ANALYTICAL COMPARISON OF HCCMEs

BY

MUMTAZ AHMED

School of Economics

International Institute of Islamic Economics

International Islamic University Islamabad

2012

1- Economics

2- Economics Rights

ANALYTICAL COMPARISON OF HCCMEs

Researcher

Mumtaz Ahmed

Reg. No. 05-SE/PhD (Et)/F04

Supervisor

Prof. Dr. Asad Zaman

School of Economics

International Institute of Islamic Economics,

International Islamic University Islamabad, Pakistan

ANALYTICAL COMPARISON OF HCCMEs

A dissertation submitted to the

International Institute of Islamic Economics,

International Islamic University Islamabad, Pakistan

In partial fulfillment of the requirements

for the award of the degree of

Doctor of Philosophy in

Econometrics

By

MUMTAZ AHMED

AUGUST, 2012

RAMZAN, 1433 Hijrah

APPROVAL SHEET

ANALYTICAL COMPARISON OF HCCMEs

BY

MUMTAZ AHMED

Reg. No: 05-SE/PhD(Et)-F04

Accepted by the International Institute of Islamic Economics, International Islamic University, Islamabad, as partial fulfillment of the requirements for the award of degree of **DOCTOR OF PHILOSOPHY** in **ECONOMETRICS**

Supervisor: (Dr. Asad Zaman)

Internal Examiner: (Dr. Hamid Hassan)

External Examiner 1: (Dr. Eatzaz Ahmed)

External Examiner 2: (Dr. Rehana Siddiqui)

Dr. Pervez Zamurrad Janjua
Head
School of Economics
International Institute of Islamic Economics
International Islamic University Islamabad

Dr. Asad Zaman
Director General
International Institute of Islamic Economics
International Islamic University Islamabad

DECLARATION

I hereby declare that this thesis, neither as a whole nor as a part thereof, has been copied out from any source. It is further declared that I have carried out this research by myself and have completed this thesis on the basis of my personal efforts under the guidance and help of my supervisor. If any part of this thesis is proven to be copied out or earlier submitted, I shall stand by the consequences. No portion of work presented in this thesis has been submitted in support of any application for any other degree or qualification in International Islamic University or any other university or institute of learning.

Mumtaz Ahmed

بسم الله الرحمن الرحيم

In the name of Allah,
the Most Beneficent,
the Most Merciful

DEDICATIONS

To my beloved Parents for all the prayers, love and support

I don't have words to say them thanks

I pray for their long life!!!

ACKNOWLEDGEMENTS

Deepest thanks to Almighty ALLAH (تبارک وتعالیٰ) for blessing me with all the creative ideas, strength and courage to complete this thesis and always answering my prayers!

After ALLAH (تبارک وتعالیٰ), I would like to thank the last Prophet Hazrat Muhammad Mustafa (صلی اللہ علیہ وآلہ وسلم) for his high moral values, excellence and perfection of his teachings and guidance (Sunnah) and being the best role model for the whole mankind!

I would take this opportunity to express my extremely humble and respectable gratitude to my thesis supervisor, Dr. Asad Zaman for providing me the best research guidance throughout my study at IIUI. With his enthusiasm, inspiration, and great efforts to explain things clearly and simply, he helped to make econometrics fun for me. Throughout my thesis-writing period, he remained much supportive and provided encouragement, best advice, best company, and lots of great ideas. He is a role model for me as a researcher as well as a teacher. He always provided me help, guidance whenever needed and highlighted my mistakes in a very supportive and explainable way and always provided comments on my thesis on priority basis in spite of his much busy schedule. This is unforgettable for me. I would have been lost, for sure, without him. I wish him best of luck for his future life. My prayers are always with him.

I would like to thank my all teachers since my childhood for believing in me and making me strong enough to reach at this place. I salute them all from bottom of my heart. May Allah bless them all always!

In addition, I acknowledge all faculty members and staff at IIIE for always being so kind and providing me good company especially Dr. Abdul Jabbar who shared many useful ideas with me and always supported me. Niaz Ali Shah, M. Zareef, Aashiq, Tahir Furkhan are special for me who always helped and supported me in all type of administrative issues. In addition, Zaheer Ahmed and Tauqir Ahmed deserve special mention as both remained with me in difficult times and were always there for encouragement and help. I wish them best of luck in their studies as well as in their future life.

I would like to thank all my colleagues (Saud Ahmed, Iftikhar Hussain, Atiq-ur-Rehman, Tanvir-ul-Islam, Javed Iqbal , Jahanzeb Malik, M. Rehman, Atif Maqbool and Tahir Shah) for all the emotional support, camaraderie, entertainment, and caring they provided.

I also thank my all other friends (too many to mention here but they know who they are!) for providing support and friendship that I needed.

I highly acknowledge the financial support provided by the higher education commission (HEC) of Pakistan to complete this PhD.

Lastly, but most importantly, I wish to thank my family, especially my beloved and respected parents for all the support, prayers and unconditional love and care since my childhood. I don't have words to say that how much I love them. In addition, I wish to thank my sisters for all the prayers and love, and brothers for always providing a very loving and supportive environment and providing me opportunities to fully focus on my studies. I would say Hats-Off to all family members for all the care, love and support.

MUMTAZ AHMED

ABSTRACT

This thesis considers the issue of evaluating heteroskedasticity consistent covariance matrix estimators (HCCME) in linear heteroskedastic regression models. Several HCCMEs are considered, namely: HC0 (White estimator), HC1 (Hinkley estimator), HC2 (Horn, Horn & Duncan estimator) and HC3 (Mackinnon & White estimator). It is well known that White estimator is biased in finite samples; see e.g. Chesher & Jewitt and Mackinnon & White. A number of simulation studies show that HC2 & HC3 perform better than HC0 over the range of situations studied. See e.g. Long & Ervin, Mackinnon & White and Cribari-Neto & Zarkos.

The existing studies have a serious drawback that they are just based on simulations and not analytical results. A number of design matrices as well as skedastic functions are used but the possibilities are too large to be adequately explored by simulations. In the past, analytical formulas have been developed by several authors for the means and the variances of different types of HCCMEs but the expression obtained are too complex to permit easy analysis. So they have not been used or analyzed to explore and investigate the relative performance of different HCCMEs. Our goal in this study is to analytically investigate the relative performance of different types of HCCMEs. One of the major contributions of this thesis is to develop new analytic formulae for the biases of the HCCMEs. These formulae permit us to use minimax type

criteria to evaluate the performance of the different HCCMEs. We use these analytical formulae to identify regions of the parameter space which provide the ranges for the best and the worst performance of different estimators. If an estimator performs better than another in the region of its worst behavior, then we can confidently expect it to be better. Similarly, if an estimator is poor in area of its best performance, than it can be safely discarded. This permits, for the first time, a sharp and unambiguous evaluation of the relative performance of a large class of widely used HCCMEs.

We also evaluate the existing studies in the light of our analytical calculations. Ad hoc choices of regressors and patterns of heteroskedasticity in existing studies resulted in ad hoc comparison. So there is a need to make the existing comparisons meaningful. The best way to do this is to focus on the regions of best and worst performance obtained by analytical formulae and then compare the HCCMEs to judge their relative performance. This will provide a deep and clear insight of the problem in hand. In particular, we show that the conclusions of most existing studies change when the patterns of heteroskedasticity and the regressor matrix is changed. By using the analytical techniques developed, we can resolve many questions:

1) Which HCCME to use
2) How to evaluate the relative performance of different HCCMEs
3) How much potential size distortion exists in the heteroskedasticity tests
4) Patterns of heteroskedasticity which are least favorable, in the sense of creating maximum bias.

Our major goal is to provide practitioners and econometricians a clear cut way to be able to judge the situations where heteroskedasticity corrections can benefit us the most and also which method must be used to do such corrections.

Our results suggest that HC2 is the best of all with lowest maximum bias. So we recommend that practitioners should use only HC2 while performing heteroskedasticity corrections.

TABLE OF CONTENTS

DEDICATIONS VII

ACKNOWLEDGEMENTS VIII

ABSTRACT IV

TABLE OF CONTENTS IV

LIST OF FIGURES VII

LIST OF TABLES VIII

CHAPTER 1: INTRODUCTION TO HETEROSKEDASTICITY CONSISTENT COVARIANCE MATRIX ESTIMATORS (HCCMES) 9

1.1: INTRODUCTION 9
1.2: OLS METHOD UNDER HETEROSKEDASTICITY 13
1.3: METHODS OF HETEROSKEDASTICITY CORRECTIONS 14
1.4: EVALUATION OF HCCMEs: SIMULATION VERSUS ANALYTICS 15
1.5: ORGANISATION OF THE STUDY 17

CHAPTER 2: LITERATURE REVIEW 19

2.1: INTRODUCTION 19
2.2: HUCME 22
2.3: HCCMEs 22
2.3.1: EICKER-WHITE (HC0) ESTIMATOR 23
2.3.2: HINKLEY (HC1) ESTIMATOR 24
2.3.3: HORN, HORN AND DUNCAN (HC2) ESTIMATOR 25
2.3.4: MACKINNON AND WHITE (HC3) ESTIMATOR 27
2.3.5: SOME ESTIMATORS NOT EVALUATED 28
2.4: COMPARISON & EVALUATION OF HCCMEs 29
2.5: OBJECTIVES OF THE STUDY 33

CHAPTER 3: THE HETEROSKEDASTIC REGRESSION MODEL 34

3.1: INTRODUCTION 34
3.2: THE BASIC REGRESSION MODEL (SINGLE REGRESSOR CASE) 34
3.3: CENTERING THE REGRESSORS 36
3.4: ORDER OF CONSISTENCY 39

CHAPTER 4: A MINIMAX ESTIMATOR 47

4.1: INTRODUCTION 47
4.2: BIAS OF EW-TYPE ESTIMATES 48
4.3: MAXIMUM BIAS 51
4.3.1: MAXIMUM BIAS WITH NORMAL REGRESSORS 54
4.4: THE MINIMAX BIAS ESTIMATOR 60
4.4.1: AN INVARIANCE CONJECTURE 64
4.4.2: THE MINIMAX VALUE OF 'A' 67
4.4.3: EVALUATION OF RELATIVE PERFORMANCE 71

CHAPTER 5: BIASES OF HCCMES 74

5.1: INTRODUCTION 74
5.2: BIAS OF GENERAL ESTIMATOR 75
5.3: BIAS OF EICKER-WHITE (HC0) ESTIMATOR 81
5.3.1: FINITE SAMPLE BIAS OF HC0 81
5.3.2: SECOND ORDER ASYMPTOTIC BIAS OF HC0 82
5.4: BIAS OF HINKLEY (HC1) ESTIMATOR 84
5.4.1: FINITE SAMPLE BIAS OF HC1 84
5.4.2: SECOND ORDER ASYMPTOTIC BIAS OF HC1 86
5.5: BIAS OF HORN, HORN AND DUNCAN (HC2) ESTIMATOR 87
5.5.1: FINITE SAMPLE BIAS OF HC2 87
5.5.2: SECOND ORDER ASYMPTOTIC BIAS OF HC2 89
5.6: BIAS OF MACKINNON AND WHITE (HC3) ESTIMATOR 92
5.6.1: FINITE SAMPLE BIAS OF HC3 92
5.6.2: SECOND ORDER ASYMPTOTIC BIAS OF HC3 96

CHAPTER 6: ASYMPTOTIC MAXIMUM BIASES OF HCCMES 99

6.1: INTRODUCTION 99

6.2: ASYMPTOTIC MAXIMUM BIAS 100
6.2.1: ASYMPTOTIC MAXIMUM BIAS WITH SYMMETRIC REGRESSORS *103*
6.2.1.1: Asymptotic Maximum Bias of Eicker-White (HC0) Estimator 105
6.2.1.2: Asymptotic Maximum Bias of Hinkley (HC1) Estimator 107
6.2.1.3: Asymptotic Maximum Bias of Horn, Horn and Duncan (HC2) Estimator 108
6.2.1.4: Asymptotic Maximum Bias of Mackinnon and White (HC3) Estimator 110
6.2.1.5: Asymptotic Maximum Bias of Minimax Estimator 111
6.2.1.6: Asymptotic Comparison of HCCMEs (Symmetric Regressors Case) 112
6.2.2: ASYMPTOTIC MAXIMUM BIAS WITH ASYMMETRIC REGRESSORS *114*
6.2.2.1: Asymptotic Maximum Bias of Eicker-White (HC0) Estimator 118
6.2.2.2: Asymptotic Maximum Bias of Hinkley (HC1) Estimator 123
6.2.2.3: Asymptotic Maximum Bias of Horn, Horn and Duncan (HC2) Estimator 128
6.2.2.4: Asymptotic Maximum Bias of Mackinnon and White (HC3) Estimator 131
6.2.2.5: Asymptotic Maximum Bias of Minimax Estimator 134
6.2.2.6: Asymptotic Comparison of HCCMEs (Asymmetric Regressors Case) 137
6.3: EXTENSION TO MULTIPLE REGRESSORS CASE 142
6.3.1: CASE 1: ORTHOGONAL REGRESSORS *142*
6.3.2: CASE 2: PRIORITIZED REGRESSORS *143*
6.3.3: CASE 3: CATEGORIZED VARIABLES WITH UNIQUE RANKING *143*
6.3.4: CASE 4: GENERAL CASE *144*
6.4: A REAL WORLD APPLICATION OF HCCMEs 146

