UNIVERSIDADE FEDERAL DE PERNAMBUCO
Centro de Ciências Exatas e da Natureza

Pós-Graduação em Matemática Computacional

Essays on heteroskedasticity

Maria da Glória Abage de Lima

Tese de Doutorado

RECIFE
30 de maio de 2008

UNIVERSIDADE FEDERAL DE PERNAMBUCO
Centro de Ciências Exatas e da Natureza

Maria da Glória Abage de Lima

Essays on heteroskedasticity

Trabalho apresentado ao Programa de Pós-Graduação em Matemática Computacional do Centro de Ciências Exatas e da Natureza da UNIVERSIDADE FEDERAL DE PERNAMBUCO como requisito parcial para obtenção do grau de Doutor em Matemática Computacional.

Orientador: *Prof. Dr. Francisco Cribari Neto*

RECIFE
30 de maio de 2008

Lima, Maria da Glória Abage de
Essays on heteroskedasticity / Maria da Glória Abage de Lima. – Recife : O Autor, 2008.
xi, 120 folhas : il., fig., tab.

Tese (doutorado) – Universidade Federal de Pernambuco. CCEN. Matemática Computacional, 2008.

Inclui bibliografia e apêndice.

1. Análise de regressão. I. Título.

519.536 CDD (22.ed.) MEI2008-064

UFPE60 ANOS

UNIVERSIDADE FEDERAL DE PERNAMBUCO
Centro de Ciências Exatas e da Natureza
Doutorado em Matemática Computacional

PARECER DA BANCA EXAMINADORA DE DEFESA DE TESE DE DOUTORADO

MARIA DA GLÓRIA ABAGE DE LIMA

"ESSAYS ON HETEROSKEDASTICITY"

A Banca Examinadora composta pelos Professores Francisco Cribari Neto (Presidente), Klaus Leite Pinto Vasconcellos, Audrey Helen Mariz de Aquino Cysneiros, todos da Universidade Federal de Pernambuco, Gauss Moutinho Cordeiro, da Universidade Federal Rural de Pernambuco, Silvia Lopes de Paula Ferrari, da Universidade de São Paulo, considera a candidata:

(✓) APROVADA COM DISTINÇÃO ( ) APROVADA ( ) REPROVADA

Secretaria do Programa de Doutorado em Matemática Computacional do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, aos 30 dias do mês de maio de 2008.

PROF. FRANCISCO CRIBARI NETO
PRESIDENTE

PROF. KLAUS LEITE PINTO VASCONCELLOS
1° EXAMINADOR

PROF AUDREY HELEN MARIZ AQUINO CYSNEIROS
2° EXAMINADOR

PROF. GAUSS MOUTINHO CORDEIRO
3° EXAMINADOR

PROF. SILVIA LOPES DE PAULA FERRARI
4ª EXAMINADOR

Av. Prof. Luis Barros Freire, s/nº, Cidade Universitária - CEP: 50670-901 - Recife - PE
FONE/FAX: (81) 2126-8412 - E-mail: ppgmc@ufpe.br - Home Page: www.ppgmc.ufpe.br

A Roberto e Juliana.

Acknowledgements

First of all, I thank God, who has always been a source of strength to me.

I also thank Roberto and Juliana. The love we share is a great source of strength and inspiration.

I am also grateful to my advisor, prof. Francisco Cribari Neto, who was always available and willing to discuss with me the progresses of the research activities that led to the present dissertation.

I thank my employer, Universidade Federal Rural de Pernambuco, for the support that allowed me to complete the Doctoral Program in Computational Mathematics at UFPE.

I also express gratitute to the following professors: Francisco Cribari Neto, Klaus Leite Pinto Vasconcellos, Audrey Helen Mariz de Aquino Cysneiros, Francisco José de Azevêdo Cysneiros and Sóstenes Lins. They have greatly contributed to my doctoral education. The comments made by professors Klaus Leite Pinto Vasconcellos and Sílvio Melo on a preliminary version of the Ph.D. dissertation during my qualifying exam were helpful and were greatly appreciated.

I am also grateful to my colleagues Graça, Andréa, Ademakson, Donald and Mirele. I have truly enjoyed our time together.

I thank my friends at the Statistics Departament at UFPE, who have always been kind to me. Special thanks go to Valéria Bittencourt, who is efficient, polite and available.

Finally, I would like to thank Angela, who has helped me with the final steps needed to complete the submission of this dissertation to UFPE.

Tudo tem o seu tempo determinado, e há tempo para todo o propósito debaixo do céu:

—ECLESIASTES (3;1)

Resumo

Esta tese de doutorado trata da realização de inferências no modelo de regressão linear sob heteroscedasticidade de forma desconhecida. No primeiro capítulo, nós desenvolvemos estimadores intervalares que são robustos à presença de heteroscedasticidade. Esses estimadores são baseados em estimadores consistentes de matrizes de covariâncias propostos na literatura, bem como em esquemas bootstrap. A evidência numérica favorece o estimador intervalar HC4. O Capítulo 2 desenvolve uma seqüência corrigida por viés de estimadores de matrizes de covariâncias sob heteroscedasticidade de forma desconhecida a partir de estimador proposto por Qian e Wang (2001). Nós mostramos que o estimador de Qian-Wang pode ser generalizado em uma classe mais ampla de estimadores consistentes para matrizes de covariâncias e que nossos resultados podem ser facilmente estendidos a esta classe de estimadores. Finalmente, no Capítulo 3 nós usamos métodos de integração numérica para calcular as distribuições nulas exatas de diferentes estatísticas de testes quasi-*t*, sob a suposição de que os erros são normalmente distribuídos. Os resultados favorecem o teste HC4.

Palavras-chave: bootstrap, correção de viés, distribuição exata de estatísticas quasi-*t*, estimadores consistentes para a matriz de covariâncias sob heteroscedasticidade, heteroscedasticidade, intervalos de confiança consistentes sob heteroscedasticidade, testes quasi-*t*.

Abstract

This doctoral dissertation addresses the issue of performing inference on the parameters that index the linear regression model under heteroskedasticity of unknown form. In the first chapter we develop heteroskedasticity-robust interval estimators. These are based on different heteroskedasticity-consistent covariance matrix estimators (HCCMEs) proposed in the literature and also on bootstrapping schemes. The numerical evidence presented favors the HC4 interval estimator. Chapter 2 develops a sequence of bias-corrected covariance matrix estimators based on the HCCME proposed by Qian and Wang (2001). We show that the Qian-Wang estimator can be generalized into a broad class of heteroskedasticity-consistent covariance matrix estimators and that our results can be easily extended to such a class of estimators. Finally, Chapter 3 uses numerical integration methods to compute the exact null distributions of different quasi-*t* test estatistics under the assumption that the errors are normally distributed. The results favor the HC4-based test.

Keywords: bias correction, bootstrap, exact distributions of quasi-*t* statistics, heteroskedasticity, heteroskedasticity-consistent covariance matrix estimators (HCCME), heteroskedasticity-consistent interval estimators (HCIE), quasi-*t* tests

Contents

List of Figures

List of Tables

CHAPTER 1

Heteroskedasticity-consistent interval estimators

1.1 Introduction

The linear regression model is commonly used by practitioners of many different fields to model the dependence of a variable of interest on a set of explanatory variables. The regression parameters are most often estimated by ordinary least squares (OLS). Under the usual assumptions, the resulting estimator is optimal in the class of unbiased and linear estimators; it is also consistent and asymptotically normal. A commonly violated assumption is that known as homoskedasticity, which states that all error variances must be the same. The OLS estimator (OLSE), however, remains unbiased, consistent and asymptotically normal when such an assumption does not hold, i.e., under heteroskedasticity. It is, thus, a valid and useful estimator. The trouble lies in the usual estimator of its covariance matrix, which becomes inconsistent and biased when the error variances are not equal. Asymptotically valid hypothesis testing inference on the regression parameters, however, requires a consistent estimator for such a covariance matrix, from which one obtains standard errors and estimated covariances. Several heteroskedasticity-consistent covariance matrix estimators (HCCMEs) were proposed in the literature. The most well known estimators are the HC0 (White, 1980), HC2 (Horn, Horn and Duncan, 1975), HC3 (Davidson and MacKinnon, 1993) and HC4 (Cribari–Neto, 2004) estimators.[1] HC0 was proposed by Halbert White in an influential *Econometrica* paper and is the most used estimator in empirical studies. White's estimator is commonly used by practitioners, especially by reseachers in economics and finance. His paper had been cited over 4,500 times by mid 2007. It is noteworthy, nonetheless, that the covariance matrix estimator proposed by Halbert White is typically considerably biased in finite-samples, especially when the data contain leverage points (Chesher and Jewitt, 1987).

As noted by Long and Ervin (2000, p. 217), given that heteroskedasticity is common in cross-sectional data, methods that correct for heteroskedasticity are essential for prudent data analysis. The most employed approach in practice, as noted above, is to use ordinary least squares estimates of the regression parameters coupled with HC0 or alternative consistent standard errors. The HC0 variants were designed to achieve superior finite sample performance. According to Davidson and MacKinnon (2004, p. 199), "these heteroskedasticity-consistent standard errors, which may also be referred to as heteroskedasticity-robust, are often enormously useful." In his econometrics textbook, Jeffrey Wooldridge writes (Wooldridge, 2000, p. 249): "In the last two decades, econometricians have learned to adjust standard errors, t, F and LM statistics so that they are valid in the presence of heteroskedasticity of unknown form.

[1] Zeileis (2004) describes a computer implementation of these estimators.

This is very convenient because it means we can report new statistics that work, regardless of the kind of heteroskedasticity present in the population."

Several authors have evaluated the finite sample behavior of HCCMEs as point estimators of the true underlying covariance matrix and also the finite sample performance of asymptotically valid tests based on such estimators. The available numerical results suggest that the HC2 estimator is the least biased (indeed, it is unbiased under homoskedasticity) and that the HC3-based test outperforms the competition in terms of size control; see, e.g., Cribari–Neto and Zarkos (1999, 2001), Cribari–Neto, Ferrari and Oliveira (2005), Long and Ervin (2000) and MacKinnon and White (1985).[2]

In this chapter we address the following question: what are the finite sample properties of heteroskedasticity-consistent interval estimators (HCIEs) constructed using OLSEs of the regression parameters and HCCMEs? We also consider weighted bootstrap-based interval estimators in which data resampling is used to obtain replicates of the parameter estimates. The bootstrap schemes we use combine the percentile with the weighted boostrap, which is robust to heteroskedaticity. Alternative bootstrap estimators based on the wild, percentile-t and (y, X) bootstrap are also considered and evaluated.

We aim at bridging a gap in the existing literature: the evaluation of finite sample interval estimation under heteroskedasticity of unknown form. As noted by Harrell (2001) in the preface of his book on regression analysis, "judging by the increased emphasis on confidence intervals in scientific journals there is reason to believe that hypothesis testing is gradually being deemphasized." In this chapter, focus is placed on confidence intervals for regression parameters when the practitioner believes that there is some form of heteroskedasticity.

Our results show that interval inference on the parameters that index the linear regression model based on the popular White (HC0) estimator can be highly misleading in small samples. In particular, the HC0 interval estimator typically displays considerable undercoverage. Overall, the best performing interval estimator – even when inference is carried out on more than one parameter, i.e., through confidence regions – is the HC4 interval estimator, which even outperforms four different bootstrap interval estimators.

The chapter unfolds as follows. Section 1.2 introduces the linear regression model and also some point estimators of the OLSE covariance matrix. HCIEs are introduced in Section 1.3. Section 1.4 contains numerical results, i.e., results from Monte Carlo investigation; they focus on the finite sample behavior of different interval estimators. Section 1.5 considers confidence intervals based on the weighted, wild, (y, X) and percentile-t bootstrapping schemes whereas Section 1.6 presents confidence regions that are asymptotically valid under heteroskedasticity of unknown form; these sections also contain numerical evidence. Finally, Section 1.7 offers some concluding remarks.

[2] Cribari–Neto, Ferrari and Cordeiro (2000) show that it is possible to obtain improved HC0 point estimators by using an iterative bias reducing scheme; see also Cribari–Neto and Galvão (2003). Godfrey (2006) argues that restricted (rather than unrestricted) residuals should be used in the HCCMEs when these are used in test statistics with the purpose of testing restrictions on regression parameters.

1.2 The model and some point estimators

The model of interest is the linear regression model, namely:

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

where y is an n-vector of observations on the dependent variable (variable of interest), X is a fixed $n \times p$ matrix of regressors ($\text{rank}(X) = p < n$), $\beta = (\beta_0, \ldots, \beta_{p-1})'$ is a p-vector of unknown regression parameters and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ is an n-vector of random errors. The following assumptions are commonly made:

A1 The model $y = X\beta + \varepsilon$ is correctly specified;

A2 $\mathbb{E}(\varepsilon_i) = 0$, $i = 1, \ldots, n$;

A3 $\mathbb{E}(\varepsilon_i^2) = \text{var}(\varepsilon_i) = \sigma_i^2$ $(0 < \sigma_i^2 < \infty)$, $i = 1, \ldots, n$;

A3' $\text{var}(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$ $(0 < \sigma^2 < \infty)$;

A4 $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0 \; \forall \, i \neq j$;

A5 $\lim_{n\to\infty} n^{-1}(X'X) = Q$, where Q is a positive definite matrix.

Under [A1], [A2], [A3] and [A4], the covariance matrix of ε is

$$\Omega = \text{diag}\{\sigma_i^2\},$$

which reduces to $\Omega = \sigma^2 I_n$ when $\sigma_i^2 = \sigma^2 > 0$, $i = 1, \ldots, n$, under [A3'] (homoskedasticity), where I_n is the $n \times n$ identity matrix.

The OLSE of β is obtained by minimizing the sum of squared errors, i.e., by minimizing

$$\varepsilon'\varepsilon = (y - X\beta)'(y - X\beta);$$

the estimator can be written in closed-form as

$$\widehat{\beta} = (X'X)^{-1}X'y.$$

Suppose [A1] holds (i.e., the model is correctly specified). It can be shown that:

i) Under [A2], $\widehat{\beta}$ is unbiased for β, i.e., $\mathbb{E}(\widehat{\beta}) = \beta$, $\forall \beta \in \mathbb{R}^p$.

ii) $\Psi_{\widehat{\beta}} = \text{var}(\widehat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$.

iii) Under [A2], [A3], [A5] and also under uniformly bounded variances, $\widehat{\beta}$ is a consistent estimator of β, i.e., $\text{plim}(\widehat{\beta}) = \beta$, where plim denotes limit in probability.

iv) Under [A2], [A3'] and [A4], $\widehat{\beta}$ is the best linear unbiased estimator of β (Gauss–Markov Theorem).

From ii), we note that under homoskedasticity $\text{var}(\widehat{\beta}) = \sigma^2(X'X)^{-1}$, which can be easily estimated as $\widehat{\text{var}}(\widehat{\beta}) = \widehat{\sigma}^2(X'X)^{-1}$, where $\widehat{\sigma}^2 = \widehat{\varepsilon}'\widehat{\varepsilon}/(n-p)$, $\widehat{\varepsilon}$ being the n-vector of OLS residuals:

$$\widehat{\varepsilon} = y - X\widehat{\beta} = \{I_n - X(X'X)^{-1}X'\}y = (I_n - H)y.$$

The matrix $H = X(X'X)^{-1}X'$ is known as the 'hat matrix', since $Hy = \widehat{y}$. Its diagonal elements assume values on the standard unit interval $(0,1)$ and add up to p, the rank of X, thus averaging p/n. It is noteworthy that the diagonal elements of H $(h_1, \ldots, h_n)$ are commonly used as measures of the leverages of the corresponding observations; indeed observations such that $h_i > 2p/n$ or $h_i > 3p/n$ are taken to be leverage points (see Davidson and MacKinnon, 1993).

When the model is heteroskedastic and Ω is known (which rarely happens), one can use the generalized least squares estimator (GLSE), which is given by

$$\widehat{\beta}_G = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

It is easy to show that

$$\mathbb{E}(\widehat{\beta}_G) = \beta,$$
$$\Psi_{\widehat{\beta}_G} = \text{var}(\widehat{\beta}_G) = (X'\Omega^{-1}X)^{-1}.$$

Note that under homoskedasticity $\widehat{\beta}_G = \widehat{\beta}$ and $\text{var}(\widehat{\beta}_G) = \text{var}(\widehat{\beta})$.

The error covariance matrix Ω, however, is usually unknown, which renders the GLSE unfeasible. A feasible estimator can be obtained by replacing Ω by a consistent estimator $\widehat{\Omega}$; the resulting estimator is the feasible least squares estimator (FGLSE):

$$\widehat{\widehat{\beta}} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y.$$

Consistent estimation of Ω, however, requires a model for the variances, such as, for instance, $\sigma_i^2 = \exp(z_i'\gamma)$, where z_i is a q-vector ($q < n$) of variables that affect the variances and γ is a q-vector of unknown parameters that can be consistently estimated. The FGLSE relies on the assumption made about the variances, which is a drawback in situations (as is oftentimes the case) where the practitioner has no information on the correct specification of the skedastic function. The main practical advantage of the OLSE relative to the FGLSE is that the former requires no such assumption.

Asymptotically valid testing inference on the components of β, the vector of unknown regression parameters, based on $\widehat{\beta}$ requires a consistent estimator for $\text{var}(\widehat{\beta})$, i.e., for the OLSE covariance matrix. Under homoskedasticity, as noted earlier, one can easily estimate $\Psi_{\widehat{\beta}}$ as

$$\widehat{\Psi}_{\widehat{\beta}} = \widehat{\text{var}}(\widehat{\beta}) = \widehat{\sigma}^2(X'X)^{-1}.$$

Under heteroskedasticity of unknown form, one can perform inferences on β based on its OLSE $\widehat{\beta}$, which is consistent, unbiased and asymptotically normal, and on a consistent estimator of its covariance matrix.

White (1980) derived a consistent estimator for $\Psi_{\widehat{\beta}}$ by noting that consistent estimation of Ω (which has n unknown parameters) is not required; one only needs to consistently estimate $X'\Omega X$ (which has $p(p+1)/2$ elements regardless of the sample size).[3] That is, one needs to

[3] For consistent covariance matrix estimation under heteroskedasticity, we shall also assume:

A6 $\lim_{n\to\infty} n^{-1}(X'\Omega X) = S$, where S is a positive definite matrix.

find $\widehat{\Omega}$ such that $\operatorname{plim}\left((X'\Omega X)^{-1}(X'\widehat{\Omega}X)\right) = I_p$. The White estimator, also known as HC0, is obtained by replacing the ith diagonal element of Ω in the expression for $\Psi_{\widehat{\beta}}$ by the ith squared OLS residual, i.e.,

$$\text{HC0} = (X'X)^{-1}X'\widehat{\Omega}_0X(X'X)^{-1},$$

where $\widehat{\Omega}_0 = \operatorname{diag}\{\widehat{\varepsilon}_i^2\}$.

White's estimator is consistent under both homoskedasticity and heteroskedasticity of unknown form. Nonetheless, it can be quite biased in finite samples, as evidenced by the numerical results in Cribari–Neto and Zarkos (1999, 2001); see also the results in Chesher and Jewitt (1987). The bias is usually negative; the White estimator is thus 'too optimistic', i.e., it tends to underestimate the true variances. Additionally, the HC0 bias is more decisive when the regression design includes leverage points. As noted by Chesher and Jewitt (1987, p. 1219), the possibility of severe downward bias in the HC0 estimator arises when there are large h_i, because the associated least squares residuals have small magnitude on average and the HC0 estimator takes small residuals as evidence of small error variances.

Based on the results in Horn, Horn and Duncan (1975), MacKinnon and White (1985) proposed a variant of the HC0 estimator: the HC2 estimator, which uses

$$\widehat{\Omega}_2 = \operatorname{diag}\{\widehat{\varepsilon}_i^2/(1-h_i)\},$$

where h_i is the ith diagonal element of the hat matrix (H). It can be shown that HC2 is unbiased under homoskedasticity.

Consistent covariance matrix estimation under heteroskedasticity can also be performed via jackknife. Indeed, the numerical evidence in MacKinnon and White (1985) favors jackknife-based inference. Davidson and MacKinnon (1993) argue that the jackknife estimator is closely approximated by the estimator obtained by replacing $\widehat{\Omega}_0$, used in HC0, by

$$\widehat{\Omega}_3 = \operatorname{diag}\{\widehat{\varepsilon}_i^2/(1-h_i)^2\}.$$

This estimator is known as HC3.

Cribari–Neto (2004) proposed a variant of the HC3 estimator known as HC4; it uses

$$\widehat{\Omega}_4 = \operatorname{diag}\{\widehat{\varepsilon}_i^2/(1-h_i)^{\delta_i}\},$$

where $\delta_i = \min\{4, h_i/\overline{h}\} = \min\{4, nh_i/p\}$ (note that $\overline{h} = n^{-1}\sum_{i=1}^n h_i = p/n$). The exponent controls the level of discounting for observation i and is given by the ratio between h_i and the average of the h_i's, $\overline{h}$, up to the truncation point set at 4. Since $0 < 1-h_i < 1$ and $\delta_i > 0$, it follows that $0 < (1-h_i)^{\delta_i} < 1$. Hence, the ith squared residual will be more strongly inflated when h_i is large relative to $\overline{h}$. This linear discounting is truncated at 4, which amounts to twice the level of discounting used by the HC3 estimator, so that $\delta_i = 4$ when $h_i > 4\overline{h} = 4p/n$.

1.3 Heteroskedasticity-consistent interval estimators

Our chief interest lies in the interval estimation of the unknown regression parameters. We shall consider HCIEs based on the OLSE $\widehat{\beta}$ and on the HC0, HC2, HC3 and HC4 HCCMEs.

Under homoskedasticity and when the errors are normally distributed, the quantity

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma}^2 c_{jj}}},$$

where c_{jj} is the jth diagonal element of $(X'X)^{-1}$, follows a t_{n-p} distribution. It is thus easy to construct exact confidence intervals for β_j, $j = 0, \ldots, p-1$.

Under heteroskedasticity, as noted earlier, the covariance matrix of the OLSE is

$$\Psi_{\widehat{\beta}} = \text{var}(\widehat{\beta}) = (X'X)^{-1} X'\Omega X (X'X)^{-1}.$$

The consistent estimators presented in the previous section are sandwich-type estimators for such a covariance matrix. In what follows, we shall use the HCk, $k = 0, 2, 3, 4$, estimators of variances and covariances. Let, for $k = 0, 2, 3, 4$,

$$\widehat{\Omega}_k = D_k \widehat{\Omega} = D_k \text{diag}\{\widehat{\varepsilon}_i^2\};$$

for HC0, $D_0 = I_n$;
for HC2, $D_2 = \text{diag}\{1/(1-h_i)\}$;
for HC3, $D_3 = \text{diag}\{1/(1-h_i)^2\}$;
for HC4, $D_4 = \text{diag}\{1/(1-h_i)^{\delta_i}\}$.

Therefore,

$$\widehat{\Psi}^{(k)}_{\widehat{\beta}} = (X'X)^{-1} X' \widehat{\Omega}_k X (X'X)^{-1}, \quad k = 0, 2, 3, 4.$$

For $k = 0, 2, 3, 4$, consider the quantity

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\Psi}^{(k)}_{jj}}},$$

where $\widehat{\Psi}^{(k)}_{jj}$ is the jth diagonal element of $\widehat{\Psi}^{(k)}_{\widehat{\beta}}$, i.e., the estimated variance of $\widehat{\beta}_j$ obtained from the estimator HCk, $k = 0, 2, 3, 4$. It follows from the asymptotic normality of $\widehat{\beta}_j$ and from the consistency of $\widehat{\Psi}^{(k)}_{jj}$ that the quantity above converges in distribution to the standard normal distribution as $n \to \infty$. It can thus be used to construct HCIEs. Let $0 < \alpha < 1/2$. A class of $(1-\alpha)\times 100\%$ (two-sided) confidence intervals for β_j, $j = 0, \ldots, p-1$, is

$$\widehat{\beta}_j \pm z_{1-\alpha/2} \sqrt{\widehat{\Psi}^{(k)}_{jj}},$$

$k = 0, 2, 3, 4$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. The next section contains numerical evidence on the finite sample performance of these HCIEs.

1.4 Numerical evaluation

The Monte Carlo evaluation uses the following linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \sigma_i \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\varepsilon_i \sim (0, 1)$ and $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0 \; \forall i \neq j$. Here,

$$\sigma_i^2 = \sigma^2 \exp\{a x_i\}$$

with $\sigma^2 = 1$. At the outset, we focus on the situation where the errors are normally distributed. We shall numerically estimate the coverage probabilities of the different HCIEs and compute the average lengths of the different intervals. The covariate values were selected as random draws from the $\mathcal{U}(0, 1)$ distribution; we have also selected such values as random draws from the Student t_3 distribution so that the regression design would include leverage points. The sample sizes are $n = 20, 60, 100$. We generated 20 values of the covariates when the sample size was $n = 20$; for the larger sample sizes, these values were replicated three and five times ($n = 60$ and $n = 100$, respectively) so that the level of heteroskedasticity, measured as

$$\lambda = \max\{\sigma_i^2\} / \min\{\sigma_i^2\}, \quad i = 1, \ldots, n,$$

remained constant as the sample size increased. We have considered the situation where the error variances are constant (homoskedasticity, $\lambda = 1$) and also two situations in which there is heteroskedasticity. Simulations under homoskedasticity were performed by setting $a = 0$. Under well balanced data (covariate values obtained as uniform random draws, no observation with high leverage), we used $a = 2.4$ and $a = 4.165$, which yielded $\lambda = 9.432$ and $\lambda = 49.126$, respectively. Under leveraged data (covariate values obtained as t_3 random draws, observations with high leverage in the data), we used $a = 0.222$ and $a = 0.386$, which yielded $\lambda = 9.407$ and $\lambda = 49.272$, respectively. Therefore, numerical results were obtained for $\lambda = 1$ (homoskedasticity), $\lambda \approx 9$ and $\lambda \approx 49$. The values of the regression parameters used in the data generation were $\beta_0 = \beta_1 = 1$. The number of Monte Carlo replications was 10,000 and all simulations were carried out using the `Ox` matrix programming language (Doornik, 2001).

The nominal coverage of all confidence intervals is $1 - \alpha = 0.95$. The standard confidence interval (OLS) used standard errors from $\widehat{\sigma}^2 (X'X)^{-1}$ and was computed as

$$\widehat{\beta}_j \quad \pm \quad t_{1-\alpha/2, n-2} \sqrt{\widehat{\sigma}^2 c_{jj}},$$

where $t_{1-\alpha/2,n-2}$ is the $1 - \alpha/2$ quantile from Student's t_{n-2} distribution. The HCIEs were computed, as explained earlier, as

$$\widehat{\beta}_j \quad \pm \quad z_{1-\alpha/2} \sqrt{\widehat{\Psi}_{jj}^{(k)}},$$

$k = 0, 2, 3, 4$ (HC0, HC2, HC3 and HC4, respectively).

Table 1.1 presents the maximal leverages in the two regression designs; the values of $2p/n$ and $3p/n$, which are often used as threshold values for identifying leverage points, are also

presented. When the covariate values are obtained as random draws from the t_3 distribution, the regression design clearly includes leverage points; the maximal leverage almost reaches $8p/n$. On the other hand, when the covariate values are selected as random draws from the standard uniform distribution, the maximal leverage does not exceed $3p/n$. By considering the two regression designs, we shall be able to investigate the finite sample performances of the different HCIEs under both balanced and unbalanced data.

Table 1.1 Maximal leverages and rule-of-thumb thresholds used to detect leverage points.

	$\mathcal{U}(0,1)$	t_3	threshold	
n	$h_{\max}$	$h_{\max}$	$2p/n$	$3p/n$
20	0.233	0.780	0.200	0.300
60	0.077	0.260	0.067	0.100
100	0.046	0.156	0.040	0.060

Table 1.2 presents the empirical coverages (cov.) and the average lengths of the different confidence intervals for β_1 (slope) under balanced regression design (no leverage point) and normal errors. The corresponding numerical results for the unbalanced regression design (leverage points in the data) are given in Table 1.3.

Table 1.2 Confidence intervals for β_1: coverages (%) and lengths; balanced design and normal errors.

		$n = 20$		$n = 60$		$n = 100$	
	interval	cov.	length	cov.	length	cov.	length
$\lambda = 1$	HC0	89.76	2.76	93.39	1.73	93.64	1.36
	HC2	91.94	3.01	93.99	1.78	94.06	1.38
	HC3	93.64	3.28	94.63	1.83	94.48	1.41
	HC4	93.21	3.23	94.47	1.82	94.39	1.40
	OLS	95.38	3.28	94.92	1.82	94.87	1.40
$\lambda \approx 9$	HC0	91.03	5.79	93.62	3.57	94.05	2.80
	HC2	92.90	6.21	94.09	3.65	94.33	2.84
	HC3	94.64	6.66	94.68	3.73	94.66	2.87
	HC4	93.86	6.43	94.33	3.69	94.41	2.85
	OLS	96.47	7.14	96.33	3.97	96.27	3.05
$\lambda \approx 49$	HC0	90.33	11.67	93.30	7.25	93.88	5.69
	HC2	92.25	12.47	93.98	7.40	94.24	5.76
	HC3	94.27	13.32	94.56	7.56	94.56	5.83
	HC4	93.24	12.76	94.22	7.45	94.32	5.78
	OLS	95.62	13.78	95.36	7.70	95.15	5.93

The figures in Table 1.2 (well balanced design) show that, under homoskedasticity ($\lambda = 1$), the HC3, HC4 and OLS confidence intervals have empirical coverages close to the nominal

level (95%) for all sample sizes. The HC0 and HC2 intervals display good coverage when the sample size is not small. Additionally, the average lengths of the HC3, HC4 and OLS confidence intervals are similar. When $\lambda = 9.432$, the HC3 e HC4 confidence intervals display coverages that are close to the nominal coverage (95%) for all sample sizes. Again, the HC0 and HC2 confidence intervals do not display good coverage when the sample size is small ($n = 20$). For instance, the empirical coverages of the HC3 and HC4 confidence intervals for β_1 when $n = 20$ are, respectively, 94.64% and 93.86%, whereas the corresponding figures for the HC0 and HC2 confidence intervals are 91.03% and 92.90%. The average lengths of all intervals increase substantially relative to the homoskedastic case. When $\lambda = 49.126$, the empirical coverages of the HC3 and HC4 confidence intervals for β_1 are close to the selected nominal level for all samples sizes. Once again, the average lengths of all intervals increased relative to the previous case.

The results reported in Table 1.3 were obtained by imposing an unbalanced regression design (there are leverage points in the data). When $\lambda = 1$ (homoskedasticity), only the HC4 interval has excess coverage when the sample size is small (98.14%); for larger sample sizes, the HC4 HCIE outperforms the other consistent interval estimators. When the strength of heteroskedasticity increases ($\lambda \approx 9$ and then $\lambda \approx 49$), the coverages of all intervals deteriorate (Table 1.3); the HC4 HCIE is the least sensitive to the increase of the level of heteroskedasticity. For example, under strong heteroskedasticity ($\lambda \approx 49$) and $n = 20$, the empirical coverage of the HC4 confidence interval for β_1 is 97.53% whereas the coverages of the HC0, HC2 and HC3 intervals are, respectively, 26.73% 56.97% and 86.62%; it is noteworthy, in particular, the dreadful coverage of the HC0 HCIE. It is also interesting to note that the average lengths of all confidence intervals are considerably smaller when the data contain leverage points relative to the well balanced regression design.

Our next goal is to evaluate the finite sample behavior of the different HCIEs under nonnormal innovations. We have considered asymmetric (exponential with unit mean) and fat-tailed (t_3) distributions for the errors, ε_i, which were generated independently and were normalized to have zero mean and unit variance. (Recall that unequal error variances are introduced by multiplying ε_i by σ_i.) Table 1.4 presents the empirical coverages and the average lengths of the different confidence intervals for the slope parameter, β_1, under leveraged data and exponentially distributed errors; similar results for fat tailed errors are presented in Table 1.5. The results in Table 1.4 (exponential errors) suggest that under homoskedasticity ($\lambda = 1$), the OLS confidence interval displays coverages that are close to the nominal level (95%) and that only the HC4 HCIE displays good finite sample coverage (when $n = 60, 100$). Under heteroskedasticity, however, no interval estimator displayed good coverage.

Table 1.5 contains results for the case where the errors follow a fat-tailed distribution (t_3); inference is performed on β_1 and there are leverage points in the data. Under homoskedasticity, the OLS and HC3 confidence intervals display coverages that are close to the expected coverage (95%); HC4 displays slight overcoverage. Under heteroskedasticity, the HC4 HCIE clearly outperforms the remaining HCIEs as far as coverage is concerned. For instance, when $n = 20$ and $\lambda \approx 49$, the HC4 interval estimator coverage was 96.91% whereas the corresponding figures for the HC0, HC2 and HC3 interval estimators were 33.24%, 59.65% and 85.99%. (Note the extremely large coverage distortion that one obtains when interval estimation is based on the

Table 1.3 Confidence intervals for β_1: coverages (%) and lengths; unbalanced design and normal errors.

		$n = 20$		$n = 60$		$n = 100$	
	interval	cov.	length	cov.	length	cov.	length
$\lambda = 1$	HC0	73.41	0.26	87.50	0.21	91.00	0.17
	HC2	84.00	0.38	90.09	0.23	92.33	0.19
	HC3	92.52	0.68	92.30	0.26	93.54	0.20
	HC4	98.14	2.84	95.49	0.33	95.59	0.22
	OLS	94.83	0.45	94.83	0.25	95.01	0.19
$\lambda \approx 9$	HC0	44.69	0.33	80.12	0.44	86.97	0.39
	HC2	66.38	0.54	84.22	0.51	89.25	0.42
	HC3	84.95	1.04	87.81	0.59	91.12	0.45
	HC4	96.84	4.53	93.47	0.78	94.35	0.53
	OLS	67.27	0.50	68.95	0.30	69.80	0.23
$\lambda \approx 49$	HC0	26.73	0.48	77.52	0.87	86.04	0.77
	HC2	56.97	0.88	82.33	1.01	88.53	0.84
	HC3	86.62	1.79	86.04	1.17	90.79	0.91
	HC4	97.53	7.96	92.50	1.58	93.99	1.08
	OLS	40.73	0.59	48.06	0.39	50.35	0.31

Table 1.4 Confidence intervals for β_1: coverages (%) and lengths; unbalanced design and skewed errors.

