of the class of covariance matrix estimators we propose here, the results relative to bias adjustment given in this section do not generalize those obtained for the Qian–Wang estimator. This is so because the orders of the corrected estimators for $i = 1, \ldots, 4$ differ from the corresponding orders when $i = 0$.

We shall now use the estimators

$$\widehat{\Psi}_{iA}^{(1)} = P \widehat{\Omega}_{iA}^{(1)} P' = P \widehat{\Omega}_i^{(1)} A_i^{-1} P', \quad i = 0, \ldots, 4,$$

to estimate the variance of linear combinations of the components in $\widehat{\beta}$. Let c be a p-vector of scalars. The estimator of $\Phi = \text{var}(c'\widehat{\beta})$ is

$$\begin{aligned}\widehat{\Phi}_{iA}^{(1)} &= c'\widehat{\Psi}_{iA}^{(1)}c \\ &= c'P\widehat{\Omega}_i^{(1)}G_iP'c \\ &= c'P[\widehat{\Omega} - D_iM^{(1)}(\widehat{\Omega})]G_iP'c \\ &= c'P\widehat{\Omega}G_iP'c - c'PD_iM^{(1)}(\widehat{\Omega})G_iP'c \\ &= c'PG_i^{1/2}\widehat{\Omega}G_i^{1/2}P'c - c'PD_i^{1/2}G_i^{1/2}M^{(1)}(\widehat{\Omega})G_i^{1/2}D_i^{1/2}P'c.\end{aligned}$$

Now let $w_i = G_i^{1/2}P'c$, $v_i = G_i^{1/2}D_i^{1/2}P'c$, $W_i = (w_iw_i')_d$ and $V_i = (v_iv_i')_d$. We then have

$$\begin{aligned}\widehat{\Phi}_{iA}^{(1)} &= w_i'\widehat{\Omega}w_i - v_i'M^{(1)}(\widehat{\Omega})v_i \\ &= \widehat{\varepsilon}'W_i\widehat{\varepsilon} - \widehat{\varepsilon}'[M^{(1)}(V_i)]\widehat{\varepsilon} \\ &= \widehat{\varepsilon}'(W_i - M^{(1)}(V_i))\widehat{\varepsilon} \\ &= \widehat{\varepsilon}'Q_{iA}^{(1)}\widehat{\varepsilon},\end{aligned}$$

where $Q_{iA}^{(1)} = W_i - M^{(1)}(V_i)$.

It is possible to write $\widehat{\Phi}_{iA}^{(1)}$ as a quadratic form in a random vector a which has zero mean and unit covariance:

$$\widehat{\Phi}_{iA}^{(1)} = a'C_{iA}^{(1)}a,$$

where $C_{iA}^{(1)} = \Omega^{1/2}(I - H)Q_{iA}^{(1)}(I - H)\Omega^{1/2}$. For simplicity of notation, let $C_{iA}^{(1)} = C_{iA}$. Following the arguments outlined in Sect. 4, we can show that

$$\text{var}(\widehat{\Phi}_{iA}^{(1)}) = \text{var}(a'C_{iA}a) = 2\text{tr}(C_{iA}^2)$$

when the errors are independent and normally distributed.

In what follows, we shall report the results of a small numerical evaluation using the same two-covariate regression model used in Sect. 5. In particular, we report the total relative biases of the bias-corrected versions of the modified HC0, HC1, HC2, HC3 and HC4 estimators; the modification consists of multiplying these estimators by A_i^{-1} so that they become unbiased under homoskedasticity. The results are displayed in Table 11. Note that $\widehat{\Psi}_{0A}^{(1)}$ is the Qian–Wang estimator $\widehat{V}^{(1)}$ (see Table 3). It is noteworthy that under well-balanced data, the total relative biases of the corrected modified HC1 through HC4 estimators are smaller than those of the Qian–Wang estimator. Under leveraged data, small sample size ($n = 20$) and heteroskedasticity ($\lambda \approx 9$ and $\lambda \approx 49$), the corrected modified HC4 estimator is considerably less biased than the corrected modified HC0 (Qian–Wang) estimator. For instance, the total relative biases of the later under $\lambda \approx 9$ and $\lambda \approx 49$ are 0.222 and 0.443, respectively, whereas the corresponding biases of the former are 0.025 and 0.025; under strong heteroskedasticity, the bias