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS 148

APPENDICES 158

APPENDIX A: ESTIMATORS NOT COVERED IN THE STUDY 158
A.1: BOOTSTRAPPED BASED HCCMEs *158*
A.1.1: NAÏVE BOOTSTRAP ESTIMATOR 159
A.1.2: JACKKNIFE ESTIMATOR (JA) 160
A.1.3: WEIGHTED BOOTSTRAP ESTIMATOR 160
A.1.4: LEVERAGED ADJUSTED BOOTSTRAP ESTIMATORS 161
A.2: ESTIMATORS BASED ON GMM *161*
A.3: ROBUST HCCMEs *162*
A.4: BIAS CORRECTED HCCMEs *163*
A.5: HIGH-LEVERAGED BASED ESTIMATORS *164*
A.5.1: CRIBARI-NETO (HC4) ESTIMATOR 164
A.5.2: CRIBARI-NETO et al. (HC5) ESTIMATOR 165

LIST OF FIGURES

Figure 4-1: Positive and Negative Biases of Minimax HCCME 57

Figure 4-2: Positive and Negative Biases of Minimax HCCME by Varying 'a' with SS =100 and K =3 63

Figure 6-1: Comparison of second order asymptotic maximum Bias of all HCCMEs over 'ρ' with M^2=100 140

LIST OF TABLES

Table 4-1: Maximum Bias with varying skewness with fixed kurtosis at Sample size 100 for different set of regressors 66

Table 6-1: Overall Maximum Bias of HCCMEs 147

Chapter 1: INTRODUCTION TO HETEROSKEDASTICITY CONSISTENT COVARIANCE MATRIX ESTIMATORS (HCCMEs)

1.1: INTRODUCTION

The linear regression model is extensively used by applied researchers. It makes up the building block of most of the existing empirical work in econometrics, statistics and economics as well as the related fields. Regardless of this reality, very little is known about the properties of statistical inferences made from this model when customary assumptions are violated. In particular, classical linear regression model requires the researchers to assume that variances of the error term are same. This assumption often violated in cross sectional data. This is called heteroskedasticity.

We quote from Stock and Watson (2003) as,

> *"At a general level, economic theory rarely gives any reason to believe that errors are homoskedastic. It is therefore prudent to assume that errors might be heteroskedastic unless if you have compelling reasons to believe otherwise".*

Using the analysis of Shepp (1965), we can prove the existence of error sequences which are heteroskedastic, but the heteroskedasticity cannot be detected even asymptotically with 100% power. i.e.

"Homoskedasticity is potentially unverifiable even with an infinite amount of data".

Thus failure to reject the null of homoskedasticity does not provide sufficient insurance against the alternative of heteroskedasticity.

DEFINITION: Distinguishability

Let $(X_1, X_2, ...)$ be a sequence of independent and identically distributed (IID) positive random variables with common distribution F. Let $Y_i = X_i / \sigma_i$ for any sequence of constants σ_i. The sequence Y_i is distinguishable from the sequence X_i, if and only if there exists a sequence of hypotheses tests, T_n of null and alternative hypotheses:

$$\begin{array}{ll} H_0 : \sigma_i = 1, & i = 1, 2, ..., n \\ H_1 : \sigma_i \neq 1, & \text{for some } i \end{array}$$

Such that size of the tests goes to zero and power goes to one as n approaches infinity, i.e.

$$\lim_{n \to \infty} P(T_n \, rejects \, H_0 / H_0) = 0$$
$$\lim_{n \to \infty} P(T_n \, rejects \, H_0 / H_1) = 1$$

Theorem 0-1:

Let $Z=\{z_1,z_2,...\}$ be a sequence of independent and identically distributed (IID) positive random variables. Let $\sigma=\{\sigma_1,\sigma_2,...\}$ be a numerical sequence, where σ_n represents the error in scaling Z_n. The sequence Z is distinguishable from $\frac{Z}{\sigma}$ if and only if $\sum\log\left(\sigma_n^2\right)=+\infty$.

Proof: Let $W=\log\left(\frac{Z}{\sigma}\right)=\log(Z)-\log(\sigma)=X-a$, where $X=\log(Z)$ and $a=\log(\sigma)$. According to Shepp (1965), [if $X=\{X_1,X_2,...\}$ is a sequence of IID random variables and $a=\{a_1,a_2,...\}$ is a numerical sequence, a_n representing the error in centering X_n. Then the sequence X is distinguishable from the sequence X-a if $\sum a_n^2=+\infty$], so using Shepp's result, we can say that the sequence $\log(Z)$ is distinguishable from $\log\left(\frac{Z}{\sigma}\right)$ if and only if $\sum\log\left(\sigma_n^2\right)=+\infty$. Since log(Z) is a monotonic transformation of Z, so the same result will hold for the sequence Z. Hence we can say that the sequence Z is distinguishable from $\frac{Z}{\sigma}$ if and only if $\sum\log\left(\sigma_n^2\right)=+\infty$. This proves the theorem.■

So, we can say that heteroskedasticity may be present but may not be detectible and one should test for heteroskedasticity to get valid estimates.

The issue of heteroskedasticity arises in cross-sectional, time series as well as in finance data. We list some important examples of situations where heteroskedasticity arises.

1) In studies of family income and expenditures, one expects that high income families' spending rate is more volatile while spending patterns of low income families is less volatile. [See Gujarati (2004), Prais & Houthakker (1955), Greene (2003, Ch. 11, p. 215) and Griffiths et al. (1993) for an example of income and food expenditure].
2) In error-learning models where individuals benefits from their previous mistakes, for example, number of typing errors reduces with the increase in time spent on typing practice. This also reduces the variation among the typing mistakes. [See Pearce-Hall model (1980) for error learning theory, Gujarati (2004, Ch 11, p. 389) and Kennedy (2003) for examples].
3) When one or more regressors in a regression model has skewed distribution, e.g; the distribution of income, wealth and education in most societies is skewed which causes heteroskedasticity. [See Gujarati (2004, Ch 11, p. 389) for details].
4) If a regression model is misspecified (i.e. an important variable is ommited) then this misspecification can cause heteroskedasticity in regression errors. [See Gujarati (2004, Ch 11, p. 391), JB Ramsay (1969) for details].
5) Outliers in the data can cause heteroskedasticity. [See Gujarati (2004, Ch 11, p. 390].

6) Incorrect data transformation (e.g. ratio or first difference transformation) can lead to heteroskedasticity. [See Hendry (1995) & Gujarati (2004, Ch 11, p. 381) for details].

7) Incorrect functional form (e.g. linear versus log-linear models) can cause heteroskedasticity. [See Hendry (1995), Kennedy (2003), JB Ramsay (1969), Joachim Zietz (2001) and Gujarati (2004, Ch 11, p. 391) for details].

1.2: OLS METHOD UNDER HETEROSKEDASTICITY

Ordinary Least Square (OLS) method is most often used to get the parameter estimates in the linear regression model. When errors in the regression model are heteroskedastic, then Ordinary Least Squares estimates of the linear parameters remain unbiased and consistent but are no longer efficient.

The customary estimate of covariance matrix estimator of the OLS parameters becomes biased and inconsistent. This means that when heteroskedasticity is overlooked, the inferences in the regression model are no longer reliable.

Most of the econometricians and statisticians while performing the analysis, report the t-stats using the wrong standard errors, i.e. OLS standard errors which assume homoscedasticity. Use of heteroskedasticity consistent standard errors can change results in the sense that significance of the regressors change, i.e. significant regressors might appear insignificant and vice-versa.

1.3: METHODS OF HETEROSKEDASTICITY CORRECTIONS

In the literature three main methods are used to handle the problem of heteroskedasticity.

The first, which is more commonly used, is to test the regression errors for heteroskedasticity. If the test does not reject the null hypothesis of homoscedasticity, then OLS analysis is used. Otherwise suitable adjustments for heteroskedasticity are made by transforming the data, in log forms, etc. This method is known as Pre-testing.

The second method is to use HCCME – heteroskedasticity corrected covariance matrix estimators. These estimators were first proposed by Eicker (1963, 1967) and introduced by White (1980) into the econometric literature.

A third method, introduced by Newey and West (1987), extends this methodology to correct standard errors for both heteroskedasticity and potential autocorrelation, called heteroskedasticity and autocorrelation consistent (HAC) estimator. For the purpose of this thesis, we consider the simplest case, where heteroskedasticity is the only misspecification. This simplicity makes analytic derivations possible. The case of dynamic models would add substantial analytical complications, and is not treated here.

There are many reasons to suspect that the second method is superior to the first. Many simulation studies including Mackinnon & White (1985), Cribari-Neto & Zarkos (1999) and Cribari-Neto et al. (2007) support this conclusion, and show that it is better to use HCCMEs rather than do a pre-test for heteroskedasticity.

1.4: Evaluation of HCCMEs: Simulation versus Analytics

Since heteroskedasticity corrections are relatively easy to implement, and provide for more robust inference, there is general agreement with the idea that we should use HCCMEs.

A quote from Wooldridge (2000, pg. 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".

However, a number of practical obstacles have hindered widespread adoption of HCCMEs. The initial proposals of Eicker and White were found to have rather large small sample biases, usually downward, which results in wrong inferences in linear regression models (See, e.g., Bera et al., 2002; Chesher and Jewitt, 1987; Cribari-Neto et al., 2000; Cribari-Neto and Zarkos, 2001; Furno, 1997). But when sample size increases, the bias shrinks which makes White estimator a consistent one. Numerous alternatives have been proposed to reduce bias, but no clear cut winner has emerged. The presence of a large number of alternative HCCMEs with widely differing small sample properties and competing claims to superiority leaves the practitioner without guidance as to what to do.

A major problem in evaluating the performance of the HCCMEs is the complexity of the analytic formulae required for their evaluation. Chesher and Jewitt (1987) made some progress in this direction by deriving analytic formulae for the exact small sample bias of some important HCCMEs. Cribari-Neto (2000) and Cribari-Neto (2004) have provided some asymptotic analytic evaluations of biases. The extreme complexity of these formulae has hindered analytic comparisons of different HCCMEs.

Since analytics have not been possible, many simulation studies of the relative performance of the different HCCME's have been made. Simulation studies suffer from a serious defect in this area – the performance of the HCCME's is directly dependent both on the sequence of regressors and on the heteroskedastic sequence. This is an extremely high dimensional space of which only a miniscule portion can be explored via simulations. It stands to reason that each HCCME will have its regions of strengths and weaknesses within this parameter space. If so, choosing the regressors and heteroskedastic alternative in different ways will lead to conflicting evaluations. This appears to be reflected in the simulation studies which arrive at differing conclusions regarding the relative strengths of the different HCCME's.

The only solution to the problems with HCCMEs is to evaluate them analytically, which is what we undertake in this study. Because of the complexity of the algebra, we restrict our attention to the case of a single regressor. The extension of results to multiple regression models is also provided at the end, but this involves significant additional complications. Thus it is useful to set out the basic methodology in the simpler context of a single regressor model.

1.5: ORGANISATION OF THE STUDY

Rest of the thesis is organized as follows:

Chapter 2 gives literature review. The first section of this chapter provides an introduction to heteroskedasticity corrections. The second section includes the discussion of the usual estimator of the OLS covariance matrix which we label heteroskedasticity-unadjusted covariance matrix estimator (HUCME). Section 3 of the same chapter provides the discussion of heteroskedasticity consistent covariance matrix estimators (HCCMEs). Here we included only those estimators which were compared in this study while the details of all other estimators are provided briefly in the appendix for interested readers. The last section of chapter 2 discusses some issues regarding comparison of HCCMEs and also presents a critique of existing studies.

Chapter 3 discusses the main heteroskedastic regression model used throughout the study and the related issues regarding consistency of the HUCME. The model considered in the thesis is one regressor case to lay the basis for more complex analysis later.