		$n = 20$		$n = 60$		$n = 100$	
	interval	cov.	length	cov.	length	cov.	length
$\lambda = 1$	HC0	76.87	0.25	87.31	0.20	89.85	0.17
	HC2	87.26	0.37	89.42	0.22	90.94	0.18
	HC3	94.47	0.66	91.27	0.25	92.23	0.20
	HC4	98.53	2.75	93.52	0.31	93.59	0.22
	OLS	94.37	0.44	94.67	0.25	95.06	0.19
$\lambda \approx 9$	HC0	44.19	0.31	74.52	0.41	81.71	0.36
	HC2	64.35	0.51	78.22	0.47	83.67	0.39
	HC3	83.09	0.97	81.69	0.54	85.43	0.43
	HC4	95.93	4.21	87.37	0.71	88.43	0.50
	OLS	71.95	0.49	68.97	0.29	68.13	0.23
$\lambda \approx 49$	HC0	25.72	0.45	70.26	0.80	79.74	0.72
	HC2	54.39	0.81	74.57	0.92	82.06	0.78
	HC3	83.54	1.63	78.68	1.06	84.00	0.85
	HC4	96.85	7.26	85.11	1.43	87.53	1.00
	OLS	40.62	0.57	45.14	0.38	46.79	0.30

Table 1.5 Confidence intervals for β_1: coverages (%) and lengths; unbalanced design and fat tailed errors.

		$n = 20$		$n = 60$		$n = 100$	
	interval	cov.	length	cov.	length	cov.	length
$\lambda = 1$	HC0	77.73	0.24	90.40	0.19	92.62	0.16
	HC2	86.73	0.35	92.63	0.21	93.87	0.17
	HC3	93.89	0.62	94.40	0.24	95.02	0.18
	HC4	98.52	2.59	97.16	0.30	96.72	0.20
	OLS	93.93	0.41	94.11	0.24	94.69	0.18
$\lambda \approx 9$	HC0	51.76	0.29	83.41	0.39	89.09	0.34
	HC2	70.47	0.47	87.50	0.44	90.99	0.36
	HC3	87.46	0.91	90.79	0.51	92.67	0.39
	HC4	97.32	3.91	95.33	0.67	95.70	0.46
	OLS	74.39	0.46	73.03	0.28	73.68	0.22
$\lambda \approx 49$	HC0	33.24	0.41	80.37	0.75	87.76	0.67
	HC2	59.65	0.74	85.17	0.87	89.98	0.72
	HC3	85.99	1.47	88.74	1.00	92.18	0.79
	HC4	96.91	6.52	94.33	1.35	95.23	0.93
	OLS	51.32	0.53	52.12	0.36	53.35	0.29

standard error proposed by Halbert White!)

Finally, we note that we have also performed simulations based on regression models with 3 and 5 regressors. The results were similar to those of the single regressor model and are not reported.

1.5 Bootstrap intervals

An alternative approach is to use data resampling to perform interval estimation; in particular, one can base inference on the bootstrap method proposed by Bradley Efron (Efron, 1979). The weighted bootstrap of Wu (1986) can be used to obtain a standard error that is asymptotically correct under heteroskedasticity of unknown form. We propose the use of the percentile bootstrap confidence interval combined with a weighted bootstrap resampling scheme. Interval inference on β_j ($j = 0, \ldots, p-1$) can be performed as follows.

S1 For each i, $i = 1, \ldots, n$, draw t_i^* randomly from a zero mean and unit variance population;

S2 Construct a bootstrap sample (y^*, X), where

$$y_i^* = x_i\widehat{\beta} + t_i^*\widehat{\varepsilon}_i / \sqrt{1-h_i},$$

x_i being the ith row of X;

S3 Compute the OLSE of β: $\widehat{\beta}^* = (X'X)^{-1}X'y^*$;

S4 Repeat steps 1 through 3 a large number of times (say, B times);

S5 The lower and upper limits of the $(1-\alpha)\times 100\%$ confidence interval for β_j $(0 < \alpha < 1/2)$ are, respectively, the $\alpha/2$ and $1-\alpha/2$ quantiles of the B bootstrap replicates $\widehat{\beta}^*_j$.

The quantity t_i^*, $i = 1,\ldots,n$, must be sampled from a population that has mean zero and variance equal to one, such as, for instance, $a_1,\ldots,a_n$, where

$$a_i = \frac{\widehat{\varepsilon}_i - \overline{\widehat{\varepsilon}}}{\sqrt{n^{-1}\sum_{i=1}^{n}(\widehat{\varepsilon}_i - \overline{\widehat{\varepsilon}})^2}}, i = 1,\ldots,n,$$

with $\overline{\widehat{\varepsilon}} = n^{-1}\sum_{i=1}^{n}\widehat{\varepsilon}_i$, which equals zero when the regression model contains an intercept. We shall call this implementation 'scheme 1', in contrast to 'scheme 2', where sampling is done from the standard normal distribution.

In what follows (Table 1.6), we shall compare, using Monte Carlo simulations, the finite sample behavior of the HCIEs described in Section 1.3 to the hybrid (percentile/weighted) bootstrap interval estimator described above. Inference is performed on the slope parameter (β_1), the number of Monte Carlo replications was 5,000 and the number of bootstrap replications was $B = 500$.

We note from Table 1.6 that the coverages and average lengths of the two bootstrap confidence intervals are similar, especially when $n = 60$ and $n = 100$. Additionally, by contrasting the results to those in Tables 1.2 and 1.3, we note that when the data do not contain leverage points the bootstrap confidence intervals behave similarly to the HC0 confidence interval; under unbalanced data, the bootstrap inference is similar to that achieved by the HC2 HCIE. Overall, the bootstrap HCIEs are outperformed by the HC4 HCIE.

We shall now consider alternative bootstrap estimators. Similar to the previous estimator, they are based on the percentile method. They are, nonetheless, obtained using different resampling schemes. The first alternative estimator employs the wild bootstrap of Liu (1988), who proposed t_i^* to be randomly selected from a population that has third central moment equal to one, in addition to zero mean and unit variance. She has shown that when this is the case, the weighted bootstrap of Wu (1986) shares the usual second order asymptotic properties of the classical bootstrap. In other words, by adding the restriction that the third central moment equals one it is possible to correct the skewness term in the Edgeworth expansion of the sampling distribution of $\mathbf{1}'\hat{\beta}$, where $\mathbf{1}$ is an n-vector of ones. Liu's wild bootstrap is implemented by sampling t_i^* in such a fashion that it equals -1 with probability 1/2 and $+1$ with the same probability (Rademacher distribution).[4] The remainder of the bootstrapping scheme described above remains unchanged.

The second alternative estimator is obtained by bootstrapping pairs instead of residuals; see, e.g., Efron and Tibshirani (1993, pp. 113–115). Here, one resamples pairs $\mathbf{z}_i = \{(x_i, y_i)\}$,

[4]The use of the Rademacher distribution in this context has been suggested by several authors; see, e.g., Flachaire (2005).

Table 1.6 Bootstrap confidence intervals for β_1: coverages (%) and lengths; balanced and unbalanced regression designs; normal errors; weighted bootstrap.

			$n = 20$		$n = 60$		$n = 100$	
bootstrap	design	λ	cov.	length	cov.	length	cov.	length
scheme 1	balanced	$\lambda = 1$	90.94	2.99	93.96	1.76	93.60	1.37
		$\lambda \approx 9$	91.94	6.16	94.28	3.61	94.22	2.81
		$\lambda \approx 49$	91.60	12.38	93.80	7.33	93.96	5.71
	unbalanced	$\lambda = 1$	84.82	0.38	89.60	0.23	90.88	0.18
		$\lambda \approx 9$	65.70	0.53	84.06	0.50	88.38	0.41
		$\lambda \approx 49$	55.50	0.86	82.02	1.00	87.50	0.83
scheme 2	balanced	$\lambda = 1$	89.56	2.93	93.54	1.76	93.78	1.37
		$\lambda \approx 9$	90.06	6.02	94.02	3.61	94.10	2.81
		$\lambda \approx 49$	89.36	12.31	93.52	7.36	93.86	5.74
	unbalanced	$\lambda = 1$	84.40	0.38	89.58	0.23	91.02	0.18
		$\lambda \approx 9$	65.02	0.53	84.54	0.51	88.84	0.41
		$\lambda \approx 49$	53.64	0.86	82.22	1.08	87.64	0.85

$i = 1,\ldots,n$. The parameter vector β is estimated using the bootstrap sample of responses $y^* = (y_1^*,\ldots,y_n^*)'$ together with the pseudo-design matrix X^* formed out of $x_1^*,\ldots,x_n^*$. This bootstrapping scheme is also known as the (y,X) bootstrap.

The simulation results for interval inference on β_1 using the two alternative bootstrap interval estimators described above, i.e., the estimators based on the wild bootstrap and on the bootstrap of pairs of observations, are presented in Table 1.7. The number of Monte Carlo and bootstrap replications are as before. The figures in Table 1.7, when contrasted with the simulation results reported in Table 1.6, show that the weighted bootstrap estimator slightly outperforms the wild bootstrap estimator when the regression design is balanced but under unbalanced regression designs it is clearly better. Indeed, when the sample size is small ($n = 20$) and heteroskedasticity is strong, the coverage of the wild bootstrap estimator can be dreadful (e.g., 21.84% when the desired coverage is 95%; $\lambda \approx 49$). The figures in Table 1.7 also show that the estimator obtained by bootstrapping pairs of observations outperforms both the weighted and wild bootstrap estimators when the regressors matrix contains leverage points. For instance, when $n = 20$, $\lambda \approx 49$ and there are observations with high leverage in the data, the empirical coverages of the weighted (scheme 1) and wild bootstrap interval estimators are 55.50% and 21.84%, respectively, whereas the empirical coverage of the interval bootstrap estimator that uses bootstrapping of pairs is 87.86%.

The last bootstrap interval estimator for β_j we consider combines weighted resampling with the percentile-*t* method. (See Efron and Tibshirani, 1993, pp. 160-162, for details on the bootstrap-*t* approach.) The estimator can be computed as follows.

S1 For each i, $i = 1,\ldots,n$, draw t_i^* randomly from a zero mean and unit variance population;

S2 Construct a bootstrap sample (y^*, X), where

$$y_i^* = x_i\widehat{\beta} + t_i^*\widehat{\varepsilon}_i/\sqrt{1-h_i},$$

x_i being the ith row of X;

S3 Compute the OLSE of β ($\widehat{\beta}^*$) and $z^* = (\widehat{\beta}_j^* - \widehat{\beta}_j)/\sqrt{\widehat{\text{var}}(\widehat{\beta}_j^*)}$, where $\sqrt{\widehat{\text{var}}(\widehat{\beta}_j^*)}$ is a heteroskedasticity-consistent standard error of $\widehat{\beta}_j^*$ for the bootstrap sample and $\widehat{\beta}_j$ is the OLSE of β_j computed from the original sample.

S4 Repeat steps 1 through 3 a large number of times (say, B times);

S5 The lower and upper limits of the $(1-\alpha)\times 100\%$ confidence interval for β_j $(0 < \alpha < 1/2)$ are, respectively, $\widehat{\beta}_j - \hat{t}^{(1-\alpha/2)}\sqrt{\widehat{\text{var}}(\widehat{\beta}_j)}$ and $\widehat{\beta}_j - \hat{t}^{(\alpha/2)}\sqrt{\widehat{\text{var}}(\widehat{\beta}_j)}$, where $\hat{t}^{(\gamma)}$ is the γ quantile $(0 < \gamma < 1)$ of the B values of z^* $(z_1^*,\ldots,z_B^*)$ and $\sqrt{\widehat{\text{var}}(\widehat{\beta}_j)}$ is the same heteroskedasticity-consistent standard error used in Step 3 (now computed, however, using the original, not the resampled, responses).

We report Monte Carlo evidence on the finite-sample behavior of the percentile-*t* bootstrap interval estimator of β_1 in Table 1.8. Three heteroskedasticity-consistent standard errors are

Table 1.7 Bootstrap confidence intervals for β_1: coverages (%) and lengths; balanced and unbalanced regression designs; normal errors; wild bootstrap and pairs bootstrap.

			$n = 20$		$n = 60$		$n = 100$	
bootstrap	design	λ	cov.	length	cov.	length	cov.	length
wild	balanced	$\lambda = 1$	89.64	2.96	93.56	1.76	93.76	1.37
		$\lambda \approx 9$	89.40	6.01	93.92	3.59	93.60	2.80
		$\lambda \approx 49$	88.48	11.85	93.10	7.24	93.34	5.67
	unbalanced	$\lambda = 1$	71.94	0.25	85.92	0.20	89.02	0.17
		$\lambda \approx 9$	40.08	0.29	77.10	0.39	84.34	0.36
		$\lambda \approx 49$	21.84	0.41	72.96	0.74	83.06	0.72
pairs	balanced	$\lambda = 1$	90.20	2.98	93.92	1.75	94.02	1.36
		$\lambda \approx 9$	90.52	6.25	94.12	3.62	93.86	2.81
		$\lambda \approx 49$	89.64	12.48	93.62	7.32	93.64	5.70
	unbalanced	$\lambda = 1$	90.00	0.64	91.04	0.26	91.72	0.19
		$\lambda \approx 9$	88.50	0.76	88.98	0.47	89.98	0.40
		$\lambda \approx 49$	87.86	0.98	87.70	0.87	89.08	0.78

used in Steps 3 and 5, namely: HC0, HC3 and HC4. t_i^*, $i = 1,\ldots,n$, has been sampled from the standard normal distribution, and, as before, $1-\alpha = 0.95$. The results show that when the data are not leveraged it does not make much difference which consistent standard error is used in the bootstrapping scheme. However, in unbalanced situations the percentile-*t* bootstrap with HC4 standard errors displays superior behavior, especially when the sample size is small. For example, when $n = 20$ and under strong heteroskedasticity, the coverages of the bootstrap-*t* confidence intervals with HC0, HC3 and HC4 standard errors are 75.12%, 84.42% and 89.36%, respectively.

Overall, the best performing bootstrap estimator is the (y, X) bootstrap estimator when the sample size is small ($n = 20$) and the percentile-*t* bootstrap estimator when the sample size is large ($n = 100$). It is noteworthy, however, that the HC4 HCCIE outperforms all bootstrap-based interval estimators.

1.6 Confidence regions

We shall now consider confidence regions that are asymptotically valid under heteroskedasticity of unknown form. To that end, we write the regression model

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

as

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \tag{1.6.1}$$

where y, X, β and ε are as described in Section 1.2, X_j and β_j are $n\times p_j$ and $p_j\times 1$, respectively, $j = 1,2$, with $p = p_1 + p_2$ such that $X = [X_1\ X_2]$ and $\beta = (\beta_1', \beta_2')'$.

The OLSE of the vector of regression coefficients in (1.6.1) is $\widehat{\beta} = (\widehat{\beta}_1', \widehat{\beta}_2')'$, where

$$\widehat{\beta}_2 = \left(R_2'R_2\right)^{-1} R_2'y,$$

with $R_2 = M_1X_2$ and $M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$. Since $\widehat{\beta}_2$ is asymptotically normal with mean vector β_2 and covariance matrix

$$V_{22} = \left(R_2'R_2\right)^{-1} R_2'\Omega R_2 \left(R_2'R_2\right)^{-1},$$

the quadratic form

$$W = \left(\widehat{\beta}_2 - \beta_2\right)' V_{22}^{-1} \left(\widehat{\beta}_2 - \beta_2\right)$$

is asymptotically $\chi^2_{p_2}$; the result still holds when V_{22} is replaced by a function of the data $\ddot{V}_{22}$ such that $\text{plim}(\ddot{V}_{22}) = V_{22}$. In particular, we can use the following consistent estimator of the covariance matrix of $\widehat{\beta}_2$:

$$\ddot{V}_{22}^{(k)} = (R_2'R_2)^{-1} R_2'\widehat{\Omega}_k R_2 (R_2'R_2)^{-1},$$

where $\widehat{\Omega}_k$, $k = 0,2,3,4$, is as defined in Section 1.2.

Table 1.8 Bootstrap confidence intervals for β_1: coverages (%) and lengths; balanced and unbalanced regression designs; normal errors; percentile-*t* bootstrap with HC0, HC3 e HC4 standard errors.

			$n = 20$		$n = 60$		$n = 100$	
standard error	design	λ	cov.	length	cov.	length	cov.	length
HC0	balanced	$\lambda = 1$	92.30	3.19	94.48	1.80	94.34	1.39
		$\lambda \approx 9$	93.04	6.45	94.58	3.67	94.34	2.83
		$\lambda \approx 49$	93.02	13.00	94.18	7.44	94.20	5.76
	unbalanced	$\lambda = 1$	83.68	0.48	89.50	0.25	91.52	0.19
		$\lambda \approx 9$	73.00	0.76	86.84	0.62	90.18	0.46
		$\lambda \approx 49$	75.12	1.39	86.86	1.34	90.46	0.96
HC3	balanced	$\lambda = 1$	92.40	3.22	94.48	1.80	94.36	1.39
		$\lambda \approx 9$	92.90	6.42	94.60	3.66	94.32	2.83
		$\lambda \approx 49$	92.94	12.90	94.14	7.44	94.18	5.76
	unbalanced	$\lambda = 1$	82.10	0.84	89.82	0.26	91.74	0.19
		$\lambda \approx 9$	78.88	1.51	87.70	0.65	90.78	0.47
		$\lambda \approx 49$	84.42	2.96	87.82	1.38	90.66	0.96
HC4	balanced	$\lambda = 1$	92.36	3.24	94.52	1.80	94.38	1.39
		$\lambda \approx 9$	92.74	6.39	94.62	3.66	94.26	2.83
		$\lambda \approx 49$	92.84	12.81	94.14	7.43	94.18	5.76
	unbalanced	$\lambda = 1$	85.06	1.63	90.52	0.27	92.20	0.20
		$\lambda \approx 9$	84.64	2.89	88.70	0.68	91.32	0.48
		$\lambda \approx 49$	89.36	5.43	88.90	1.41	90.92	0.97

Let $0 < \alpha < 1$ and let $\chi^2_{p_2,\alpha}$ be such that

$$\Pr(\chi^2_{p_2} < \chi^2_{p_2,\alpha}) = 1 - \alpha;$$

that is, $\chi^2_{p_2,\alpha}$ is the $1-\alpha$ upper quantile of the $\chi^2_{p_2}$ distribution. Also, let

$$\ddot{W}^{(k)} = \left(\widehat{\beta}_2 - \beta_2\right)' \left(\ddot{V}^{(k)}_{22}\right)^{-1} \left(\widehat{\beta}_2 - \beta_2\right).$$

Thus, the $100(1-\alpha)\%$ confidence region for β_2 is given by the set of values of β_2 such that

$$\ddot{W}^{(k)} < \chi^2_{p_2,\alpha}. \tag{1.6.2}$$

In what follows we shall numerically evaluate the finite sample performance of the different confidence regions. The regression model used in the simulation is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, \ldots, n,$$

where ε_i is a zero mean normally distributed error which is free of serial correlation.

The covariate values were, as before, selected as random draws from the standard uniform (well balanced design) and t_3 (unbalanced) distributions. The number of Monte Carlo replications was 10,000, the sample sizes considered were $n = 20, 60, 100$ and $1 - \alpha = 0.95$; simulations were performed under both homoskedasticity and heteroskedasticity. The reported coverages correspond to the percentage of replications in which (1.6.2) holds when inference is performed on β_1 and β_2 (jointly). In order to contrast the finite sample performances of confidence regions ('joint') and confidence intervals (for β_1 and β_2 separately), individual coverages are also reported.

Table 1.9 contains the maximal leverages of the two regression designs together with the usual thresholds used in the detection of leverage points. We note that the design in which the covariate values were selected as draws from a Student t distribution includes observations with very high leverage, unlike the other design (standard uniform draws).

Table 1.9 Maximal leverages and rule-of-thumb thresholds used to detect leverage points; two covariates

	$\mathcal{U}(0,1)$	t_3	threshold	
n	$h_{\max}$	$h_{\max}$	$2p/n$	$3p/n$
20	0.261	0.858	0.30	0.45
60	0.087	0.286	0.10	0.15
100	0.052	0.172	0.06	0.09

Table 1.10 contains the coverages of the different confidence regions ('joint') and also of the individual confidence intervals for $1 - \alpha = 0.95$. We note, at the outset, that the joint coverages are always smaller than the individual ones, the difference being larger when the regression

Table 1.10 Confidence regions for β_1 and β_2: coverages (%); balanced and unbalanced designs; normal errors. The coverages of the individual confidence intervals are also reported.

			HC0			HC2			HC3			HC4		
n	design	λ	joint	β_1	β_2	joint	β_1	β_2	joint	β_1	β_2	joint	β_1	β_2
20	balanced	$\lambda = 1$	82.95	88.20	87.58	87.48	91.47	90.96	91.46	94.18	93.44	89.15	92.77	92.03
		$\lambda \approx 9$	84.38	89.74	88.57	88.95	92.71	91.73	92.41	95.27	94.59	90.01	93.72	92.82
		$\lambda \approx 49$	84.31	89.53	89.77	89.21	92.54	93.13	92.81	95.20	95.54	90.47	93.54	94.20
	unbalanced	$\lambda = 1$	64.10	77.51	85.50	80.07	88.20	90.75	91.20	95.94	95.64	95.75	99.32	99.37
		$\lambda \approx 9$	35.27	49.27	79.23	61.09	69.10	86.66	84.27	89.10	94.57	95.00	98.54	99.17
		$\lambda \approx 49$	19.72	29.36	68.19	50.21	52.86	80.18	84.36	87.98	94.45	95.56	98.50	99.23
60	balanced	$\lambda = 1$	91.73	93.32	93.13	92.70	94.17	94.06	93.77	95.00	94.66	93.13	94.53	94.32
		$\lambda \approx 9$	91.79	93.75	93.37	92.81	94.59	94.14	93.80	95.20	94.70	93.01	94.84	94.33
		$\lambda \approx 49$	91.68	93.49	93.42	92.91	94.35	94.19	93.89	94.97	94.97	93.14	94.52	94.41
	unbalanced	$\lambda = 1$	84.70	89.60	92.43	88.08	92.05	93.69	90.86	93.82	94.69	94.00	96.37	95.48
		$\lambda \approx 9$	78.79	82.09	90.72	83.01	86.07	92.68	87.19	89.12	94.54	92.15	94.30	96.35
		$\lambda \approx 49$	76.97	78.11	87.24	81.76	82.95	90.44	86.24	86.81	92.80	91.66	92.90	95.87
100	balanced	$\lambda = 1$	93.34	93.68	94.10	93.96	94.29	94.50	94.56	94.70	94.99	94.14	94.49	94.69
		$\lambda \approx 9$	93.07	93.57	94.37	93.76	93.99	94.84	94.35	94.54	95.29	94.01	94.10	94.97
		$\lambda \approx 49$	93.08	93.56	94.48	93.74	94.04	95.00	94.37	94.50	95.37	93.90	94.21	95.10
	unbalanced	$\lambda = 1$	88.82	91.85	92.74	90.70	93.04	93.52	92.59	94.39	94.19	94.30	96.14	94.67
		$\lambda \approx 9$	85.60	87.40	92.57	88.23	89.75	93.53	90.50	91.81	94.57	93.39	94.90	95.87
		$\lambda \approx 49$	85.09	85.53	90.58	87.73	88.25	92.24	90.18	90.55	93.78	93.09	94.18	95.85

design is unbalanced. It can also be seen that when $n = 20$ the HC0 and HC2 regions and intervals can display severe undercoverage. For instance, the HC0 (HC2) confidence region coverage when $n = 20$, the data includes observations with high leverage and heteroskedasticity is strong is less than 20% (approximately 50%), which is much smaller than the nominal coverage (95%); the corresponding HC3 and HC4 confidence regions coverages, in contrast, are 84.36% and 95.56%. Overall, the results show that the HC4 confidence region outperforms the competition, especially under leveraged data. The HC3 confidence region is competitive when the regression design is well balanced.

1.7 Concluding remarks

It is oftentimes desirable to perform asymptotically correct inference on the parameters that index the linear regression model under heteroskedasticity of unknown form. Different variance and covariance estimators have been proposed in the literature, and numerical evidence on the finite sample performance of these point estimators and associated hypothesis tests are available. In this chapter, we have considered and numerically evaluated the finite sample behavior of a class of heteroskedasicity-consistent interval/region estimators. The numerical evaluation was carried out under both homoskedasticity and heteroskedasticity; regression designs both without and with high leverage observations were considered. The results show that interval estimation based on the popular White (HC0) estimator can be quite misleading when the sample size is not large. Overall, the results favor the HC4 interval estimator, which displayed much more reliable finite sample behavior than the HC0 and HC2 interval estimators, and even outperformed its HC3 counterpart.

Bootstrap interval estimation was also considered. Four bootstrap interval estimators were described and evaluated, namely: weighted, wild, pairs and percentile-t. The best performing bootstrap estimators were the pairs bootstrap estimator and that obtained using the percentile-t method, the former displaying superior behavior when the sample size was small and the latter being superior for larger sample sizes (100 observations, in the case of our numerical exercise). It is also noteworthy that the wild bootstrap interval estimator displayed poor coverage under leveraged data, its exact coverage being over four times smaller than the desired coverage in an extreme situation (small sample size, strong heteroskedasticity, leveraged data).

Based on the results in this chapter, we encourage practitioners to perform interval inference in linear regressions using the HC4 interval estimator.

CHAPTER 2

Bias-adjusted covariance matrix estimators

2.1 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 chapter 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, Ferrari and Cordeiro (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 chapter unfolds as follows. Section 2.2 introduces the linear regression model and some heteroskedasticity-consistent covariance matrix estimators. In Section 2.3 we derive a sequence of consistent estimators for the covariance matrix of the ordinary least squares estimator. We do so by defining a sequential bias correcting scheme which is initialized at the estimator proposed by Qian and Wang (2001). In Section 2.4 we obtain estimators for the variance of linear combinations of the elements in the vector of ordinary least squares estimators. Results from a numerical evaluation are presented in Section 2.5; these are exact, not Monte Carlo results. Two empirical applications that use real data are presented and discussed in Section 2.6. In Section 2.7 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, Section 2.8 offers some concluding remarks.

2.2 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., $\Omega = \text{cov}(\varepsilon) = \text{diag}\{\sigma_i^2\}$.

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 heteroskedaticity, 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.[1]

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

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

[1] In what follows, we shall omit the order subscript when denoting the identity matrix; the order must be implicitly understood.

where $\widehat{\Omega} = \text{diag}\{\widehat{\varepsilon}_i^2\}$. 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) HC1 $= P\widehat{\Omega}_1P' = PD_1\widehat{\Omega}P'$, where $D_1 = (n/(n-p))I$;

(ii) (Horn, Horn and Duncan, 1975) HC2 $= P\widehat{\Omega}_2P' = PD_2\widehat{\Omega}P'$, where

$$D_2 = \text{diag}\{1/(1-h_i)\};$$

(iii) (Davidson and MacKinnon, 1993) HC3 $= P\widehat{\Omega}_3P' = PD_3\widehat{\Omega}P'$, where

$$D_3 = \text{diag}\{1/(1-h_i)^2\};$$

(iv) (Cribari–Neto, 2004) HC4 $= P\widehat{\Omega}_4P' = PD_4\widehat{\Omega}P'$, where

$$D_4 = \text{diag}\{1/(1-h_i)^{\delta_i}\}, \quad \delta_i = \min\{4, nh_i/p\}.$$

As noted earlier, the HC0 estimator is considerably biased in samples of small to moderate sizes. Cribari–Neto, Ferrari and Cordeiro (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, Ferrari and Cordeiro (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

$$B_{\widehat{\Omega}^{(k)}}(\Omega) = (-1)^k M^{(k+1)}(\Omega), \tag{2.2.1}$$

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

$$\widehat{\Psi}^{(k)} = P\widehat{\Omega}^{(k)}P'. \tag{2.2.2}$$

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

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

$k = 1, 2, \ldots$.

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 $O(n^{-2})$. [2]

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 $K = (H)_d = \text{diag}\{h_i\}$, 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

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

where

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

and

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

The Qian–Wang estimator can be written as

$$\widehat{V}^{(1)} = PD^{(1)}P'. \tag{2.3.1}$$

At the outset, we shall show that the estimator in (2.3.1) is a bias corrected version of the estimator proposed by Halbert White except for an additional correction factor. Note that

$$\begin{aligned} d_i &= (\widehat{\varepsilon}_i^2 - \widehat{b}_i)g_{ii} \\ &= (\widehat{\varepsilon}_i^2 - C_i'\widehat{\Omega}C_i + 2\widehat{\varepsilon}_i^2 C_i'C_i)g_{ii}. \end{aligned} \tag{2.3.2}$$

[2] The results in Cribari–Neto, Ferrari and Cordeiro (2000) were generalized to HC0–HC3 by Cribari–Neto and Galvão (2003).

The bias corrected estimator in (2.2.2) obtained using $k = 1$ (one-step correction) can be written as $\widehat{\Psi}^{(1)} = P\widehat{\Omega}^{(1)}P'$, where

$$\begin{aligned}\widehat{\Omega}^{(1)} &= \widehat{\Omega} - M^{(1)}(\widehat{\Omega}) \\ &= \widehat{\Omega} - \{H\widehat{\Omega}(H-2I)\}_d \\ &= \text{diag}\{\widehat{\varepsilon}_i^2 - C_i'\widehat{\Omega}C_i + 2\widehat{\varepsilon}_i^2 h_i\}. \end{aligned} \tag{2.3.3}$$

Since $h_i = C_i'C_i$, it is easy to see that (2.3.2) equals the ith diagonal element of $\widehat{\Omega}^{(1)}$ in (2.3.3), 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 (2.3.1) 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

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

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

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

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

$$B_{D^{(k)}}(\Omega) = (-1)^{k-1} M^{(k-1)}(\Omega)(G - I) + (-1)^k M^{(k+1)}(\Omega)G, \tag{2.3.4}$$

$k = 1, 2, \ldots$.

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

$$\widehat{V}^{(k)} = PD^{(k)}P' \tag{2.3.5}$$

being the kth order bias corrected estimator of Ψ. The bias of $\widehat{V}^{(k)}$ follows from (2.3.4) and (2.3.5):

$$B_{\widehat{V}^{(k)}}(\Omega) \quad = \quad P[B_{D^{(k)}}(\Omega)]P'. \tag{2.3.6}$$

We shall now obtain the order of the bias in (2.3.6). To that end, we make the same assumptions on the matrices X, P, H and Ω as in Section 2.2. We saw in (2.3.4) 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, Ferrari and Cordeiro (2000) and

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

as in (2.2.1). 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 Section 2.2.

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, Ferrari and Cordeiro (2000); see Section 2.2. (Recall, however, that $k = 1$ here yields the unmodified Qian–Wang estimator, which is in itself a correction to White's estimator.)

2.4 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

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

The kth order corrected estimator of our sequence of bias corrected estimators, given in (2.3.5), 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,

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

$k = 1, 2, \ldots$.

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,

$$\widehat{\Phi}^{(1)}_{QW} = \widehat{\varepsilon}'W\widehat{\varepsilon} - \sum_{s=1}^{n} \widehat{\alpha}_s w_s^2. \tag{2.4.1}$$

Given that

$$\widehat{\alpha}_s = \sum_{t=1}^{n} h_{st}^2 \widehat{\varepsilon}_t^2 - 2h_{ss}\widehat{\varepsilon}_s^2, \tag{2.4.2}$$

where h_{st} denotes the (s,t) element of H, the summation in (2.4.1) 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 (2.4.2) 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 (2.4.1) 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,

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

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,

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

It can also be shown that

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

More generally,

$$\begin{aligned} \widehat{\Phi}^{(k)}_{QW} &= c'\widehat{\Psi}^{(k)}_{QW}c \\ &= \widehat{\varepsilon}'Q^{(k)}\widehat{\varepsilon}, \quad k = 1, 2, \ldots, \end{aligned} \tag{2.4.3}$$

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, Ferrari and Cordeiro (2000) have shown that the HC0 variance estimator of $c'\widehat{\beta}$ is given by

$$\begin{aligned} \widehat{\Phi}^{(k)}_{W} &= c'\widehat{\Psi}^{(k)}c \\ &= \widehat{\varepsilon}'Q^{(k)}_{W}\widehat{\varepsilon}, \quad k = 0, 1, 2, \ldots, \end{aligned} \tag{2.4.4}$$

where $Q^{(k)}_{W} = \sum_{j=0}^{k}(-1)^j M^{(j)}(B)$. 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 (2.4.3) reduces to (2.4.4) for $k \geq 1$.

We shall now write the quadratic form in (2.4.3) as a quadratic form in a vector of uncorrelated, zero mean and unit variance random variates.

We have seen in Section 2.2 that $\widehat{\varepsilon} = (I - H)y$. We can then write

$$\begin{aligned} \widehat{\Phi}^{(k)}_{QW} &= \widehat{\varepsilon}'Q^{(k)}\widehat{\varepsilon} \\ &= y'(I - H)Q^{(k)}(I - H)y \\ &= y'\Omega^{-1/2}\Omega^{1/2}(I - H)Q^{(k)}(I - H)\Omega^{1/2}\Omega^{-1/2}y \\ &= z'C^{(k)}_{QW}z, \end{aligned} \tag{2.4.5}$$

where $C^{(k)}_{QW} = \Omega^{1/2}(I - H)Q^{(k)}(I - H)\Omega^{1/2}$ 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 $\text{cov}(z) = \text{cov}(\Omega^{-1/2}y) = I$.

Note that

$$\theta'C^{(k)}_{QW} = \beta'X'\Omega^{-1/2}\Omega^{1/2}(I - H)Q^{(k)}(I - H)\Omega^{1/2} = \beta'X'(I - H)Q^{(k)}(I - H)\Omega^{1/2}.$$

Since $X'(I - H) = 0$, then $\theta'C^{(k)}_{QW} = 0$. Hence, equation (2.4.5) can be written as

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

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

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

(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

$$\mathrm{var}(\widehat{\Phi}_{QW}^{(k)}) = d'\Lambda d + 2\mathrm{tr}(C_{QW}^2), \tag{2.4.6}$$

where d is a column vector formed out of the diagonal elements of C_{QW}, $\mathrm{tr}(C_{QW})$ is the trace of C_{QW} and $\Lambda = \mathrm{diag}\{\gamma_i\}$, where $\gamma_i = (\mu_{4i} - 3\sigma_i^4)/\sigma_i^4$ is the excess of kurtosis of the ith error. When the errors are independent and normally distributed, $\gamma_i = 0$. Thus, $\Lambda = 0$ and (2.4.6) simplifies to

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

For the sequence of corrected HC0 estimators, one obtains (Cribari–Neto, Ferrari and Cordeiro, 2000)

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

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

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 independent and normally distributed with $\mathbb{E}(\varepsilon_i) = 0$ and $\mathrm{var}(\varepsilon_i) = \exp(a x_{i2})$, $i = 1, \ldots, n$. 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 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 2.1 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 log-normal 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 2.1 Maximal leverages for the two regression designs.