Table 11 Total relative biases for the corrected estimators $\widehat{\Psi}_{iA}^{(1)}$, $i = 0, 1, \ldots, 4$, which are unbiased under homoskedasticity

Covariates	n	λ	$\widehat{\Psi}_{0A}^{(1)}$	$\widehat{\Psi}_{1A}^{(1)}$	$\widehat{\Psi}_{2A}^{(1)}$	$\widehat{\Psi}_{3A}^{(1)}$	$\widehat{\Psi}_{4A}^{(1)}$
$\mathcal{U}(0, 1)$	20	1	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.033	0.027	0.022	0.009	0.012
		≈ 49	0.044	0.036	0.030	0.015	0.024
	40	1	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.007	0.006	0.004	0.001	0.002
		≈ 49	0.010	0.008	0.006	0.003	0.005
	60	1	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.003	0.002	0.002	0.000	0.001
		≈ 49	0.004	0.003	0.003	0.001	0.002
LN(0, 1)	20	1	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.222	0.208	0.157	0.085	0.025
		≈ 49	0.443	0.413	0.306	0.156	0.025
	40	1	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.034	0.029	0.017	0.002	0.037
		≈ 49	0.069	0.060	0.037	0.004	0.073
	60	1	0.000	0.000	0.000	0.000	0.000
		≈ 9	0.014	0.012	0.007	0.002	0.018
		≈ 49	0.029	0.025	0.014	0.003	0.035

The values of the covariates were selected as random uniform and lognormal draws

of the corrected modified HC4 estimator is nearly 18 times smaller than that of the Qian–Wang estimator.

Finally, we shall revisit the empirical application that uses data on per capita spending on public schools (Sect. 6). The standard errors obtained using two of the estimators proposed in this section (corrected using up to three iterations) are given in Table 12; these results are to be compared to those in Table 7. In particular, we present standard errors obtained from the HC3 (Davidson and MacKinnon 1993) and HC4 (Cribari-Neto 2004) estimators and also heteroskedasticity-robust standard errors from our modified versions of these estimators and their first three corrected variants. We note that the standard errors given here are larger than those obtained using White's estimator and its corrected versions, and also larger than the Qian–Wang (uncorrected and corrected) counterparts in the presence of leverage points (cases 1 and 2). In particular, note the standard errors of $\widehat{\beta}_3$ when the data contain all 50 observations (three iterations of the bias correcting scheme): 1,487.68 and 1,545.93 (1,167.94 and 1,357.25 for the two third-order corrected standard errors in Table 7). That is, the new standard errors suggest that there is even more uncertainty involved in the estimation of β_3 than the standard errors reported in the previous section. (Recall that $\widehat{\beta}_3 = 1{,}587.042$.) As noted earlier, a scatterplot shows a satisfactory linear scatter except for a single high leverage point: Alaska. The standard errors reported in Table 12 signal that the estimation of the quadratic income effect is highly uncertain,

Table 12 Standard errors; modified and corrected estimators: $\widehat{\Psi}_{3A}^{(i)}$ and $\widehat{\Psi}_{4A}^{(i)}$, $i = 1, \ldots, 4$; first application