Chapter 4 provides the analytical apparatus which permits us to introduce and derive a new minimax estimator that has substantially smaller bias than the standard Eicker-White. The minimax properties of the same are discussed as well.

Chapter 5 presents results for the bias of general estimator which takes HC0, HC1, HC2 and HC3 as its special cases and analytically provides their finite samples as well as asymptotic bias.

Chapter 6 includes results regarding maximum bias of all HCCMEs. The bias formulae simplify substantially for the case of symmetric regressors, so this is treated separately from the case of asymmetric regressors. The chapter concludes with a comparison of the HCCMEs based on the maximum bias. The last section of the same chapter extends the simple regression model to multiple regressors' case and related issues are discussed.

Chapter 7 provides the conclusion and further recommendations.

At the end, an appendix is provided, which includes the discussion of all other estimators which have not been taken into account in our thesis for interested readers.

Chapter 2: LITERATURE REVIEW

2.1: INTRODUCTION

It is well known that Ordinary Least Squares (OLS) estimates are inefficient though they remain unbiased and consistent under heteroskedasticity. The covariance matrix of the OLS estimates becomes inconsistent as well as biased because it is based on the assumption that regression errors are homoskedastic. Thus there is a need to get correct estimates of the covariance matrix of OLS estimates in the presence of heteroskedasticity to make valid statistical inferences.

The usual methodology to deal with heteroskedasticity, practiced by the researchers and practitioners, is to use OLS estimate of the regression parameters which is unbiased and consistent but not efficient along with the covariance matrix which is consistent regardless of whether variances are same or not. This strategy introduces heteroskedasticity consistent covariance matrix estimators, commonly known as HCCMEs.

There is a large literature on how to get a consistent covariance matrix estimator of OLS estimates of regression parameters under heteroskedasticity. A number of attempts have been made in this direction, which are the main source of literature regarding heteroskedasticity corrections.

There are two main approaches being used to find the variance covariance matrix estimator of OLS estimates of true parameters in linear regression model.

a. Finding HCCMEs by modifying the original Eicker and White estimator
b. Finding HCCMEs by employing bootstrap methodology

We will discuss only the first approach here which is the focus of the present study. The detail of second approach is given briefly in the appendix A for the interested readers.

For more clarity and ease, it is useful to introduce the forthcoming discussion in the context of a standard heteroskedastic regression model.

Consider a linear regression model,

EQ: 2-1 $$y = X\beta + \varepsilon$$

Where, y is the T x 1 vector of dependent variable, X is T x K matrix of regressors, β is the K x 1 matrix of unknown parameters and ε is the T x 1 vector of unobservable errors with mean zero and covariance matrix, Σ. i.e. E(ε) = 0 and Cov (ε) = Σ, where, Σ is a diagonal matrix, i.e. $\Sigma = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_T^2)$.The regression model in ***EQ: 2-1*** is a general heteroskedastic model.

For later use, define the OLS estimator of parameter β as, $\beta = (X'X)^{-1} X'y$. Let 'e' be the vector of OLS residuals, defined as, $e = y - X\beta$. Then it is easily calculated that the true covariance matrix of OLS estimator (β) under heteroskedasticity is:

$$\Omega = (X'X)^{-1} X'\Sigma X(X'X)^{-1}$$

All estimators of the covariance matrix (Ω) to be discussed are based on replacing Σ by some estimate, Σ. We now list the main estimators which we will analyze in our study. Our main concern is the analysis of four most popular HCCMEs which are mostly used by the software packages. These include, HC0 introduced by Eicker-White (1980), HC1 suggested by Hinkley (1977), HC2 proposed by Horn, Horn & Duncan (1975) and finally HC3 introduced by Mackinnon and White (1985). There are two more HCCMEs in this sequence which are concerned with the high-leveraged observations in the design matrix. We will not discuss them since our focus in this thesis is the *balanced* regressors case; where there are no outliers in the design matrix [See ***EQ: 5-19*** for details].

We discuss each of the above mentioned estimators briefly in the following subsections.

2.2: HUCME

Recall that $\beta = (X'X)^{-1} X'y$ and that the OLS residuals are $e = y - X\beta$. The usual OLS estimate of regression error variance is:- $s^2 = \dfrac{e'e}{T-K}$

The simplest estimator of Ω ignores heteroskedasticity and estimates Σ by $\hat{\Sigma}_{OLS} = s^2 I_T$. We will call this the HUCME: Heteroskedasticity Unadjusted Covariance Matrix Estimator. Then the usual estimator of Ω simplifies to:

$$\Omega_{OLS} = s^2 (X'X)^{-1}$$

This estimator of the covariance matrix is inconsistent (See **Theorem 0-1** for details). This means that confidence intervals based on it will be wrong, even in very large samples. Consequently, we will make wrong decisions regarding the significance or otherwise of the regressors.

2.3: HCCMEs

To resolve the problem of the inconsistency of the HUCME, a large number of heteroskedasticity-consistent covariance matrix estimators (HCCMEs) have been introduced in the literature. The first of these is the Eicker-White (HC0) estimator. We will also discuss three other estimators below. These estimators have an algebraic structure which permits analytic analysis by our methods. There are many other estimators which cannot be analyzed so easily; some of these alternatives are discussed in Appendix A for the sake of completeness. The sections below will introduce and discuss estimators we plan to analyze in this thesis.

2.3.1: EICKER-WHITE (HC0) ESTIMATOR

Literature of heteroskedasticity consistent covariance matrix estimators begins with the influential paper by Eicker (1963, 1967), who introduced the first heteroskedasticity-consistent covariance matrix estimator (HCCME) in statistics literature. This estimator consistently estimates the covariance matrix of OLS estimator in the presence of arbitrary heteroskedasticity in error variances. His estimator was generalized by White (1980) to cover many types of dynamic models used by econometricians. We will call it the Eicker-White (EW) throughout this study to acknowledge the priority of Eicker who was the first to introduce this idea. The novelty of Eicker-White lies in the possibility of finding a consistent estimator for the OLS covariance matrix even though the heteroskedastic error variances cannot themselves be consistently estimated, and form an infinite dimensional nuisance parameter. EW estimator has been labeled HC0 in the literature and is consistent under both homoskedasticity and heteroskedasticity of unknown form. This estimator is commonly used to construct quasi-t statistics and asymptotic normal critical values. The quasi-t statistic based on this estimator displays a tendency to over reject the null hypothesis when it is true; i.e. these tests are typically too liberal in finite samples. [See Cribari-Neto & Zarkos (2004), Cribari-Neto et al. (2007)].

The main idea of the Eicker-White estimator is to replace the unknown variances by the squares of the OLS residuals. i.e., Eicker-White estimator (HC0) estimates the unknown Σ by $\hat{\Sigma}$, where, it replaces the unknown variances in Σ by the OLS squared residuals. Thus estimated covariance matrix has the form,

$$\Omega_{HC0} = (X'X)^{-1}(X'\hat{\Sigma}_{HC0}X)(X'X)^{-1}$$

Where,

$$\hat{\Sigma}_{HC0} = diag(e_1^2, e_2^2, \cdots, e_T^2)$$

Here, e^2's are the squares of the OLS residuals, $e = y - X\beta$.

Using the OLS squared residuals to estimate the unknown variances was a good starting point and opens a new area of research for the researchers. Since Eicker, many alternatives have been proposed in the literature. Of these, the ones analyzed in this thesis are discussed below.

2.3.2 HINKLEY (HC1) ESTIMATOR

Since the average of the OLS squared residual is a biased estimate of the true variances, Hinkley proposed an alternative to Eicker's estimator in 1977. Instead of dividing by 'T', Hinkley's estimator, known as HC1, divides by (T-K). Hence the degree of freedom adjustment done by Hinkley gave another heteroskedasticity consistent estimator which should be superior to HC0.

Hinkley estimator can be written as: $\Omega_{HC1} = (X'X)^{-1}(X'\hat{\Sigma}_{HC1}X)(X'X)^{-1}$

Where, $$\hat{\Sigma}_{HC1} = \frac{T}{T-K} diag\left(e_t^2\right), \qquad t = 1,2,...,T$$

2.3.3: HORN, HORN AND DUNCAN (HC2) ESTIMATOR

Horn, Horn and Duncan (1975) proposed another alternative to Eicker's estimator. They suggested that under homoskedasticity, the ratio of expected value of OLS squared residuals and the discounting term is equal to true variance. The discounting term is (1-h_{tt}), where h_{tt} is the t^{th} entry of the Hat matrix, $H = X(X'X)^{-1}X'$. Their estimator is known as HC2 in the literature. To motivate the HC2, we proceed as follows:

Let $e = y - X\hat{\beta}$ be the OLS residual vector.

Note that we can write 'e' as $e = My = M\varepsilon$, where $M = I - H$

Consider $Cov(e) = Cov(M\varepsilon) = MCov(\varepsilon)M' = M\Sigma M' = (I-H)\Sigma(I-H)'$

Since M=I-H is symmetric, so, $Cov(e) = (I-H)\Sigma(I-H)$

Note that the above is the covariance matrix in case of heteroskedasticity.

But in case of homoscedasticity, i.e, $\Sigma = \sigma^2 I$

$Cov(e) = \sigma^2 (I - H)$. Here we made use of the fact that M is idempotent, i.e. $M^2=M$.

Hence in case of homoscedasticity,

$$Var(e_t) = E(e_t^2) = (1 - h_{tt})\sigma_t^2$$

This can be written as $E\left(\dfrac{e_t^2}{1-h_{tt}}\right) = \sigma_t^2$

From above expression, we can see that $\dfrac{e_t^2}{1-h_{tt}}$ is an unbiased estimator of true variances in case of homoscedasticity. This property need not hold under heteroskedasticity; that is why, Horn, Horn and Duncan called it a 'nearly' unbiased estimator. Dividing OLS squared residuals by the corresponding entries of the discounting term, $(1-h_{tt})$, gives the Horn, Horn and Duncan HCCME (HC2) which is an unbiased estimator under homoskedasticity.

HC2 estimate of the covariance matrix is:

$$\Omega_{HC2} = (X'X)^{-1}(X'\hat{\Sigma}_{HC2}X)(X'X)^{-1}$$

Where, $$\hat{\Sigma}_{HC2} = diag\left(\frac{e_t^2}{1-h_{tt}}\right),\ t = 1, 2, \ldots, T$$

2.3.4: MACKINNON AND WHITE (HC3) ESTIMATOR

The fourth estimator was suggested by Mackinnon and White (HC3). Their objective was to improve HC2 by dividing each of the squared residual by the square of the discounting term, $(1-h_{tt})$, $t = 1,2,...,T$, where, h_{tt} is the t-th entry of the Hat matrix, H, $H = X(X'X)^{-1}X'$.

HC3 estimator is given by:

$$\Omega_{HC3} = (X'X)^{-1}(X'\hat{\Sigma}_{HC3}X)(X'X)^{-1}$$

Where,

$$\hat{\Sigma}_{HC3} = diag\left(\frac{e_t^2}{(1-h_{tt})^2}\right), \qquad t = 1,2,...,T$$

Dividing by the $(1\text{-}h_{tt})^2$ leads to over-correcting the OLS residuals, but if the regression model is heteroskedastic, observations with large variances will tend to influence the estimates heavily, and they will therefore tend to have residuals that are too small. Thus this estimator may be attractive if large variances are associated with large values of h_{tt}. HC3 is proved to be a close approximation to Jackknife estimator [See Efron, 1979] by Mackinnon and White (1985).

2.3.5: SOME ESTIMATORS NOT EVALUATED

Later in 1997, a robust version of HCCMEs was introduced by Furno (1997) and she showed that small sample bias can be reduced by following her estimators. These estimators are also covered in detail in the appendix A of the study. These estimators are too complex to allow for analytic analysis, and hence we do not consider them in the present study.

Cribari-Neto et al. (2000) introduced bias corrected versions of the HC0 and then Cribari-Neto and Galvao (2003) generalized Cribari-Neto et al. (2000) idea to give the bias corrected versions of HC1, HC2 and HC3 along with HC0. These are also provided in the appendix for the interested readers, as they are not directly relevant with the present study.

During the last decade the focus of research shifted to the estimators which work well when the regression design contains high influential observations. This leads to the development of two new HCCMEs for the high-leveraged regression designs. These are known as HC4 proposed by Cribari-Neto (2004) and HC5 suggested by Cribari-Neto et al. (2007). These are also not covered in the current study since this study deals with the balanced regressors. But their detailed versions are provided in the appendix A for the interested readers.

In our study we are considering only HC0 to HC3 based estimators because the analysis of these four estimators requires substantial analytical work. The analysis of HC4 and HC5 can be covered in a later study.