	$\mathcal{U}(0,1)$	LN(0,1)	threshold	
n	$h_{\max}$	$h_{\max}$	$2p/n$	$3p/n$
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 2.2 presents the total relative bias of the OLS variance estimator.[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 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 2.2 Total relative bias of the OLS variance estimator.

	$\mathcal{U}(0,1)$			LN(0,1)		
n	$\lambda = 1$	$\lambda \approx 9$	$\lambda \approx 49$	$\lambda = 1$	$\lambda \approx 9$	$\lambda \approx 49$
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

Table 2.3 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 $\widehat{V}4^{(1)}$).[4] 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

[3] The bias of the OLS covariance matrix estimator is given by $(n-p)^{-1}\mathrm{tr}\{\Omega(I-H)\}(X'X)^{-1} - P\Omega P'$; here, $p = 3$.

[4] Note that, following the notation used in Sections 2.2 and 2.3, HC04 and $\widehat{V}4^{(1)}$, for example, correspond to $\widehat{\Psi}^{(4)}$ and $\widehat{V}^{(5)}$, respectively.

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 2.4 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 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, Ferrari and Cordeiro (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.[5] The results are presented in Table 2.5. 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 is of almost 22 times when the covariate values are selected as random lognormal draws).[6] The corrections to the HC0 estimator proposed by Cribari-Neto, Ferrari and Cordeiro (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 zero mean, are uncorrelated, and each ε_i has variance $\sigma_i^2 = \exp\{ax_i\}$. 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 2.6. 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.

[5]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).

[6]Note that these figures are *not* relative.

Table 2.3 Total relative biases. The values of the covariates were selected as random uniform and lognormal draws.

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

Table 2.4 Square roots of the total relative mean squared errors. The values of the covariates were selected as random uniform and lognormal draws.

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

Table 2.5 Maximal biases. The values of the covariates were selected as random uniform and lognormal draws.

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

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 chapter 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. We note, however, that such estimates are oftentimes used for performing inferences on the regression parameters. We have run a 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 errors are independent and normally distributed with zero mean and variance $\sigma_i^2 = \exp(a x_{i1})$. The interest lies in the test of $\mathcal{H}_0 : \beta_3 = 0$ versus $\mathcal{H}_1 : \beta_3 \neq 0$. 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, 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 $n \geq 60$.[7]

[7]The wild bootstrap performed considerably better under less extreme leverage, e.g., when $h_{\max} < 4p/n$.

Table 2.6 Maximal biases, $n = 40$, 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, $3p/n = 0.150$.

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

2.6 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 U.S. 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 $\widehat{\beta}_1 = 832.91$, $\widehat{\beta}_2 = -1834.20$ and $\widehat{\beta}_3 = 1587.04$. 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, Mississipi and Washington, D.C. (their leverage measures are 0.651, 0.200 and 0.208, respectively; note that $3p/n = 0.180$).

In Table 2.7 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 Mississipi were removed from data ($n = 47$). Table 2.8 contains information on the detection of leverage points in these four situations.

The figures in Table 2.7 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 1348.36 (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, Mississipi 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 ($\widehat{\beta}_3 = 1587.04$); this standard error is highly biased in favor of the quadratic model specification. The Qian–Wang standard error equals 1348.36, 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: 1385.77.

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 2.7 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	1243.04	1495.05	1638.07	1741.22	1824.42	2011.74	1960.72	1983.10	2023.45	2066.01
	$\widehat{\beta}_3$	519.08	829.99	1001.78	1098.54	1167.94	1223.77	1348.36	1314.92	1330.15	1357.25	1385.77
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
	$\widehat{\beta}_2$	1064.0	936.92	1039.39	1104.93	1156.01	1196.63	1243.19	1220.43	1243.39	1265.96	1284.65
	$\widehat{\beta}_3$	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
	$\widehat{\beta}_2$	1419.9	1394.09	1465.84	1473.92	1473.28	1471.89	1482.49	1473.60	1471.90	1470.92	1470.34
	$\widehat{\beta}_3$	942.71	949.41	1001.46	1008.06	1008.04	1007.28	1013.03	1008.16	1007.27	1006.71	1006.36
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
	$\widehat{\beta}_2$	1647.6	1699.02	1797.21	1814.12	1817.45	1818.01	1816.07	1817.34	1817.98	1818.05	1818.00
	$\widehat{\beta}_3$	1085.1	1140.63	1209.57	1221.72	1224.14	1224.56	1222.82	1224.02	1224.53	1224.59	1224.56

Table 2.8 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 ($h_{\max}$) 0.208 0.200	0.12	0.18	3.62
2	49	0.562 ($h_{\max}$) 0.250	0.122	0.184	3.05
3	48	0.312 ($h_{\max}$) 0.197	0.125	0.187	1.67
4	47	0.209 ($h_{\max}$)	0.128	0.191	1.09

Table 2.9 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 becomes well balanced when these two observations are removed from the sample ($n = 18$).

Table 2.10 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 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.

Table 2.9 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 2.10 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

2.7 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 Section 2.2. 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 = $\widehat{\Psi}_0 = P\widehat{\Omega}_0P' = PD_0\widehat{\Omega}P'$, where $D_0 = I$. In Section 2.2 we have presented some variants of HC0, namely:

(i) HC1 = $\widehat{\Psi}_1 = P\widehat{\Omega}_1P' = PD_1\widehat{\Omega}P'$, $D_1 = (n/(n-p))I$;

(ii) HC2 = $\widehat{\Psi}_2 = P\widehat{\Omega}_2P' = PD_2\widehat{\Omega}P'$, $D_2 = \text{diag}\{1/(1-h_i)\}$;

(iii) HC3 = $\widehat{\Psi}_3 = P\widehat{\Omega}_3P' = PD_3\widehat{\Omega}P'$, $D_3 = \text{diag}\{1/(1-h_i)^2\}$;

(iv) HC4 = $\widehat{\Psi}_4 = P\widehat{\Omega}_4P' = PD_4\widehat{\Omega}P'$, $D_4 = \text{diag}\{1/(1-h_i)^{\delta_i}\}$ 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

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

Note that

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

and

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

As we have done in Section 2.2, we can write $\widehat{\Psi}_i^{(1)} = P\widehat{\Omega}_i^{(1)}P'$, where

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

Thus,[8]

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

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

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

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

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):

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

Hence, the estimator

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

is unbiased:

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

It is noteworthy that the Qian–Wang estimator given in Section 2.3 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 $A_0 = \{I+HKH-2KK\}_d$.[9]

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

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

[9] In Section 2.2, $\widehat{\Omega}_0^{(1)}$ was denoted as $\widehat{\Omega}^{(1)}$.

Hence,

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

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 Sections 2.3 and 2.4 can be easily extended to the more general class of estimators considered here.

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

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

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

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

Its bias can be expressed as

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

After k iterations of the bias correcting scheme we obtain

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

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

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

$k = 1, 2, \ldots$.

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

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

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 Section 2.3 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 $\Phi = \mathrm{var}(c'\widehat{\beta})$ 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 $w_i = G_i^{1/2}P'c$, $v_i = G_i^{1/2}D_i^{1/2}P'c$, $W_i = (w_i w_i')_d$ and $V_i = (v_i v_i')_d$. 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 Section 2.4, 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 Section 2.5. 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 2.11. Note that $\widehat{\Psi}_{0A}^{(1)}$ is the Qian–Wang estimator $\widehat{V}^{(1)}$ (see Table 2.3). 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 heteroskedatisticity, the bias 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 (Section 2.6). The standard errors obtained using two of the estimators proposed in this section (corrected using up to three iterations) are given in Table 2.12; these results are to be compared to those in Table 2.7. 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 their 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): 1487.68 and 1545.93 (1167.94

Table 2.11 Total relative biases for the corrected estimators $\widehat{\Psi}^{(1)}_{iA}$, $i = 0,1,\ldots,4$, which are unbiased under homoscedasticity. The values of the covariates were selected as random uniform and lognormal draws.

covariates	n	λ	$\widehat{\Psi}^{(1)}_{0A}$	$\widehat{\Psi}^{(1)}_{1A}$	$\widehat{\Psi}^{(1)}_{2A}$	$\widehat{\Psi}^{(1)}_{3A}$	$\widehat{\Psi}^{(1)}_{4A}$
$\mathcal{U}(0,1)$	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

Table 2.12 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	$\widehat{\beta}_1$	1095.00	836.07	811.58	810.32	816.41	3008.01	877.89	850.95	845.81	848.29
	$\widehat{\beta}_2$	2975.41	2270.31	2204.41	2201.27	2217.96	8183.19	2384.47	2311.75	2297.97	2304.82
	$\widehat{\beta}_3$	1995.24	1522.06	1478.41	1476.47	1487.68	5488.93	1598.76	1550.44	1541.32	1545.93
2	$\widehat{\beta}_1$	594.80	485.52	483.52	485.60	487.75	1239.75	506.35	509.48	507.75	506.03
	$\widehat{\beta}_2$	1630.15	1330.58	1325.49	1331.55	1337.73	3414.20	1389.70	1397.94	1393.26	1388.60
	$\widehat{\beta}_3$	1103.03	899.90	896.69	901.00	905.35	2320.83	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
	$\widehat{\beta}_2$	1593.62	1473.01	1470.92	1470.21	1469.92	1688.73	1455.63	1465.90	1467.64	1468.54
	$\widehat{\beta}_3$	1087.41	1007.94	1006.71	1006.29	1006.11	1150.05	997.58	1003.71	1004.73	1005.27
4	$\widehat{\beta}_1$	707.15	668.18	667.81	667.69	667.65	725.74	668.14	667.69	667.57	667.57
	$\widehat{\beta}_2$	1925.44	1819.43	1818.44	1818.10	1817.99	1980.52	1819.39	1818.12	1817.77	1817.79
	$\widehat{\beta}_3$	1297.35	1225.53	1224.85	1224.63	1224.55	1337.81	1225.55	1224.65	1224.40	1224.41

and 1357.25 for the two third-order corrected standard errors in Table 2.7). 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 $\widehat{\beta}_3 = 1587.042$.) As noted earlier, a scatterplot shows a satisfactorily linear scatter except for a single high leverage point: Alaska. The standard errors reported in Table 2.12 signal that the estimation of the quadratic income effect is highly uncertain, 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 5488.93 to 1598.76.

2.8 Concluding remarks

In this chapter 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 the 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.

CHAPTER 3

Inference under heteroskedasticity: numerical evaluation

3.1 Introduction

The linear regression model is commonly used by practitioners to model the relationship between a variable of interest and a set of explanatory or independent variables. In particular, the mean of the response is a linear function of a finite number of regression parameters, each multiplying a different covariate. The mean of the dependent variable (variable of interest) is thus affected by other variables. It is commonly assumed, however, that the conditional variance of the response is constant. If we view the response as the sum of a linear predictor (which involves unknown parameters and covariates) and a zero mean unobservable error term, that amounts to assuming that all errors share the same variance, which is known as homoskedasticity. In many applications, however, the errors are heteroskedastic, i.e., their variances are not constant. The ordinary least squares estimator (OLSE) of the vector of regression parameters remains unbiased, consistent and asymptotically normal under unequal error variances. Nevertheless, its usual covariance matrix estimator, from which we obtain standard errors for the regression parameter estimates, is no longer valid. The standard practice is to base inference on standard errors obtained from a heteroskedasticity-consistent covariance matrix estimator (HCCME) which has the property of being consistent regardless of whether homoskedasticity holds; indeed, the estimator is asymptotically correct under heteroskedasticity of unknown form. The most commonly used HCCME was proposed by Halbert White in an influential and highly cited paper published nearly 30 years ago (White, 1980). White's estimator, also known as HC0, can be, however, quite biased in samples of typical sizes; see, e.g., Cribari–Neto (2004), Cribari–Neto and Zarkos (1999, 2001), Long and Ervin (2000) and MacKinnon and White (1985). In particular, substantial downward bias can occur for regression designs containing points of high leverage (Chesher and Jewitt, 1987). The use of White's variance estimator may thus lead one to find spurious relationships between the variable of interest and other variables.

A few variants of the White (HC0) estimator were proposed in the literature. They include the HC1 (Hinkley, 1977), HC2 (Horn, Horn and Duncan, 1975) and HC3 (Davidson and MacKinnon, 1993) estimators. The Monte Carlo evidence in Long and Ervin (2000) favors HC3-based inference. According to the authors (p. 223), "for samples less than 250, HC3 should be used." They "recommend that HC3-based tests should be used routinely for testing individual coefficients in the linear regression model."

Four new promissing HCCMEs were recently proposed, namely: the $\widehat{V}_1$ and $\widehat{V}_2$ estimators

of Qian and Wang (2001), the HC4 estimator of Cribari–Neto (2004) and the HC5 estimator of Cribari–Neto et al. (2007). A nice feature of Qian and Wang's $\widehat{V}_1$ and $\widehat{V}_2$ estimators is that they are unbiased under homoskedasticity. It is also possible to show that, under heteroskedasticity, the bias of $\widehat{V}_1$ converges to zero faster than that of HC0. Using Monte Carlo simulations, Cribari–Neto (2004) showed that hypothesis testing based on HC4 can even outperform inference obtained from a computationally intensive double bootstrapping scheme. Cribari–Neto et al. (2007) argue that HC5-based inference should be preferred when the data contain strong leverage points.

Standard errors that deliver asymptotically correct inference even when the errors of the model do not share the same variance are extremely useful in applications. Davidson and MacKinnon (2004, p. 199) note that "these heteroskedasticity-consistent standard errors, which may also be referred to as heteroskedasticity-robust, are often enormously useful." Jeffrey Wooldridge agrees (Wooldridge, 2000, p. 249): "In the last two decades, econometricians have learned to adjust standard errors, t, F and LM statistics so that they are valid in the presence of heteroskedasticity of unknown form. This is very convenient because it means we can report new statistics that work, regardless of the kind of heteroskedasticity present in the population." We add, nonetheless, that practitioners should be careful when basing their inferences on HCCMEs since the associated tests may display unreliable behavior in finite samples. It is important to use heteroskedasticity-robust tests that are reliable in samples of typical sizes.

The chief goal of this chapter is to use numerical integration methods to perform an exact (not Monte Carlo) evaluation of the finite-sample behavior of tests based on the four recently proposed heteroskedaticity-correct standard errors ($\widehat{V}_1$, $\widehat{V}_2$, HC4 and HC5). HC0- and HC3-based inferences are included in the analysis as benchmarks. Additionally, our results shed light on the choice of constants used in the definition of $\widehat{V}_2$ and HC5. They also show that HC4-based inference can be considerably more reliable than that based on alternative HCCMEs since the null distribution of this test statistic is typically better approximated by the limiting null distribution (from which we obtain critical values for the test) than those of the alternative test statistics. [1]

The chapter unfolds as follows. Section 3.2 introduces the model and some heteroskedasticity-robust standard errors. In Section 3.3 we show how HCCMEs can be used in the variance estimation of a linear combination of the regression parameter estimates. Section 3.4 shows that by assuming that the errors are normally distributed it is possible to write the test statistics as ratios of quadratic forms in a vector of uncorrelated standard normal variates, which allows us to use the numerical integration method proposed by Imhof (1961); see Farebrother (1990) for details on this algorithm. The first numerical evaluation is performed in Section 3.5; here, we focus on inferences based on HC0, HC3, HC4 and $\widehat{V}_1$. In Section 3.6, we present the HCCME $\widehat{V}_2$, write it in matrix form and show how it can be used in the variance estimation of a linear combination of regression parameter estimates. We note that this estimator is indexed by a real constant, a, and that Qian and Wang (2001) propose using $a = 2$ in order to minimize its mean squared error. The numerical evaluation in Section 3.7 focuses on $\widehat{V}_1$ and $\widehat{V}_2$; in particular, it sheds some light on the choice of a when the interest lies in hypothesis testing inference.

[1] We focus on quasi-t tests. For joint tests on more than one parameter, see Cai and Hayes (2008) and Godfrey (2006).

In Section 3.8, we turn to the HC5 HCCME proposed by Cribari–Neto et al. (2007), which is numerically evaluated against the HC3 and HC4 estimators. We also address the issue of selecting the value of a constant that indexes the HC5 HCCME. Finally, Section 3.9 concludes the chapter.

3.2 The model and some heteroskedasticity-robust standard errors

The model of interest is the linear regression model:

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

where y is an n-vector of responses, ε is an n-vector of random errors, X is a full column rank $n \times p$ matrix of regressors ($\text{rank}(X) = p < n$) and $\beta = (\beta_0, \ldots, \beta_{p-1})'$ is a p-vector of unknown regression coefficients. Each error term ε_t, $t = 1, \ldots, n$, has zero mean and variance $0 < \sigma_t^2 < \infty$; the errors are also assumed to be uncorrelated, i.e., $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0 \ \forall t \neq s$. Hence, $\text{cov}(\varepsilon) = \Omega = \text{diag}\{\sigma_t^2\}$.

The ordinary least squares estimator of β is obtained from the minimization of the sum of squared errors and is available in closed form: $\widehat{\beta} = (X'X)^{-1}X'y$. It is unbiased and its covariance matrix can be written as $\text{cov}(\widehat{\beta}) = \Psi = P\Omega P'$, where $P = (X'X)^{-1}X'$. Under homoskedasticity, $\sigma_t^2 = \sigma^2 > 0 \ \forall t$, and thus, $\Psi = \sigma^2(X'X)^{-1}$.

When all errors share the same variance, the OLSE $\widehat{\beta}$ is the best linear unbiased estimator of β. Under heteroskedasticity, however, it is no longer efficient, but it remains unbiased, consistent and asymptotically normal.

In order to perform hypothesis testing inference on the regression parameters it is necessary to estimate Ψ, the covariance matrix of $\widehat{\beta}$. When all errors share the same variance, Ψ can be easily estimated as

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

where $\widehat{\sigma}^2 = (y - X\widehat{\beta})'(y - X\widehat{\beta})/(n-p) = \widehat{\varepsilon}'\widehat{\varepsilon}/(n-p)$ is an unbiased estimator of the common error variance. Here,

$$\widehat{\varepsilon} = y - \widehat{y} = (I - H)y = My, \tag{3.2.1}$$

$H = X(X'X)^{-1}X'$ being an $n \times n$ symmetric and idempotent matrix and $M = I - H$, where I is the n-dimensional identity matrix. (H is known as the 'hat matrix' since $Hy = \widehat{y}$.) The diagonal elements of H ($h_1, \ldots, h_n$) assume values in the standard unit interval $(0, 1)$ and their sum (the rank of H) equals p, so that $\overline{h} = n^{-1}\sum_{t=1}^{n} h_t = p/n$. It is noteworthy that h_t is frequently used as a measure of the leverage of the tth observation and that observations such that $h_t > 2p/n$ or $h_t > 3p/n$ are taken to be leverage points; see, e.g., Davidson and MacKinnon (1993).

Our interest lies in the estimation of Ψ in situations where the error variances are not taken to be constant, i.e., we wish to estimate the covariance matrix of $\widehat{\beta}$ given by $(X'X)^{-1}X'\Omega X(X'X)^{-1}$ in a consistent fashion regardless of whether homoskedasticity holds. White (1980) has observed that Ψ can be consistently estimated as long as $X'\Omega X$ is consistently estimated; that is, it is not necessary to obtain a consistent estimator for Ω (which has n unknown elements), it is only necessary to consistently estimate $X'\Omega X$ (which has $p(p+1)/2$ unknown elements,

regardless of the sample size). White (1980) then proposed the following estimator for Ψ:

$$\text{HC0} = \widehat{\Psi}_0 = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1} = P\widehat{\Omega}P' = PE_0\widehat{\Omega}P',$$

where $\widehat{\Omega} = \text{diag}\{\widehat{\varepsilon}_t^2\}$ and $E_0 = I$.

White's estimator (HC0) is consistent under both homoskedasticity and heteroskedasticity of unknown form: $\text{plim}(\widehat{\Psi}_0\Psi^{-1}) = I_p$, where I_p denotes the p-dimensional identity matrix and plim denotes limit in probability. Nonetheless, it can be substantially biased in small samples. In particular, HC0 is typically 'too optimistic', i.e., it tends to underestimate the true variances in finite samples; thus, the associated tests (i.e., tests whose statistics employ HC0) tend to be liberal. The problem is more severe when the data include leverage points; see, e.g., Chesher and Jewitt (1987).

A few variants of HC0 were proposed in the literature. They include finite-sample corrections in the estimation of Ω and are given by:

i (Hinkley, 1977) $\text{HC1} = \widehat{\Psi}_1 = P\widehat{\Omega}_1P' = PE_1\widehat{\Omega}P'$, where $E_1 = (n/(n-p))I$;

ii (Horn, Horn and Duncan, 1975) $\text{HC2} = \widehat{\Psi}_2 = P\widehat{\Omega}_2P' = PE_2\widehat{\Omega}P'$, where

$$E_2 = \text{diag}\{1/(1-h_t)\};$$

iii (Davidson and MacKinnon, 1993) $\text{HC3} = \widehat{\Psi}_3 = P\widehat{\Omega}_3P' = PE_3\widehat{\Omega}P'$, where

$$E_3 = \text{diag}\{1/(1-h_t)^2\};$$

iv (Cribari–Neto, 2004) $\text{HC4} = \widehat{\Psi}_4 = P\widehat{\Omega}_4P' = PE_4\widehat{\Omega}P'$, where

$$E_4 = \text{diag}\{1/(1-h_t)^{\delta_t}\}, \quad \delta_t = \min\{4, nh_t/p\}.$$

Additionally, Qian and Wang (2001) proposed an alternative estimator for $\text{cov}(\widehat{\beta})$, which we shall denote as $\widehat{V}_1$. It was obtained by bias-correcting HC0 and then modifying the resulting estimator so that it becomes unbiased under homoskedasticity. Let $C_t = X(X'X)^{-1}x_t'$, $t = 1, \ldots, n$, i.e., C_t denotes the tth column of H (hat matrix); here, x_t is the tth row of X. Also, let

$$D_1 = \text{diag}\{d_{1t}\} = \text{diag}\{(\widehat{\varepsilon}_t^2 - \widehat{b}_t)g_{tt}\},$$

where

$$g_{tt} = (1 + C_t'KC_t - 2h_t^2)^{-1}$$

and

$$\widehat{b}_t = C_t'(\widehat{\Omega} - 2\widehat{\varepsilon}_t^2 I)C_t;$$

here, $K = (H)_d$, i.e., $K = \text{diag}\{h_t\}$.

Their estimator is then given by $\widehat{V}_1 = PD_1P'$. We note that D_1 can be expressed in matrix form as

$$D_1 = [\widehat{\Omega} - \{H\widehat{\Omega}(H - 2I)\}_d]G,$$

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

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

Let c be a given p-vector of constants. We write the variance of a linear combination of the elements of $\widehat{\beta}$ as

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

We can estimate Ψ using HCi, $i = 0,\ldots,4$, to obtain the following estimator for Φ:

$$\widehat{\Phi}_i = c'\widehat{\Psi}_i c = c'P\widehat{\Omega}_i P'c = c'PE_i\widehat{\Omega}P'c, \quad i = 0,\ldots,4.$$

Let

$$V_i = (v_i v_i')_d, \tag{3.3.1}$$

where $v_i = E_i^{1/2}P'c$, $i = 0,\ldots,4$. We can then write

$$\widehat{\Phi}_i = v_i'\widehat{\Omega}v_i$$

and, since $\widehat{\Omega} = (\widehat{\varepsilon}\widehat{\varepsilon}')_d$, it is possible to show that

$$\widehat{\Phi}_i = \widehat{\varepsilon}'V_i\widehat{\varepsilon}, \quad i = 0,\ldots,4.$$

It is then clear that $\widehat{\Phi}_i$ can be written as a quadratic form in the vector of residuals, which have zero mean and are correlated. We shall write $\widehat{\Phi}_i$ as a quadratic form in a random vector z of zero mean and unit covariance. Following Cribari–Neto, Ferrari and Cordeiro (2000), it is possible to write

$$\widehat{\Phi}_i = z'G_i z,$$

where $\mathbb{E}[z] = 0$, $\mathrm{cov}(z) = I$ and

$$G_i = \Omega^{1/2}(I-H)V_i(I-H)\Omega^{1/2}.$$

Consider now covariance matrix estimation using the estimator proposed by Qian and Wang (2001): $\widehat{\Phi}_{QW_1} = c'\widehat{\Psi}_{QW_1}c = c'\widehat{V}_1 c$. Hence,

$$\widehat{\Phi}_{QW_1} = c'\widehat{V}_1 c = c'PD_1P'c,$$

where, as before, $D_1 = [\widehat{\Omega} - \{H\widehat{\Omega}(H-2I)\}_d]G$.

Let A be an $n\times n$ diagonal matrix and let $M^{(1)}(A) = \{HA(H-2I)\}_d$. Therefore,

$$D_1 = \widehat{\Omega}G - M^{(1)}(\widehat{\Omega})G.$$

Also, let $w = G^{1/2}P'c$ and $W = (ww')_d$. It then follows that

$$\widehat{\Phi}_{QW_1} = w'\widehat{\Omega}w - w'M^{(1)}(\widehat{\Omega})w.$$

Since $w'\widehat{\Omega}w = \widehat{\varepsilon}'W\widehat{\varepsilon}$ and $w'M^{(1)}(\widehat{\Omega})w = \widehat{\varepsilon}'M^{(1)}(W)\widehat{\varepsilon}$ (see Cribari-Neto, Ferrari and Cordeiro, 2000), then

$$\widehat{\Phi}_{QW_1} = \widehat{\varepsilon}'[W - M^{(1)}(W)]\widehat{\varepsilon} = \widehat{\varepsilon}'V_{QW_1}\widehat{\varepsilon},$$

where

$$V_{QW_1} = W - M^{(1)}(W). \tag{3.3.2}$$

We shall now write $\widehat{\Phi}_{QW_1}$ as a quadratic form in a zero mean and unit covariance random vector. It can be shown, after some algebra, that

$$\widehat{\Phi}_{QW_1} = z' G_{QW_1} z,$$

where $\mathbb{E}[z] = 0$, $\mathrm{cov}(z) = I$ and

$$G_{QW_1} = \Omega^{1/2}(I-H)V_{QW_1}(I-H)\Omega^{1/2}.$$

3.4 Approximate inference using quasi-t tests

We shall now consider quasi-t test statistics based on standard errors obtained from the HCCMEs described in Section 3.2. The interest lies in testing the null hypothesis $\mathcal{H}_0 : c'\beta = \eta$ against a two-sided alternative hypothesis, where c is a given p-vector and η is a given scalar.

The quasi-t statistic given by

$$t = \frac{c'\widehat{\beta} - \eta}{\sqrt{\widehat{\mathrm{var}}(c'\widehat{\beta})}},$$

where $\sqrt{\widehat{\mathrm{var}}(c'\widehat{\beta})}$ is a standard error obtained from one of the HCCMEs described in this chapter, does not have, under the null hypothesis, a Student t distribution. Nonetheless, it is easy to show that, under $\mathcal{H}_0$, the limiting distribution of t is $\mathcal{N}(0,1)$. As a consequence, the limiting null distribution of t^2 is χ^2_1.

Note that

$$\widehat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta+\varepsilon) = \beta + (X'X)^{-1}X'\varepsilon.$$

Thus, when $\varepsilon \sim \mathcal{N}(0,\Omega)$,

$$\widehat{\beta} = \beta + (X'X)^{-1}X'\Omega^{1/2}z,$$

where $z \sim \mathcal{N}(0,I)$, and we can thus write t^2 as a ratio of two quadratic forms in a Gaussian zero mean and unit covariance random vector. The numerator of t^2 can be written as

$$\begin{aligned}
(c'\widehat{\beta}-\eta)^2 &= \{c'\beta + c'(X'X)^{-1}X'\Omega^{1/2}z - \eta\}'\{c'\beta + c'(X'X)^{-1}X'\Omega^{1/2}z - \eta\} \\
&= \{(c'\beta-\eta) + c'(X'X)^{-1}X'\Omega^{1/2}z\}'\{(c'\beta-\eta) + c'(X'X)^{-1}X'\Omega^{1/2}z\} \\
&= (c'\beta-\eta)'(c'\beta-\eta) + 2(c'\beta-\eta)c'(X'X)^{-1}X'\Omega^{1/2}z \\
&+ z'\Omega^{1/2}X(X'X)^{-1}cc'(X'X)^{-1}X'\Omega^{1/2}z.
\end{aligned}$$

In Section 3.3 we wrote $\widehat{\Phi} = \widehat{\mathrm{var}}(c'\widehat{\beta})$ as a quadratic form in a zero mean and unit covariance random vector for five HCCMEs:

(i) $\widehat{\Phi}_i = z'G_i z$, where $G_i = \Omega^{1/2}(I-H)V_i(I-H)\Omega^{1/2}$, for estimators HC$i$, $i = 0,\ldots,4$;

(ii) $\widehat{\Phi}_{QW_1} = z'G_{QW_1}z$, where $G_{QW_1} = \Omega^{1/2}(I-H)V_{QW_1}(I-H)\Omega^{1/2}$ for $\widehat{V}_1$.

(Note that V_i and V_{QW_1} are given in (3.3.1) and (3.3.2), respectively.)

Hence,

$$t^2 = \frac{z'Rz}{z'G_{(\cdot)}z} + \frac{(c'\beta-\eta)'(c'\beta-\eta)+2(c'\beta-\eta)c'(X'X)^{-1}X'\Omega^{1/2}z}{z'G_{(\cdot)}z}, \tag{3.4.1}$$

where $R = \Omega^{1/2}X(X'X)^{-1}cc'(X'X)^{-1}X'\Omega^{1/2}$, $G_{(\cdot)} = G_i$, $i = 0,\ldots,4$, for HCi, and $G_{(\cdot)} = G_{QW_1}$ for $\widehat{V}_1$.

When $c'\beta = \eta$, the second term on the right hand side of (3.4.1) vanishes and, as a result,

$$\Pr(t^2 \leq \gamma | c'\beta = \eta) = \Pr_0(z'Rz/z'G_{(\cdot)}z \leq \gamma), \tag{3.4.2}$$

where $\Pr_0$ denotes 'probability under the null hypothesis'.

In the next section, we shall use Imhof's (1961) numerical integration algorithm to compute the exact null distribution function of t^2. The algorithm allows the evaluation of probabilities of ratios of quadratic forms in a vector of normal variates. To that end, we shall add the assumption that the errors are normally distributed, i.e., we shall assume that $\varepsilon_t \sim \mathcal{N}(0,\sigma_t^2), t = 1,\ldots,n$. The numerical evaluation will follow from comparing the exact null distributions of the test statistics obtained using different heteroskedasticity-robust standard errors and the asymptotic null distribution (χ_1^2) used in the test.

3.5 Exact numerical evaluation

We shall now use Imhof's (1961) numerical integration algorithm to evaluate (3.4.2), i.e., to compute the exact null distributions of different quasi-t test statistics t^2 (test statistics that are based on different standard errors). These exact distributions shall be compared to the null limiting distribution (χ_1^2) from which critical values are obtained. All numerical evaluations were carried out using the `0x` matrix programming language (Doornik, 2001). We report results for different values of γ.

The following regression model was used in the evaluation:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1,\ldots,n,$$

where ε_t, $t = 1,\ldots,n$, is normally distributed with mean zero and variance $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$; also, $\mathbb{E}[\varepsilon_t\varepsilon_s] = 0\ \forall t \neq s$. We use

$$\lambda = \max\{\sigma_t^2\}/\min\{\sigma_t^2\}$$

as a measure of the heteroskedasticity strength. When the errors are homoskedastic, it follows that $\lambda = 1$; on the other hand, the larger is λ, the stronger the heteroskedasticity.

The null hypothesis under test is $\mathcal{H}_0 : \beta_1 = 0$, i.e., $\mathcal{H}_0 : c'\beta = \eta$ with $c' = (0,1)$ and $\eta = 0$. The test statistic is given by

$$t^2 = \widehat{\beta}_1^2/\widehat{\text{var}}(\widehat{\beta}_1),$$

where $\widehat{\text{var}}(\widehat{\beta}_1)$ is the $(2,2)$ element of a given HCCME.