Case	HC3	$\widehat{\Psi}_{3A}^{(1)}$	$\widehat{\Psi}_{3A}^{(2)}$	$\widehat{\Psi}_{3A}^{(3)}$	$\widehat{\Psi}_{3A}^{(4)}$	HC4	$\widehat{\Psi}_{4A}^{(1)}$	$\widehat{\Psi}_{4A}^{(2)}$	$\widehat{\Psi}_{4A}^{(3)}$	$\widehat{\Psi}_{4A}^{(4)}$
1										
$\widehat{\beta}_1$	1,095.00	836.07	811.58	810.32	816.41	3,008.01	877.89	850.95	845.81	848.29
$\widehat{\beta}_2$	2,975.41	2,270.31	2,204.41	2,201.27	2,217.96	8,183.19	2,384.47	2,311.75	2,297.97	2,304.82
$\widehat{\beta}_3$	1,995.24	1,522.06	1,478.41	1,476.47	1,487.68	5,488.93	1,598.76	1,550.44	1,541.32	1,545.93
2										
$\widehat{\beta}_1$	594.80	485.52	483.52	485.60	487.75	1,239.75	506.35	509.48	507.75	506.03
$\widehat{\beta}_2$	1,630.15	1,330.58	1,325.49	1,331.55	1,337.73	3,414.20	1,389.70	1,397.94	1,393.26	1,388.60
$\widehat{\beta}_3$	1,103.03	899.90	896.69	901.00	905.35	2,320.83	941.13	946.55	943.40	940.26
3										
$\widehat{\beta}_1$	577.11	531.42	530.54	530.25	530.13	613.29	524.21	528.47	529.19	529.57
$\widehat{\beta}_2$	1,593.62	1,473.01	1,470.92	1,470.21	1,469.92	1,688.73	1,455.63	1,465.90	1,467.64	1,468.54
$\widehat{\beta}_3$	1,087.41	1,007.94	1,006.71	1,006.29	1,006.11	1,150.05	997.58	1,003.71	1,004.73	1,005.27
4										
$\widehat{\beta}_1$	707.15	668.18	667.81	667.69	667.65	725.74	668.14	667.69	667.57	667.57
$\widehat{\beta}_2$	1,925.44	1,819.43	1,818.44	1,818.10	1,817.99	1,980.52	1,819.39	1,818.12	1,817.77	1,817.79
$\widehat{\beta}_3$	1,297.35	1,225.53	1,224.85	1,224.63	1,224.55	1,337.81	1,225.55	1,224.65	1,224.40	1,224.41

since it seems to be mostly driven by a single point (Alaska). It is also noteworthy that the HC4 estimator seems to be largely positively biased (in the opposite direction of HC0) and that iteration of the bias-correcting scheme coupled with the proposed modification yields standard errors more in line with what one would expect based on the remaining estimates; e.g., the standard error of $\widehat{\beta}_3$ is reduced from 5,488.93 to 1,598.76.

8 Concluding remarks

In this paper we derived a sequential bias correction scheme to the heteroskedasticity-consistent covariance matrix estimator proposed by Qian and Wang (2001). The corrections are such that the order of the bias decreases as we move along the sequence. The numerical evidence showed that the gain in precision can be substantial when one uses the adjusted versions of the estimator. It has also been shown that the corrected Qian–Wang estimators are typically less biased than their respective corrected HC0 (White) estimators. We have also proposed a general class of heteroskedasticity-consistent covariance matrix estimators which generalizes the Qian–Wang estimator. We have shown that the sequential bias adjustment proposed for the Qian–Wang estimator can be easily extended to the more general class of estimators we have proposed.

Acknowledgments The first author acknowledges research grants from CNPq. The authors thank two anonymous referees for comments and suggestions.

References

Cagan, P. (1974). *Common stock values and inflation: the historical record of many countries.* Boston: National Bureau of Economic Research.

Chesher, A., Jewitt, I. (1987). The bias of a heteroskedasticity consistent covariance matrix estimator. *Econometrica*, *55*, 1217–1222.

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

Cribari-Neto, F., Zarkos, S. G. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews*, *18*, 211–228.

Cribari-Neto, F., Zarkos, S. G. (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. *Journal of Statistical Computation and Simulation*, *68*, 391–411.

Cribari-Neto, F., Ferrari, S. L. P., Cordeiro, G. M. (2000). Improved heteroskedasticity-consistent covariance matrix estimators. *Biometrika*, *87*, 907–918.

Davidson, R., MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.

Ferrari, S. L. P., Cribari-Neto, F. (1998). On bootstrap and analytical bias corrections. *Economics Letters*, *58*, 7–15.

Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics and Data Analysis*, *49*, 361–376.

Greene, W. H. (1997). *Econometric analysis*, 3rd edn. Upper Saddle River: Prentice-Hall.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics*, *19*, 285–292.

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.

Long, J. S., Ervin, L. H. (2000). Using heteroskedasticity-consistent standard errors in the linear regression model. *American Statistician*, *54*, 217–224.

MacKinnon, J. G., Smith, A. A. (1998). Approximate bias corrections in econometrics. *Journal of Econometrics*, *85*, 205–230.

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

Qian, L., Wang, S. (2001). Bias-corrected heteroscedasticity robust covariance matrix (sandwich) estimators. *Journal of Statististical Computation and Simulation*, *70*, 161–174.

Rao, C. R. (1973). *Linear statistical inference and its applications*, 2nd edn. New York: Wiley.

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