An alternative stream of research is to use bootstrap based methods to find the covariance matrix of OLS estimator. Efron (1979) proposed this method the first time, called naïve bootstrap. One of the bootstrapped based estimators is the Jackknife (JA), (See Appendix A.1.2). Mackinnon & White (1985) showed that HC3 is a close approximation to Jackknife (JA). We are evaluating HC3 along with rival estimators (HC0, HC1 and HC2) in the current study, so we can safely say that we are considering some of the bootstrapped based estimators. The focus of our study is on developing analytical methods for evaluations of bias. Most bootstrap estimators are simulation based and hence cannot be evaluated analytically. Therefore they are excluded from this study.

2.4: COMPARISON & EVALUATION OF HCCMEs

Mackinnon and White (1985) compared the performance of HCCMEs (HC0, HC1, HC2 and HC3) using extensive Monte-Carlo simulations, and showed that, Eicker-White (HC0) estimator is downward biased in finite samples. The Monte-Carlo results favored HC3 on the basis of size distortions. Our analytics supports this conclusion, showing that HC0 can have very large biases. Later, Chesher & Jewitt (1987) developed a formula for the bias of the original Eicker-White HCCME (HC0) and suggested that it is always biased downward when regression design contains high leveraged observations. They also gave expressions for the lower and upper bounds for the proportionate bias of the HC0 estimator. Our formulae are for the simple bias and not the proportional ones. In addition, Chesher & Jewitt (1987) used the ratio of maximum and minimum variance to represent the degree of heteroskedasticity but in our case, the minimum variance is zero and maximum variance is bounded by putting an upper bound on variances (U), so in our case the ratio of maximum to minimum variance is infinity. Hence our bias formulae are not directly comparable with the one obtained by Chesher & Jewitt (1987). Further research is required to compare both results and we leave it open for future researchers.

Orhan (2000) in his PhD Thesis analytically calculated the bias of Eicker-White and Horn, Horn and Duncan estimator, when the regression model contains only one regressor and regressors are standardized to have mean zero and variance unity. He also compared the biases of different HCCMEs (OLS, White, Hinkley's estimator, Horn, Horn & Duncan estimator, Jackknife estimator, Maximum likelihood (ML) estimator, Bias Corrected estimator, Bootstrap estimator, Pre OLS and the James Stein estimator) using a number of different criteria (Chi-Square loss, Entropy loss ,Quadratic loss and the t-loss). He used three real world data sets to do the comparison. His results are conflicting and are data specific. His main finding is that ML should be preferred.

Biases of different estimators vary with the configuration of unknown heteroskedasticity. Our analytical formulae permit us to calculate the least favorable configurations of heteroskedasticity which generate the maximum bias. This allows to evaluate and rank estimators on the basis of their worst case biases. Our findings suggest that HC2 estimator proposed by Horn, Horn and Duncan has least maximum bias as compared to all other estimators, namely, HC0, HC1 and HC3. Actually the results provided by Orhan (2000) are data specific, and we know that if we change the design matrix or the skedastic function, the results get changed. The same thing happened in Orhan (2000). Our findings suggest that HC2 should be used if we are comparing HC0 to HC3. But since our results did not cover other estimator [Jackknife estimator, Maximum likelihood (ML) estimator, Bias Corrected estimator, Bootstrap estimator, and the James Stein estimator], so we cannot say about their performance compared to HC2.

Cribari-Neto and Zarkos (1999) using Monte-Carlo analysis judged the performance of HC0 to HC3 HCCMEs. Their results favored HC2 estimator when the evaluation criterion is bias. These findings are consistent with our study.

Scot Long & Ervin (2000) performed Monte-Carlo simulations by considering a number of design matrices and the error structures to compare various HCCMEs using size distortion as the deciding criteria. Their results favored HC3 against its rivals and they suggested that one should use HC3 when sample size is less than 250. Our findings suggest that the performance of HC3 is better than HC0 but its performance is very poor as compared to HC1 and HC2. Although Long & Ervin (2000) used a number of design matrices as well as the error structures, but they missed many other combinations of regressors and the skedastic sequences. So they arrived at the wrong conclusion due to simulation based results.

The finite sample behavior of three alternative estimators (HC1, HC2 and HC3) is found to be better than that of Eicker-White (HC0) estimator because these estimators already incorporate small sample corrections. Many simulation studies suggested that these estimators are better than HC0, e.g., Mackinnon and White (1985), Davidson and Mackinnon (1993), Cribari-Neto and Zarkos (1999).Our results indicate the same.

Since the above studies are based on simulations, and not the analytics, so they come up with different conclusions regarding the performance of HCCMEs, e.g. Mackinnon and White (1985) stressed to use HC3, Cribari-Neto and Zarkos (1999) favored HC2, Long and Ervin (2000) suggested HC3, Cribari-Neto (2004) advocated HC4 and Cribari-Neto et al. (2007) provided some evidence for HC5, all studies use size distortion as the deciding criteria. The different conclusions are due to the fact that the performance of HCCMEs depend on the structure of the design matrix as well as the skedastic function, and since simulations cannot take into account all the combinations of regressors and skedastic functions, so the question of comparing HCCMEs is not answerable using simulations and can be well captured with the help of analytical results, which we provide in this thesis. Using analytics, we gave exact expressions for the maximum positive and negative biases of all HCCMEs; this allows us to find the least favorable cases for each HCCME. Now if an HCCME is found to perform well in its worst performance region, then surely this will perform better in other areas. So the only way to compare HCCMEs is the analytics and not the simulations. That is why we provide analytical results for the comparison.

2.5: OBJECTIVES OF THE STUDY

In this section, we provide the main objectives of the study.

Our goal in this study is to analytically investigate the relative performance of four most popular HCCMEs (HC0, HC1, HC2 and HC3) which are mostly used by Software Packages.

a) To do the comparison, we developed, for the first time in literature, the analytical formulae for the biases of HCCMEs. In particular, we gave exact expressions for the maximum positive and negative biases of all HCCMEs

b) Using the analytical formulae developed, we used Minimax Criteria to evaluate the performance of HCCMEs. In particular, we identified the regions of parameter space which provide the ranges for the worst performance of each HCCME. This allows to evaluate and rank estimators on the basis of their worst case biases. If an HCCME is found to perform well in its worst performance region, then surely it will perform better in other areas.

c) This permits, for the first time, a sharp and unambiguous evaluation of the relative performance of a large class of widely used HCCMEs.

Our major goal is to provide practitioners and econometricians a clear cut way to be able to judge the situations where heteroskedasticity corrections can benefit us the most and also which method must be used to do such corrections.

Chapter 3: THE HETEROSKEDASTIC REGRESSION MODEL

3.1: Introduction

In this chapter we present our basic regression model and the related definitions. The bias of OLS estimate of the variance of error term is derived by re-parameterizing the linear regression model. In addition, the issue of consistency of OLS estimate of variances of error term has been explained explicitly using analytical formulae for the bias.

3.2: The Basic Regression Model (Single Regressor Case)

In this section, we set out the basic model and definitions required to state our results. We will consider a linear regression model with a single regressor, *x*, and t=1,2,...,T observations

EQ: 0-1
$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T)'$ be the T x 1 vector of errors. We assume that E(ε) = 0, but allow for heteroskedasticity by assuming that $Cov(\varepsilon) = \Sigma$, where Σ is a diagonal matrix: $\Sigma = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_T^2)$.Let $\beta = (\beta_1, \beta_2)'$ be the 2 x 1 vector of regression coefficients. As usual, we can define vector y and matrix of regressors, X, to write the model in matrix form:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix}, y = X\beta + \varepsilon$$

The OLS estimate of the coefficient β is:

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

The covariance matrix of OLS estimates of β is:

$$\Omega = (X'X)^{-1} X'\Sigma X (X'X)^{-1}$$

The main objective of our interest is the estimation of Ω_{22}, the variance of the OLS estimator of β_2. This will determine the significance of the regressor x.

3.3: Centering the Regressors

Let $\bar{x} = (1/T)\sum x_t$ be the average of the regressors, as usual. An important issue which has not received attention in the literature is that estimates of the variance Ω_{22} have different properties if the model is re-parameterized as follows:

EQ: 0-2 $$y_t = \beta_1' + \beta_2(x_t - \bar{x}) + \varepsilon_t, \qquad \text{where,} \qquad \beta_1' = \beta_1 + \beta_2\bar{x}$$

Heteroskedasticity corrections are different in model provided in ***EQ: 0-2***, where the regressors have been centered, from the original model. There are many reasons to prefer model in ***EQ: 0-2*** over model in ***EQ: 0-1*** when making heteroskedasticity corrections and this is the approach we will follow throughout this thesis. Recall that the overall F statistic for the regression evaluates the coefficients of the regressors for significance while removing the constant term from considerations. Exactly in the same way, it is preferable to assess the significance of the regressor, x_t, after removing the portion of it collinear with the constant term. The second reason for preferring model in ***EQ: 0-2*** is that the analysis substantially simpler and offers formulae which are much easier to interpret and understand. The third reason is that the formulae for the covariance matrix are substantially simpler, and correspond to simple intuition about the problem. This will be discussed in detail later, after we present a formula for Ω_{22}.

Lemma 1: The variance of the OLS estimate $\hat{\beta}_2$ with heteroskedasticity is:

EQ: 0-3
$$\Omega_{22} = \frac{\left\{\sum_{t=1}^{T}\sigma_t^2 x_t^2 - \bar{x}\sum_{t=1}^{T}\sigma_t^2 x_t\right\} - \bar{x}\left\{\sum_{t=1}^{T}\sigma_t^2 x_t - \bar{x}\sum_{t=1}^{T}\sigma_t^2\right\}}{\left\{\sum_{t=1}^{T}(x_t - \bar{x})^2\right\}^2}$$

Proof: Both the derivation and interpretation of this result becomes much simpler if we introduce an artificial ordered pair of random variables (Z,V) which takes one of the T possible values (x_t, σ_t^2) with equal probability (1/T) for each outcome. In terms of these random variables, the formula for Ω_{22} can be written in a much more revealing form:

EQ: 0-4
$$\Omega_{22} = \frac{\mathrm{Cov}(VZ, Z) - \mathrm{Cov}(V, Z)EZ}{T\mathrm{Var}(Z)^2}$$

To get to this result, we compute the matrices entering the formula for $\Omega = (X'X)^{-1} X'\Sigma X \left(X'X\right)^{-1}$ as follows.

EQ: 0-5
$$X'X = T\begin{pmatrix} 1 & EZ \\ EZ & EZ^2 \end{pmatrix}, (X'X)^{-1} = \frac{1}{T\mathrm{Var}(Z)}\begin{pmatrix} EZ^2 & -EZ \\ -EZ & 1 \end{pmatrix},$$
$$X'\Sigma X = T\begin{pmatrix} EV & EVZ \\ EVZ & EVZ^2 \end{pmatrix}$$

Multiplying through, rearranging terms, and applying the formula $Cov(X,Y) = EXY - (EX)(EY)$ leads to the following expressions for the entries of the matrix Ω:

EQ: 0-6

$$\Omega_{11} = \frac{1}{T\text{Var}(Z)^2}\left[EZ^2\left\{\text{Cov}(VZ,Z) - \text{Cov}(V,Z^2)\right\} - EZ\left\{\text{Cov}(VZ^2,Z) - \text{Cov}(VZ,Z^2)\right\}\right]$$

EQ: 0-7 $$\Omega_{12} = \frac{1}{T\text{Var}(Z)^2}\left[\text{Cov}(V,Z)EZ^2 - \text{Cov}(VZ,Z)EZ\right]$$

EQ: 0-8 $$\Omega_{22} = \frac{1}{T\text{Var}(Z)^2}\left[\text{Cov}(VZ,Z) - \text{Cov}(V,Z)EZ\right]$$

This proves the lemma.■

Let $W = (Z - EZ)/\sqrt{Var(Z)}$ be the standardization of the random variable Z. The following lemma shows how the formula for Ω_{22} simplifies in the model with centered regressors:

Lemma 2: If the regressors have mean 0, so that EZ=0, then

$$\Omega_{22} = \frac{Cov(VZ,Z)}{T\text{Var}(Z)^2} = \frac{(EVZ^2)}{T\text{Var}(Z)^2} = \frac{EVW^2}{T\text{Var}(Z)}$$

If variances, V, and the squared standardized regressors W^2 are uncorrelated, then Ω_{22}, the quantity we wish to estimate, is proportional to $EVW^2 = (EV)(EW^2) = EV$, or the average variance. This will be properly estimated by usual OLS based estimates

(HUCME) which ignore heteroskedasticity. On the other hand, in model given in ***EQ: 0-1***, this condition does not suffice. In model given in ***EQ: 0-1***, standard estimates are unbiased only if the sequence of variances is uncorrelated BOTH with 'X' and with X^2, (See **Theorem 0-2** below).