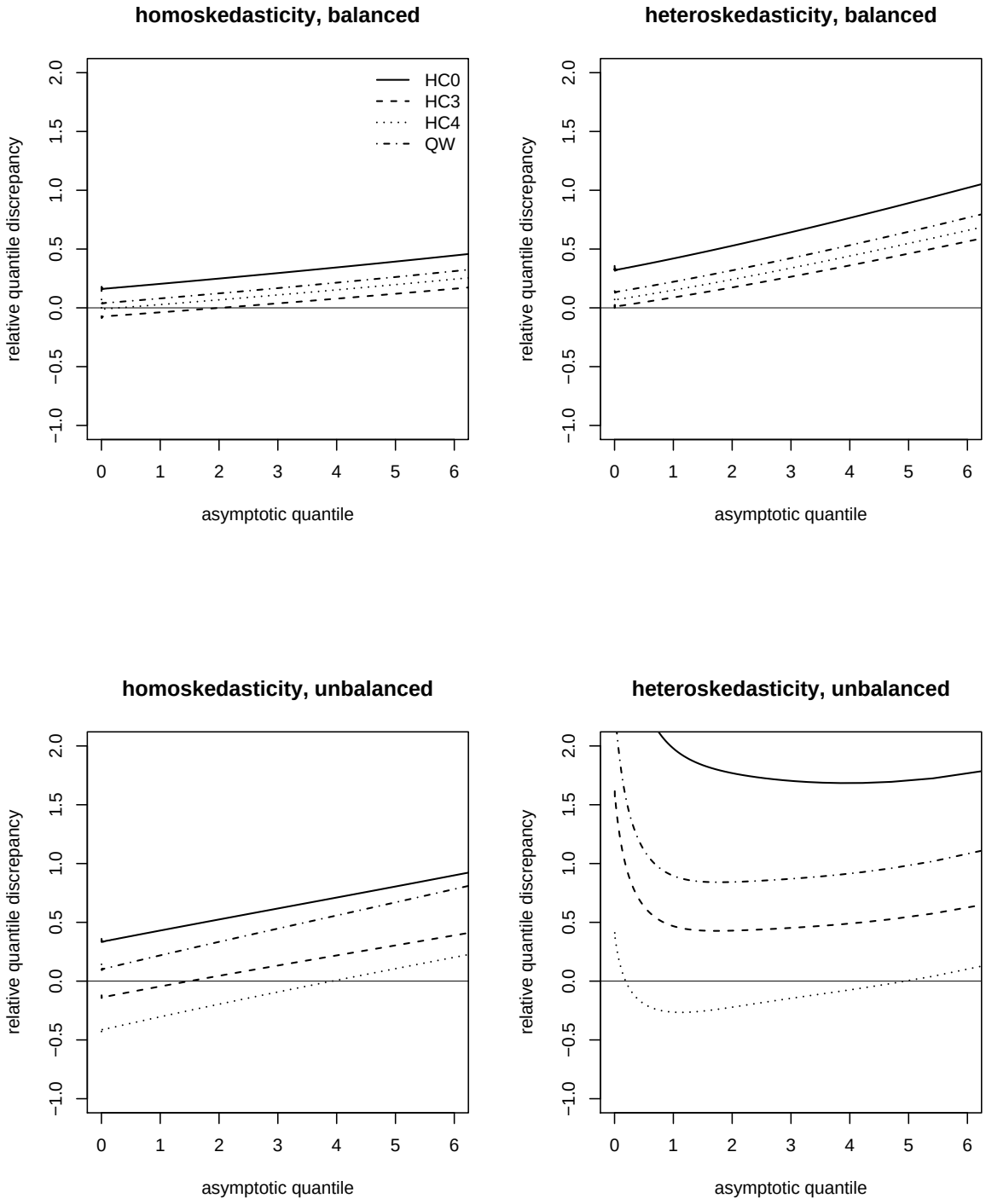

Figure 3.1 Relative quantile discrepancy plots, $n = 25$: balanced and unbalanced regression designs, homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 100$).

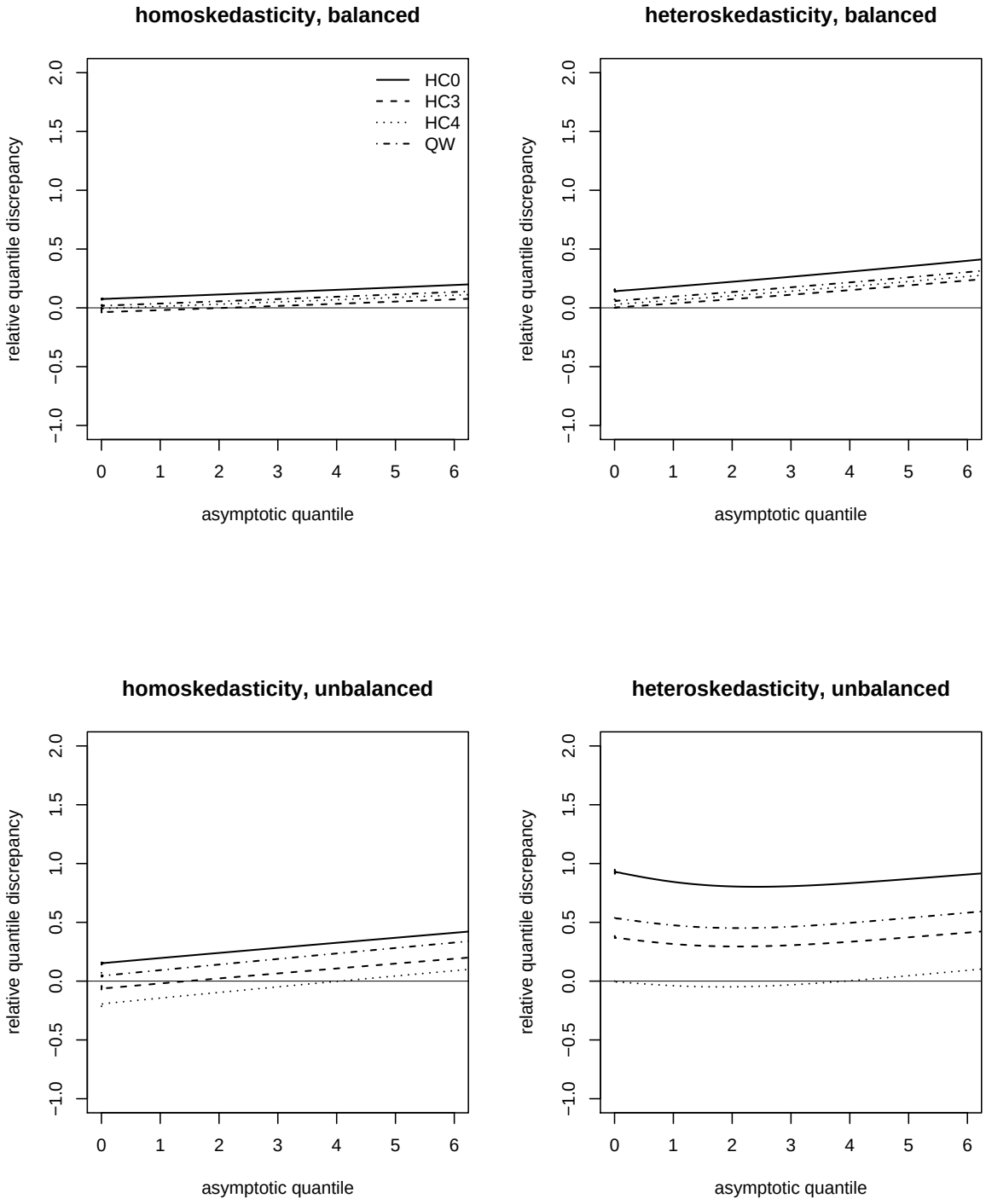

Figure 3.2 Relative quantile discrepancy plots, $n = 50$: balanced and unbalanced regression designs, homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 100$).

Table 3.1 Maximal leverages for the two regression designs.

	$\mathcal{U}(0,1)$	t_3	threshold	
n	$h_{\max}$	$h_{\max}$	$2p/n$	$3p/n$
25	0.143	0.350	0.16	0.24
50	0.071	0.175	0.08	0.12

We set the sample size at $n = 25$ and then replicate the covariate values to obtain a sample of 50 observations. We consider two regression designs, namely: (i) without a leverage point (regressors randomly generated from $\mathcal{U}(0,1)$), and (ii) with leverage points (regressors randomly generated from t_3); see Table 3.1.

Figures 3.1 and 3.2 plot the relative quantile discrepancies versus the corresponding asymptotic quantiles for $n = 25$ and $n = 50$, respectively. Relative quantile discrepancy is defined as the difference between exact quantiles and asymptotic quantiles divided by the latter. The closer to zero the relative quantile discrepancy, the better the approximation of the exact null distribution of the test statistic by the limiting χ^2_1 distribution. (All panels include a horizontal reference line indicating no relative quantile discrepancy.) We present results for test statistics that use HC0, HC3, HC4 and $\widehat{V}_1$ (QW) standard errors under both homoskedasticity and heteroskedasticity and for the two regression designs (balanced and unbalanced, i.e., without and with leverage points). When the error variances are not constant, $\lambda \approx 100$ (the largest standard deviation is approximately 10 times larger than the smallest one).

We note from Figures 3.1 and 3.2 that the HC0-based test is the worst performing test in all situations; its exact null distribution is poorly approximated by the limiting null distribution, more so under heteroskedasticity, especially when the regression design is unbalanced (leverage points in the data). In the top two panels (balanced design), the HC3-based test is the best performing test, closely followed by HC4. In the bottom panels (unbalanced design, the most critical situation), the HC4 test clearly outperforms all other tests; HC3 is the runner up, followed by $\widehat{V}_1$ and, finally, by HC0. It is noteworthy that, when the sample size is small ($n = 25$) and the data are leveraged and heteroskedastic (Figure 3.1, lower right panel), the HC0 ($\widehat{V}_1$ and HC3) test statistic 0.95 quantile is nearly three (2 and 1.5) times larger than the asymptotic 0.95 quantile (3.841). As a consequence, the test can be expected to be quite liberal at the 5% nominal level.

Table 3.2 presents the computed probabilities $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ for the different test statistics for values of γ given by the 0.90, 0.95 and 0.99 quantiles of the limiting null distribution (χ^2_1) ($n = 50$). The closer the corresponding probabilities are to 0.90, 0.95 and 0.99, the better the approximation at these quantiles. When the data include no leverage point, there are no noticeable differences among the tests. In the unbalanced regression design, however, the differences in the computed probabilities can be large. For instance, in the presence of high leverage observations and under heteroskedasticity, the distribution functions of the HC0-, HC3-, HC4- and $\widehat{V}_1$-based test statistics evaluated at 3.841 (the 0.95 quantile of χ^2_1) are 0.855, 0.914, 0.950, 0.896, respectively.

We shall now perform a numerical evaluation using real (not simulated) data. In particular,

Table 3.2 $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ for $n = 50$; γ equals the 0.90, 0.95 and 0.99 quantiles of χ^2_1; test statistics based on four different standard errors.

		balanced				unbalanced			
λ	Pr	HC0	HC3	HC4	$\widehat{V}_1$	HC0	HC3	HC4	$\widehat{V}_1$
1	0.90	0.880	0.898	0.893	0.889	0.859	0.892	0.909	0.874
	0.95	0.933	0.947	0.943	0.940	0.916	0.940	0.951	0.927
	0.99	0.982	0.987	0.986	0.984	0.973	0.982	0.986	0.977
≈ 100	0.90	0.862	0.885	0.880	0.875	0.777	0.852	0.906	0.828
	0.95	0.918	0.935	0.932	0.928	0.855	0.914	0.950	0.896
	0.99	0.973	0.981	0.979	0.978	0.944	0.972	0.986	0.964

we use the data in Greene (1997, Table 12.1, p. 541).[2] The variable of interest (y) is the per capita spending on public schools and the independent variables, x and x^2, are the per capita income by state in 1979 in the United States and its square; income is scaled by 10^{-4}.[3] The regression model is

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t, \quad t = 1, \ldots, 50.$$

The errors are uncorrelated, each ε_t being normally distributed with zero mean and variance $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$. The interest lies in the test of $\mathcal{H}_0 : \beta_2 = 0$. When $\alpha_1 = \alpha_2 = 0$, then $\lambda = 1$ (homoskedasticity); when $\alpha_1 = 0$ and $\alpha_2 = 4.6$, then $\lambda \approx 50$ (heteroskedasticity). The covariate values were replicated twice and three times to yield samples of size 100 and 150, respectively. Note that by replicating the covariate values to form larger samples we guarantee that the strength of heteroskedasticity remains unchanged as n grows.

Table 3.3 presents the leverage points in the three cases we consider. We work with the complete data set (case 1), we remove the observation with the largest h_t, Alaska, thus reducing the sample size to 49 observations (case 2) and, finally, we remove all three leverage points (case 3). In the latter case, a new (and mild) leverage point emerges.

Figure 3.3 presents relative quantile discrepancy plots for three sample sizes, namely: $n = 50, 100, 150$ (case 1). Again, the evaluation was carried out under both equal and unequal error variances, and the tests considered are those whose statistics employ standard errors obtained from HC0, HC3, HC4 and $\widehat{V}_1$.

When $n = 50$ and all errors share the same variance ($\lambda = 1$), the exact null distribution of the HC3-based test statistic is better approximated by the limiting null distribution (χ^2_1) than those of the competing statistics; overall, the HC4 test is the second best performing test. We note that the test based on $\widehat{V}_1$ is not uniformly better than that based on HC0; the latter displays superior behavior for asymptotic quantiles in excess of (approximately) 4. As the sample size increases (to 100 and then to 150), the relative quantile discrepancies of all test statistics shrink toward zero.

[2] The original source of the data is the U.S. Department of Commerce.

[3] Wisconsin has been dropped from the data set since it had missing data, and Washington D.C. was included, hence $n = 50$.

Table 3.3 Leverage measures, thresholds for detecting leverage points and ratio between $h_{\max}$ and $3p/n$; education data.

case	n	h_t	$2p/n$	$3p/n$	$h_{\max}/(3p/n)$
1	50	0.651 ($h_{\max}$) 0.208 0.200	0.12	0.18	3.62
2	49	0.562 ($h_{\max}$) 0.250	0.122	0.184	3.05
3	47	0.209 ($h_{\max}$)	0.128	0.191	1.09

Under heteroskedasticity, the finite-sample behavior of all tests deteriorate, since the exact distributions of the test statistics become more poorly approximated by the limiting null distribution, from which critical values are obtained. Overall, the HC4 test is the best performing test, especially at the 3.841 quantile (5% nominal level). The HC3 test comes in second place. We also note that the exact distributions of the test statistics that use standard errors from HC0 and $\widehat{V}_1$ are very poorly approximated by χ^2_1 when $n = 50$. Indeed, their exact quantiles can be over five times larger than the corresponding χ^2_1 quantiles!

Table 3.4 contains the computed probabilities $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ for the different test statistics at the 0.95 asymptotic quantile ($\gamma = 3.841$) under both homoskedasticity and heteroskedasticity and for cases 1, 2 and 3 ($n = 50, 49, 47$, respectively). The closer these probabilities are to 0.95, the better the approximation used in the test. In cases 1 and 2 (leveraged data) and under equal error variances, the HC3 test outperforms the competition, being followed by HC4. It is also noteworthy the poor behavior of the HC0 and $\widehat{V}_1$ tests. In case 3 (and with $\lambda = 1$), the probabilities obtained using HC3- and HC4-based tests are quite close to the desired probability (0.95). Under heteroskedasticity, HC4 is clearly the best performing test. We note the dreadful behavior of the HC0 and $\widehat{V}_1$ tests when the error variances are not constant and $n = 50$: the computed probabilities are around 0.61 and 0.73, respectively, whereas the desired figure would be 0.95!

Table 3.4 $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ using data on public spending on education, $n = 50, 49$ and 47 (cases 1, 2, and 3, respectively); γ equals the 0.95 quantile of χ^2_1; test statistics based on four different standard errors.

	$\lambda = 1$			$\lambda \approx 50$		
statistic	$n = 50$	$n = 49$	$n = 47$	$n = 50$	$n = 49$	$n = 47$
HC0	0.8593	0.8747	0.9235	0.6113	0.6971	0.8830
HC3	0.9410	0.9408	0.9484	0.8549	0.8943	0.9284
HC4	0.9789	0.9744	0.9497	0.9528	0.9593	0.9465
$\widehat{V}_1$	0.8758	0.8817	0.9354	0.7286	0.7748	0.9059

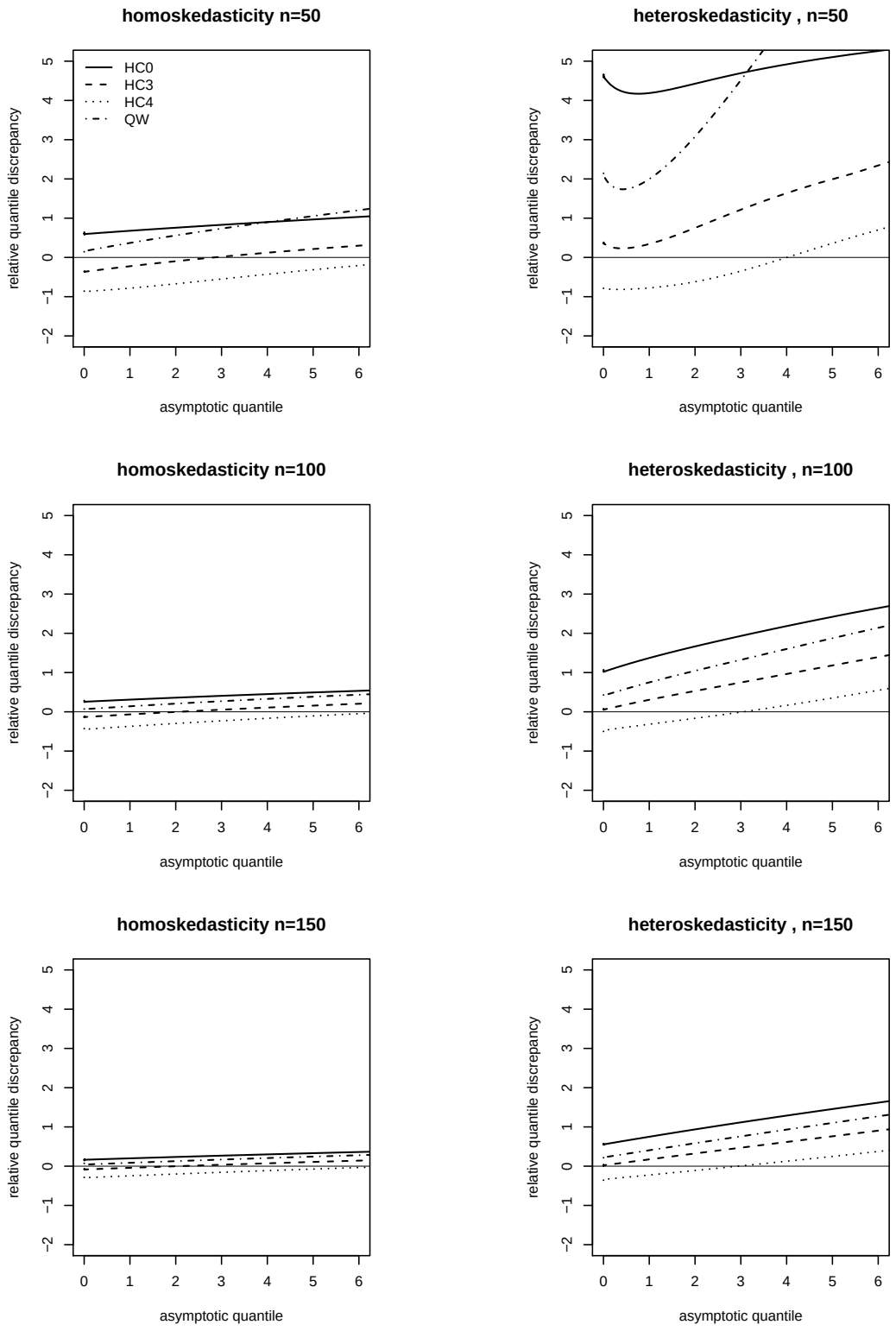

Figure 3.3 Relative quantile discrepancy plots using data on public spending on education, $n = 50, 100, 150$: homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 50$).

3.6 An alternative standard error

Qian and Wang (2001) have proposed an additional alternative HCCME, which we shall denote as $\widehat{V}_2$. It is given by $\widehat{V}_2 = PD_2P'$, where

$$D_2 = \text{diag}\{d_{2t}\} = \text{diag}\{\widehat{\varepsilon}_t^2 + \widehat{\sigma}^2 h_t\} = \widehat{\Omega} + \widehat{\sigma}^2 (H)_d = \widehat{\Omega} + \widehat{\sigma}^2 K;$$

as before, $K = (H)_d$, i.e., K is obtained by setting all nondiagonal elements of H (the hat matrix) equal to zero.

We will show that, like their other estimator ($\widehat{V}_1$), this HCCME is unbiased under equal error variances. It follows from (3.2.1) that

$$\mathbb{E}[\widehat{\varepsilon}] = \mathbb{E}[My] = \mathbb{E}[M(X\beta + \varepsilon)] = \mathbb{E}[M\varepsilon] = 0,$$

since $MX = (I - H)X = 0$. Given that $\widehat{\Omega} = \text{diag}\{\widehat{\varepsilon}_t^2\} = (\widehat{\varepsilon}\widehat{\varepsilon}')_d$, and that

$$\begin{aligned} \mathbb{E}[\widehat{\varepsilon}\widehat{\varepsilon}'] &= \text{cov}(\widehat{\varepsilon}) + \mathbb{E}[\widehat{\varepsilon}]\mathbb{E}[\widehat{\varepsilon}'] \\ &= \text{cov}(M\varepsilon) \\ &= M\Omega M, \end{aligned}$$

we have

$$\mathbb{E}[\widehat{\Omega}] = \mathbb{E}[(\widehat{\varepsilon}\widehat{\varepsilon}')_d] = \{(I - H)\Omega(I - H)\}_d.$$

Note that

$$\mathbb{E}[\widehat{\Omega}] = \{(I - H)\Omega(I - H)\}_d = \{H\Omega H - 2H\Omega + \Omega\}_d = \{H\Omega(H - 2I)\}_d + \Omega = M^{(1)}(\Omega) + \Omega,$$

where $M^{(1)}(\Omega) = \{H\Omega(H - 2I)\}_d$. Under homoskedasticity,

$$\begin{aligned} \mathbb{E}[\widehat{\Omega}] &= M^{(1)}(\sigma^2 I) + \sigma^2 I \\ &= \{H\sigma^2 I(H - 2I)\}_d + \sigma^2 I \\ &= \sigma^2\{HH - 2H\}_d + \sigma^2 I \\ &= \sigma^2\{-H\}_d + \sigma^2 I \\ &= -\sigma^2 K + \sigma^2 I \\ &= \sigma^2(I - K). \end{aligned}$$

Using the definition of D_2, it follows that

$$\begin{aligned} \mathbb{E}[D_2] &= \mathbb{E}[\widehat{\Omega}] + \mathbb{E}[\widehat{\sigma}^2]K \\ &= \sigma^2(I - K) + \sigma^2 K \\ &= \sigma^2 I. \end{aligned}$$

Thus, when all errors share the same variance,

$$\widehat{V}_2 = PD_2P'$$

is unbiased for Ψ. We note that $\widehat{V}_2$ is a modified version of HC0; the modification is such that the estimator becomes unbiased under homoskedasticity.

Based on $\widehat{V}_2$ the authors defined a family of HCCMEs indexed by the n-vector $f = (f_1, \ldots, f_n)'$ by making

$$d_{2t}(f_t) = f_t\widehat{\varepsilon}_t^2 + \widehat{\sigma}^2\{1 - f_t(1-h_t)\}, \quad t = 1, \ldots, n. \tag{3.6.1}$$

Here,

$$D_2(f) = \text{diag}\{d_{2t}(f_t)\} = A\widehat{\Omega} + \widehat{\sigma}^2(I - A\Lambda),$$

where

$$A = \text{diag}\{f_t\} \tag{3.6.2}$$

and

$$\Lambda = \text{diag}\{1 - h_t\} = I - K. \tag{3.6.3}$$

When the error variances are all equal,

$$\begin{aligned}
\mathbb{E}[D_2(f)] &= \mathbb{E}[A\widehat{\Omega} + \widehat{\sigma}^2(I - A(I-K))] \\
&= A\mathbb{E}[\widehat{\Omega}] + \mathbb{E}[\widehat{\sigma}^2](I - A(I-K)) \\
&= A\sigma^2(I-K) + \sigma^2(I - A + AK) \\
&= \sigma^2 I.
\end{aligned}$$

That is, the family of estimators

$$\widehat{V}_2(f) = PD_2(f)P'$$

is unbiased under homoskedasticity for any choice of f that depends only on the regressors. Note that

(i) when $f = (1, \ldots, 1)'$, we obtain $\widehat{V}_2$;

(ii) when $f = (0, \ldots, 0)'$, we obtain the OLSE of Ψ used under homoskedasticity;

(iii) when $f = ((1-\widehat{\varepsilon}_1^2/\widehat{\sigma}^2)/(1-h_1-\widehat{\varepsilon}_1^2/\widehat{\sigma}^2), \ldots, (1-\widehat{\varepsilon}_n^2/\widehat{\sigma}^2)/(1-h_n-\widehat{\varepsilon}_n^2/\widehat{\sigma}^2))'$, we obtain HC0;[4]

(iv) when $f = (1/(1-h_1), \ldots, 1/(1-h_n))'$, we obtain HC2.

In order to simplify the notation, we shall hereafter denote $D_2(f)$ by D_2 and $\widehat{V}_2(f)$ by $\widehat{V}_2$.

To achieve a reduction in the variability induced by the presence of leverage points in the data, Qian and Wang (2001) suggested using

$$f_t = 1 - ah_t, \quad t = 1, \ldots, n, \tag{3.6.4}$$

in (3.6.1), where a is a real constant. Their suggestion is to use $a = 2$ when the goal is to reduce the mean squared error (MSE), and to use a smaller value of a (even zero) when the chief concern is bias. We shall denote this estimator for a given value of a by $\widehat{V}_2(a)$.

[4]Note that, here (HC0), c depends on y through $\widehat{\sigma}^2$, and it is not possible to establish the unbiasedness of the resulting estimator.

In Section 3.3 we have shown that the estimator of Φ, the variance of $c'\widehat{\beta}$, that uses $\widehat{V}_1$ is given by $\widehat{\Phi}_{QW_1} = \widehat{\varepsilon}'[W - M^{(1)}(W)]\widehat{\varepsilon} = \widehat{\varepsilon}' V_{QW_1}\widehat{\varepsilon}$, where $V_{QW_1} = W - M^{(1)}(W)$. By writing $\widehat{\Phi}_{QW_1}$ as a quadratic form in a vector of uncorrelated, zero mean and unit variance random variables we obtained $\widehat{\Phi}_{QW_1} = z'G_{QW_1}z$, where $\mathbb{E}[z] = 0$, $\text{cov}(z) = I$ and

$$G_{QW_1} = \Omega^{1/2}(I - H)V_{QW_1}(I - H)\Omega^{1/2}.$$

When $\widehat{V}_2 = \widehat{\Psi}_{QW_2} = PD_2P'$ is used in the estimation of Φ, we obtain

$$\widehat{\Phi}_{QW_2} = c'\widehat{\Psi}_{QW_2}c = c'PD_2P'c,$$

where $D_2 = A\widehat{\Omega} + \widehat{\sigma}^2(I - A\Lambda)$; A and Λ are as defined in (3.6.2) and (3.6.3), respectively.

Let $L = (n - p)^{-1}(I - A\Lambda)$. Then,[5]

$$D_2 = \widehat{\varepsilon}'\widehat{\varepsilon}L + A\widehat{\Omega}.$$

Let $\ell = L^{1/2}P'c$, $v^* = A^{1/2}P'c$ and $V^* = (v^*v^{*'})_d$. Note that

$$v^{*'}\widehat{\Omega}v^* = v^{*'}[(\widehat{\varepsilon}\widehat{\varepsilon}')_d]v^* = \widehat{\varepsilon}'[(v^*v^{*'})_d]\widehat{\varepsilon} = \widehat{\varepsilon}'V^*\widehat{\varepsilon}$$

and

$$\ell'(\widehat{\varepsilon}'\widehat{\varepsilon}I)\ell = \widehat{\varepsilon}'[\ell'\ell I]\widehat{\varepsilon}.$$

Therefore,

$$\begin{aligned}
\widehat{\Phi}_{QW_2} &= c'PD_2P'c \\
&= c'P[\widehat{\varepsilon}'\widehat{\varepsilon}L + A\widehat{\Omega}]P'c \\
&= c'P[L^{1/2}(\widehat{\varepsilon}'\widehat{\varepsilon}I)L^{1/2} + A^{1/2}\widehat{\Omega}A^{1/2}]P'c \\
&= c'PL^{1/2}(\widehat{\varepsilon}'\widehat{\varepsilon}I)L^{1/2}P'c + c'PA^{1/2}\widehat{\Omega}A^{1/2}P'c \\
&= \ell'(\widehat{\varepsilon}'\widehat{\varepsilon}I)\ell + v^{*'}\widehat{\Omega}v^* \\
&= \widehat{\varepsilon}'(\ell'\ell I)\widehat{\varepsilon} + \widehat{\varepsilon}'V^*\widehat{\varepsilon} \\
&= \widehat{\varepsilon}'(\ell'\ell I + V^*)\widehat{\varepsilon} \\
&= \widehat{\varepsilon}'V_{QW_2}\widehat{\varepsilon},
\end{aligned}$$

where $V_{QW_2} = \ell'\ell I + V^*$.

We then write $\widehat{\Phi}_{QW_2}$ as a quadratic form in a vector of uncorrelated, zero mean and unit variance random variables (z) as

$$\widehat{\Phi}_{QW_2} = z'G_{QW_2}z,$$

where

$$G_{QW_2} = \Omega^{1/2}(I - H)V_{QW_2}(I - H)\Omega^{1/2}.$$

[5]Recall that $\widehat{\sigma}^2 = (n - p)^{-1}\widehat{\varepsilon}'\widehat{\varepsilon}$.

3.7 A numerical evaluation of quasi-t tests based on $\widehat{V}_1$ and $\widehat{V}_2$

At the outset, we shall use numerical integration to determine the optimal value of a in (3.6.4) for hypothesis testing inference. The following regression model is used:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, \ldots, n,$$

where each $\varepsilon_t, t = 1, \ldots, n$, is normally distributed with zero mean and variance $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$; the errors are uncorrelated, i.e., $\mathbb{E}[\varepsilon_t \varepsilon_s] = 0 \; \forall t \neq s$. As in Section 3.5, 25 covariate values were obtained and replicated once to yield $n = 50$, and two regression designs are used: balanced and unbalanced, as described in Table 3.1. The interest lies in testing $\mathcal{H}_0 : \beta_1 = 0$, i.e., $\mathcal{H}_0 : c'\beta = \eta$ with $c' = (0, 1)$ and $\eta = 0$, using the following test statistic:

$$t^2 = \widehat{\beta}_1^2 / \widehat{\text{var}}(\widehat{\beta}_1).$$

As noted earlier, its limiting null distribution is χ_1^2.

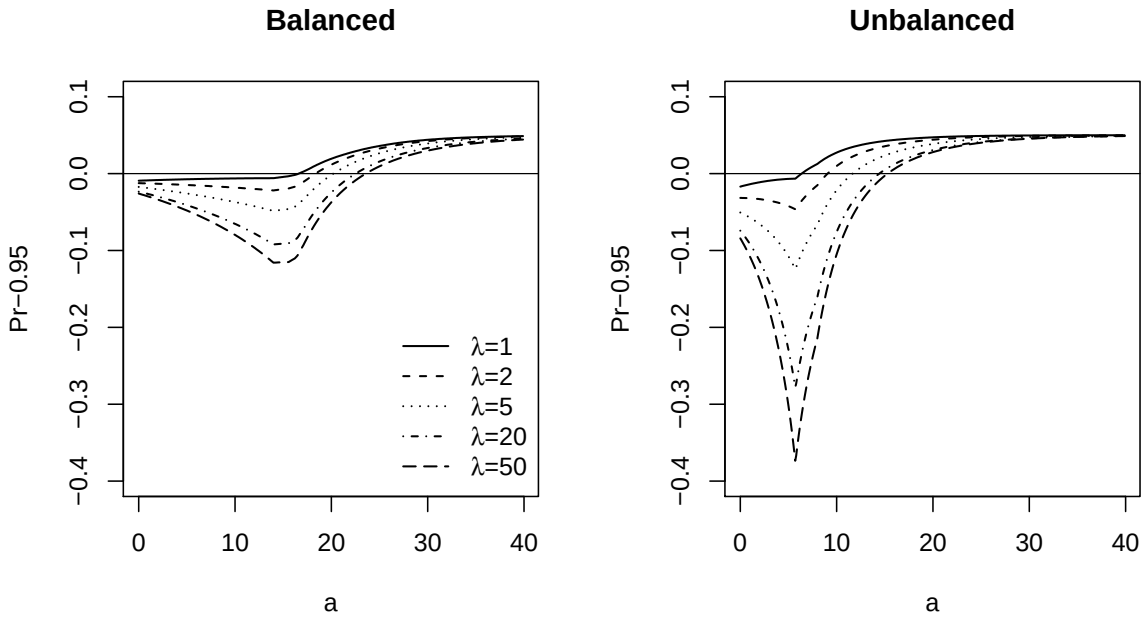

Figure 3.4 $\Pr(t^2 \leq \gamma \mid c'\beta = \eta) - \Pr(\chi_1^2 \leq \gamma)$, for $n = 50$, using $\widehat{V}_2(a)$ for different values of a, γ being the 0.95 quantile of χ_1^2 (3.841), balanced and unbalanced regression designs, different levels of heteroskedasticity.

Figure 3.4 presents plots of the differences between $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, where γ is the 0.95 quantile of χ_1^2, and 0.95, the nominal (asymptotic) probability. We note that:

i The value of a has great impact on the quality of the approximation used in the test;

ii The optimal value of a depends on the heteroskedasticity strength and also on whether the data contain high leverage observations;

iii The differences of the two probabilities (exact and asymptotic) are not monotonic in a;

iv In balanced regression designs it is best to use $a = 0$, whereas under leveraged data one should use $a \approx 15$.

v In the presence of high leverage observations, the optimal value of a for hypothesis testing inference is quite different from that proposed by Qian and Wang (2001).[6]

Figure 3.5 contains relative quantile discrepancies plots for test statistics based on $\widehat{V}_1$ and $\widehat{V}_2(a)$ for $a = 0, 2, 10, 15$. We note from Figure 3.5 that:

i The null distributions of all test statistics are well approximated by the limiting null distribution (χ^2_1) when all errors share the same variance and the regression design is balanced;

ii In the absence of leverage points and under heteroskedasticity, the null distributions of the test statistics based on the following HCCMEs are well approximated by the limiting null distribution (χ^2_1): $\widehat{V}_1$, $\widehat{V}_2(0)$ and $\widehat{V}_2(2)$;

iii Under homoskedasticity and leveraged data, the quantiles of the test statistics based on $\widehat{V}_1$, $\widehat{V}_2(0)$ and $\widehat{V}_2(2)$ are similar;

iv Under heteroskedasticity and leveraged data, the best performing test is that based on $\widehat{V}_2(a)$ with $a = 15$, which is clearly superior to the alternative tests.

Table 3.5 presents the computed probabilities $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ at the 0.90, 0.95 and 0.99 asymptotic (χ^2_1) quantiles for test statistics based on HC3, HC4, $\widehat{V}_1$ and $\widehat{V}_2(a)$ with $a = 0, 2$ e 15 when $\lambda = 1$, $\lambda \approx 50$ and $\lambda \approx 100$. The figures in this table suggest that:

i Under homoskedasticity and balanced data, all probability discrepancies are small, i.e., the computed exact probabilities are close to their nominal counterparts (0.90, 0.95 and 0.99);

ii When the errors have different variances and the regression design is balanced (no leverage point), the following estimators yield exact probabilities that are close to the respective asymptotic probabilities: HC3, HC4, $\widehat{V}_1$ and $\widehat{V}_2(a)$ with $a = 0$ and $a = 2$ (with slight advantage to HC3 and HC4);

iii When the error variances are constant and the data are leveraged, the only estimator that yields exact probabilities that are not close to the asymptotic ones is $\widehat{V}_2(15)$, the smallest probability discrepancies being those of HC4;

[6] Recall that these authors proposed using $a = 2$ for MSE minimization.