The more stringent condition is needed because X is correlated with the constant. This shows that conditions for consistency of the HCCME are simpler, easier to fulfill and make more intuitive sense in the model with the centered regressors. This gives us a third reason to prefer model with centered regressors for making heteroskedasticity corrections.

3.4: Order of Consistency

White (1980) motivates the introduction of his heteroskedasticity corrected covariance matrix estimates by stating that "It is well known that the presence of heteroskedasticity … leads to inconsistent covariance matrix estimates". This is true only after altering the covariance matrix being estimated by rescaling it to have a positive definite limit. It is worthwhile to spell out this technicality.

Note that model given in ***EQ: 0-1*** and ***EQ: 0-2*** coincide when the regressors, x_t have mean 0. We will henceforth work with model in ***EQ: 0-1*** under this assumption, which is equivalent to assuming that EZ=0.

With heteroskedasticity, the variance of the OLS estimate $\hat{\beta}_2$ with centered regressors is $\Omega_{22,T} = (EVW^2)/(T\mathrm{Var}(Z))$ from Lemma 2. Both V and W^2 are strictly positive sequences. Under reasonable assumptions on the sequence of regressors and variances (e.g. both are stationary, or ergodic) both EVW^2 and $\mathrm{Var}(Z)$ will have finite non-zero asymptotic values. Thus $\Omega_{22,T}$ will decline to zero. On first blush, a reasonable definition for consistency for a sequence of estimators $\hat{\Omega}_{22,T}$ would appear to be:

$$\operatorname*{plim}_{T\to\infty}\left(\hat{\Omega}_{22,T} - \Omega_{22,T}\right) = 0$$

Here plim is the probability limit, the standard weak convergence concept used for defining consistency. However, with this definition, the usual estimator of OLS covariance (see ***EQ: 0-11***) is consistent, even though it does not take heteroskedasticity into account. Both the estimator and the quantity being estimated converge to zero, and so the limiting difference is zero. This does not appear to be a satisfactory definition because any sequence converging to zero is consistent, even if it has nothing to do with the problem at hand. The following definition from Akahira and Takeuchi (1981) takes care of the problem.

Definition: A sequence of estimates of $\Omega_{22,T}$ is k-th order consistent if

EQ: 0-9
$$\operatorname*{plim}_{T\to\infty} T^k\left(\hat{\Omega}_{22,T} - \Omega_{22,T}\right) = 0$$

Then we can easily check that the usual HUCME for OLS is zero-order consistent but not first order consistent. Without explicit mention, the literature on the topic adopts first order consistency as the right definition of consistency. For example, Theorem 3 of White (1980) rescales the covariance matrix so that it is asymptotically positive definite, so as to show inconsistency of the usual estimates. With this refined notion of consistency, it is possible to characterize conditions for consistency of the HUCME in standard regression models as follows.

Theorem 0-1: In the model of ***EQ: 0-2***, after centering the regressor, the HUCME based variance estimate $\hat{\Omega}_{22}^{\text{OLS}}$ of Ω_{22} which ignores heteroskedasticity is k-th order consistent if and only if

EQ: 0-10
$$\lim_{T\to\infty} T^{k-1} \frac{\sqrt{\text{Var}(V)\left(EW^4-1\right)}}{\text{Var}(Z)} \text{Corr}\left(V,W^2\right)=0$$

Proof: Let $\hat{\beta}=(X'X)^{-1}X'y$ be the OLS estimates and $e=y-X\hat{\beta}$ be the OLS residuals. Then the standard HUCME of the OLS estimates is: $\Omega^{OLS}{}'=\sigma^2(X'X)^{-1}$, where $\sigma^2=e'e/(T-2)$ With centered regressors, the (2,2) entry of the $(X'X)^{-1}$ matrix is $\left[T\,\text{Var}(Z)\right]^{-1}$. It follows that the (2,2) entry of the HUCME is:

EQ: 0-11 $\overset{OLS}{\Omega_{22}} = e'e / \left[T(T-2)\text{Var}(Z) \right]$

So, the bias of HUCME for the variance of $\hat{\beta}_2$ is:

EQ: 0-12 $B_{22}^{OLS} = E\left(\overset{OLS}{\Omega_{22}}\right) - \Omega_{22} = \frac{1}{\text{TVar}(Z)}\left(\frac{1}{T-2}Ee'e - EVW^2\right)$

Note that $EVW^2 = \text{Cov}(V, W^2) + (EV)(EW^2) = \text{Cov}(V, W^2) + EV$.

Substituting into the previous expression yields:

EQ: 0-13 $B_{22}^{OLS} = \frac{1}{\text{TVar}(Z)}\left(\frac{1}{T-2}Ee'e - EV\right) - \frac{\text{Cov}(V, W^2)}{T\text{Var}(Z)}$

To evaluate the bias, we need to calculate $Ee'e$, which is done below:

Lemma 3: The expected value of the sum of squared residuals is:

$$Ee'e = \sum_{t=1}^{T} \sigma_t^2 - \frac{1}{T}\sum_{t=1}^{T}\left(1 + w_t^2\right)\sigma_t^2 = (T-1)EV - EVW^2$$

Proof: Let $H = X(X'X)^{-1}X'$. It follows that

$$E(e'e) = E\left\{\text{tr}\left(\varepsilon'(I-H)\varepsilon\right)\right\} = \text{tr}E\left((I-H)\varepsilon\varepsilon'\right) = \text{tr}(I-H)\Sigma$$

Substituting the values $h_{tt} = \frac{1}{T\text{Var}(Z)}\left(EZ^2 + x_t^2\right) = \frac{1}{T}\left(1 + \frac{1}{\text{Var}(Z)}x_t^2\right) = \frac{1}{T}\left(1 + w_t^2\right)$

and $\Sigma = diag\left(\sigma_t^2\right), i = 1, 2, \ldots, T$ leads to the lemma. ■

It follows that

$$\left(Ee'e/[T-2]\right)-EV=\left(EV-EVW^{2}\right)/(T-2)=-\text{Cov}(V,W^{2})/(T-2)$$

Substituting into ***EQ: 0-13***, we get:

EQ: 0-14 $$B_{22}^{OLS}=-\frac{\text{Cov}\left(V,W^{2}\right)}{\text{TVar}(Z)}\left(\frac{1}{T-2}+1\right)$$

Note that, $$\text{Corr}\left(V,W^{2}\right)=\frac{\text{Cov}\left(V,W^{2}\right)}{\sqrt{\text{Var}\left(V\right)\text{Var}\left(W^{2}\right)}}$$

Also, $\text{Var}\left(W^{2}\right)=EW^{4}-1,\qquad \left(\therefore EW^{2}=1\right)$

So, we have;

EQ: 0-15 $$B_{22}^{OLS}=-\frac{\sqrt{\text{Var}\left(V\right)\left(EW^{4}-1\right)}\text{Corr}\left(V,W^{2}\right)}{\text{TVar}(Z)}\left(\frac{1}{T-2}+1\right)$$

We can write it as,

EQ: 0-16 $$T\,B_{22}^{OLS}=-\frac{\sqrt{\text{Var}\left(V\right)\left(EW^{4}-1\right)}\text{Corr}\left(V,W^{2}\right)}{\text{Var}(Z)}\left(\frac{1}{T-2}+1\right)$$

Taking limit as 'T' approaches infinity on both sides of **EQ: 0-16**, leads to required result.■

Remark 1: When V and X^2 are not correlated, then V and W^2 are also not correlated. It follows that $Cov(V,W^2) = EVW^2 - (EV)(EW^2) = 0$. In this case, from **Lemma 2** above, we see that

$$\Omega_{22} = \frac{EVW^2}{T\text{Var}(Z)} = \frac{(EV)(EW^2)}{\text{TVar}(Z)} = \frac{EV}{\text{TVar}(Z)} \quad \textit{EQ: 0-17}$$

This is exactly the expression for the variance as occurs in the case of homoskedasticity when each σ_t^2 is replaced by the average value EV of all the variances. This means that when V and W^2 are uncorrelated, this model is equivalent to a homoskedastic model for the purpose of estimating variance of β_2. This is why the usual variance estimate which ignores heteroskedasticity succeeds under this condition.

Remark 2: The leading case is where both the heteroskedastic sequence of variances and the sequence of regressors is stationary. In this case, a necessary and sufficient condition for first order consistency of the HUCME is that the correlation between the variances and the squared regressors is asymptotically zero. Higher order consistency requires this correlation to go to zero at a suitably fast rate. However, if the regressors are non-stationary and/or have a deterministic trend, Var(Z) can go to infinity and result in consistency of the HUCME even when variances are correlated with squared regressors. This consistency can be offset if Var(V) (which is a measure of heteroskedasticity) increases to infinity, and/or EW^4 (which measures the Kurtosis of the regressors) increases to infinity. If the product of these two factors also goes to infinity sufficiently fast, HUCME will again be inconsistent.

Remark 3: A more complex condition for higher order consistency of OLS obtains in the original model, without centering the regressors. Essentially, this requires correlation between the heteroskedastic variance sequence and both 'X' and X^2 to go to zero. The required condition is provided in the following theorem:

Theorem 0-2: Higher order consistency of OLS in original model (***EQ: 0-1***), without centered regressors, is given by:

EQ: 0-18

$$\lim_{T\to\infty} T^{k-1} \frac{2(EZ)\sqrt{\mathrm{Var}(Z)Var(V)}\,\mathrm{Corr}(Z,V)-\sqrt{\mathrm{Var}(Z^2)Var(V)}\,\mathrm{Corr}(Z^2,V)}{\{\mathrm{Var}(Z)\}^2}=0$$

Proof: A direct and intuitive way to prove the theorem is to replace w in **EQ: 0-14** by:

$$W=(Z-EZ)/\sqrt{Var(Z)}$$

Note that, **EQ: 0-14** can be written as,

$$B_{22}^{OLS} \doteq \frac{\left((EW^2)(EV)-EW^2V\right)}{\mathrm{TVar}(Z)}\left(\frac{1}{T-2}+1\right)$$

Replacing the value of W, leads to:

$$B_{22}^{OLS} = \frac{2(EZ)\{EZV-(EZ)(EV)\}-\{EZ^2V-(EZ^2)(EV)\}}{\mathrm{T}\{\mathrm{Var}(Z)\}^2}\left(\frac{1}{T-2}+1\right)$$

Writing in terms of covariance form, we get:

$$B_{22}^{OLS} \doteq \frac{2(EZ)Cov(Z,V)-Cov(Z^2,V)}{\mathrm{T}\{\mathrm{Var}(Z)\}^2}\left(\frac{1}{T-2}+1\right)$$

Now converting covariances into correlations, we have:

EQ: 0-19

$$B_{22}^{OLS} = \frac{2(EZ)\sqrt{Var(Z)Var(V)}Corr(Z,V) - \sqrt{Var(Z^2)Var(V)}Corr(Z^2,V)}{\mathrm{T}\{\mathrm{Var}(Z)\}^2}\left(\frac{1}{T-2}+1\right)$$

We can write it as:

EQ: 0-20

$$T\,B_{22}^{OLS} = \frac{2(EZ)\sqrt{Var(Z)Var(V)}Corr(Z,V) - \sqrt{Var(Z^2)Var(V)}Corr(Z^2,V)}{\{\mathrm{Var}(Z)\}^2}\left(\frac{1}{T-2}+1\right)$$

Taking limit as 'T' approaches infinity on both sides leads to required result.■

Chapter 4: A MINIMAX ESTIMATOR

4.1: Introduction

EW (HC0) estimator and Hinkley (HC1) estimator pre-multiplies OLS squared residuals by '1' and 'T/(T-2) respectively. In order to evaluate these, it is convenient to introduce a class of estimators which multiplies the squared residuals by some constant. This class includes both HC0 and HC1. We show that the maximum bias of this class of estimators can be evaluated analytically. This permits us to find a best estimator within this class. The Minimax estimator is the one which minimizes the maximum bias. We compute this estimator and show that it has substantially smaller bias compared to both HC0 and HC1.