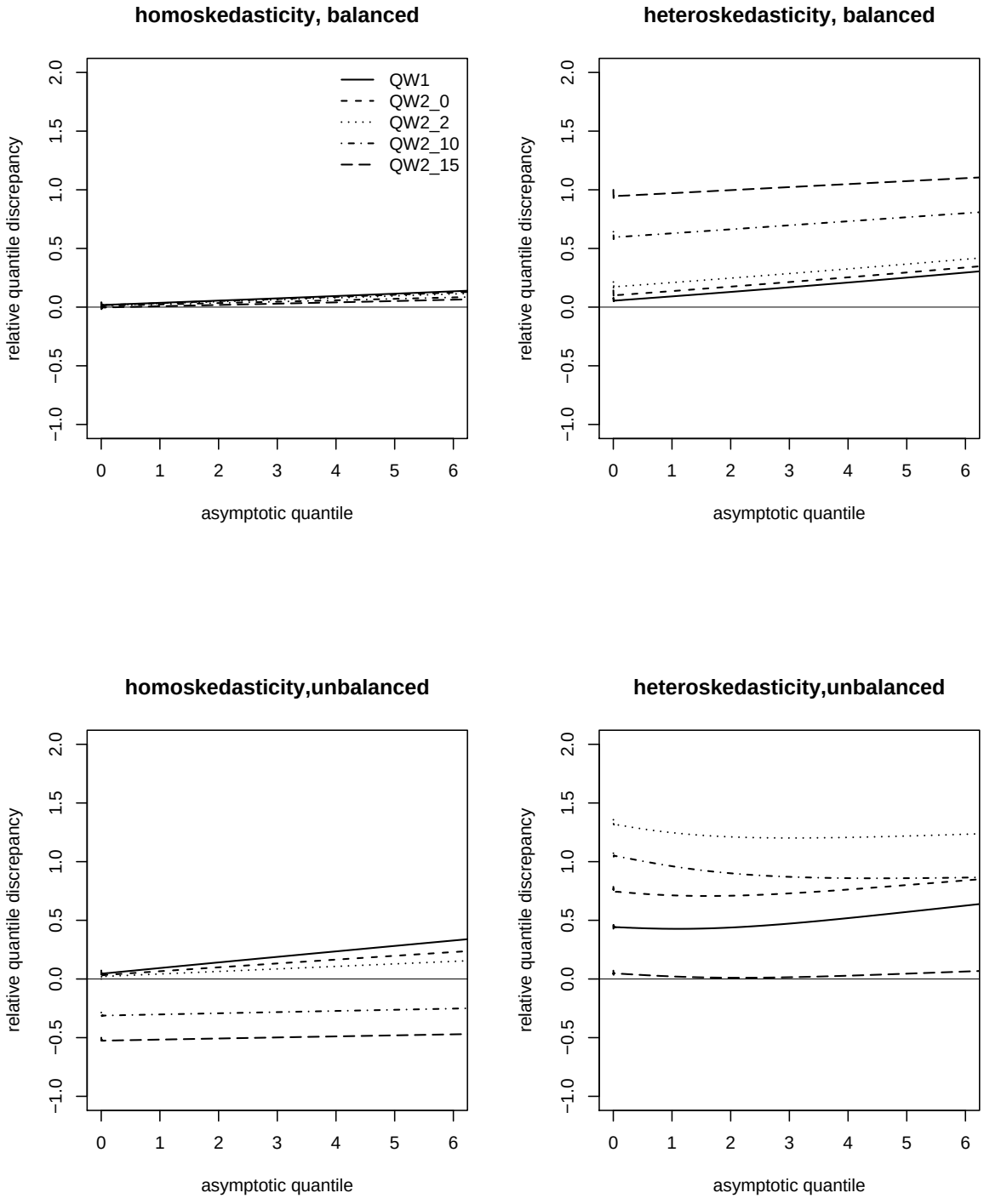

Figure 3.5 Relative quantile discrepancy plots, $n = 50$: balanced and unbalanced regression designs, homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 50$), using estimators $\widehat{V}_1$(QW1) and $\widehat{V}_2(a)$: $a = 0$ (QW2_0), $a = 2$ (QW2_2), $a = 10$ (QW2_10) and $a = 15$ (QW2_15).

Table 3.5 $\Pr(t^2 \le \gamma \mid c'\beta = \eta)$ for $n = 50$; γ equals the 0.90, 0.95 and 0.99 quantiles of χ^2_1; test statistics based on HC3, HC4, $\widehat{V}_1$, $\widehat{V}_2(0)$, $\widehat{V}_2(2)$ and $\widehat{V}_2(15)$.

		balanced						unbalanced					
λ	Pr	HC3	HC4	$\widehat{V}_1$	$\widehat{V}_2(0)$	$\widehat{V}_2(2)$	$\widehat{V}_2(15)$	HC3	HC4	$\widehat{V}_1$	$\widehat{V}_2(0)$	$\widehat{V}_2(2)$	$\widehat{V}_2(15)$
1	0.90	0.898	0.893	0.889	0.890	0.891	0.896	0.892	0.909	0.874	0.881	0.887	0.977
	0.95	0.947	0.943	0.940	0.941	0.942	0.946	0.940	0.951	0.927	0.933	0.939	0.992
	0.99	0.987	0.986	0.984	0.985	0.985	0.987	0.982	0.986	0.977	0.981	0.984	0.999
$\approx$ 50	0.90	0.885	0.881	0.876	0.869	0.858	0.757	0.853	0.903	0.830	0.792	0.729	0.898
	0.95	0.936	0.932	0.929	0.924	0.916	0.835	0.913	0.947	0.895	0.866	0.812	0.947
	0.99	0.981	0.980	0.978	0.976	0.973	0.929	0.971	0.984	0.962	0.948	0.917	0.987
$\approx$ 100	0.90	0.885	0.880	0.875	0.868	0.856	0.742	0.852	0.906	0.828	0.784	0.711	0.893
	0.95	0.935	0.932	0.928	0.923	0.914	0.821	0.914	0.950	0.896	0.861	0.798	0.945
	0.99	0.981	0.979	0.978	0.975	0.972	0.920	0.972	0.986	0.964	0.947	0.910	0.986

iv Under heteroskedasticity and leveraged data, the estimators HC4 and $\widehat{V}_2(15)$ yield the best performing tests, i.e., the exact probabilities of the test statistics that use HC4 and $\widehat{V}_2(15)$ variance estimates are considerably closer to their nominal counterparts than those associated with the competing test statistics.

We now move to a second numerical evaluation, which uses the data on public spending on education described earlier. The model used in the numerical exercise was

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t, \quad t = 1, \ldots, 50,$$

the skedastic function being as described in Section 3.5. We have used numerical integration to compute the exact quantiles of the following test statistics: HC3, HC4, $\widehat{V}_1$ and $\widehat{V}_2(a)$ with $a = 0, 2, 15$. The sample sizes were $n = 50, 100$ and computations were carried under both homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 50$).

We note from Figure 3.6, which contains the relevant quantile discrepancy plots, that under equal error variances the smallest relative quantile discrepancies are those of the test statistics based on $\widehat{V}_2(0)$ and $\widehat{V}_2(2)$. Under heteroskedasticity, however, the best performing tests are $\widehat{V}_2(15)$ e HC4.

3.8 Yet another heteroskedasticity-consistent standard error: HC5

In Section 3.2, we presented four HCCMEs in which the tth squared residual is divided by $(1-h_t)^{\delta_t}$; $\delta_t = 0$ for HC0, $\delta_t = 1$ for HC2, $\delta_t = 2$ for HC3 and $\delta_t = \min\{4, nh_t/p\}$ for HC4. These adjustments aim at 'inflating' the squared residuals according to their respective leverage measures, which are obtained from the hat matrix.

Cribari–Neto, Souza and Vasconcellos (2007) have proposed the HC5 estimator, which is given by

$$\text{HC5} = P\widehat{\Omega}_5 P' = PE_5\widehat{\Omega}P',$$

where $E_5 = \text{diag}\{1/\sqrt{(1-h_t)^{\delta_t}}\}$ and

$$\delta_t = \min\left\{\frac{nh_t}{p}, \max\left\{4, \frac{nkh_{\max}}{p}\right\}\right\},$$

with $h_{\max} = \max\{h_1, \ldots, h_n\}$. Here, $0 < k < 1$ is a constant; the authors suggested using $k = 0.7$ based on pilot Monte Carlo simulation results. It is noteworthy that $h_{\max}$ may now affect the discount terms of all squared residuals, and not only of the corresponding squared residual.

We note that HC5 can also be used in the variance estimation of a linear combination of the regression parameter estimates in $\widehat{\beta}$. Here,

$$\widehat{\Phi}_5 = \widehat{\varepsilon}' V_5 \widehat{\varepsilon},$$

where $V_5 = (v_5 v_5')_d$ and $v_5 = E_5^{1/2} P' c$. It is possible to write $\widehat{\Phi}_5$ as a quadratic form in a vector z of zero mean and unit covariance as $\widehat{\Phi}_5 = z' G_5 z$, where $G_5 = \Omega^{1/2}(I-H)V_5(I-H)\Omega^{1/2}$.

We shall now use Imhof's (1961) numerical integration algorithm to obtain the exact null distributions of HC5-based test statistics and evaluate the first order χ^2_1 approximation used in

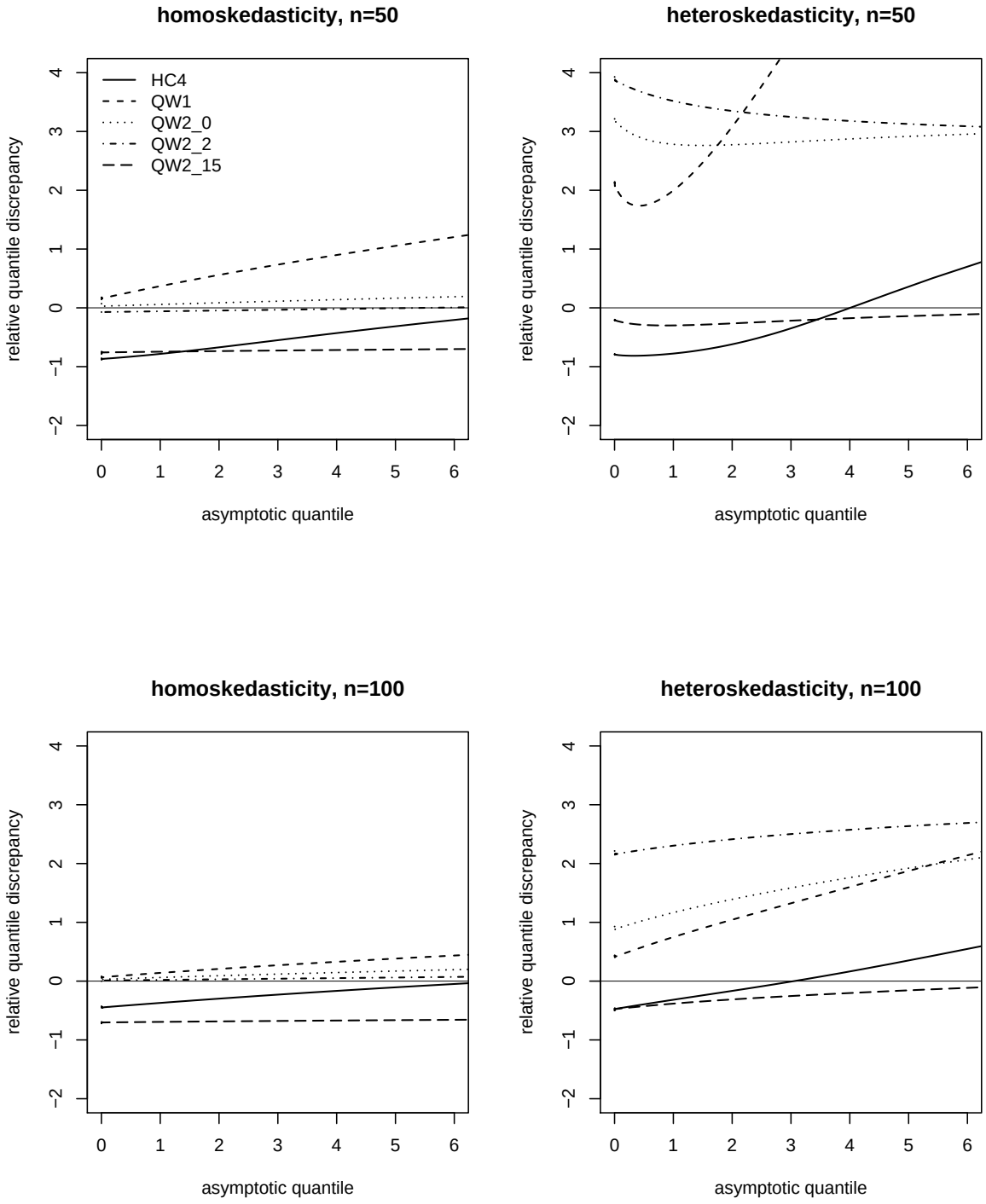

Figure 3.6 Relative quantile discrepancy plots using data on public spending on education, $n = 50, 100$: homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 50$).

the test. In the evaluation, we shall also consider HC3- and HC4-based tests for benchmarking. The regression model used is

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t, \quad t = 1, \ldots, n.$$

Here, $\varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$, where $\sigma_t^2 = \exp(\alpha_1 x_{1t} + \alpha_2 x_{2t})$, $t = 1, \ldots, n$; also, $\mathbb{E}[\varepsilon_t \varepsilon_s] = 0 \; \forall t \neq s$. The null hypothesis under test is $\mathcal{H}_0 : c'\beta = \eta$, with $c' = (0, 0, 1)$ and $\eta = 0$, and the test statistic is

$$t^2 = \widehat{\beta}_2^2 / \widehat{\text{var}}(\widehat{\beta}_2),$$

where $\widehat{\text{var}}(\widehat{\beta}_2)$ is a heteroskedasticity-consistent variance estimate. The sample size is $n = 50$; each covariate value is replicated once when $n = 100$. There are two regression designs, namely: balanced (covariate values randomly obtained from the standard uniform distribution) and unbalanced (covariate values randomly obtained from the standard lognormal distribution); see Table 3.6.

Table 3.6 Maximal leverages for the two regression designs.

	$\mathcal{U}(0,1)$	$\mathcal{LN}(0,1)$	threshold	
n	$h_{\max}$	$h_{\max}$	$2p/n$	$3p/n$
50	0.114	0.648	0.12	0.18
100	0.057	0.324	0.06	0.09

Figure 3.7 presents relative quantile discrepancy plots under homoskedaticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 50$) for $n = 50, 100$. When the regression design is balanced (no leverage point), the null distributions of all three test statistics are very well approximated by the limiting null chi-squared distribution. When data contain high leverage observations, however, the quality of the approximations deteriorate, and the HC4- and HC5-based tests are clearly superior to the HC3-based test, especially at the quantile of main interest (the 0.95 asymptotic quantile, which is the 5% critical value: 3.841). It is noteworthy that under unequal error variances, unbalaced regression design and $n = 50$, the exact 0.95 quantile of the HC3-based test statistic is over twice the corresponding asymptotic quantile!

Table 3.7 contains the computed probabilities $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, $\gamma = 3.841$ (0.95 asymptotic quantile), for HC3-, HC4- and HC5-based test statistics. We present results for homoskedasticity and two different levels of heteroskedasticity. The figures in Table 3.7 show that the computed probabilities are close to their nominal (asymptotic) counterparts when the data are free from high leverage observations. When the regression design is unbalanced, however, the HC4 and HC5 computed probabilities are closer to the respective desired levels than those computed using the HC3 HCCME, except under homoskedasticity and at the 10% nominal level (0.90 nominal probability); HC5 very slightly outperforms HC4 at the 5% nominal level (0.95 nominal probability).

The next numerical evaluation uses the data on public spending on education. As before, the skedastic function is $\exp(\alpha_1 x_t + \alpha_2 x_t^2)$, $t = 1, \ldots, n$. Here, $\alpha_1 = \alpha_2 = 0$ yields $\lambda = 1$ (homoskedasticity) and $\alpha_1 = 0, \alpha_2 = 3.8$ yields $\lambda \approx 25$ when all observations are used ($n = 50$).

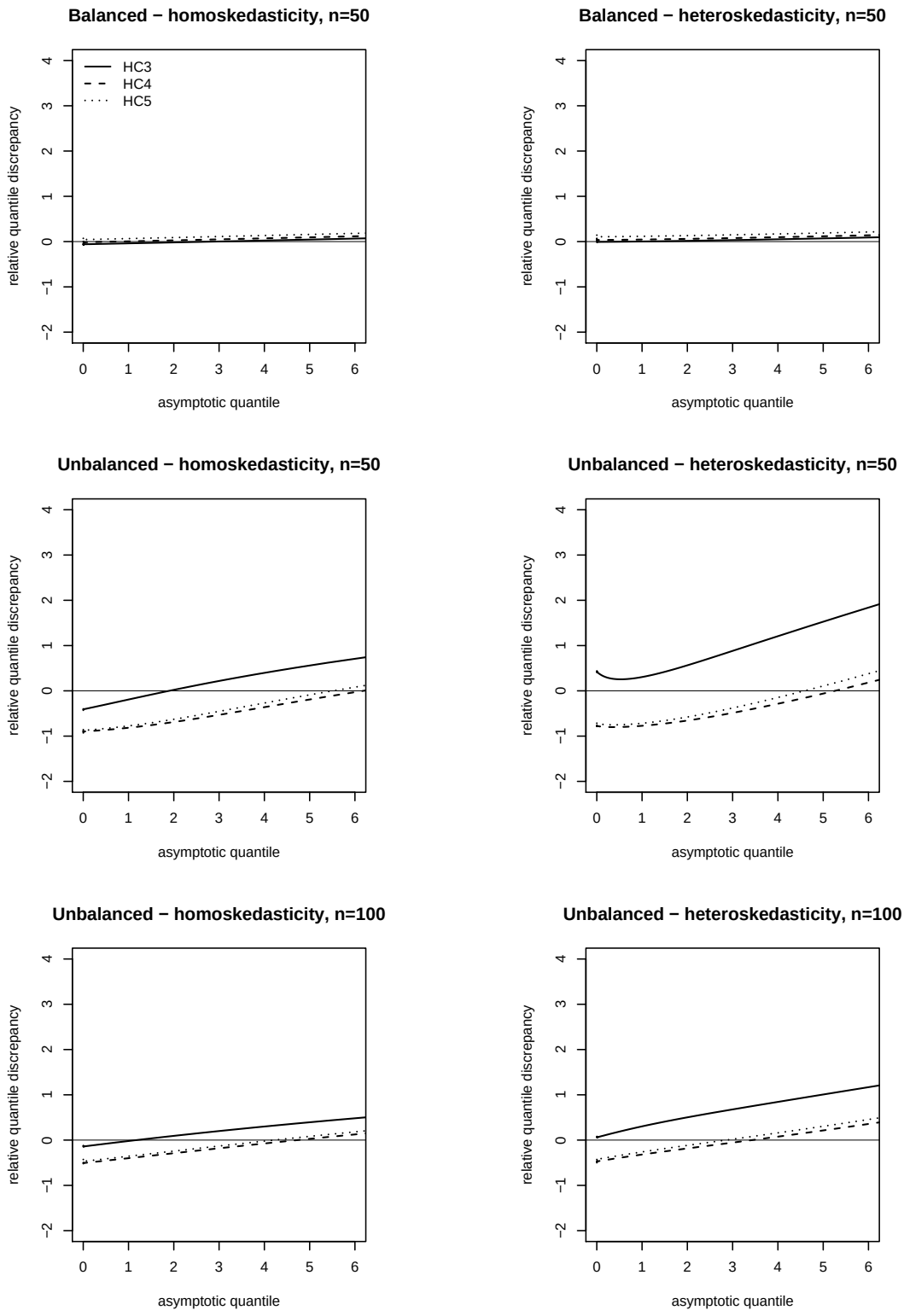

Figure 3.7 Relative quantile discrepancy plots: balanced and unbalanced regression designs, homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 50$), using estimators HC3, HC4 and HC5.

Table 3.7 $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ for $n = 50$; γ equals the 0.90, 0.95 and 0.99 quantiles of χ^2_1; test statistics based on HC3, HC4 and HC5.

		balanced			unbalanced		
λ	Pr	HC3	HC4	HC5	HC3	HC4	HC5
1	0.90	0.901	0.893	0.883	0.882	0.957	0.950
	0.95	0.948	0.943	0.936	0.922	0.973	0.968
	0.99	0.987	0.985	0.983	0.966	0.989	0.986
≈ 25	0.90	0.896	0.889	0.878	0.823	0.948	0.938
	0.95	0.945	0.940	0.932	0.873	0.965	0.957
	0.99	0.986	0.984	0.981	0.933	0.983	0.979
≈ 50	0.90	0.895	0.888	0.877	0.820	0.952	0.942
	0.95	0.945	0.940	0.932	0.872	0.967	0.960
	0.99	0.986	0.984	0.981	0.932	0.984	0.980

When the three leverage points are removed from the data ($n = 47$), we use $\alpha_1 = 0, \alpha_2 = 7.3$ to obtain $\lambda \approx 25$. Relative quantile plots are presented in Figure 3.8 (homoskedasticity and heteroskedasticity, $n = 47, 50, 100$). It is clear that under equal error variances and unbalanced design ($n = 50$), of all computed distributions, the null distribution of the HC3-based test statistic is the one best approximated by χ^2_1 (the approximate distribution used in the test). Under heteroskedasticity and leveraged data, nonetheless, the HC4 and HC5 tests display superior finite sample behavior (HC5 slightly better at the 0.95 nominal quantile) relative to HC3. When the regression design is balanced ($n = 47$), the three tests display similar behavior, regardless of whether the error variances are equal, the null distributions of the HC3 and HC4 test statistics being slightly better approximated by the limiting null distribution than that of HC5. It is also noteworthy that the relative quantile discrepancy plots under leveraged data, heteroskedasticity and $n = 50$ in Figures 3.7 are 3.8 are very similar.

Table 3.8 contains, for the data on public spending on education, the computed probabilities $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ for the HC3-, HC4- and HC5-based test statistics, where γ equals the 0.95 χ^2_1 quantile (3.841). We present results for cases 1 and 3 ($n = 50$ and $n = 47$, respectively). Under homoskedasticity and leveraged data ($n = 50$), the best performing test is that whose test statistic uses the HC3 standard error. When the regression design is unbalanced and the error variances are not constant, however, the HC4 and HC5 computed probabilities are considerably closer to 0.95 than those computed using the HC3-based test statistic. For instance, the computed probabilities for the HC4- and HC5-based statistics were respectively equal to 0.956 and 0.947 (0.953 and 0.943) when $\lambda \approx 25$ ($\lambda \approx 50$).

Next, we shall use numerical integration to evaluate the impact of the value of k (usually set at $k = 0.7$) on the limiting null approximation used in the HC5 test. The evaluation is based on a simple regression model; the errors are uncorrelated, each error having zero mean and variance $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$. The sample size is $n = 50$ and the covariate values are selected as random draws from the standard lognormal distribution. We consider two regressions designs,

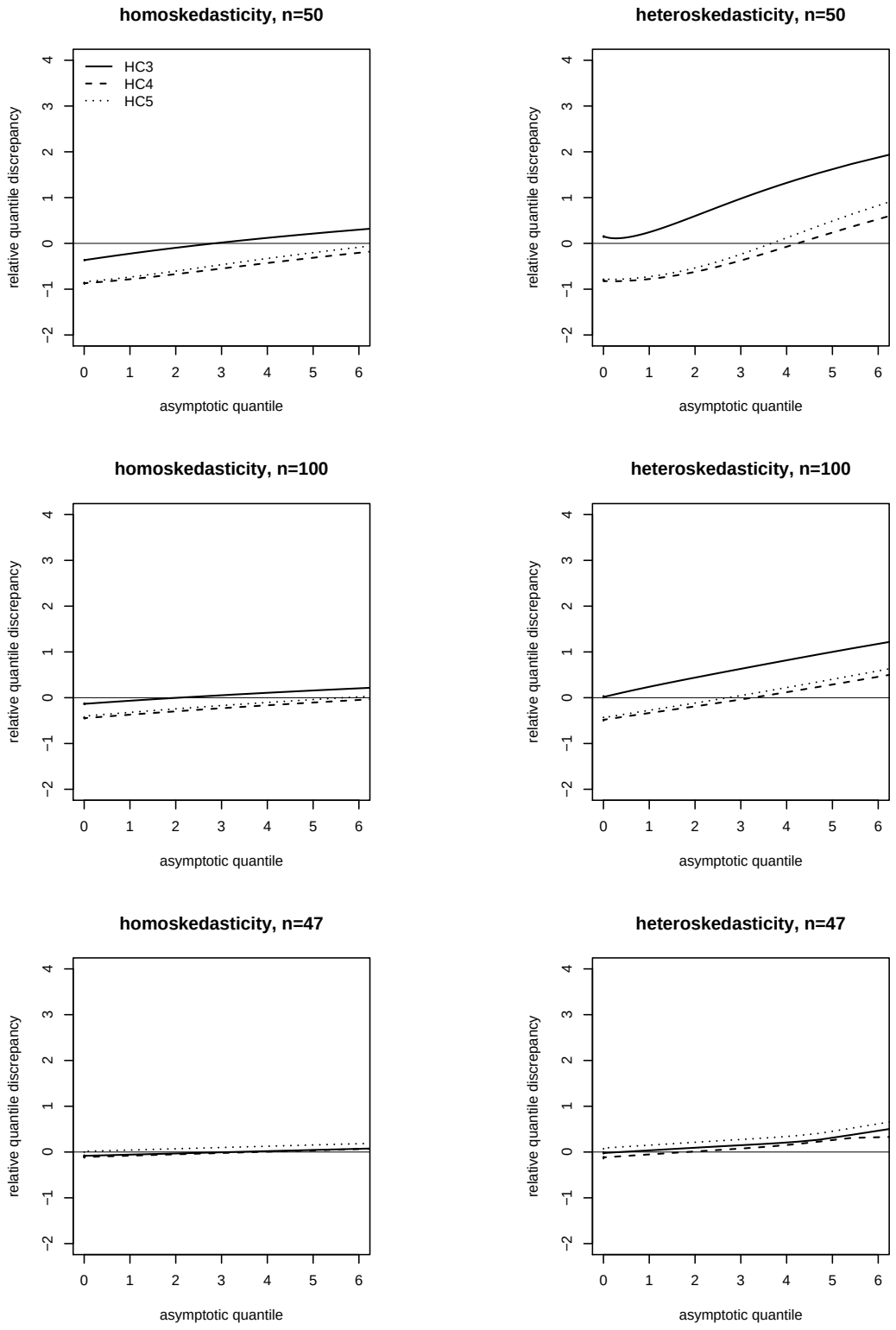

Figure 3.8 Relative quantile discrepancy plots using data on public spending on education: equal and unequal error variances, balanced ($n = 47$) and unbalanced ($n = 50$) regression designs, using estimators HC3, HC4 and HC5.

Table 3.8 $\Pr(t^2 \le \gamma \mid c'\beta = \eta)$ using data on public spending on education, $n = 50$ and 47 (cases 1 and 3, respectively); γ equals the 0.95 quantile of χ^2_1; test statistics based on three different standard errors.

	$\lambda = 1$		$\lambda \approx 25$		$\lambda \approx 50$	
statistic	$n = 50$	$n = 47$	$n = 50$	$n = 47$	$n = 50$	$n = 47$
HC3	0.941	0.948	0.867	0.931	0.855	0.928
HC4	0.979	0.950	0.956	0.937	0.953	0.946
HC5	0.973	0.937	0.947	0.917	0.943	0.912

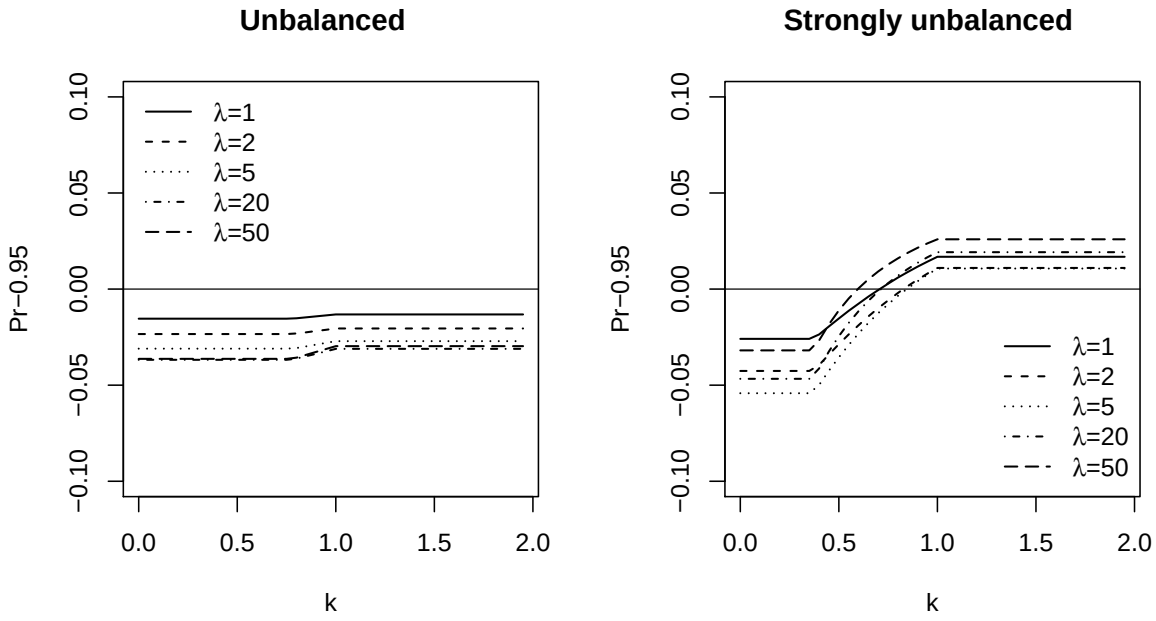

Figure 3.9 $\Pr(t^2 \le \gamma \mid c'\beta = \eta) - \Pr(\chi^2_1 \le \gamma)$, for $n = 50$, using HC5 with different values of k, γ being the 0.95 quantile of χ^2_1 (3.841), unbalanced and strongly unbalanced regression designs, different levels of heteroskedasticity.

namely: unbalanced ($h_{\max}/(3p/n) = 1.71$) and strongly unbalanced ($h_{\max}/(3p/n) = 3.58$). The results of the numerical evaluation are graphically displayed in Figure 3.9, which contains plots of the differences between $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, where γ is the 0.95 quantile of χ^2_1, and 0.95, the nominal (asymptotic) probability. The probability discrepancies are plotted against k. We note that in the unbalanced situation the value of k has little impact on the quality of the first order asymptotic approximation used in the test. However, in the strongly unbalanced design, values of k between 0.6 and 0.8 yield the best approximations. As a consequence, these results suggest that 0.7, the value of k suggested by Cribari–Neto, Souza and Vasconcellos (2007), is indeed a good choice.

In Figure 3.10 we present the same probability discrepancies displayed in Figure 3.9 but now using the data on public spending on education. Again, values of k between 0.6 and 0.7 seem to be a good choice for the HC5-based test.

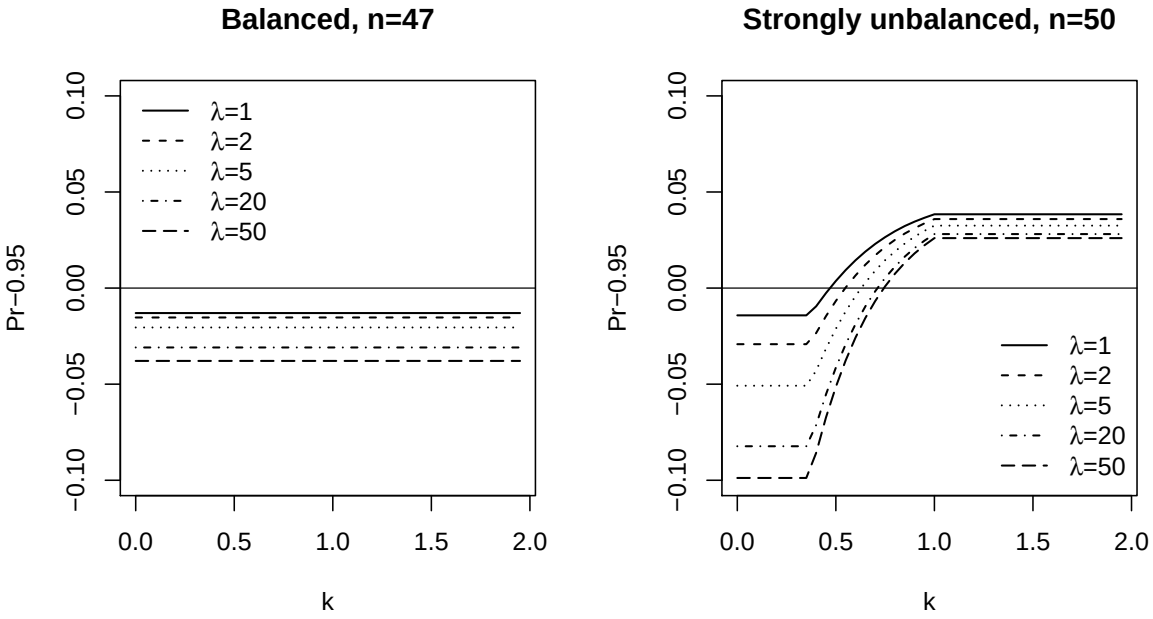

Figure 3.10 $\Pr(t^2 \leq \gamma \mid c'\beta = \eta) - \Pr(\chi^2_1 \leq \gamma)$, using HC5 with different values of k, γ being the 0.95 quantile of χ^2_1 (3.841), data on public spending on education, unbalanced ($n = 47$) and strongly unbalanced ($n = 50$) regression designs, different levels of heteroskedasticity.