4.2: Bias of EW-type Estimates

We will now derive analytical expressions for the bias of a class of estimators which includes the Eicker-White, as well as the Hinkley bias-corrected version of the HCCME. Consider estimators of true covariance matrix having the form:

EQ: 4-1 $$\hat{\Omega}(\alpha) = (X'X)^{-1} X'\left(\alpha\hat{\Sigma}\right)X(X'X)^{-1}$$

Where, α is any positive scalar, and $\hat{\Sigma} = \text{diag}\left(e_1^2, \cdots, e_T^2\right)$ with e_t^2 is the square of the t-th OLS residual. Note that if $\alpha = 1$, we have EW estimator of true covariance matrix and if $\alpha = \frac{T}{T-2}$, we have Hinkley's (1977) estimator of the same. In this section, we provide analytical expressions for the bias of $\hat{\Omega}_{22}(\alpha)$, the variance of $\hat{\beta}_2$ under heteroskedasticity.

As before, it is convenient to work with the artificial random variable (V, Z) which takes each of the 'T' possible values (σ_t^2, x_t) for $t = 1, 2, \cdots, T$ with equal probability $1/T$. We assume that the regressors have been centered, so that EZ=0 and EZ^2=Var(Z). Standardize 'Z' by introducing $W = Z / \sqrt{Var(Z)}$, and note that EW=0 and $Var(W) = EW^2 = 1$. According to **Lemma 3**, the true variance of the OLS estimate $\hat{\beta}_2$ of the slope parameter β_2 is given by:

EQ: 4-2 $$\Omega_{22} = \frac{EW^2V}{T\left(\text{Var}(Z)\right)}$$

The HCCME of the variance of slope parameter is:

EQ: 4-3 $$\hat{\Omega}_{22}(\alpha) = \frac{\alpha}{T^2\text{Var}(Z)}\sum_{t=1}^{T} w_t^2 e_t^2$$

The following theorem gives the bias of this HCCME.

Theorem 4-1: The bias $\text{B}_{22}(\alpha) = E\hat{\Omega}_{22}(\alpha) - \Omega_{22}$ of the HCCME for the variance of slope parameter is:

EQ: 4-4

$$B_{22} = \frac{1}{T^3\text{Var}(Z)}\sum_{t=1}^{T}\left[\alpha + \left(2\alpha EW^3\right)w_t + \left\{T(\alpha-1) + \alpha\left(EW^4 - 2\right)\right\}w_t^2 - 2\alpha w_t^4\right]\sigma_t^2$$

Proof: From the expressions for $\hat{\Omega}_{22}$ and Ω_{22} given earlier [See ***EQ: 4-2*** and ***EQ: 4-3***], we get,

EQ: 4-5 $$B_{22} = E\hat{\Omega}_{22} - \Omega_{22} = \frac{1}{T^2 Var(Z)}\sum_{t=1}^{T} w_t^2\left[\alpha E\left(e_t^2\right) - \sigma_t^2\right]$$

Before proceeding, we need $E\left(e_t^2\right)$, which is given by following lemma:

Lemma 4: With centered regressors, the expected value of OLS squared residuals is given by:

EQ: 4-6 $$E\left(e_t^2\right)=\sigma_t^2+\frac{1}{T}\left[EV-2\sigma_t^2-2w_t^2\sigma_t^2+2w_tEWV+w_t^2EW^2V\right]$$

Proof: The OLS residuals are $e=y-X\hat{\beta}=(I-H)\varepsilon$, where, $H=X\left(X'X\right)^{-1}X'$ is the 'hat matrix' as before. Using the standardized regressors $w_t=x_t/\text{Var}(Z)$, we can calculate the (i,j) entry of H to be:

$$H_{ij}=\frac{1}{T\text{Var}(Z)}\left(EZ^2+x_ix_j\right)=\frac{1}{T}\left(1+\frac{1}{\text{Var}(Z)}x_ix_j\right)=\frac{1}{T}\left(1+w_iw_j\right)$$

Now note that $e_t=\varepsilon_t-\sum_{j=1}^{T}\text{h}_{tj}\varepsilon_j=\left(1-h_{tt}\right)\varepsilon_t-\sum_{\substack{j=1\\ j\neq t}}^{T}\text{h}_{tj}\varepsilon_j$.

Since E e=0, and the ε's are independent, the variance of 'e' is the sum of the variances. This can be explicitly calculated as follows:

$$\begin{aligned}E\left(e_t^2\right)&=(1-h_{tt})^2\sigma_t^2+\sum_{j=1,j\neq t}^{T}h_{tj}^2\sigma_j^2=(1-2h_{tt})\sigma_t^2+\sum_{j=1}^{T}h_{tj}^2\sigma_j^2\\&=\left(1-\frac{2}{T}\left(1+w_t^2\right)\right)\sigma_t^2+\sum_{j=1}^{T}\left(\frac{1}{T}\left(1+w_tw_j\right)\right)^2\sigma_j^2\end{aligned}$$

From this it follows that:

$$E\left(e_t^2\right)=\sigma_t^2-\frac{2}{T}\sigma_t^2-\frac{2}{T}w_t^2\sigma_t^2+\frac{1}{T}EV+\frac{1}{T}w_t^2EW^2V+\frac{2}{T}w_t\left(EWV\right)$$

This is easily translated into the expression given in the Lemma.■

Substituting the expression of the

Lemma 4 in ***EQ: 4-5*** above, and noting that $EW^{j}V = (1/T)\sum w_t^j \sigma_t^2$ leads to the expression given in Theorem.■

4.3: Maximum Bias

Having analytical expressions for the bias allow us to calculate the configuration of variances which leads to the maximum bias. In this section we characterize this least favorable form of heteroskedasticity, and the associated maximum bias. We first re-write the expression for bias in a form that permits easy calculations of the required maxima.

Define polynomial p(λ) as

EQ: 4-7 $$p(\lambda) = \alpha + \left(2\alpha EW^3\right)\lambda + \left(T(\alpha-1) + \alpha\left(EW^4 - 2\right)\right)\lambda^2 - 2\alpha\lambda^4$$

From the expression for bias given in Theorem of the previous section, we find that

EQ: 4-8 $$B_{22} = \frac{1}{T^3 \mathrm{Var}(Z)} \sum_{t=1}^{T} p(w_t)\sigma_t^2$$

If the variances are unconstrained, then the bias can be arbitrarily large, so we assume some upper bound on the variances: $\forall t : \sigma_t^2 \leq U$. Under this assumption, we proceed to derive the largest possible second order bias for the class of EW-type estimators under consideration. Since the expression is not symmetric, and the maximum positive bias may differ from the maximum negative bias, we give expressions for both in our preliminary result below.

Theorem 4-2: Let B^+ and B^- be the maximum possible positive and negative biases of the EW-type estimators $\hat{\Omega}_{22}(\alpha)$ defined in ***EQ: 4-1*** above. These are given by:

EQ: 4-9 $$B^+ = \max_{\sigma_t^2 \leq U} B_{22} = \frac{1}{T^3 \mathrm{Var}(Z)} \sum_{t=1}^{T} \max\left(p(w_t), 0\right) U$$

EQ: 4-10 $$B^- = \min_{\sigma_t^2 \leq U} B_{22} = \frac{1}{T^3 \mathrm{Var}(Z)} \sum_{t=1}^{T} \min\left(p(w_t), 0\right) U$$

Proof: Note that, maximum positive and negative biases can be found by maximizing and minimizing the same with respect to variances, i.e.,

$$B^+ = \max_{\sigma_t^2} B_{22} = \frac{1}{T^3 Var(Z)} \max_{\sigma_t^2} \sum_{t=1}^{T} \left(p(w_t)\sigma_t^2\right)$$

$$B^- = \min_{\sigma_t^2} B_{22} = \frac{1}{T^3 Var(Z)} \min_{\sigma_t^2} \sum_{t=1}^{T} \left(p(w_t)\sigma_t^2\right)$$

Where, $p(w_t) = \alpha + \left(2\alpha EW^3\right)w_t + \left(T\left(\alpha - 1\right) + \alpha\left(EW^4 - 2\right)\right)w_t^2 - 2\alpha w_t^4$

In order to maximize $\sum_{t=1}^{T}\left(p(w_t)\sigma_t^2\right)$ with respect to variances (σ_t^2)

i.e, $\max_{\sigma_t^2}\sum_{t=1}^{T}\left(p(w_t)\sigma_t^2\right)$.

Note that, we have to maximize a sum of linear functions. Each term in the sum can be maximized separately with respect to variances (σ_t^2). i.e. $\max_{\sigma_t^2}\left(p(w_t)\sigma_t^2\right)$.

Since, $\forall t : \sigma_t^2 \le U$, so to maximize a linear function, we have to set variances (σ_t^2) to its maximum possible value 'U' when the coefficient is positive, and its minimum possible value '0' when the coefficient is negative. Since, we have sum of such terms, so maximizing each term separately leads to:

$$\mathcal{B}^+ = \max_{\sigma_t^2 \le U} B_{22} = \frac{1}{T^3 \mathrm{Var}(Z)} \sum_{t=1}^{T} \max\left(p(w_t), 0\right) U$$

This is the required result.

A similar analysis can be done to get minimum bias, which is also the maximum negative bias. We replace σ_t^2 by the maximizing value 'U', if the coefficient is negative and by minimizing value '0' when the coefficient is positive. This leads to the following equation for maximum negative bias:

$$\mathcal{B}^- = \min_{\sigma_t^2 \le U} B_{22} = \frac{1}{T^3 \mathrm{Var}(Z)} \sum_{t=1}^{T} \min\left(p(w_t), 0\right) U$$

This proves the theorem. ∎

We now try to obtain more explicit characterizations of these maxima and minima. It turns out that the case where the regressors are normally distributed offers significant simplifications in analytic expressions, so we first consider this case.

4.3.1: Maximum Bias with Normal Regressors

Under the assumption that the regressors x are i.i.d. normal, we derive analytical formulae for the approximate large sample maximum bias $\mathcal{B} = \max\left(\mathcal{B}^{+}, -\mathcal{B}^{-}\right)$. In large samples, the skewness EW^3 should be approximately zero, while the kurtosis, EW^4 should be approximately 3.

Making these asymptotic approximations, the polynomial p(λ) simplifies to:

EQ: 4-11 $$p(\lambda) = \alpha + \left(T\left(\alpha - 1\right) + \alpha\right)\lambda^2 - 2\alpha\lambda^4$$

In large samples, reasonable HCCME's will have $\alpha \approx 1$, so it is convenient to re-parameterize by setting $\alpha = 1 + a/T$, where 'a' is a positive constant. Evaluation of the expressions for bias requires separating values of w for which p(w)>0 from those for which p(w)<0. This is easily done since p(w) is a quadratic in w^2.

Lemma 5: p(w)>0 if and only if $-\sqrt{r} < w < +\sqrt{r}$, where r is the unique positive root of the quadratic $p(w^2)$. In large samples, this root is:

EQ: 4-12

$$r = \frac{1 + a + \sqrt{8 + (1 + a)^2}}{4}$$

Proof: From ***EQ: 4-11***, the quadratic equation in w^2 is given by:

$$p(w) = \alpha + \left(T\left(\alpha - 1\right) + \alpha\right)w^2 - 2\alpha w^4$$

Where, $\alpha = 1 + a/T$

So, the quadratic can be written as:

$$\left(1 + \frac{a}{T}\right) + \left(1 + a + \frac{a}{T}\right)w^2 - 2\left(1 + \frac{a}{T}\right)w^4 = 0$$

Since this is quadratic in w^2, so its positive root (since r<0 is not possible) can be written as:

$$r = \frac{\left(1 + a + \frac{a}{T}\right) + \sqrt{\left(1 + a + \frac{a}{T}\right)^2 + 8\left(1 + \frac{a}{T}\right)^2}}{4\left(1 + \frac{a}{T}\right)}$$

Taking limit as 'T' approaches infinity, leads to required result.■

This permits a more explicit characterization of the minimum and maximum biases derived earlier. In this framework, a=0 corresponds to the Eicker-White estimator, while the Hinkley bias correction amounts to setting a=2 (Since, $\alpha = T/(T-2) \approx 1 + 2/T$).

In order to calculate the bias functions explicitly, we need to specify the sequence of regressors. We first consider the case of normal regressors, which permits certain simplifications. Other cases are considered later.

The following Theorem gives the relationship between the maximum bias and the parameter 'a'.

Theorem 4-3: Suppose the regressor sequence is i.i.d. Normal. Let ϕ and Φ be the density and cumulative distribution function of a standard normal random variable. Then the maximum positive and negative biases of the estimator $\hat{\Omega}_{22}(\alpha)$ in large samples can be written as:

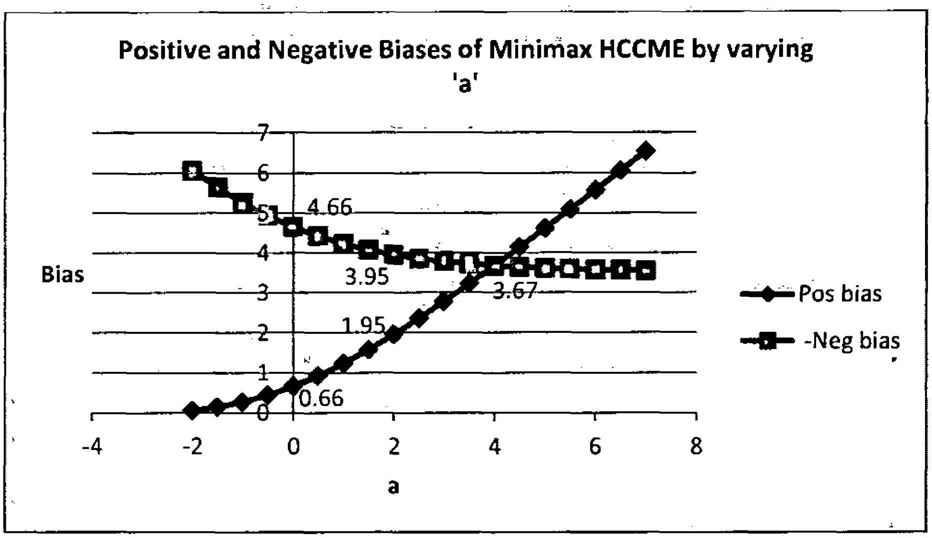
EQ: 4-13 $$\mathcal{B}^{+}(\mathrm{a}) = 2\{2r - \mathrm{a} + 5\}\sqrt{r}\,\phi\left(\sqrt{r}\right) + 2(\mathrm{a}-4)\Phi\left(\sqrt{r}\right) - \mathrm{a} + 4$$

EQ: 4-14 $$-\mathcal{B}^{-}(\mathrm{a}) = 2\{2r - \mathrm{a} + 5\}\sqrt{r}\,\phi\left(\sqrt{r}\right) + 2(\mathrm{a}-4)\Phi\left(\sqrt{r}\right) - 2\mathrm{a} + 8$$

Remark 1: The maximum bias functions are plotted in **Figure 4-1** below. Recall that the maximum bias is obtained by setting heteroskedasticity to the worst possible configuration, which makes the bias as large as possible. Maximal positive and negative biases require different configurations of heteroskedasticity, which is why they are plotted separately. Overall maximum bias is the maximum of these two functions. The point of intersection of these two curves is the place where this maximum bias is the lowest possible – the minimax bias. As shown in **Figure 4-1** below, the two maximum bias functions intersect at a=4, as can easily be verified analytically from the formulae above (See **EQ: 4-15** below). Note that this minimax bias estimator with a=4 improves substantially on the Eicker-White estimator with a=0. Figure also gives the positive and negative biases of HC0, HC1 and Minimax estimator in numerical form. Note that at 'a=0', the positive bias of HC0 is 0.66 while negative bias of the same is 4.66 in absolute form. Similarly, the positive and absolute negative biases of HC1 are 1.95 and 3.95 respectively at 'a=2' while Minimax estimator has same value (3.67) of positive and negative biases which occurs at the intersection 'a=4'.

Figure 4-1: Positive and Negative Biases of Minimax HCCME

(Normal Distribution Case)



Note: Pos and Neg denote positive and negative biases respectively.

Proof: Let $I(W > r)$ be the indicator function taking values '1' and '0' according to whether or not the indicated inequality holds. The assumption of normality of regressors means that W can be treated as a random variable with a standard normal distribution. We have the following large sample approximations for the terms in the polynomial

$$p(w_t) = \alpha + bw_t^2 - 2\alpha w_t^4$$

$$\alpha = 1, \quad b = 1 + \text{a}$$

Maximum positive bias as a function of 'a' is given by:

$$\text{B}^+(\text{a}) \approx P\left(W^2 < r\right) + (1 + \text{a})EW^2 \text{ I}\left\{W^2 < r\right\} - 2EW^4 \text{ I}\left\{W^2 < r\right\}$$

Where;

$$P\left(W^2 < r\right) = P\left(|W| < \sqrt{r}\right) = \Phi\left(\sqrt{r}\right) - \Phi\left(-\sqrt{r}\right)$$

$$EW^2 \text{ I}\left\{W^2 < r\right\} = -2\sqrt{r}\phi\left(\sqrt{r}\right) + \Phi\left(\sqrt{r}\right) - \Phi\left(-\sqrt{r}\right)$$

$$EW^4 \text{ I}\left\{W^2 < r\right\} = -2r^{3/2}\phi\left(\sqrt{r}\right) - 6\sqrt{r}\,\phi\left(\sqrt{r}\right) + 3\left(\Phi\left(\sqrt{r}\right) - \Phi\left(-\sqrt{r}\right)\right)$$

These expressions are obtained by evaluating the integrals of the normal density via integration by parts.

Putting the values of $P\left(W^2 < r\right), EW^2 \text{ I}\left\{W^2 < r\right\}, EW^4 \text{ I}\left\{W^2 < r\right\}$ and making use of the fact that $\Phi\left(-\sqrt{r}\right) = 1 - \Phi\left(\sqrt{r}\right)$, we get:

$$\mathcal{B}^{+}(\mathrm{a}) = 2\{2r - \mathrm{a} + 5\}\sqrt{r}\,\phi(\sqrt{r}) + 2(\mathrm{a}-4)\Phi(\sqrt{r}) - \mathrm{a} + 4$$

Similarly, maximum negative bias as a function of 'a' is:

$$\mathrm{B}^{-}(\mathrm{a}) \approx P(W^2 > r) + (1+a)EW^2\ \mathrm{I}\{W^2 > r\} - 2EW^4\ \mathrm{I}\{W^2 > r\}$$

Where,

$$P(W^2 > r) = P(|W| > \sqrt{r}) = P(W < -\sqrt{r}) + P(W > \sqrt{r}) = 1 + \Phi(-\sqrt{r}) - \Phi(\sqrt{r})$$

$$EW^2\ \mathrm{I}\{W^2 > r\} = 2\sqrt{r}\phi(\sqrt{r}) + 2\left[1 - \Phi(\sqrt{r})\right]$$

$$EW^4\ \mathrm{I}\{W^2 > r\} = \left[2r^{3/2}\phi(\sqrt{r}) + 6\sqrt{r}\phi(\sqrt{r}) + 6\left(1 - \Phi(\sqrt{r})\right)\right]$$

Again, putting the values of $P(W^2 > r), EW^2\ \mathrm{I}\{W^2 > r\}, EW^4\ \mathrm{I}\{W^2 > r\}$ and making use of the fact $\Phi(-\sqrt{r}) = 1 - \Phi(\sqrt{r})$, we get:

$$-\mathcal{B}^{-}(\mathrm{a}) = 2\{2r - \mathrm{a} + 5\}\sqrt{r}\,\phi(\sqrt{r}) + 2(\mathrm{a}-4)\Phi(\sqrt{r}) - 2\mathrm{a} + 8$$

This proves the theorem. ■

Now solving, $\mathcal{B}^{+}(\mathrm{a}) = -\mathcal{B}^{-}(\mathrm{a})$, we get:

EQ: 4-15 $\mathrm{a} = 4 = 3 + 1 = kurtosis + 1$

As we will see, the optimal value of 'a' depends on the kurtosis of the regressors, so the above decomposition clarifies the relation between the kurtosis (which is 3 for standard normal) and the minimax value of 'a'. This also confirms the discussion provided in **Remark 1** in **Theorem 4-3.** ■

In next section, we provide a new minimax bias estimator which minimizes the maximum bias.

4.4: The Minimax Bias Estimator

Given any particular sequence of regressors (x_t), our formulae above permit calculation of an optimal value of 'a' – the one for which the maximum bias is the lowest possible. This may be called the minimax value of 'a'. The bias functions themselves depend on the skewness, kurtosis, as well as other characteristics of the sequence of regressors, as indicative asymptotic calculations for the normal regressor case in the previous section show (See **Theorem 4-3** above).

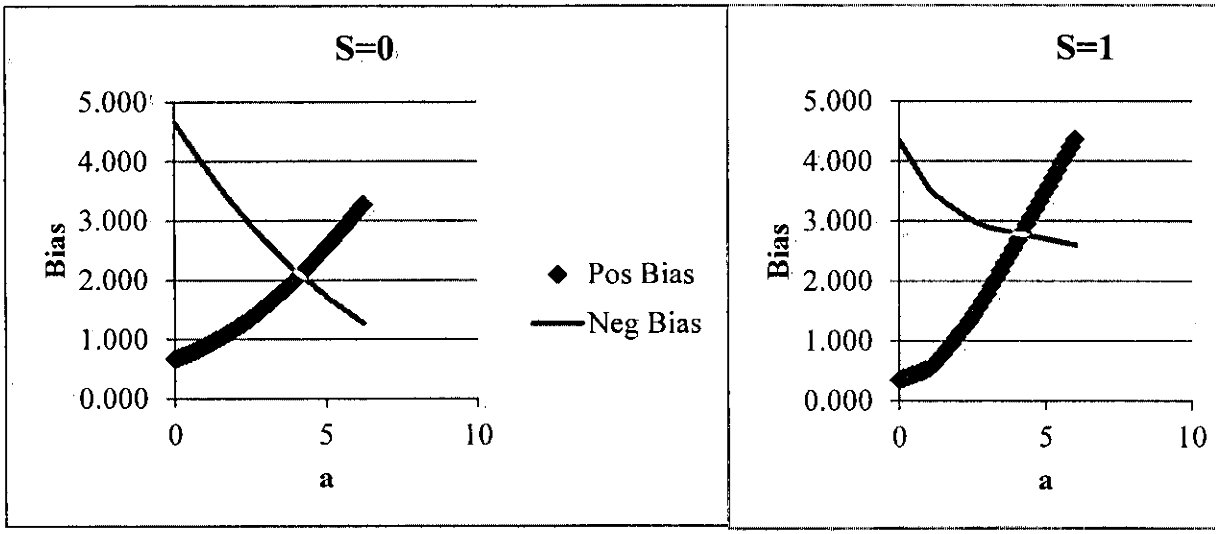
To check this for other regressor sequences except normal ones, we generated several sequence of regressors for a fixed sample size 'T' and a fixed value of kurtosis 'K' but by varying skewness and calculated maximum and minimum biases. The object of this exercise was to evaluate the dependency of the minimax value of 'a' upon the regressor sequence. To our surprise the value of minimax 'a' came out the same regardless of any value of skewness but it is found to be dependent only on 'T' and kurtosis (K) of the regressors. We now provide some details of these calculations.

We generated several sequence of regressors with matching kurtosis, i.e., first we generated one random sequence with kurtosis equal to 2 and calculated maximum positive and negative bias functions for this sequence of regressors. The minimax value of 'a' was calculated by setting these to be equal. Then we changed the sequence of regressors in such a way that the new sequence has exactly the same kurtosis, i.e., 2 and we again calculated maximum positive and negative bias functions, and the minimax value of 'a'. To our surprise, the value of 'a' came out exactly the same as was calculated from the first sequence with matching kurtosis. To further confirm, we generated several sequences with same kurtosis, and found the same minimax value of 'a'. Along the same lines, we generated other sequences with kurtosis equal to 3, 4, ... etc. and found that the minimax value of 'a' depends only on kurtosis and sample size. Also sequences of regressors with varying degrees of skewness but with matching kurtosis yield the same results; although the bias functions were different, the minimax value of 'a' remained the same.

To save space, we are only providing here results of two samples of sizes 100 and kurtosis equal to 3. But first sample has skewness measure equal to 0 while the second sample has skewness equal to '1'. For each set up, we calculated the maximum positive and negative biases of Minimax HCCME by varying 'a'. Our results indicate that the minimax value of 'a' is same (a=4.166) for both samples. The details are provided in **Figure 4-2** below.

The same value of 'a' emerged with samples having same kurtosis and sample sizes. This provides heuristic support for our invariance conjecture, formally stated in **Section 4.4.1**: below, that minimax value of 'a' depends only on the sample size and kurtosis of the regressors and actual sequence of regressors does not matter.

Figure 4-2: Positive and Negative Biases of Minimax HCCME by Varying 'a' with SS =100 an



Note: K and S denote Kurtosis and Skewness measures respectively, whereas, Pos and Neg denote pos respectively and SS represents Sample Size.

Due to the complexity of the relation between regressors and the minimax value of 'a', we were unable to establish this result analytically, and hence we leave it as a conjecture stated below.