Table 3.9 contains the computed probabilities $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, where γ is the 0.95 χ^2_1 quantile, for HC4- and HC5-based statistics (the latter obtained using different values of k) in a two-covariate regression model where the covariate values are obtained as random draws from the $\mathcal{LN}(0,1)$ distribution with $n = 50$, under homoskedasticity and strong heteroskedasticity. Also, $h_{\max}/(3p/n) \approx 3.60$, so there is a strong leverage point in the data. The figures in Table 3.9 show that, overall, the best χ^2 approximation for the HC5-based test statistic takes place

when $k = 0.6, 0.7$. Using these values of k the HC5 test slightly outperforms the HC4 test.

Table 3.9 $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ for $n = 50$; γ equals the 0.95 quantile of χ^2_1; test statistics based on HC4 and HC5 (different values of k) standard errors.

test statistic	$\lambda = 1$	$\lambda \approx 50$
HC5 ($k = 0.5$)	0.942	0.917
HC5 ($k = 0.6$)	0.956	0.943
HC5 ($k = 0.7$)	0.968	0.960
HC5 ($k = 0.8$)	0.976	0.970
HC5 ($k = 0.9$)	0.981	0.977
HC5 ($k = 1.0$)	0.986	0.983
HC4	0.973	0.967

We shall now report the results of a 10,000 replication Monte Carlo experiment where the null rejection probabilities of HC4 and HC5 (different values of k) tests are computed. The model used in the evaluation is

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t, \quad t = 1, \ldots, 50.$$

Table 3.10 Null rejection rates of HC4 and HC5 quasi-t tests; HC5 standard errors are computed using different values of k; nominal level: 5%.

	$h_{\max}/(3p/n)$			
	3.60		1.14	
test statistic	$\lambda = 1$	$\lambda \approx 50$	$\lambda = 1$	$\lambda \approx 50$
HC5 ($k = 0.5$)	5.42	8.47	7.20	8.09
HC5 ($k = 0.6$)	4.04	5.88	7.20	8.09
HC5 ($k = 0.7$)	2.83	4.25	7.20	8.09
HC5 ($k = 0.8$)	2.17	3.23	7.20	8.09
HC5 ($k = 0.9$)	1.64	2.44	7.20	8.09
HC5 ($k = 1.0$)	1.28	1.82	7.20	8.09
HC4	2.40	3.56	5.72	6.37

The errors are uncorrelated, each ε_t being normally distributed with zero mean and variance $\sigma_t^2 = \exp(\alpha_1 x_{1t} + \alpha_2 x_{2t})$. The covariate values were selected as random draws from the standard lognormal distribution. Simulations were performed under both homoskedasticity ($\lambda = 1$) and heteroskedasticity ($\lambda \approx 50$) and we consider two different settings in which the values of $h_{\max}/(3p/n)$ are 3.60 and 1.14. The interest lies in testing $\mathcal{H}_0 : \beta_2 = 0$ against $\mathcal{H}_1 : \beta_2 \neq 0$. Data generation was carried out using $\beta_0 = \beta_1 = 1$ and $\beta_2 = 0$. The quasi-t test statistics considered employ HC4 and HC5 ($k = 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$) standard errors.

Table 3.10 presents the null empirical rejection rates at the nominal level $\alpha = 0.05$. All entries are percentages. When leverage is strong ($h_{\max}/(3p/n) \approx 3.60$), $k = 0.6$ overall yields the best HC5 test. When leverage is weak, however, the HC5 test is outperformed by HC4 regardless the value of k.

3.9 Concluding remarks

We have considered the issue of making inference on the parameters that index the linear regression model under heteroskedasticity of unknown form. We have numerically evaluated the first order asymptotic approximation used in quasi-t tests. Our evaluation did *not* rely on Monte Carlo simulations.[7] Rather, we assumed normality, wrote the test statistics as ratios of quadratic forms in a vector of uncorrelated standard normal variates, and used Imhof's (1961) numerical integration algorithm to compute the exact distribution functions of the different test statistics. We have included in the analysis test statistics that use standard errors obtained from the widely used HCCME proposed by Halbert White (HC0) and from HC3, an often praised HCCME (e.g., Long and Ervin, 2000). In addition, we included alternative test statistics based on several recently proposed heteroskedasticity-robust standard errors, namely: HC4 (Cribari–Neto, 2004), $\widehat{V}_1$ and $\widehat{V}_2$ (Qian and Wang, 2001), and HC5 (Cribari–Neto, Souza and Vasconcellos, 2007). We have also made use of numerical integration methods to shed light on the choice of constants used in the definitions of $\widehat{V}_2$ and HC5. Our main findings can be outlined as follows:

i Under equal error variances and in balanced regression designs, the null distributions of all test statistics are typically well approximated by the limiting null distribution, from which we obtain critical values for the tests;

ii Under homoskedasticity and leveraged data, the best first order asymptotic approximations are those for the HC4- and HC5-based tests;

iii Under heteroskedasticity and in balanced regression designs, the tests based on HC3, HC4, HC5, $\widehat{V}_1$ and $\widehat{V}_2(a)$ with $a = 0$ and $a = 2$ appear to be reliable in finite samples;

iv The best performing tests under heteroskedasticity and leveraged data are those whose test statistics use standard errors from HC4, HC5 and $\widehat{V}_2(15)$.

Since in practice we have no knowledge of the strength of heteroskedasticity and leverage points are often present in the data, we recommend the use of HC4-based tests or HC5-based tests with $k = 0.6$ or $k = 0.7$ when leverage is very intense.

[7] We only used Monte Carlo simulation once in our analysis: to estimate the null rejection rates of different HC5 tests. All other numerical evaluations were carried out in exact fashion by using a numerical integration method.

CHAPTER 4

Conclusions

The object of interest of this doctoral dissertation was the linear regression model. The assumption that all error variances are equal (homoskedasticity) is commonly violated in regression analysis that use cross-sectional data. It is thus important to develop and evaluate inference strategies that are robust to heteroskedasticity. This was our main motivation.

At the outset, we have proposed different heteroskedasticity-consistent interval estimators (HCIEs) for the regression parameters. They are based on variance and covariance estimators that are asymptotically correct under heteroskedasticity of unknown form and also under equal error variances. We have also considered bootstrap-based interval estimators. Our numerical evaluation revealed that the HC4 HCIE outperforms all other interval estimators, including those that employ bootstrap resampling.

We then moved to point estimation of variances and covariances. We considered a heteroskedasticity-consistent covariance matrix estimator (HCCME) proposed by L. Qian and S. Wang in 2001, which is a modified version of the well known White estimator. We have obtained a sequence of bias-adjusted estimators in which the biases vanish at faster rates as we move along the sequence. We have also generalized the Qian–Wang estimator, and obtained alternative sequences of improved estimators. Our numerical results have shown that the proposed bias-adjusting schemes can be quite effective in small samples.

Finally, we addressed the issue of performing hypothesis testing inference in the linear regression model under heteroskedasticity of unknown form. We have added the Gaussianity assumption and used a numerical integration algorithm to compute the exact distribution functions of different quasi-t test statistics, which were then compared to the respective limiting null distribution. To that end, we have shown that such statistics can be written as ratios of quadratic forms in standard normal (Gaussian) random vectors. We focused on test statistics that use four recently proposed heteroskedasticity-robust standard errors. Two of them employ constants that are chosen in an *ad hoc* manner, and our results have shed light on their optimal values. Overall, our numerical evaluations favored the HC4-based test.

CHAPTER 5

Resumo do Capítulo 1

5.1 Introdução

O modelo de regressão linear é comumente utilizado para modelar a relação entre uma variável de interesse e um conjunto de variáveis explicativas. Quando dados de corte transversal são usados, a suposição de que os erros do modelo têm a mesma variância (homoscedasticidade) é freqüentemente violada. Uma prática comum é usar o estimador de mínimos quadrados ordinários (EMQO) em conjunção com um estimador consistente de sua matriz de covariâncias quando o interesse reside em inferências por testes de hipóteses. Nesse capítulo, o nosso foco recai sobre inferências através de estimadores intervalares. Nós propomos e avaliamos numericamente diferentes estimadores intervalares robustos à presença de heteroscedasticidade, inclusive estimadores baseados em esquemas de reamostragem de bootstrap.

5.2 O modelo e alguns estimadores pontuais

O modelo de interesse é o modelo de regressão linear

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

onde y é um vetor de observações de dimensão n na variável dependente, X é uma matriz fixa $n \times p$ de regressores (posto$(X) = p < n$), $\beta = (\beta_0, \ldots, \beta_{p-1})'$ é um p-vetor de parâmetros desconhecidos e $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ é um vetor n-dimensional de erros aleatórios. As suposições a seguir são comumente feitas:

A1 O modelo $y = X\beta + \varepsilon$ está corretamente especificado;

A2 $\mathbb{E}(\varepsilon_i) = 0$, $i = 1, \ldots, n$;

A3 $\mathbb{E}(\varepsilon_i^2) = \text{var}(\varepsilon_i) = \sigma_i^2$ $(0 < \sigma_i^2 < \infty)$, $i = 1, \ldots, n$;

A3' $\text{var}(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$ $(0 < \sigma^2 < \infty)$;

A4 $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0 \; \forall \, i \neq j$;

A5 $\lim_{n \to \infty} n^{-1}(X'X) = Q$, onde Q é uma matrix positiva definida.

Sob [A1], [A2], [A3] and [A4], a matriz de covariâncias de ε é

$$\Omega = \text{diag}\{\sigma_i^2\},$$

que se reduz a $\Omega = \sigma^2 I_n$ quando $\sigma_i^2 = \sigma^2 > 0$, $i = 1, \ldots, n$, i.e.,sob [A3'] (homoscedasticidade), onde I_n é a matriz identidade de ordem n.

O EMQO of β é obtido minimizando a soma de quadrados dos erros, i.e., minimizando

$$\varepsilon'\varepsilon = (y - X\beta)'(y - X\beta);$$

o estimador pode ser escrito em forma fechada como

$$\widehat{\beta} = (X'X)^{-1}X'y.$$

Suponha que [A1] é verdadeira. Pode ser mostrado que:

i) Sob [A2], $\widehat{\beta}$ é não viesado para β, i.e., $\mathbb{E}(\widehat{\beta}) = \beta \ \forall \beta \in I\!R^p$.

ii) $\Psi_{\widehat{\beta}} = \text{var}(\widehat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$.

iii) Sob [A2], [A3], [A5] e também considerando as variâncias uniformemente limitadas, $\widehat{\beta}$ é um estimador consistente para β, i.e., $\text{plim}(\widehat{\beta}) = \beta$, onde plim denota limite em probabilidade.

iv) Sob [A2], [A3'] e [A4], $\widehat{\beta}$ é o melhor estimador linear não-viesado de β (Teorema de Gauss–Markov) .

De ii), notamos que sob homoscedasticidade $\text{var}(\widehat{\beta}) = \sigma^2(X'X)^{-1}$, que pode ser facilmente estimado como $\widehat{\text{var}}(\widehat{\beta}) = \widehat{\sigma}^2(X'X)^{-1}$, onde $\widehat{\sigma}^2 = \widehat{\varepsilon}'\widehat{\varepsilon}/(n-p)$ e $\widehat{\varepsilon}$ é o vetor de resíduos do ajuste por mínimos quadrados:

$$\widehat{\varepsilon} = y - X\widehat{\beta} = \{I_n - X(X'X)^{-1}X'\}y = (I_n - H)y.$$

A matriz $H = X(X'X)^{-1}X'$ é chamada de 'matriz chapéu', uma vez que $Hy = \widehat{y}$. Seus elementos diagonais assumem valores no intervalo $(0, 1)$ e somam p, o posto de X, sendo portanto sua média p/n. Observe-se que os elementos diagonais de H $(h_1, \ldots, h_n)$ são comumente usados como medidas de alavancagem das correspondentes observações; observações tais que $h_i > 2p/n$ ou $h_i > 3p/n$ são consideradas pontos de alavanca (veja Davidson and MacKinnon, 1993).

Sob heteroscedasticidade, quando conhecemos a matriz Ω ou uma função cedástica que nos permita estimar Ω, podemos estimar β utilizando o estimador de mínimos quadrados generalizado (EMQG) ou o estimador de mínimos quadrados generalizado viável (EMQGV). Para realizar inferências sobre β que sejam válidas assintoticamente sob heteroscedasticidade, costuma-se utilizar o EMQO de β, que permanece consistente, não-viesado e assintoticamente normal, conjuntamente com um estimador consistente de sua matriz de covariâncias.

White (1980) derivou um estimador consistente para $\Psi_{\widehat{\beta}}$ observando que não é necessário estimar consistentemente Ω (que tem n parâmetros desconhecidos); é necessário apenas estimar consistentemente $X'\Omega X$ (que tem $p(p+1)/2$ elementos distintos independentemente do tamanho da amostra). [1] Necessita-se apenas encontrar $\widehat{\Omega}$ tal que $\text{plim}\left((X'\Omega X)^{-1}(X'\widehat{\Omega}X)\right) = I_p$.

[1] Para estimar consistentemente a matriz de covariâncias, sob heteroscedasticidade, também supomos:

A6 $\lim_{n\to\infty} n^{-1}(X'\Omega X) = S$, onde S é uma matriz positiva definida.

O estimador de White, também conhecido como HC0, é obtido substituindo o i-ésimo elemento diagonal de Ω na expressão para $\Psi_{\widehat{\beta}}$ pelo quadrado do i-ésimo resíduo de mínimos quadrados, i.e.,

$$\text{HC0} = (X'X)^{-1}X'\widehat{\Omega}_0 X(X'X)^{-1},$$

onde $\widehat{\Omega}_0 = \text{diag}\{\widehat{\varepsilon}_i^2\}$.

O estimador de White é consistente sob homoscedasticidade e sob heteroscedasticidade de forma desconhecida. Entretanto, ele pode ser bastante viesado em amostras finitas, como evidenciado pelos resultados numéricos obtidos por Cribari–Neto e Zarkos (1999, 2001). Adicionalmente, a magnitude do viés é inflacionada quando o desenho de regressão contém pontos de alavanca.

Algumas variantes do HC0 foram propostas na literatura. Essas variantes incluem correções para amostras finitas na estimação de Ω e são dadas por:

i (Hinkley, 1977) HC1 $= \widehat{\Psi}_1 = P\widehat{\Omega}_1 P' = PD_1\widehat{\Omega}P'$, onde $D_1 = (n/(n-p))I$;

ii (Horn, Horn e Duncan, 1975) HC2 $= \widehat{\Psi}_2 = P\widehat{\Omega}_2 P' = PD_2\widehat{\Omega}P'$, onde

$$D_2 = \text{diag}\{1/(1-h_i)\};$$

iii (Davidson e MacKinnon, 1993) HC3 $= \widehat{\Psi}_3 = P\widehat{\Omega}_3 P' = PD_3\widehat{\Omega}P'$, onde

$$D_3 = \text{diag}\{1/(1-h_i)^2\};$$

iv (Cribari–Neto, 2004) HC4 $= \widehat{\Psi}_4 = P\widehat{\Omega}_4 P' = PD_4\widehat{\Omega}P'$, onde

$$D_4 = \text{diag}\{1/(1-h_i)^{\delta_i}\}, \quad \delta_i = \min\{4, nh_i/p\}.$$

5.3 Estimadores intervalares consistentes sob heteroscedasticidade

Nosso principal interesse consiste em obter intervalos de confiança para os parâmetros desconhecidos do modelo de regressão.

Consideramos 'heteroskedasticity-consistent interval estimators' (HCIEs) baseados em $\widehat{\beta}$ (EMQO) e nos 'heteroskedasticity-consistent covariance matrix estimators' (HCCMEs) HC0, HC2, HC3 e HC4. Sob homoscedasticidade e quando os erros são normalmente distribuídos, a quantidade

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\sigma}^2 c_{jj}}},$$

onde c_{jj} é o j-ésimo elemento diagonal de $(X'X)^{-1}$, tem distribuição t_{n-p}, sendo então fácil construir intervalos de confiança exatos para β_j, $j = 0, \ldots, p-1$.

Sob heteroscedasticidade, a matriz de covariâncias do EMQO é

$$\Psi_{\widehat{\beta}} = \text{var}(\widehat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

Os estimadores consistentes apresentados anteriormente são estimadores do tipo sanduíche para a matriz de covariâncias. A seguir usaremos os estimadores HCk, $k = 0,2,3,4$. Seja, para $k = 0,2,3,4$,

$$\widehat{\Omega}_k = D_k\widehat{\Omega} = D_k\text{diag}\{\widehat{\varepsilon}_i^2\};$$

para HC0, $D_0 = I_n$;
para HC2, $D_2 = \text{diag}\{1/(1-h_i)\}$;
para HC3, $D_3 = \text{diag}\{1/(1-h_i)^2\}$;
para HC4, $D_4 = \text{diag}\{1/(1-h_i)^{\delta_i}\}$.
Portanto,

$$\widehat{\Psi}_{\widehat{\beta}}^{(k)} = (X'X)^{-1}X'\widehat{\Omega}_k X(X'X)^{-1}, \quad k = 0,2,3,4.$$

Para $k = 0,2,3,4$, considere a quantidade

$$\frac{\widehat{\beta}_j - \beta_j}{\sqrt{\widehat{\Psi}_{jj}^{(k)}}},$$

onde $\widehat{\Psi}_{jj}^{(k)}$ é o j-ésimo elemento diagonal de $\widehat{\Psi}_{\widehat{\beta}}^{(k)}$, i.e., a variância estimada de $\widehat{\beta}_j$ obtida do estimador HCk, $k = 0,2,3,4$. Segue-se da normalidade assintótica of $\widehat{\beta}_j$ e da consistência de $\widehat{\Psi}_{jj}^{(k)}$ que a quantidade acima converge em distribuição à distribuição normal padrão quando $n \to \infty$. Portanto, essa quantidade pode ser usada para construir HCIEs. Seja $0 < \alpha < 1/2$. Uma classe de intervalos de confiança com cobertura de $(1-\alpha)\times 100\%$ para β_j, $j = 0,\ldots,p-1$, é

$$\widehat{\beta}_j \pm z_{1-\alpha/2}\sqrt{\widehat{\Psi}_{jj}^{(k)}},$$

$k = 0,2,3,4$, onde $z_{1-\alpha/2}$ é o quantil $1-\alpha/2$ da distribuição normal padrão.

5.4 Avaliação numérica

A avaliação de Monte Carlo usa o seguinte modelo de regressão linear:

$$y_i = \beta_0 + \beta_1 x_i + \sigma_i\varepsilon_i, \quad i = 1,\ldots,n,$$

onde $\varepsilon_i \sim (0,1)$ e $\mathbb{E}(\varepsilon_i\varepsilon_j) = 0\ \forall i \neq j$. Aqui,

$$\sigma_i^2 = \sigma^2\exp\{ax_i\}$$

com $\sigma^2 = 1$. São considerados erros normais bem como erros exponenciais (com média um) e erros obtidos de uma distribuição de caudas pesadas (t_3). Estimaremos numericamente as probabilidades de cobertura dos diferentes HCIEs e calcularemos seus comprimentos médios. Os valores da covariável foram selecionados aleatoriamente da distribuição $\mathcal{U}(0,1)$ bem como da distribuição t_3; no segundo caso, o desenho de regressão contém pontos de alavanca. Os tamanhos de amostra são $n = 20,60,100$. Nós geramos 20 valores da covariável quando o

tamanho da amostra é $n = 20$; para tamanhos maiores esses valores foram replicados três e cinco vezes ($n = 60$ e $n = 100$, respectivamente) de modo que o nível de heteroscedasticidade, medido por

$$\lambda = \max\{\sigma_i^2\}/\min\{\sigma_i^2\}, \quad i = 1, \ldots, n,$$

permanece constante quando o tamanho da amostra aumenta. Consideramos o caso homoscedástico ($\lambda = 1$) e dois níveis de heteroscedasticidade ($\lambda \approx 9$ e $\lambda \approx 49$). Para gerar os valores de y utilizamos $\beta_0 = \beta_1 = 1$. O número de réplicas de Monte Carlo foi 10,000 e todas as simulações foram realizadas utilizando a linguagem de programação 0x (Doornik, 2001).

A cobertura nominal de todos os intervalos é $1 - \alpha = 0.95$. O intervalo de confiança padrão (MQO) usou erros padrão de $\widehat{\sigma}^2(X'X)^{-1}$ e foi calculado como

$$\widehat{\beta}_j \quad \pm \quad t_{1-\alpha/2,n-2}\sqrt{\widehat{\sigma}^2 c_{jj}},$$

onde $t_{1-\alpha/2,n-2}$ é o quantil $1-\alpha/2$ da distribuição t_{n-2}. Os HCIEs foram calculados como

$$\widehat{\beta}_j \quad \pm \quad z_{1-\alpha/2}\sqrt{\widehat{\Psi}_{jj}^{(k)}},$$

$k = 0, 2, 3, 4$ (HC0, HC2, HC3 e HC4, respectivamente).

Os resultados encontram-se apresentados nas Tabelas 1.2, 1.3, 1.4 e 1.5 do Capítulo 1. Observa-se que os intervalos obtidos com o estimador HC0 podem ser bastante ruins quando o tamanho da amostra não é grande. Os resultados favorecem o estimador intervalar HC4, que apresenta comportamento mais confiável do que aqueles apresentados pelos intervalos obtidos com HC0 e HC2 e mesmo com HC3.

5.5 Intervalos bootstrap

Um enfoque alternativo é usar reamostragem dos dados para obter intervalos de confiança; em particular utilizaremos o método bootstrap proposto por Bradley Efron (Efron, 1979). O bootstrap ponderado de Wu (1986) pode ser usado para se obter erros-padrão assintoticamente corretos sob heteroscedasticidade de forma desconhecida. Nós propomos utilizar o intervalo de confiança bootstrap percentil combinado com um esquema de reamostragem baseado em bootstrap ponderado. A construção dos intervalos para β_j ($j = 0, \ldots, p-1$) pode ser feita como se segue.

S1 Para cada i, $i = 1, \ldots, n$, obtenha t_i^* aleatoriamente de uma população com média zero e variância unitária;

S2 Construa uma amostra bootstrap (y^*, X), onde

$$y_i^* = x_i\widehat{\beta} + t_i^*\widehat{\varepsilon}_i/\sqrt{1-h_i},$$

sendo x_i a i-ésima linha de X;

S3 Calcule o EMQO de β: $\widehat{\beta}^* = (X'X)^{-1}X'y^*$;

S4 Repita os passos 1 a 3 um grande número de vezes (por exemplo, B vezes);

S5 Os limites inferior e superior do intervalo com $(1-\alpha)\times 100\%$ de confiança para β_j $(0 < \alpha < 1/2)$ são, respectivamente, os quantis $\alpha/2$ e $1-\alpha/2$ das B réplicas bootstrap $\widehat{\beta}^*_j$.

A quantidade t^*_i, $i = 1,\ldots,n$, deve ser amostrada aleatoriamente de uma população com média zero e variância unitária.

Comparamos, usando simulação Monte Carlo, o comportamento em amostras finitas dos HCIEs descritos anteriormente com o estimador intervalar híbrido (percentil/ponderado) descrito acima (Tabela 1.6). Inferências são realizadas sobre o parâmetro β_1, o número de réplicas Monte Carlo foi $5,000$ e o número de replicações bootstrap foi $B = 500$. Da comparação da Tabela 1.6 com as Tabelas 1.2, 1.3, 1.4 e 1.5 do Capítulo 1, concluímos que os intervalos HC4 apresentam melhor desempenho que os intervalos bootstrap.

Foram utilizados mais três estimadores bootstrap alternativos, que também são baseados no método percentil mas usam diferentes esquemas de reamostragem. O primeiro estimador alternativo emprega o bootstrap selvagem de Liu (1988); o segundo estimador, chamado de (y,X) bootstrap, utiliza, em vez de resíduos, reamostragem de pares $(x_i,y_i), i = 1,\ldots,n$; e o último estimador intervalar bootstrap utilizado combina reamostragem ponderada com o método percentil t (ver Efron e Tibshirani, 1993, pp.160–162). As taxas estimadas de cobertura e as amplitudes médias destes intervalos bootstrap estão nas Tabelas 1.7 e 1.8. Entre os intervalos bootstrap, o melhor desempenho quando $n = 20$ é do intervalo (y,X) bootstrap e o melhor desempenho quando o tamanho da amostra é grande $(n = 100)$ cabe ao intervalo bootstrap percentil t. Note, todavia, que os intervalos HC4 têm melhor desempenho do que todos os intervalos bootstrap.

5.6 Regiões de confiança

Nesta seção consideramos a obtenção de regiões de confiança assintoticamente válidas sob heteroscedasticidade de forma desconhecida. Escrevemos o modelo de regressão

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

como

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon, \tag{5.6.1}$$

onde y, X, β e ε são como descritos na Seção 5.2, X_j e β_j são $n\times p_j$ e $p_j\times 1$, respectivamente, $j = 1,2$, com $p = p_1 + p_2$ tal que $X = [X_1\ X_2]$ e $\beta = (\beta_1',\beta_2')'$.

O EMQO do vetor dos coeficientes de regressão em 5.6.1 é $\widehat{\beta} = (\widehat{\beta}_1',\widehat{\beta}_2')'$, onde

$$\widehat{\beta}_2 = \left(R_2'R_2\right)^{-1}R_2'y,$$

com $R_2 = M_1X_2$ e $M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$. Dado que $\widehat{\beta}_2$ é assintoticamente normal com média β_2 e matriz de covariâncias

$$V_{22} = \left(R_2'R_2\right)^{-1}R_2'\Omega R_2\left(R_2'R_2\right)^{-1},$$

a forma quadrática

$$W = \left(\widehat{\beta}_2 - \beta_2\right)' V_{22}^{-1} \left(\widehat{\beta}_2 - \beta_2\right)$$

é assintoticamente $\chi^2_{p_2}$. Este resultado permanece verdadeiro quando V_{22} é substituído por uma função dos dados $\ddot{V}_{22}$ tal que plim$(\ddot{V}_{22}) = V_{22}$. Em particular, nós podemos usar o seguinte estimador consistente da matriz de covariância de $\widehat{\beta}_2$:

$$\ddot{V}_{22}^{(k)} = (R_2'R_2)^{-1} R_2' \widehat{\Omega}_k R_2 (R_2'R_2)^{-1},$$

onde $\widehat{\Omega}_k$, $k = 0, 2, 3, 4$, é como definido na Seção 5.2.

Seja $0 < \alpha < 1$ e seja $\chi^2_{p_2,\alpha}$ tal que

$$\Pr(\chi^2_{p_2} < \chi^2_{p_2,\alpha}) = 1 - \alpha;$$

isto é, $\chi^2_{p_2,\alpha}$ é o quantil $1 - \alpha$ da distribuição $\chi^2_{p_2}$. Considere também

$$\ddot{W}^{(k)} = \left(\widehat{\beta}_2 - \beta_2\right)' \left(\ddot{V}_{22}^{(k)}\right)^{-1} \left(\widehat{\beta}_2 - \beta_2\right).$$

Então, a região com $100 \times (1 - \alpha)\%$ de confiança para β_2 é dada pelo conjunto de valores de β_2 tal que

$$\ddot{W}^{(k)} < \chi^2_{p_2,\alpha}. \tag{5.6.2}$$

Avaliaremos numericamente os desempenhos em amostras finitas das diferentes regiões de confiança. O modelo de regressão usado na simulação é

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, \ldots, n,$$

onde ε_i é um erro distribuído normalmente com média zero.

Os valores das covariáveis foram obtidos como realizações da distribuição uniforme padrão (desenho balanceado) e da distribuição t_3 (desenho não-balanceado). O número de réplicas de Monte Carlo foi 10,000, os tamanhos amostrais considerados foram $n = 20, 60, 100$ e $1 - \alpha = 0.95$. As simulações foram realizadas sob homoscedasticidade e sob heteroscedasticidade. As coberturas apresentadas correspondem às percentagens das réplicas nas quais (5.6.2) se verifica quando são realizadas inferências conjuntas sobre β_1 e β_2. A fim de comparar os desempenhos das regiões de confiança ('joint') e dos intervalos de confiança (para β_1 e β_2 separadamente) as coberturas individuais são também apresentadas na Tabela 1.10.

Observamos inicialmente na Tabela 1.10 que as coberturas conjuntas são sempre menores que as individuais. No geral, observamos que a região de confiança HC4 é a que tem melhor desempenho, especialmente quando os dados são não-balanceados. A região de confiança HC3 é competitiva quando o desenho de regressão é balanceado.

CHAPTER 6

Resumo do Capítulo 2

6.1 Introdução

A suposição de homoscedasticidade, quando são usados dados transversais nos modelos de regressão, é freqüentemente violada. Erros-padrão consistentes para o estimador de mínimos quadrados ordinários (EMQO) podem ser obtidos seguindo o enfoque proposto por White (1980). No entanto, estes erros-padrão são bastante viesados em amostras pequenas. Qian e Wang (2001), propuseram um estimador menos viesado que o proposto por White. Neste capítulo, nós definimos uma seqüência de estimadores ajustados pelo viés a partir do estimador proposto por Qian e Wang, melhorando sua precisão. Mostramos ainda, que o estimador de Qian e Wang pode ser generalizado para uma classe mais ampla de estimadores consistentes sob heteroscedasticidade e que os resultados obtidos para o estimador de Qian e Wang podem ser facilmente estendidos a esta classe de estimadores.

6.2 O modelo e estimadores da matriz de covariâncias

Considere toda a seção dois do Capítulo 4 para definição do modelo e dos estimadores HC0, HC1, HC2, HC3 e HC4.

Como observamos no Capítulo 4, o estimador HC0 é consideravelmente viesado em amostras de tamanho pequeno a moderado. Cribari–Neto, Ferrari and Cordeiro (2000) obtiveram variantes do estimador HC0 usando um mecanismo iterativo de correção de viés. A cadeia de estimadores foi obtida corrigindo HC0, em seguida corrigindo o estimador resultante e assim por diante.

Seja $(A)_d$ a matriz diagonal obtida fazendo os elementos não-diagonais da matriz quadrada A iguais a zero. Note que $\widehat{\Omega} = (\widehat{\varepsilon\varepsilon'})_d$. Então,

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

dado que $(I-H)X = 0$. Segue-se que $\mathbb{E}(\widehat{\Omega}) = \{(I-H)\Omega(I-H)\}_d$ e $\mathbb{E}(\widehat{\Psi}) = P\mathbb{E}(\widehat{\Omega})P'$. Portanto, os vieses de $\widehat{\Omega}$ e $\widehat{\Psi}$ como estimadores de Ω e Ψ são

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

e

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

respectivamente.

Cribari–Neto, Ferrari and Cordeiro (2000) definem o estimador corrigido pelo viés

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

Este estimador pode ser por sua vez corrigido pelo viés:

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

e assim sucessivamente. Após k iterações do esquema de correção obtém-se

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

Considere a seguinte função recursiva de uma matriz diagonal A $n \times n$:

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

onde $M^{(0)}(A) = A$, $M^{(1)}(A) = \{HA(H - 2I)\}_d$, e H é como definido anteriomente. Podemos então escrever $B_{\widehat{\Omega}}(\Omega) = M^{(1)}(\Omega)$. Por indução, pode ser mostrado que o estimador de k-ésima ordem corrigido pelo viés e seu respectivo viés podem ser escritos como

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

e

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

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

Define-se então uma seqüência de estimadores corrigidos pelo viés para a matriz de covariâncias como $\{\widehat{\Psi}^{(k)}, k = 1, 2, \ldots\}$, onde

$$\widehat{\Psi}^{(k)} = P\widehat{\Omega}^{(k)}P'. \tag{6.2.1}$$

O viés de $\widehat{\Psi}^{(k)}$ é

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

$k = 1, 2, \ldots$.

Mostra-se que $B_{\widehat{\Psi}^{(k)}}(\Omega) = O(n^{-(k+2)})$, isto é, o viés do estimador corrigido k vezes é de ordem $O(n^{-(k+2)})$, enquanto que o viés do estimador de White é $O(n^{-2})$.

6.3 Uma nova classe de estimadores ajustados pelo viés

Um estimador alternativo foi proposto por Qian e Wang (2001). Veremos que este estimador é uma variante do estimador HC0 corrigido pelo viés. Seja $K = (H)_d = \text{diag}\{h_1, \ldots, h_n\}$, i.e., K é a matriz diagonal que contém as medidas de alavancagem, e seja $C_i = X(X'X)^{-1}x_i'$ a i-ésima coluna da matriz H, onde x_i é a i-ésima linha de X, $i = 1, \ldots, n$.

Seguindo Qian and Wang (2001), definimos

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

onde

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

e

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

O estimador de Qian–Wang é definido como

$$\widehat{V}^{(1)} = PD^{(1)}P'. \tag{6.3.1}$$

Mostraremos inicialmente que o estimador em (6.3.1) é uma versão corrigida pelo viés do estimador proposto por White a menos de um fator de correção. Note que

$$\begin{aligned} d_i &= (\widehat{\varepsilon}_i^2 - \widehat{b}_i)g_{ii} \\ &= (\widehat{\varepsilon}_i^2 - C_i'\widehat{\Omega}C_i + 2\widehat{\varepsilon}_i^2 C_i'C_i)g_{ii}. \end{aligned} \tag{6.3.2}$$

Fazendo $k = 1$ (correção de primeira ordem), o estimador corrigido pelo viés em (6.2.1) pode ser escrito como $\widehat{\Psi}^{(1)} = P\widehat{\Omega}^{(1)}P'$, onde

$$\begin{aligned} \widehat{\Omega}^{(1)} &= \widehat{\Omega} - M^{(1)}(\widehat{\Omega}) \\ &= \widehat{\Omega} - \{H\widehat{\Omega}(H - 2I)\}_d \\ &= \text{diag}\{\widehat{\varepsilon}_i^2 - C_i'\widehat{\Omega}C_i + 2\widehat{\varepsilon}_i^2 h_i\}. \end{aligned} \tag{6.3.3}$$

Dado que $h_i = C_i'C_i$, é fácil ver que (6.3.2) é igual ao i-ésimo elemento da diagonal de $\widehat{\Omega}^{(1)}$ em (6.3.3), a menos da multiplicação por g_{ii}. Então,

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

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

Qian and Wang (2001) mostraram que $\widehat{V}^{(1)}$ é não-viesado para Ψ sob homoscedasticidade.

Utilizando o esquema de correção pelo viés descrito anteriormente, mostra-se que a forma geral do estimador de Qian-Wang corrigido k vezes é

$$\widehat{V}^{(k)} = PD^{(k)}P', \tag{6.3.4}$$

com

$$\begin{aligned} D^{(k)} &= 1_{(k>1)} \times M^{(0)}(\widehat{\Omega}) + 1_{(k>2)} \times \sum_{j=1}^{k-2} (-1)^j M^{(j)}(\widehat{\Omega}) \\ &+ \sum_{j=k-1}^{k} (-1)^j M^{(j)}(\widehat{\Omega}) G, \end{aligned} \tag{6.3.5}$$

$k = 1, 2, \ldots$, onde $1_{(\cdot)}$ é a função indicadora. Seu viés é

$$\begin{aligned} B_{D^{(k)}}(\Omega) &= (-1)^{k-1} M^{(k-1)}(\Omega)(G - I) \\ &+ (-1)^k M^{(k+1)}(\Omega) G, \end{aligned} \tag{6.3.6}$$

$k = 1, 2, \ldots$.

Definimos então a seqüência $\{\widehat{V}^{(k)}, k = 1, 2, \ldots\}$ de estimadores para Ψ. O viés de $\widehat{V}^{(k)}$ segue de (6.3.6) e (6.3.4):

$$B_{\widehat{V}^{(k)}}(\Omega) = P[B_{D^{(k)}}(\Omega)]P'. \tag{6.3.7}$$

Mostra-se que a ordem dos vieses em (6.3.7) é

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

o que leva a

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

Portanto, a ordem do viés do estimador de Qian-Wang de ordem k é a mesma do estimador de White de mesma ordem como em Cribari–Neto, Ferrari e Cordeiro (2000); ver Seção 6.2. (Lembre, entretanto, que $k = 1$ aqui, indica o estimador de Qian-Wang não-modificado, o qual é uma correção de primeira ordem do estimador de White).

6.4 Estimação da variância de combinações lineares dos elementos de $\widehat{\beta}$

Seja c um p-vetor de constantes tal que $c'\widehat{\beta}$ é uma combinação linear dos elementos de $\widehat{\beta}$. Definimos

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

O estimador corrigido de k-ésima ordem, dado em (6.3.4), é

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

e, dessa forma,

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

é o elemento de ordem k de uma seqüência de estimadores ajustados pelo viés para Φ, onde $D^{(k)}$, $k = 1, 2, \ldots$, é definido na equação (6.3.5).

Lembramos que quando $k = 1$ nós obtemos o estimador de Qian–Wang. Usando este estimador, nós obtemos

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

onde

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

Seja $W = (ww')_d$, onde $w = G^{1/2}P'c$. Podemos escrever

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

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

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

onde $\widehat{\alpha}_s$ é o elemento de ordem s da diagonal de $M^{(1)}(\widehat{\Omega}) = \{H\widehat{\Omega}(H - 2I)\}_d$ e w_s é o elemento de ordem s do vetor w. Então,

$$\widehat{\Phi}_{QW}^{(1)} = \widehat{\varepsilon}'W\widehat{\varepsilon} - \sum_{s=1}^{n}\widehat{\alpha}_s w_s^2. \tag{6.4.1}$$

Mostra-se que a equação (6.4.1) pode ser escrita em forma matricial como

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

Mais geralmente, podemos escrever

$$\widehat{\Phi}_{QW}^{(k)} = \widehat{\varepsilon}'Q^{(k)}\widehat{\varepsilon}, \quad k = 1, 2, \ldots, \tag{6.4.2}$$

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

Cribari–Neto, Ferrari e Cordeiro (2000) mostraram que o estimador da variância de $c'\widehat{\beta}$ usando HC0 é dado por

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

onde $Q_W^{(k)} = \sum_{j=0}^{k}(-1)^j M^{(j)}(B)$.

Escrevendo a forma quadrática em (6.4.2) como uma forma quadrática num vetor de variáveis aleatórias não-correlacionadas de média zero e variância unitária, temos

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

onde $C_{QW}^{(k)} = \Omega^{1/2}(I-H)Q^{(k)}(I-H)\Omega^{1/2}$ é uma matriz simétrica $n\times n$ e a é tal que $\mathbb{E}(a)=0$ e $\mathrm{cov}(a)=I$. No que se segue escreveremos $C_{QW}^{(k)}$ simplesmente como C_{QW} para simplificar a notação. Mostra-se que, quando os erros são independentes,

$$\mathrm{var}(\widehat{\Phi}_{QW}^{(k)}) = d'\Lambda d + 2\mathrm{tr}(C_{QW}^2), \tag{6.4.3}$$

onde d é um vetor coluna formado pelos elementos diagonais de C_{QW}, $\mathrm{tr}(C_{QW})$ é o traço de C_{QW} e $\Lambda = \mathrm{diag}\{\gamma_i\}$, onde $\gamma_i = (\mu_{4i} - 3\sigma_i^4)/\sigma_i^4$ é o excesso de curtose do i-ésimo erro. Quando os erros são normalmente distribuídos, $\gamma_i = 0$. Então, $\Lambda = 0$ e (6.4.3) se torna

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

Para a seqüência de estimadores HC0 corrigidos, obtemos (Cribari–Neto, Ferrari e Cordeiro, 2000)

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

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

6.5 Resultados numéricos

Utilizaremos as expressões exatas dos vieses e das variâncias das combinações lineares de $\widehat{\beta}$ para avaliar numericamente a eficiência das correções para amostras finitas dos estimadores de White (HC0) e Qian–Wang ($\hat{V}^{(1)}$). Calcularemos também as raízes quadradas dos erros quadráticos médios e os vieses maximais dos diferentes estimadores.

O modelo usado na avaliação numérica é

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1,\ldots,n,$$

onde $\varepsilon_1,\ldots,\varepsilon_n$ são erros não-correlacionados com $\mathbb{E}(\varepsilon_i)=0$ e $\mathrm{var}(\varepsilon_i) = \exp(\gamma x_{i2})$, $i=1,\ldots,n$. Nós usamos diferentes valores de γ a fim de variar o grau de heteroscedasticidade, medido através de $\lambda = \max\{\sigma_i^2\}/\min\{\sigma_i^2\}$, $i=1,\ldots,n$. Os tamanhos de amostra considerados foram $n = 20, 40, 60$. Para $n=20$, os valores das covariáveis x_{i2} e x_{i3} foram obtidos como realizações aleatórias das distribuições uniforme padrão $\mathcal{U}(0,1)$ e lognormal padrão LN$(0,1)$; sob o último desenho os dados contêm pontos de alavanca. Os vinte valores das covariáveis foram replicados duas e três vezes para os tamanhos de amostra 40 e 60, respectivamente. Isto foi feito para que o grau de heteroscedasticidade (λ) não mudasse com n.

As Tabelas 2.2 e 2.3 apresentam o viés relativo total do estimador da variância quando usamos os estimadores MQO, HC0 e suas quatro correções e o estimador de Qian-Wang ($\hat{V}^{(1)}$) e suas correções até quarta ordem. O viés relativo total é definido como a soma dos valores absolutos dos vieses relativos individuais; o viés relativo é a diferença entre a variância estimada de $\widehat{\beta}_j$ e a correspondente variância verdadeira dividida pela variância verdadeira, $j=0,1,2$. Observa-se que a correção proposta para o estimador de Qian-Wang é bem efetiva quando $n=20$, desenho não-balanceado e sob heteroscedasticidade, apresentando redução de até 23 vezes no viés relativo total quando $\lambda \approx 49$.

A Tabela 2.4 contém a raiz quadrada do erro quadrático médio total, que é definido como a soma dos erros quadráticos médios individuais standartizados pelas correspondentes variâncias verdadeiras. Nota-se que os valores correspondentes aos estimadores de Qian-Wang corrigidos são aproximadamente iguais aos equivalentes corrigidos do estimador HC0, especialmente quando $n = 40, 60$.

Determinamos também a combinação linear dos estimadores dos parâmetros de regressão cuja variância estimada apresenta viés máximo. Estes vieses maximais são dados pelos autovalores maximais das matrizes (dos valores absolutos) dos vieses.[1] Os resultados estão apresentados na Tabela 2.5. Nota-se que a correção iterativa que propusemos para melhorar o desempenho do estimador de Qian-Qang em amostras pequenas pode ser bem efetiva em alguns casos. Por exemplo, quando $n = 20, \lambda \approx 49$ e os valores das covariáveis foram selecionados da distribuição uniforme, o viés maximal do estimador de Qian-Wang foi reduzido de 0.285 a 0.012 (aproximadamente 24 vezes) após quatro iterações do esquema de ajuste pelo viés.

Foi realizado um experimento Monte Carlo com 10000 réplicas, a fim de avaliar o desempenho, em amostras finitas, de testes quasi-t baseados nos estimadores HC0 e de Qian-Wang e suas versões corrigidas até quarta ordem. Utilizou-se um modelo de regressão com duas variáveis regressoras $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$, $i = 1, \ldots, n$, erros normais com média zero e variância $\sigma_i^2 = \exp(a x_{i1})$. Os valores das covariáveis foram obtidos da distribuição t_3, onde aparecem pontos de alavanca. Para $n = 20$, $\lambda \approx 49$ e $\alpha = 0.05$, as taxas de rejeição estimadas para os testes HC0 e suas correções até quarta ordem foram respectivamente 17.46%, 16.20%, 18.31%, 18.71% e 15.97%; os valores correspondentes para os testes Qian–Wang e suas correções até quarta ordem foram 11.66%, 7.07%, 6.44%, 5.87% e 5.71%. Os testes baseados nos estimadores corrigidos de Qian–Wang apresentaram também melhor desempenho que o teste de Qian–Wang quando $\lambda = 1$ (15.28% para o teste de Qian–Wang e 8.35%, 7.59%, 7.04% e 6.58% para os testes baseados nos erros-padrão corrigidos) e quando $\lambda \approx 9$ (12.50% para o teste de Qian–Wang e 6.86%, 6.25%, 5.93% e 5.60% para os testes baseados nos estimadores corrigidos). Vemos então que as correções propostas podem conduzir a inferências mais precisas além de estimadores pontuais menos viesados.

Realizamos também simulações utilizando o bootstrap selvagem para obter um valor crítico para testes quasi-t baseados em HC3. Como sugerido por Flachaire (2005), o esquema de reamostragem foi realizado utilizando a distribuição de Rademacher. O número de réplicas de Monte Carlo foi 5000 e 500 réplicas bootstrap para cada réplica de Monte Carlo. As taxas de rejeição para o nível nominal 5%, $n = 20$, valores da covariável obtidos da distribuição t_3 e $\lambda = 1$, $\lambda \approx 9$ e $\lambda \approx 49$ foram 17.62%, 14.76% e 11.34%, respectivamente. Observamos que o bootstrap selvagem funcionou bem no caso balanceado para todos os tamanhos amostrais. No caso não-balanceado, os resultados foram satisfatórios apenas para $n \geq 60$. Observamos que o bootstrap selvagem teve melhor desempenho quando os dados possuíam pontos com nível moderado de alavancagem, e.g., quando $h_{\max} < 4p/n$.

[1] Lembramos que se A é uma matriz simétrica, então $\max_c c'Ac/c'c$ é igual ao maior autovalor de A; veja, e.g., Rao (1973, p. 62).

6.6 Ilustrações empíricas

São apresentadas duas aplicações empíricas utilizando dados reais que apresentam pontos de alta alavancagem. Foram calculados os erros-padrão dos estimadores dos parâmetros dos modelos utilizando os estimadores HC0 e $\hat{V}^{(1)}$ e suas respectivas versões corrigidas. Observou-se que na presença de pontos de alta alavancagem o estimador de Qian-Wang e suas versões corrigidas pelo viés apresentam erros-padrão maiores que os obtidos com o estimador HC0 e suas correções. Quando os pontos de alavanca são removidos, observa-se que os erros-padrão obtidos com as correções de quarta ordem de HC0 e $\hat{V}^{(1)}$ são semelhantes. Os valores de $h_{\max}$ para os casos considerados nos dois exemplos podem ser vistos nas Tabelas 2.8 e 2.9 do Capítulo 2. Os erros padrão obtidos para os diversos estimadores estão nas Tabelas 2.7 e 2.10.

6.7 Uma generalização do estimador de Qian–Wang

Nesta seção mostraremos que o estimador de Qian–Wang pode ser obtido corrigindo pelo viés o estimador HC0 e em seguida modificando este estimador para que seja não-viesado sob homoscedasticidade.

Mostraremos também que este enfoque pode ser aplicado às variantes de HC0: HC1, HC2, HC3 e HC4. Dessa forma, todos os resultados que obtivemos podem ser facilmente estendidos às versões modificadas das variantes de HC0.

Inicialmente, note que o estimador de White HC0 pode ser escrito como HC0 $= \widehat{\Psi}_0 = P\widehat{\Omega}_0 P' = PD_0\widehat{\Omega}P'$, onde $D_0 = I$. As variantes de HC0 podem também ser representadas de forma semelhante:

(i) HC1 $= \widehat{\Psi}_1 = P\widehat{\Omega}_1 P' = PD_1\widehat{\Omega}P'$, $D_1 = (n/(n-p))I$;

(ii) HC2 $= \widehat{\Psi}_2 = P\widehat{\Omega}_2 P' = PD_2\widehat{\Omega}P'$, $D_2 = \text{diag}\{1/(1-h_i)\}$;

(iii) HC3 $= \widehat{\Psi}_3 = P\widehat{\Omega}_3 P' = PD_3\widehat{\Omega}P'$, $D_3 = \text{diag}\{1/(1-h_i)^2\}$;

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

No que se segue denotaremos estes estimadores como HCi, $i = 0,1,2,3,4$. Foi mostrado que

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

Note que

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

e

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

Como vimos na Seção 6.2, podemos escrever $\widehat{\Psi}_i^{(1)} = P\widehat{\Omega}_i^{(1)}P'$, onde

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

Então,[2]

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

Quando $\Omega = \sigma^2 I$ (homoscedasticidade), segue-se que

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

onde $A = (I - K) + D_i\{K + HKH - 2KK\}_d$. Portanto, o estimador

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

é não-viesado:

$$\begin{aligned}
\mathbb{E}(\widehat{\Psi}_{iA}^{(1)}) &= \mathbb{E}(P\widehat{\Omega}_i^{(1)}A^{-1}P') \\
&= P\sigma^2 AA^{-1}P' \\
&= P\sigma^2 IP' \\
&= P\Omega P' \\
&= \Psi.
\end{aligned}$$

Notamos então que o estimador de Qian-Wang é um caso particular de $\widehat{\Psi}_{iA}^{(1)}$ quando $i = 0$, i.e., quando $D_0 = I$. De fato,

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

onde $\widehat{\Omega}_0^{(1)} = \widehat{\Omega} - M^{(1)}(\widehat{\Omega})$ e $A = \{I + HKH - 2KK\}_d$.[3]

O viés de $\widehat{\Psi}_{iA}^{(1)}$ sob heteroscedasticidade é

$$B_{\widehat{\Psi}_{iA}^{(1)}}(\Omega) = P[B_{\widehat{\Omega}_{iA}^{(1)}}(\Omega)]P',$$

onde

$$B_{\widehat{\Omega}_{iA}^{(1)}}(\Omega) = \Omega(A^{-1} - I) + (I - D_i)M^{(1)}(\Omega)A^{-1} - D_i M^{(2)}(\Omega)A^{-1}.$$

Apresentamos então uma expressão em forma fechada para o viés da classe de estimadores que consideramos nesta seção. Em particular, ela pode ser usada para corrigir estes estimadores.

[2] Lembramos que $\mathbb{E}(\widehat{\Omega} - \Omega) = M^{(1)}(\Omega)$ e que $M^{(1)}(M^{(1)}(\Omega)) = M^{(2)}(\Omega)$.

[3] Na Seção 6.2, $\widehat{\Omega}_0^{(1)}$ foi denotado como $\widehat{\Omega}^{(1)}$.

É importante notar que todos os resultados obtidos nas Seções 6.3 e 6.4 podem ser facilmente estendidas para a classe mais geral que consideramos aqui.

Vamos obter agora uma sequência de estimadores ajustados pelo viés a partir do estimador modificado

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

for $i = 1,\ldots,4$. (O caso $i = 0$ já foi contemplado quando corrigimos pelo viés o estimador de Qian–Wang. Fazendo $D_0 = I$, os resultados abaixo coincidirão com os obtidos para $\widehat{V}^{(1)}$). Seja $G_i = A_i^{-1}$.

O estimador ajustado uma vez pelo viés é

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

Seu viés pode ser expresso como

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

Após k iterações do esquema de correção pelo viés, obtemos

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

$k = 1,2,\ldots$. O viés deste estimador é dado por

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

$k = 1,2,\ldots$.

Podemos agora definir uma seqüência $\{\widehat{\Psi}_{iA}^{(k)}, k = 1,2,\ldots\}$ de estimadores ajustados pelo viés para Ψ, onde

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

é o estimador corrigido pelo viés de ordem k de Ψ e seu viés é

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

Em seguida obtemos a ordem do viés de $\widehat{\Omega}_{iA}^{(k)}$ para $k = 1, 2, \ldots$. Mostramos que $B_{\widehat{\Omega}_{iA}^{(k)}}(\Omega)$ é de ordem $O(n^{-k})$ o que implica em que $\widehat{\Psi}_{iA}^{(k)} = O(n^{-(k+1)})$, $i = 1, \ldots, 4$. Fazendo $k = 1$, vemos que a ordem dos vieses dos estimadores propostos nesta seção é maior que a ordem de $\widehat{\Psi}_{0A}^{(1)}$ (o estimador de Qian-Wang), que mostramos ser $O(n^{-3})$. Então, mesmo que o estimador de Qian–Wang seja um caso particular da classe de estimadores que propusemos aqui, os resultados relativos à ordem dos estimadores, não generaliza os resultados obtidos para o estimador de Qian–Wang (caso $i = 0$).

Usando os estimadores

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

para estimar a variância de $\Phi = c'\widehat{\beta}$, obtemos

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

quando os erros são independentes e normalmente distribuídos.

Finalmente, apresentamos os resultados de uma pequena avaliação numérica (usando o mesmo modelo e situações do experimento da Seção 6.5) onde calculamos o viés relativo total das versões corrigidas pelo viés dos estimadores HC0, HC1, HC2, HC3 e HC4 modificados; a modificação consiste em multiplicar estes estimadores por A^{-1} de modo a torná-los não-viesados sob homoscedasticidade. Os resultados estão apresentados na Tabela 2.11. Note que $\widehat{\Psi}_{0A}^{(1)}$ é o estimador de Qian–Wang $\widehat{V}^{(1)}$ (veja a Tabela 2.3). Observe que quando os dados são balanceados, os vieses relativos totais dos estimadores modificados HC1 até HC4 são menores que os obtidos com o estimador de Qian–Wang. Quando os dados são não-balanceados, pequeno tamanho de amostra ($n = 20$) e heteroscedasticidade ($\lambda \approx 9$ e $\lambda \approx 49$), o estimador HC4 modificado é consideravelmente menos viesado que o estimador HC0 modificado (estimador de Qian–Wang).

CHAPTER 7

Resumo do Capítulo 3

7.1 Introdução

Neste capítulo avaliaremos o desempenho, em amostras finitas, de testes sobre os parâmetros de modelos de regressão baseados em vários erros-padrão consistentes sob heteroscedasticidade. Como a presença de heteroscedasticidade é freqüente quando são usados dados transversais na análise de modelos de regressão, é importante verificar os desempenhos de testes quasi-*t* quando os erros-padrão são obtidos através de vários HCCMEs (heteroskedasticity-consistent covariance matrix estimators). Nosso principal objetivo é utilizar métodos de integração numérica para realizar uma avaliação exata (ao invés de usar simulação de Monte Carlo) do comportamento em amostras finitas de testes baseados em quatro erros-padrão assintoticamente corretos sob heteroscedasticidade recentemente propostos ($\hat{V}^{(1)}$, $\hat{V}^{(2)}$, HC4 e HC5). Nossos resultados também sugerem escolhas de constantes a serem usadas nas definições de $\hat{V}^{(2)}$ e HC5.

7.2 O modelo e alguns erros-padrão robustos sob heteroscedasticidade

O modelo de interesse é o modelo de regressão linear

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

onde y é um vetor de dimensão n de observações sobre a variável dependente, X é uma matriz fixa $n \times p$ de regressores (posto$(X) = p < n$), $\beta = (\beta_0, \ldots, \beta_{p-1})'$ é um p-vetor de parâmetros desconhecidos e $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ é um vetor n-dimensional de erros aleatórios. Cada erro ε_t, $t = 1, \ldots, n$, tem média zero e variância $0 < \sigma_t^2 < \infty$; os erros são não-correlacionados de modo que $\text{cov}(\varepsilon) = \Omega = \text{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$.

O estimador de mínimos quadrados ordinário (EMQO) de β é obtido da minimização da soma de quadrados dos erros e tem a forma: $\widehat{\beta} = (X'X)^{-1}X'y$. $\widehat{\beta}$ é não-viesado e sua matriz de covariâncias pode ser escrita como $\text{cov}(\widehat{\beta}) = \Psi = P\Omega P'$, onde $P = (X'X)^{-1}X'$. Sob homoscedasticidade, $\sigma_t^2 = \sigma^2 > 0 \; \forall t$ e, então, $\Psi = \sigma^2(X'X)^{-1}$.

Quando todos os erros têm a mesma variância, o EMQO $\widehat{\beta}$ é o melhor estimador linear não-viesado de β. Sob heteroscedasticidade, entretanto, $\widehat{\beta}$ deixa de ser eficiente, mas permanece não-viesado, consistente e assintoticamente normal.

A fim de realizar testes de hipóteses sobre os parâmetros de regressão é necessário estimar Ψ, a matriz de covariâncias de $\widehat{\beta}$. Sob homoscedasticidade, Ψ pode ser facilmente estimado como

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

onde $\widehat{\sigma}^2 = (y - X\widehat{\beta})'(y - X\widehat{\beta})/(n-p) = \widehat{\varepsilon}'\widehat{\varepsilon}/(n-p)$ é um estimador não-viesado da variância comum. Aqui,

$$\widehat{\varepsilon} = y - \widehat{y} = (I - H)y = My,$$

$H = X(X'X)^{-1}X'$ é uma matriz $n \times n$ simétrica e idempotente e $M = I - H$, onde I é a matriz identidade n-dimensional.

A matriz $H = X(X'X)^{-1}X'$ é chamada de 'matriz chapéu', uma vez que $Hy = \widehat{y}$. Seus elementos diagonais assumem valores no intervalo $(0, 1)$ e somam p, o posto de X, sendo portanto sua média igual a p/n. Observe-se que os elementos diagonais de H $(h_1, \ldots, h_n)$ são comumente usados como medidas de alavancagem das correspondentes observações; observações tais que $h_i > 2p/n$ ou $h_i > 3p/n$ são consideradas pontos de alavanca (veja Davidson e MacKinnon, 1993).

Nosso interesse reside na estimação de Ψ quando as variâncias dos erros não são constantes, isto é, desejamos estimar a matriz de covariâncias de $\widehat{\beta}$, dada por $(X'X)^{-1}X'\Omega X(X'X)^{-1}$, de forma consistente, independentemente do modelo ser ou não homoscedástico. White (1980) mostrou que Ψ pode ser consistentemente estimado através do seguinte estimador:

$$\text{HC0} = \widehat{\Psi}_0 = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1} = P\widehat{\Omega}_0P' = PE_0\widehat{\Omega}P',$$

onde $\widehat{\Omega} = \text{diag}\{\widehat{\varepsilon}_t^2\}$ e $E_0 = I$.

O estimador de White (HC0) é consistente sob homoscedasticidade e sob heteroscedasticidade de forma desconhecida. Entretanto, ele pode ser bastante viesado em amostras pequenas. Em particular, HC0 é tipicamente 'muito otimista', i.e., tende a subestimar a verdadeira variância em amostras finitas; os testes associados (i.e., testes cujas estatísticas empregam HC0) tendem assim a ser liberais. O problema é mais severo quando os dados incluem pontos de alavanca; veja, Chesher e Jewitt (1987).

Algumas variantes do estimador HC0 foram propostas na literatura. Elas incluem correções para amostras finitas na estimação de Ω e são dadas por:

i (Hinkley, 1977) $\text{HC1} = \widehat{\Psi}_1 = P\widehat{\Omega}_1P' = PE_1\widehat{\Omega}P'$, onde $E_1 = (n/(n-p))I$;

ii (Horn, Horn e Duncan, 1975) $\text{HC2} = \widehat{\Psi}_2 = P\widehat{\Omega}_2P' = PE_2\widehat{\Omega}P'$, onde

$$E_2 = \text{diag}\{1/(1-h_t)\};$$

iii (Davidson e MacKinnon, 1993) $\text{HC3} = \widehat{\Psi}_3 = P\widehat{\Omega}_3P' = PE_3\widehat{\Omega}P'$, onde

$$E_3 = \text{diag}\{1/(1-h_t)^2\};$$

iv (Cribari–Neto, 2004) $\text{HC4} = \widehat{\Psi}_4 = P\widehat{\Omega}_4P' = PE_4\widehat{\Omega}P'$, onde

$$E_4 = \text{diag}\{1/(1-h_t)^{\delta_t}\}, \quad \delta_t = \min\{4, nh_t/p\}.$$

Adicionalmente, Qian e Wang (2001) propuseram um estimador alternativo para $\text{cov}(\widehat{\beta})$, que denotaremos como $\widehat{V}_1$. Ele foi obtido corrigindo HC0 pelo viés e modificando o estimador resultante de modo a torná-lo não-viesado sob homoscedasticidade.

Seja $C_t = X(X'X)^{-1}x_t'$, $t = 1,\ldots,n$, i.e., C_t a t-ésima coluna de H (matriz chapéu); aqui, x_t é a t-ésima linha de X. Seja também

$$D_1 = \text{diag}\{d_{1t}\} = \text{diag}\{(\widehat{\varepsilon}_t^2 - \widehat{b}_t)g_{tt}\},$$

onde

$$g_{tt} = (1 + C_t'KC_t - 2h_t^2)^{-1}$$

e

$$\widehat{b}_t = C_t'(\widehat{\Omega} - 2\widehat{\varepsilon}_t^2 I)C_t;$$

aqui, $K = (H)_d$, i.e., $K = \text{diag}\{h_t\}$.

O estimador de Qian e Wang é $\widehat{V}_1 = PD_1P'$. Observamos que D_1 pode ser expresso em forma matricial como

$$D_1 = [\widehat{\Omega} - \{H\widehat{\Omega}(H - 2I)\}_d]G,$$

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

7.3 Estimação da variância de combinações lineares dos elementos de $\widehat{\beta}$

Seja c um dado p-vetor de constantes. Escrevemos a variância de uma combinação linear dos elementos de $\widehat{\beta}$ como

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

Podemos estimar Ψ usando HCi, $i = 0,\ldots,4$, obtendo assim o seguinte estimador de Φ:

$$\widehat{\Phi}_i = c'\widehat{\Psi}_i c = c'P\widehat{\Omega}_i P'c = c'PE_i\widehat{\Omega}P'c, \quad i = 0,\ldots,4.$$

Seja

$$V_i = (v_i v_i')_d, \tag{7.3.1}$$

onde $v_i = E_i^{1/2}P'c$, $i = 0,\ldots,4$.

Seguindo Cribari–Neto, Ferrari e Cordeiro (2000), podemos escrever

$$\widehat{\Phi}_i = z'G_i z,$$

onde $\mathbb{E}[z] = 0$, $\text{cov}(z) = I$ e

$$G_i = \Omega^{1/2}(I - H)V_i(I - H)\Omega^{1/2}.$$

Considere agora a estimação da matriz de covariâncias a partir do estimador proposto por Qian e Wang (2001): $\widehat{\Phi}_{QW_1} = c'\widehat{\Psi}_{QW_1}c = c'\widehat{V}_1 c$. Portanto,

$$\widehat{\Phi}_{QW_1} = c'\widehat{V}_1 c = c'PD_1P'c,$$

onde, como já definimos, $D_1 = [\widehat{\Omega} - \{H\widehat{\Omega}(H - 2I)\}_d]G$.

Seja A uma matriz diagonal de ordem $n \times n$ e seja $M^{(1)}(A) = \{HA(H - 2I)\}_d$. Portanto,

$$D_1 = \widehat{\Omega}G - M^{(1)}(\widehat{\Omega})G.$$

Sejam também $w = G^{1/2}P'c$ e $W = (ww')_d$. Segue-se que

$$\widehat{\Phi}_{QW_1} = w'\widehat{\Omega}w - w'M^{(1)}(\widehat{\Omega})w.$$

Mostra-se que

$$\widehat{\Phi}_{QW_1} = \widehat{\varepsilon}'[W - M^{(1)}(W)]\widehat{\varepsilon} = \widehat{\varepsilon}'V_{QW_1}\widehat{\varepsilon},$$

onde

$$V_{QW_1} = W - M^{(1)}(W). \tag{7.3.2}$$

Escrevendo $\widehat{\Phi}_{QW_1}$ como uma forma quadrática num vetor aleatório de média zero e matriz de covariâncias unitária obtemos

$$\widehat{\Phi}_{QW_1} = z'G_{QW_1}z,$$

onde $\mathbb{E}[z] = 0$, $\text{cov}(z) = I$ e

$$G_{QW_1} = \Omega^{1/2}(I - H)V_{QW_1}(I - H)\Omega^{1/2}.$$

7.4 Inferência usando testes quasi-t

Consideraremos agora testes quasi-t baseados nos erros-padrão obtidos dos HCCMEs descritos na Seção 7.2. Desejamos testar a hipótese nula $\mathcal{H}_0 : c'\beta = \eta$ contra a alternativa bilateral, onde c é um dado p-vetor e η é um dado escalar.

A estatística quasi-t dada por

$$t = \frac{c'\widehat{\beta} - \eta}{\sqrt{\widehat{\text{var}}(c'\widehat{\beta})}},$$

onde $\sqrt{\widehat{\text{var}}(c'\widehat{\beta})}$ é um erro padrão obtido de um dos HCCMEs descritos na Seção 7.2, não tem, sob a hipótese nula, distribuição t de Student. Entretanto, é fácil mostrar que, sob $\mathcal{H}_0$, a distribuição limite de t é $\mathcal{N}(0,1)$. Como conseqüência, a distribuição limite de t^2 sob $\mathcal{H}_0$ é χ^2_1.

Note que

$$\widehat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon.$$

Então, quando $\varepsilon \sim \mathcal{N}(0, \Omega)$,

$$\widehat{\beta} = \beta + (X'X)^{-1}X'\Omega^{1/2}z,$$

onde $z \sim \mathcal{N}(0, I)$, e podemos então escrever t^2 como quociente de duas formas quadráticas num vetor aleatório normal de média zero e covariância unitária. O numerador de t^2 pode ser escrito como

$$\begin{aligned}
(c'\widehat{\beta} - \eta)^2 &= \{c'\beta + c'(X'X)^{-1}X'\Omega^{1/2}z - \eta\}'\{c'\beta + c'(X'X)^{-1}X'\Omega^{1/2}z - \eta\} \\
&= \{(c'\beta - \eta) + c'(X'X)^{-1}X'\Omega^{1/2}z\}'\{(c'\beta - \eta) + c'(X'X)^{-1}X'\Omega^{1/2}z\} \\
&= (c'\beta - \eta)'(c'\beta - \eta) + 2(c'\beta - \eta)c'(X'X)^{-1}X'\Omega^{1/2}z \\
&+ z'\Omega^{1/2}X(X'X)^{-1}cc'(X'X)^{-1}X'\Omega^{1/2}z.
\end{aligned}$$

Na Seção 7.3 escrevemos $\widehat{\Phi} = \widehat{\text{var}}(c'\widehat{\beta})$ como uma forma quadrática num vetor aleatório de média zero e covariância unitária para seis HCCMEs:

(i) $\widehat{\Phi}_i = z'G_i z$, onde $G_i = \Omega^{1/2}(I-H)V_i(I-H)\Omega^{1/2}$, para os estimadores HC$i$, $i = 0,\ldots,4$;

(ii) $\widehat{\Phi}_{QW_1} = z'G_{QW_1}z$, onde $G_{QW_1} = \Omega^{1/2}(I-H)V_{QW_1}(I-H)\Omega^{1/2}$ para $\widehat{V}_1$.

Note que V_i e V_{QW_1} são definidos em (7.3.1) e (7.3.2), respectivamente.

Portanto,

$$t^2 = \frac{z'Rz}{z'G_{(\cdot)}z} + \frac{(c'\beta-\eta)'(c'\beta-\eta)+2(c'\beta-\eta)c'(X'X)^{-1}X'\Omega^{1/2}z}{z'G_{(\cdot)}z}, \tag{7.4.1}$$

onde $R = \Omega^{1/2}X(X'X)^{-1}cc'(X'X)^{-1}X'\Omega^{1/2}$, $G_{(\cdot)} = G_i$, $i = 0,\ldots,4$, para HCi, e $G_{(\cdot)} = G_{QW_1}$ para $\widehat{V}_1$.

Quando $c'\beta = \eta$, o segundo termo do lado direito de (7.4.1) desaparece e, como resultado,

$$\Pr(t^2 \leq \gamma | c'\beta = \eta) = \Pr_0(z'Rz/z'G_{(\cdot)}z \leq \gamma), \tag{7.4.2}$$

onde $\Pr_0$ denota 'probabilidade sob a hipótese nula'.

Na próxima seção, usaremos o algoritmo de integração numérica de Imhof (1961) para calcular a função de distribuição exata, sob $\mathcal{H}_0$, de t^2. O algorítmo permite o cálculo de probabilidades de quocientes de formas quadráticas em um vetor de variáveis normais. Portanto, adicionaremos a suposição de que os erros são normalmente distribuídos, i.e., assumiremos que $\varepsilon_t \sim \mathcal{N}(0,\sigma_t^2), t = 1,\ldots,n$. Na avaliação comparamos as distribuições exatas, sob a hipótese nula, de estatísticas de teste que usam diferentes erros-padrão robustos sob heteroscedasticidade com a distribuição nula assintótica (χ_1^2) usada no teste.

7.5 Avaliação numérica exata

Os cálculos numéricos realizados para obter (7.4.2) usando o algoritmo de Imhof (1961) foram feitos utilizando a linguagem de programação matricial `0x` (Doornik, 2001). Os resultados serão apresentados para diferentes valores de γ.

O seguinte modelo de regressão foi usado na avaliação:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1,\ldots,n,$$

onde ε_t, $t = 1,\ldots,n$, é normalmente distribuído com média zero e variância $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$; adicionalmente, $\mathbb{E}[\varepsilon_t\varepsilon_s] = 0\ \forall t \neq s$. Usamos

$$\lambda = \max\{\sigma_t^2\}/\min\{\sigma_t^2\}$$

como medida do nível de heteroscedasticidade. Quando os erros são homoscedásticos, $\lambda = 1$; por outro lado, quanto maior o valor de λ, maior a intensidade da heteroscedasticidade.

A hipótese nula a ser testada é $\mathcal{H}_0 : \beta_1 = 0$, i.e., $\mathcal{H}_0 : c'\beta = \eta$ com $c' = (0,1)$ e $\eta = 0$. A estatística de teste é dada por

$$t^2 = \widehat{\beta}_1^2 / \widehat{\text{var}}(\widehat{\beta}_1),$$

onde $\widehat{\text{var}}(\widehat{\beta}_1)$ é o elemento $(2,2)$ de um HCCME.

Inicialmente usamos $n = 25$ e então replicamos os valores da covariável para obter uma amostra de 50 observações. Consideramos dois desenhos de regressão: (i) sem pontos de alavanca (regressores gerados aleatoriamente da distribuição $\mathcal{U}(0,1)$), e (ii) com pontos de alavanca (regressores gerados aleatoriamente da distribuição t_3); veja Tabela 3.1.

Nas Figuras 3.1 e 3.2 são apresentados graficamente os valores das discrepâncias quantílicas relativas versus o correspondente quantil assintótico para $n = 25$ e $n = 50$. A discrepância quantílica relativa é definida como a diferença entre o quantil exato (estimado por simulação) e o quantil assintótico dividido pelo quantil assintótico. Quanto mais próximo de zero estiverem os valores, melhor a aproximação da distribuição nula exata da estatística de teste pela distribuição limite χ^2_1. (Em todas as figuras incluímos uma linha de referência horizontal indicando discrepância nula). Apresentamos resultados para estatísticas de teste que usam erros-padrão HC0, HC3, HC4 e $\widehat{V}_1$ (QW) sob homoscedasticidade e heteroscedasticidade e para os dois desenhos de regressão. Observamos nas Figuras 3.1 e 3.2 que os testes baseados no estimador HC0 apresentam o pior desempenho e que os testes baseados em HC4 apresentam o melhor desempenho (seguido dos testes HC3, $\widehat{V}_1$ e, finalmente, HC0) no caso mais crítico de desenho não-balanceado e forte nível de heteroscedasticidade. A Tabela 3.2 apresenta (para $n = 50$) as probabilidades $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ para diferentes testes estatísticos e valores de γ dados pelos quantis 0.90, 0.95 e 0.99 da distribuição nula assintótica χ^2_1. Quanto mais próximos dos valores 0.90, 0.95 e 0.99 estiverem as probabilidades calculadas, melhor será a aproximação baseada na distribuição limite.

Fizemos também uma avaliação numérica usando dados reais obtidos de Greene (1997, Tabela 12.1, p. 541) e que apresentam pontos de alta alavancagem. A variável de interesse (y) é o gasto per capita em escolas públicas e as variáveis independentes, x e x^2, são a renda per capita por estado em 1979 nos Estados Unidos e seu quadrado. O modelo de regressão usado foi

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + \varepsilon_t, \quad t = 1, \ldots, 50.$$

Os erros são não-correlacionados, cada ε_t sendo normalmente distribuído com média zero e variância $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$. Trabalhamos sob homoscedasticidade ($\lambda = 1$) e sob heteroscedasticidade ($\lambda \approx 50$). Os valores das covariáveis foram replicados duas e três vezes fornecendo amostras de tamanhos 100 e 150, respectivamente.

A Tabela 3.3 apresenta os pontos de alavanca para os três casos estudados que variaram de acordo com a inclusão ou não dos pontos de alavanca da amostra não-replicada. A Figura 3.3 apresenta as discrepâncias quantílicas relativas para os três tamanhos de amostras (50, 100, 150) considerando a situação em que as variâncias dos erros são iguais e quando são diferentes, empregando erros-padrão obtidos dos estimadores HC0, HC3, HC4 e $\widehat{V}_1$.

Quando $n = 50$ e $\lambda = 1$, a distribuição exata da estatística HC3 é melhor aproximada pela distribuição assintótica (χ^2_1) que as das estatísticas baseadas nos demais estimadores. Sob hete-

roscedasticidade, os comportamentos de todos os testes deterioram, sendo que o teste HC4 tem o melhor desempenho, especialmente no quantil 3.841 (quantil nominal 95%). O teste HC3 vem em segundo lugar enquanto que os testes HC0 e $\widehat{V}_1$ apresentam desempenhos bastante fracos.

A Tabela 3.4 contém as probabilidades $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ para diferentes estatísticas de teste no quantil assintótico 0.95 ($\gamma = 3.841$).

7.6 Um erro-padrão alternativo

Qian e Wang (2001) propuseram um HCCME alternativo, que denotaremos por $\widehat{V}_2$, definido como $\widehat{V}_2 = PD_2P'$, onde

$$D_2 = \text{diag}\{d_{2t}\} = \text{diag}\{\widehat{\varepsilon}_t^2 + \widehat{\sigma}^2 h_t\} = \widehat{\Omega} + \widehat{\sigma}^2 (H)_d = \widehat{\Omega} + \widehat{\sigma}^2 K.$$

Pode ser mostrado que, como o outro estimador de Qian e Wang ($\widehat{V}_1$), este HCCME é não-viesado quando as variâncias dos erros são iguais. Notamos que $\widehat{V}_2$ é uma versão modificada de HC0; a modificação é tal que o estimador torna-se não-viesado sob homoscedasticidade.

Baseados em $\widehat{V}_2$ os autores definiram uma família de HCCMEs indexada pelo vetor n-dimensional $f = (f_1, \ldots, f_n)'$ fazendo

$$d_{2t}(f_t) = f_t \widehat{\varepsilon}_t^2 + \widehat{\sigma}^2 \{1 - f_t(1 - h_t)\}, \quad t = 1, \ldots, n. \tag{7.6.1}$$

Aqui,

$$D_2(f_t) = \text{diag}\{d_{2t}(f_t)\} = A\widehat{\Omega} + \widehat{\sigma}^2 (I - A\Lambda),$$

onde

$$A = \text{diag}\{f_t\} \tag{7.6.2}$$

e

$$\Lambda = \text{diag}\{1 - h_t\} = I - K. \tag{7.6.3}$$

Mostra-se também que, sob homoscedasticidade, esta família de estimadores é não-viesada para qualquer escolha de f que dependa apenas dos regressores.

A fim de simplificar a notação, denotaremos de agora em diante $D_2(f_t)$ por D_2 e $\widehat{V}_2(f_t)$ por $\widehat{V}_2$.

Para reduzir a variabilidade induzida pela presença de pontos de alavanca, Qian e Wang sugerem usar

$$f_t = 1 - ah_t, \quad t = 1, \ldots, n, \tag{7.6.4}$$

em (7.6.1), onde a é uma constante real. Sua sugestão é utilizar $a = 2$ quando o objetivo é reduzir o erro quadrático médio (EQM) e utilizar um valor menor de a (mesmo zero) quando se deseja reduzir o viés. Denotaremos este estimador para um dado valor de a por $\widehat{V}_2(a)$.

Como foi feito na Seção 7.3, podemos obter um estimador da variância de uma combinação linear de $\widehat{\beta}$ usando $\widehat{V}_2$. Temos

$$\widehat{\Phi}_{QW_2} = c'\widehat{\Psi}_{QW_2}c = c'PD_2P'c,$$

onde $D_2 = A\widehat{\Omega} + \widehat{\sigma}^2(I - A\Lambda)$; A e Λ são como definidas em (7.6.2) e (7.6.3), respectivamente.

Seja $L = (n-p)^{-1}(I - A\Lambda)$. Então,[1]

$$D_2 = \widehat{\varepsilon}'\widehat{\varepsilon}L + A\widehat{\Omega}.$$

Sejam $\ell = L^{1/2}P'c$, $v^* = A^{1/2}P'c$ and $V^* = (v^*v^{*'})_d$. Podemos escrever

$$\widehat{\Phi}_{QW_2} \quad = \quad \widehat{\varepsilon}'V_{QW_2}\widehat{\varepsilon},$$

onde $V_{QW_2} = \ell'\ell I + V^*$.

Escrevendo $\widehat{\Phi}_{QW_2}$ como uma forma quadrática num vetor de variáveis aleatórias (z) não-correlacionadas, com média zero e variância unitária, obtemos

$$\widehat{\Phi}_{QW_2} = z'G_{QW_2}z,$$

onde

$$G_{QW_2} = \Omega^{1/2}(I - H)V_{QW_2}(I - H)\Omega^{1/2}.$$

7.7 Avaliação numérica de testes quasi-t baseada em $\widehat{V}_1$ e $\widehat{V}_2$

Inicialmente usaremos integração numérica para determinar o valor ótimo de a em (7.6.4) para testar hipóteses. Foi usado o modelo

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, \ldots, n.$$

Usando a estatística t^2 (equação (7.4.1)) testaremos $\mathcal{H}_0 : \beta_1 = 0$. A Figura 3.4 apresenta as diferenças entre $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, onde γ é o quantil 0.95 de χ^2_1, e 0.95, a probabilidade nominal assintótica. Notamos, entre outras coisas, que quando o desenho de regressão é balanceado é melhor usar $a = 0$, e se o desenho tem pontos fortes de alavanca é melhor utilizar $a \approx 15$.

A Figura 3.5 apresenta graficamente as discrepâncias quantílicas relativas de estatísticas de teste baseadas em $\widehat{V}_1$ e $\widehat{V}_2(a)$ para $a = 0, 2, 10, 15$. Observamos que, na ausência de pontos de alavanca e sob heteroscedasticidade, as distribuições nulas das estatísticas baseadas em $\widehat{V}_1$, $\widehat{V}_2(0)$ e $\widehat{V}_2(2)$ são bem aproximadas pela distribuição limite χ^2_1. No entanto, quando os dados são não-balanceados e sob heteroscedasticidade o melhor desempenho é do teste baseado em $\widehat{V}_2(a)$ com $a = 15$.

A Tabela 3.5 contém as probabilidade $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ calculadas nos quantis assintóticos 0.90, 0.95 e 0.99 para estatísticas de teste baseadas em HC3, HC4, $\widehat{V}_1$ e $\widehat{V}_2(a)$ com $a = 0, 2$ e 15 quando $\lambda = 1$, $\lambda \approx 50$ e $\lambda \approx 100$. Sob heteroscedasticidade e dados não-balanceados, nota-se que os estimadores HC4 e $\widehat{V}_2(15)$ conduzem aos testes com melhor desempenho.

Utilizamos novamente os dados de Greene como na Seção 7.5. Notamos da Figura 3.6, que contém as discrepâncias quantílicas relativas, que sob iguais variâncias dos erros, as menores discrepâncias são aquelas das estatísticas de teste baseadas em $\widehat{V}_2(0)$ e $\widehat{V}_2(2)$. Sob heteroscedasticidade, entretanto, os testes com melhor desempenho são $\widehat{V}_2(15)$ e HC4.

[1] Observe que $\widehat{\sigma}^2 = (n-p)^{-1}\widehat{\varepsilon}'\widehat{\varepsilon}$.

7.8 Um outro erro-padrão consistente sob heteroscedasticidade: HC5

Cribari–Neto, Souza e Vasconcellos (2007) propuseram o estimador HC5, que é dado por

$$HC5 = P\widehat{\Omega}_5 P' = PE_5\widehat{\Omega}P',$$

onde $E_5 = \text{diag}\{1/\sqrt{(1-h_t)^{\delta_t}}\}$ e

$$\delta_t = \min\left\{\frac{nh_t}{p}, \max\left\{4, \frac{nkh_{\max}}{p}\right\}\right\},$$

com $h_{\max} = \max\{h_1, \ldots, h_n\}$. Aqui, $0 < k < 1$ é uma constante; os autores sugerem usar $k = 0.7$ baseados em resultados de simulação de Monte Carlo.

Podemos também usar HC5 na estimação da variância de combinações lineares de $\widehat{\beta}$. Aqui, seguindo o procedimento da Seção 7.3, temos

$$\widehat{\Phi}_5 = \widehat{\varepsilon}' V_5 \widehat{\varepsilon},$$

onde $V_5 = (v_5 v_5')_d$ e $v_5 = E_5^{1/2} P' c$. Escrevendo $\widehat{\Phi}_5$ como uma forma quadrática num vetor z com média zero e matriz de covariâncias unitária obtemos $\widehat{\Phi}_5 = z' G_5 z$, onde $G_5 = \Omega^{1/2}(I-H)V_5(I-H)\Omega^{1/2}$.

Usaremos o algoritmo de Imhof (1961) para obter a distribuição nula exata das estatísticas de teste baseadas em HC5 e assim avaliar a aproximação de primeira ordem χ^2_1 usada no teste. Utilizaremos também resultados relativos aos testes HC3 e HC4 como referência. O modelo de regressão usado é

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t, \quad t = 1, \ldots, n.$$

Aqui, $\varepsilon_t \sim \mathcal{N}(0, \sigma_t^2)$, onde $\sigma_t^2 = \exp(\alpha_1 x_{1t} + \alpha_2 x_{2t})$, $t = 1, \ldots, n$; além disso, $\mathbb{E}[\varepsilon_t \varepsilon_s] = 0\ \forall t \neq s$. A hipótese nula é $\mathcal{H}_0 : c'\beta = \eta$, com $c' = (0, 0, 1)$ e $\eta = 0$, e a estatística de teste é

$$t^2 = \widehat{\beta}_2^2 / \widehat{\text{var}}(\widehat{\beta}_2),$$

onde $\widehat{\text{var}}(\widehat{\beta}_2)$ é um estimador consistente da variância. O tamanho da amostra é $n = 50$; cada valor da covariável foi replicado uma vez para obter $n = 100$. Há dois desenhos de regressão: balanceado (valores da covariável obtidos aleatoriamente da distribuição uniforme padrão) e não-balanceado (valores da covariável obtidos aleatoriamente da distribuição lognormal padrão). Veja Tabela 3.6.

A Figura 3.7 apresenta as discrepâncias quantílicas relativas. Quando os dados contêm pontos de alavanca e sob heteroscedasticidade, observamos que os testes HC4 e HC5 são notadamente superiores ao teste HC3, especialmente no entorno do quantil de maior interesse (quantil 0.95 da distribuição assintótica).

A Tabela 3.7 contém os valores das probabilidades $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, $\gamma = 3.841$ (quantil assintótico 0.95), para estatísticas de teste baseadas nos estimadores HC3, HC4 e HC5, sob homoscedasticidade e dois diferentes níveis de heteroscedasticidade. Os valores na Tabela 3.7 mostram que as probabilidades calculadas assumem valores próximos aos valores nominais

(assintóticos) quando os dados são balanceados. Quando o desenho é não-balanceado, entretanto, as probabilidades calculadas usando HC4 e HC5 estão mais próximas dos respectivos níveis desejados que as probabilidades calculadas usando HC3.

A próxima avaliação numérica usa os dados de educação (descritos na Seção 7.5), $\lambda = 1$ e $\lambda \approx 25$, para os casos $n = 50$ e $n = 47$. Os gráficos das discrepâncias quantílicas relativas são apresentados na Figura 3.8. Observamos que sob homoscedasticidade e desenho não-balanceado a distribuição nula da estatística de teste baseada em HC3 é a que é melhor aproximada pela distribuição χ^2_1. Sob heteroscedasticidade e na presença de pontos de alavanca, entretanto, os testes HC4 e HC5 apresentam comportamento superior com relação a HC3.

Para os dados de educação, a Tabela 3.8 apresenta os valores das probabilidades $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$ para as estatísticas baseadas nos estimadores HC3, HC4 e HC5, onde γ é o quantil 0.95 da distribuição χ^2_1.

Usaremos agora integração numérica para avaliar o impacto dos valores de k (usualmente tomado como 0.7) sobre a qualidade da aproximação oriunda da distribuição nula limite quando usamos testes HC5. A avaliação é baseada num modelo de regressão simples, erros não-correlacionados com média zero e variância $\sigma_t^2 = \exp(\alpha_1 x_t + \alpha_2 x_t^2)$. O tamanho da amostra é $n = 50$ e as covariáveis são selecionadas como realizações aleatórias da distribuição lognormal padrão. Consideraremos dois desenhos de regressão: não-balanceado ($h_{\max}/(3p/n) = 1.71$) e fortemente não-balanceado ($h_{\max}/(3p/n) = 3.58$).

Os resultados desta avaliação numérica são graficamente apresentados na Figura 3.9, que contém valores das diferenças entre $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, onde γ é o quantil 0.95 da distribuição χ^2_1, 0.95 sendo a probabilidade nominal assintótica. As discrepâncias entre as probabilidades são representadas graficamente contra k. Notamos que no caso não-balanceado o valor de k tem impacto pequeno sobre a qualidade da aproximação. Entretanto, no desenho fortemente não-balanceado, valores de k entre 0.6 e 0.8 conduzem a melhores aproximações. Como conseqüência, estes resultados sugerem que 0.7, o valor de k sugerido por Cribari–Neto, Souza e Vasconcellos (2007), é de fato uma boa escolha.

Na Figura 3.10 apresentamos as mesmas discrepâncias entre as probabilidades apresentadas na Figura 3.9 usando agora os dados de gastos públicos com educação. Novamente os valores de k entre 0.6 e 0.7 parecem ser uma boa escolha para os testes HC5.

A Tabela 3.9 contém as probabilidades $\Pr(t^2 \leq \gamma \mid c'\beta = \eta)$, onde γ é o quantil 0.95 da distribuição χ^2_1, para testes baseados em HC4 e HC5 (neste caso usando diferentes valores de k) usando um modelo de regressão com duas covariáveis cujos valores são obtidos da distribuição $\mathcal{LN}(0,1)$, com $n = 50$, sob homoscedasticidade e forte heteroscedasticidade. Aqui, $h_{\max}/(3p/n) \approx 3.60$, de modo que há forte alavancagem nestes dados. Os valores obtidos mostram que obtemos as melhores aproximações com relação à distribuição assintótica usando testes baseados em HC5, quando $k = 0.6, 0.7$.

Apresentaremos agora os resultados de 10,000 réplicas de Monte Carlo onde foram calculadas as probabilidades de rejeição de testes HC4 e HC5 (para diferentes valores de k). O modelo usado foi

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t, \quad t = 1, \ldots, 50.$$

Os erros são não-correlacionados, cada ε_t sendo normalmente distribuído com média zero e variância $\sigma_t^2 = \exp(\alpha_1 x_{1t} + \alpha_2 x_{2t})$. Os valores das covariáveis foram selecionados como rea-

lizações aleatórias da distribuição lognormal padrão. Usamos $\lambda = 1$ e $\lambda \approx 50$ e consideramos duas situações diferentes nas quais os valores de $h_{\max}/(3p/n)$ são 3.60 e 1.14. O interesse reside em testar $\mathcal{H}_0 : \beta_2 = 0$ contra $\mathcal{H}_1 : \beta_2 \neq 0$. As estatísticas de teste consideradas empregam erros-padrão HC4 e HC5 ($k = 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$).

A Tabela 3.10 apresenta as taxas de rejeição empíricas, sob a hipótese nula, ao nível nominal $\alpha = 0.05$. Quando o nível de alavancagem é forte ($h_{\max}/(3p/n) \approx 3.60$), $k = 0.6$ conduz ao teste HC5 com melhor desempenho. Quando a alavancagem é fraca, o teste HC5 é superado pelo HC4 independentemente do valor de k.

Chapter 8

Conclusões

O objeto de interesse desta tese foi o modelo de regressão linear. A suposição de que todos os erros têm variâncias iguais (homoscedasticidade) é comumente violada em análises de regressão que utilizam dados de corte transversal. Dessa forma é importante desenvolver e avaliar estratégias de inferência que sejam robustas à presença de heteroscedasticidade. Esta foi nossa principal motivação.

Inicialmente propusemos diferentes estimadores intervalares consistentes sob heteroscedasticidade (HCIEs) para os parâmetros do modelo de regressão linear. Eles são baseados em estimadores da matriz de covariâncias que são assintoticamente corretos sob heteroscedasticidade de forma desconhecida e também quando os erros têm a mesma variância. Nós consideramos também estimadores intervalares baseados em esquemas bootstrap. Nossas avaliações numéricas revelaram que os HCIE baseados no estimador HC4 apresentam o melhor desempenho, superando inclusive os estimadores intervalares que empregam esquemas de reamostragem bootstrap.

Em seguida, transferimos o foco para a obtenção de estimadores pontuais para variâncias e covariâncias. Nós consideramos um estimador consistente sob heteroscedasticidade para a matriz de covariâncias (HCCME) proposto por L. Qian e S. Wang em 2001, que é uma versão modificada do conhecido estimador de White (HC0). Nós obtivemos uma seqüência de estimadores ajustados por viés na qual os vieses dos estimadores diminuem à medida que avançamos na seqüência. Adicionalmente, generalizamos o estimador de Qian e Wang e obtivemos seqüências alternativas de estimadores melhorados. Nossos resultados numéricos mostraram que o esquema de ajuste por viés pode ser bastante eficaz em amostras pequenas.

Por fim, utilizamos testes de hipóteses para realizar inferências no modelo de regressão linear sob heteroscedasticidade de forma desconhecida. Adicionamos ao modelo a suposição de normalidade dos erros e usamos um algoritmo de integração numérica para calcular as funções de distribuição nulas exatas de diferentes estatísticas quasi-t, que foram então comparadas à distribuição limite sob a hipótese nula. Para isto, mostramos que tais estatísticas de teste podem ser escritas como quocientes de formas quadráticas em vetores aleatórios com distribuição normal padrão. Demos ênfase a estatísticas de teste que usam quatro erros-padrão recentemente propostos. Dois deles empregam constantes que são escolhidas de forma *ad hoc*, e nossos resultados sugeriram valores ótimos para estas constantes. Nossas avaliações numéricas favoreceram os testes baseados em erros-padrão HC4.

Appendix A

O algoritmo de Imhof

A.1 Algoritmo de Imhof

Seja $Q = v'Av$, em que A é uma matriz simétrica $n \times n$ dada e v é um vetor $n \times 1$ de variáveis aleatórias normalmente distribuídas com média $\mu = (\mu_1, \ldots, \mu_n)'$ e variância Ω. O problema a ser resolvido é calcular a probabilidade

$$\Pr(Q < x), \tag{A.1.1}$$

onde x é um escalar.

Se Ω é não-singular, através de uma transformação linear não-singular (Scheffé (1957) p. 418) podemos expressar Q na forma

$$Q = \sum_{r=1}^{m} \lambda_r \chi^2(h_r; \delta_r^2). \tag{A.1.2}$$

Os λ_r, $r = 1, \ldots, m$, são os valores próprios distintos de $A\Omega$, com h_r representando suas respectivas ordens de multiplicidade ($\Sigma_{r=1}^{m} h_r = n$), δ_r são certas combinações lineares das componentes do vetor μ e $\chi^2(h_r; \delta_r^2)$ são variáveis χ^2 independentes com h_r graus liberdade e parâmetro de não-centralidade δ_r^2.

A função característica de Q é

$$\phi(t) = \prod_{r=1}^{m} (1 - 2it\lambda_r)^{-\frac{1}{2}h_r} \exp\left(i \sum_{r=1}^{m} \frac{t\lambda_r \delta_r^2}{1 - 2it\lambda_r} \right). \tag{A.1.3}$$

Em Imhof (1961) vemos que a função de distribuição acumulada $F(x)$ de Q pode ser obtida pela integração numérica de uma fórmula de inversão. Tal fórmula foi derivada explicitamente por Gil-Pelaez (1951)

$$F(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{1}{t} \mathrm{Im}\{\exp(-itx)\phi(t)\} \mathrm{d}t,$$

onde $\mathrm{Im}\{q\}$ representa a parte imaginária de q, $i = \sqrt{(-1)}$ e $\phi(t)$ é a função característica de Q dada em (A.1.3). Usando algumas relações relativas a números complexos (ver Imhof (1961)) e fazendo a transformação $2t = u$, obtém-se

$$\Pr(Q < x) = \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \frac{\sin\theta(u)}{u\rho(u)} \mathrm{d}u, \tag{A.1.4}$$

onde

$$\theta(u) = \frac{1}{2}\sum_{r=1}^{m}[h_r \tan^{-1}(\lambda_r u) + \delta_r^2 \lambda_r u(1+\lambda_r^2 u^2)^{-1}] - \frac{1}{2}xu,$$

$$\rho(u) = \prod_{r=1}^{m}(1+\lambda_r^2 u^2)^{\frac{1}{4}h_r} \exp\left\{\frac{1/2\sum_{r=1}^{m}(\delta_r \lambda_r u)^2}{(1+\lambda_r^2 u^2)}\right\}.$$

A.2 Caso particular

Consideraremos o caso particular em que $\mu = 0$ e $\Omega = I_n$. Diagonalizaremos a matriz A utilizando sua forma *canônica sob similaridade ortogonal* e denotemos por λ_k e u_k, $k = 1, \ldots, n$, respectivamente, os autovalores e autovetores de A. Então,

$$Au_k = \lambda_k u_k, \quad k = 1, \ldots, n. \tag{A.2.1}$$

Chamando $U = (u_1 \quad u_2 \quad \ldots \quad u_n)$ a matriz dos n autovetores e sendo $D = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$ a matriz diagonal formada a partir dos autovalores, podemos representar o conjunto de equações (A.2.1) por

$$AU = UD.$$

Como a matriz A é simétrica, a matriz U é não-singular, de modo que $D = U^{-1}AU$, i.e., a matriz A é diagonalizável e a matriz diagonal obtida é formada pelos autovalores de A. Adicionalmente, sabemos que os autovetores de uma matriz simétrica são ortogonais entre si. Normalizando esses vetores obtemos U como uma matriz ortogonal de modo que $UU' = I_n$ e $D = U'AU$, i.e., $A = UDU'$. Dessa maneira, podemos reescrever a equação (A.1.1) em função do vetor $w = U'v$, que é normalmente distribuído com média $E[w] = U'\mu = 0$ e $\text{cov}(w) = U'U = I_n$, obtendo

$$\begin{aligned}
\Pr(v'Av < x) &= \Pr(v'UDU'v < x) \\
&= \Pr(w'Dw < x) \\
&= \Pr\Big(\sum \lambda_k w_k^2 < x\Big) \\
&= \Pr\Big(\sum \lambda_k \chi_1^2 < x\Big) \\
&= \Pr(Q < x).
\end{aligned} \tag{A.2.2}$$

Dessa forma, em (A.2.2) obtemos a expressão (A.1.1) para o caso em que a variável v tem média zero e matriz de covariâncias igual a I_n.

A.3 Função `ProbImhof`

A função `ProbImhof(x,A,B,m,S)`, apresentada a seguir e escrita na linguagem de programação Ox (Doornik (2001)) por Peter Boswijk (University of Amsterdam), avalia numericamente a probabilidade do quociente de formas quadráticas $(z'Az)/(z'Bz)$ de variáveis normais

ser menor ou igual do que um escalar $x > 0$, sendo A e B matrizes quadradas de ordem n, m sendo o vetor de médias da variável normal z e S sendo sua matriz de covariâncias. Se $B = 0$, então a distribuição de $z'Az$ é calculada.

Na função `QuanImhof(p,A,B,m,S)`, p é a probabilidade para a qual o quantil correspondente é calculado.

Utilizamos a função `ProbImhof` para avaliar a probabilidade

$$\Pr(t^2 \leq \gamma | c'\beta = \eta) = \Pr_{H_0}\left(\frac{z'Rz}{z'G_{(.)}z} \leq \gamma\right),$$

onde a variável z tem média zero e matriz de covariâncias I_n. A função `ProbImhof` faz a seguinte transformação:

$$\Pr_{H_0}\left(\frac{z'Rz}{z'G_{(.)}z} \leq \gamma\right) = \Pr(z'Rz - z'G_{(.)}z\gamma < 0) = \Pr(z'(R - \gamma G_{(.)})z < 0).$$

No código abaixo, a função `imhof_mgf(u)` calcula a expressão do integrando em (A.1.4) e `QAGI(imhof_mgf, 0, 1, &result, &abserr)` calcula a integral em (A.1.4).

```
// The function ProbImhof(x,A,B,m,S) calculates the cumulative
//distribution function, evaluated at x, of the ratio of quadratic
//forms, (z'Az)/(z'Bz), in a normal random vector z with mean vector
// m and variance matrix S.
//If B=0, then the distribution of z'Az is computed.
//
//ProbImhof(x,A,B,S)
// x  in: scalar, x-value at which distribution is evaluated;
// A  in: nxn matrix (is transformed into symmetric (A+A')/2);
// B  in: nxn matrix (is transformed into symmetric (B+B')/2 or 0;
// m  in: nx1 mean vector;
// S  in: nxn positive definite variance matrix (pd is not checked).
//
//Return value
// Returns the probability P[(z'Az)<=x]          if B==0 and
//                         P[(z'Az)/(z'Bz)<=x]   if B<>0
//with z~N[m,S].
//
//QuanImhof(p,A,B,m,S)
// p  in: scalar, probability at which quantile is evaluated;
// returns x;

#include <oxstd.h>
#include <oxfloat.h>
#include <quadpack.h>
```

```
static decl s_l, s_d, s_c;
const decl QUANT_MAXIT =200;
const decl QUANT_EPS   =1e-8;

static imhof_mgf(const u)
{
decl eps=0.5*(sumc(atan(s_l*u)+(s_d .^2) .*s_l*u ./(1+((s_l*u).^2)))
-s_c*u);
decl gam=prodc(((1+((s_l*u).^2)).^0.25).*exp(0.5*((s_d .*s_l*u).^2)
./(1+(s_l*u).^2)));
return (sin(eps)/(u*gam));
}

ProbImhof(const x, const A, const B, const m, const S)
{
decl Q, V, result, abserr;
decl P=choleski(S);
if(B==0)
{
Q = P'((A+A')/2)*P;
eigensym(Q,&s_l,&V);
s_c=x;
}
else
{
Q= A - B*x;
Q=(P'((Q+Q')/2)*P);
eigensym(Q,&s_l,&V);
s_c=0;
}
s_d=V'solvelu(P, 0, 0, unit(rows(P))) *m;
s_l=s_l';
s_d=selectifr(s_d,s_l);
s_l=selectifr(s_l,s_l);
QAGI(imhof_mgf,0,1,&result,&abserr);
return(0.5 - result/M_PI);
}

QuanImhof(const p, const A, const B, const m, const S)
{
decl i, pa, pb, xa, xb, w, diff, pn, xn;

if (p<=0)
```

```
    return 0;
//find an initial bracket

pb=xb=0.0;
do
{
    pa=pb;xa=xb;
xb=(xb+1)*2;
pb=ProbImhof(xb, A, B, m, S);
}while(pb<p);

for (i=0; ;++i)
{
diff=pb-pa;
w=diff > 0.01 ? 0.5 : (pb-p) /diff;
xn=xa*w+xb*(1-w);
pn=ProbImhof(xn, A, B, m, S);
if(pn<p)
    xa=xn,pa=pn;
else
    xb=xn,pb=pn;
if(pb - p <QUANT_EPS)
    return xb;
else if(p - pa < QUANT-EPS )
    return xa;
if(i>=QUANT-MAXIT)
    return .NaN;
}
}
```

Bibliography

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

[2] Cai, L.; Hayes, A. F. (2008). A new test of linear hypotheses in OLS regression under heteroscedasticity of unknown form. *Journal of Educational and Behavioral Statistics*, 33, 21–40.

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

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

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

[6] Cribari–Neto, F.; Ferrari, S.L.P.; Oliveira, W.A.S.C. (2005). Numerical evaluation of tests based on different heteroskedasticity-consistenty covariance matrix estimators. *Journal of Statistical Computation and Simulation*, 75, 611–628.

[7] Cribari–Neto, F.; Galvão, N.M.S. (2003). A class of improved heteroskedasticity-consistent covariance matrix estimators. *Communications in Statistics, Theory and Methods*, 32, 1951–1980.

[8] Cribari-Neto, F.; Souza, T.C.; Vasconcellos, K.L.P. (2007). Inference under heteroskedasticity and leveraged data. *Communications in Statistics, Theory and Methods*, 36, 1877-1888. Errata forthcoming.

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

[10] 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.

[11] Davidson, R.; MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

[12] Davidson, R.; MacKinnon, J.G. (2004). *Econometric Theory and Methods*. New York: Oxford University Press.

[13] Doornik, J. (2001). *Ox: an Object-oriented Matrix Programming Language*, 4th ed. London: Timberlake Consultants & Oxford: `http://www.doornik.com`.

[14] Efron, B. (1979). Bootstrapping methods: another look at the jackknife. *Annals of Statistics*, 7, 1–26.

[15] Efron, B.; Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. London: Chapman & Hall.

[16] Farebrother, R.W. (1990). The distribution of a quadratic form in normal variables. *Applied Statistics*, 39, 294–309.

[17] Ferrari, S.L.P.; Cribari–Neto, F. (1998). On bootstrap and analytical bias correction. *Economics Letters*, 58, 7–15.

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

[19] Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, 38, 481-482.

[20] Greene, W.H. (1997). *Econometric Analysis*, 3rd ed. Upper Saddle River: Prentice Hall.

[21] Godfrey, L.G. (2006). Tests for regression models with heteroskedasticity of unknown form. *Computational Statistics and Data Analysis*, 50, 2715–2733.

[22] Harrell, Jr., F.E. (2001). *Regression Modeling Strategies*. New York: Springer.

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

[24] 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.

[25] Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48, 419–426.

[26] Liu, R.Y. (1988). Bootstrap procedure under some non i.i.d. models. *Annals of Statistics*, 16, 1696–1708.

[27] Long, J.S.; Ervin, L.H. (2000). Using heteroskedasticity-consistent standard errors in the linear regression model. *American Statistician*, 54, 217–214.

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

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

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

[31] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.

[32] Scheffé, H. (1959). *The Analysis of Variance*, Wiley.

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

[34] Wooldridge, J.M. (2000). *Introductory Econometrics: a Modern Approach*. Cincinnati: South-Western College Publishing.

[35] Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, 14, 1261–1295.

[36] Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, 11:10.

This volume has been typeset in LaTeXwith the UFPEThesis class (www.cin.ufpe.br/~paguso/ufpethesis).