4.4.1: An Invariance Conjecture

The object of this section is to state a conjecture about the minimax value of 'a' in the class of estimators defined in ***EQ: 4-4*** above. We recapitulate the basic definitions to make this section self-contained.

Consider a regression model $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$, with $E\varepsilon_t = 0$, and $\mathrm{Var}(\varepsilon_t) = \sigma_t^2$. Introduce an artificial random variable (V,Z) which takes one of T possible values (σ_t^2, x_t) for $t = 1, 2, \cdots, T$ with equal probabilities $1/T$. Define standardized regressors $w_t = (x_t - \bar{x}) / \sqrt{\mathrm{Var}(Z)}$, where, $\bar{x} = (1/T)\sum x_t$, and $\mathrm{Var}(Z) = (1/T)\sum (x_t - \bar{x})^2$ as usual. Let $K = \frac{1}{T}\sum_{t=1}^{T} w_t^4$ be the kurtosis of the standardized sequence of regressors.

Then the variance of the OLS estimate of β_2 is Ω_{22} explicitly given in ***EQ: 0-11*** earlier.

The class of estimators of this variance under consideration is defined as:

$$\hat{\Omega}_{22}(\alpha) = \frac{\alpha}{T^2 \mathrm{Var}(Z)} \sum_{t=1}^{T} w_t^2 e_t^2$$

Chapter 5: BIASES OF HCCMEs

5.1: Introduction

In order to compare the performances of the other two HCCMEs namely, Horn, Horn and Duncan (HC2) estimator and Mackinnon and White (HC3) estimator, with that of Eicker-White (HC0), Hinkley (HC1) and Minimax estimator, we will now calculate the maximum biases of HC2 and HC3. For our basic model and notations, refer to **Section 3.2**: earlier.

The bias of any estimator Ω of Ω is given by:

EQ: 5-1

$$Bias\left(\Omega\right)=E\left(\Omega\right)-\Omega$$

The following sections are devoted to get the biases of each of the above mentioned HCCMEs including OLS estimator.

Before deriving expressions for the bias of each HCCME, we introduce a general estimator which takes each of the existing HCCMEs as a special case of it. The detail of this general estimator is given in subsequent sections.

5.2: Bias of General Estimator

Consider estimators of true covariance matrix having the form:

EQ: 5-2 $\hat{\Omega}(A_i) = (X'X)^{-1}X'\left(\hat{\Sigma}_i\right)X(X'X)^{-1}, \qquad \hat{\Sigma}_i = A_i\hat{\Sigma}, \qquad i = 0,1,2,3.$

Where, $\hat{\Sigma} = \text{diag}\left(e_1^2, \cdots, e_T^2\right)$ with e_t^2 is the square of the t-th OLS residual.

Note that,

$A_0 = I$ gives White's (HC0) estimator

$A_1 = \left(\dfrac{T}{T-2}\right)I$ gives Hinkley's (HC1) estimator

$A_2 = diag\left\{\dfrac{1}{1-h_{tt}}\right\}$ gives Horn, Horn & Duncan's (HC2) estimator

$A_3 = diag\left\{\dfrac{1}{(1-h_{tt})^2}\right\}$ gives Mackinnon & White's (HC3) estimator

Also note that, $A_{0t} = 1, \forall t$, $A_{1t} = \frac{T}{T-2}, \forall t$ and let $A_{2t} = \frac{1}{1-h_{tt}}$ and $A_{3t} = \frac{1}{(1-h_{tt})^2}$ are the (t,t) diagonal entries of A_2 and A_3 respectively and h_{tt} is the t^{th} entry of Hat matrix, $H = X(X'X)^{-1}X'$

Note that the above form of HCCMEs has been taken from Cribari-Neto and Galvao (2003).

In this section, we provide analytical expressions for the bias of $\hat{\Omega}_{22}(A_i)$, the variance of $\hat{\beta}_2$ under heteroskedasticity.

As before, it is convenient to work with the artificial random variable (V_T, Z_T) which takes each of the 'T' possible values (σ_t^2, x_t) for $t = 1, 2, \cdots, T$ with equal probability $\frac{1}{T}$. We assume that the regressors have been centered, so that $EZ_T = 0$ and $EZ_T^2 = Var(Z_T)$. Standardize Z_T by introducing $W_T = \frac{Z_T}{\sqrt{Var(Z_T)}}$, and note that $EW_T = 0$ and $Var(W_T) = EW_T^2 = 1$.

Note that our artificial random variables V_T, Z_T and W_T depend on sample size 'T'. We introduce EV_T, EZ_T, EZ_T^2, EW_T and EW_T^2 to indicate that expectation is being taken for these random variables at sample size 'T'. Further note that we will drop the subscript 'T' for convenience in situations where the dependence on 'T' is not relevant.

According to **Lemma 2** of **Chapter 3:**, the true variance of the OLS estimate $\hat{\beta}_2$ of the slope parameter β_2 is given by:

EQ: 5-3
$$\Omega_{22} = \frac{EW^2V}{T\left(\mathrm{Var}(Z)\right)}$$

For HCCMEs of the type under discussion, the variance of the estimate of the slope parameter is:

EQ: 5-4
$$\hat{\Omega}_{22}(A_i) = \frac{1}{T^2\mathrm{Var}(Z)}\sum_{t=1}^{T} w_t^2 A_{it} e_t^2$$

The following theorem gives the biases of these HCCME.

Theorem 5-1: The bias $\mathrm{B}_{22}(A_i) = E\hat{\Omega}_{22}(A_i) - \Omega_{22}, i = 0,1,2,3$, of the HCCMEs for the variance of slope parameter is:

EQ: 5-5
$$\begin{aligned}\left\{T^2 Var(Z)\right\}\mathrm{B}_{22}(A_i) &= \frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{1}{T}\sum_{t=1}^{T} w_t^2 A_{it}\right) + 2\left(\frac{1}{T}\sum_{t=1}^{T} w_t^3 A_{it}\right) w_t\right]\sigma_t^2 \\ &+ \frac{1}{T}\sum_{t=1}^{T}\left[\left\{T(A_{it}-1) + \left(\frac{1}{T}\sum_{t=1}^{T} w_t^4 A_{it}\right) - 2A_{it}\right\} w_t^2 - 2A_{it} w_t^4\right]\sigma_t^2\end{aligned}$$

Remark: Using this expression, we can easily show that under usual assumptions, this bias of all these HCCMEs is of second order. This means that $\lim_{T\to\infty}\left\{\mathrm{T}\mathrm{Var}(Z)\mathrm{B}_{22}(A_i)\right\} = 0$, while, $\lim_{T\to\infty}\left\{\mathrm{T}^2\mathrm{Var}(Z)\mathrm{B}_{22}(A_i)\right\}$ is non-zero and $\lim_{T\to\infty}\left\{\mathrm{T}^3\mathrm{Var}(Z)\mathrm{B}_{22}(A_i)\right\} = \infty$

Proof: From the expressions for $\hat{\Omega}_{22}$ and Ω_{22} given earlier [see ***EQ: 5-3*** and ***EQ: 5-4***], we get,

EQ: 5-6 $$B_{22}(A_i) = E\hat{\Omega}_{22}(A_i) - \Omega_{22} = \frac{1}{T^2 Var(Z)} \sum_{t=1}^{T} w_t^2 \left[A_{it} E(e_t^2) - \sigma_t^2 \right]$$

Before proceeding, we need $E(e_t^2)$, which is given by following lemma:

Lemma 6: With centered regressors, the expected value of OLS squared residuals is given by:

EQ: 5-7 $$E(e_t^2) = \sigma_t^2 + \frac{1}{T}\left[EV - 2\sigma_t^2 - 2w_t^2\sigma_t^2 + 2w_t EWV + w_t^2 EW^2V \right]$$

Proof: The OLS residuals are $e = y - X\hat{\beta} = (I - H)\varepsilon$, where, $H = X(X'X)^{-1}X'$ is the 'hat matrix' as before. Using the standardized regressors $w_t = x_t / \text{Var}(Z)$, we can calculate the (i,j) entry of H to be:

$$H_{ij} = \frac{1}{T\text{Var}(Z)}\left(EZ^2 + x_i x_j\right) = \frac{1}{T}\left(1 + \frac{1}{\text{Var}(Z)} x_i x_j\right) = \frac{1}{T}\left(1 + w_i w_j\right)$$

Now note that $e_t = \varepsilon_t - \sum_{j=1}^{T} h_{tj}\varepsilon_j = (1 - h_{tt})\varepsilon_t - \sum_{\substack{j=1 \\ j\neq t}}^{T} h_{tj}\varepsilon_j$.

Since E e=0, and the ε's are independent, the variance of e is the sum of the variances. This can be explicitly calculated as follows:

$$E\left(e_t^2\right)=(1-h_{tt})^2\sigma_t^2+\sum_{j=1,j\neq t}^{T}h_{tj}^2\sigma_j^2=(1-2h_{tt})\sigma_t^2+\sum_{j=1}^{T}h_{tj}^2\sigma_j^2$$

$$=\left(1-\frac{2}{T}\left(1+w_t^2\right)\right)\sigma_t^2+\sum_{j=1}^{T}\left(\frac{1}{T}\left(1+w_tw_j\right)\right)^2\sigma_j^2$$

From this it follows that:

$$E\left(e_t^2\right)=\sigma_t^2-\frac{2}{T}\sigma_t^2-\frac{2}{T}w_t^2\sigma_t^2+\frac{1}{T}EV+\frac{1}{T}w_t^2EW^2V+\frac{2}{T}w_t\left(EWV\right)$$

Re-arranging terms, we get;

$$E\left(e_t^2\right)=\sigma_t^2+\frac{1}{T}\left[EV-2\sigma_t^2-2w_t^2\sigma_t^2+2w_tEWV+w_t^2EW^2V\right]$$

This proves the lemma.■

Now, substituting the expression of the **Lemma 6** in ***EQ: 5-6*** above, we get:

$$\mathrm{B}_{22}(A_t)=\frac{1}{T^2Var\left(Z\right)}\sum_{t=1}^{T}w_t^2\left[A_{tt}\left\{\sigma_t^2+\frac{1}{T}\left[EV-2\sigma_t^2-2w_t^2\sigma_t^2+2w_tEWV+w_t^2EW^2V\right]\right\}-\sigma_t^2\right]$$

Simplification leads to:

$$\left\{T^2Var\left(Z\right)\right\}\mathrm{B}_{22}(A_t)=\sum_{t=1}^{T}w_t^2A_{tt}\sigma_t^2+\frac{1}{T}\left(EV\right)\sum_{t=1}^{T}w_t^2A_{tt}-\frac{2}{T}\sum_{t=1}^{T}w_t^2A_{tt}\sigma_t^2-\frac{2}{T}\sum_{t=1}^{T}w_t^4A_{tt}\sigma_t^2$$
$$+\frac{2}{T}\left(EWV\right)\sum_{t=1}^{T}w_t^3A_{tt}+\frac{1}{T}\left(EW^2V\right)\sum_{t=1}^{T}w_t^4A_{tt}-\sum_{t=1}^{T}w_t^2\sigma_t^2$$

Re-arranging terms, we have:

$$\left\{T^2Var(Z)\right\}\mathrm{B}_{22}(A_i)=(EV)\left(\frac{1}{T}\sum_{t=1}^{T}w_t^2A_{it}\right)+2(EWV)\left(\frac{1}{T}\sum_{t=1}^{T}w_t^3A_{it}\right)+\frac{1}{T}\left(T\sum_{t=1}^{T}w_t^2A_{it}\sigma_t^2\right)$$
$$-\frac{1}{T}\left(T\sum_{t=1}^{T}w_t^2\sigma_t^2\right)+\left(EW^2V\right)\left(\frac{1}{T}\sum_{t=1}^{T}w_t^4A_{it}\right)-\frac{2}{T}\sum_{t=1}^{T}w_t^2A_{it}\sigma_t^2-\frac{2}{T}\sum_{t=1}^{T}w_t^4A_{it}\sigma_t^2$$

Writing in summation form,

$$\left\{T^2Var(Z)\right\}\mathrm{B}_{22}(A_i)=\frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{1}{T}\sum_{t=1}^{T}w_t^2A_{it}\right)+2\left(\frac{1}{T}\sum_{t=1}^{T}w_t^3A_{it}\right)w_t\right]\sigma_t^2$$
$$+\frac{1}{T}\sum_{t=1}^{T}\left[\left\{T(A_{it}-1)+\left(\frac{1}{T}\sum_{t=1}^{T}w_t^4A_{it}\right)-2A_{it}\right\}w_t^2-2A_{it}w_t^4\right]\sigma_t^2$$

This leads to the expression in **Theorem 5-1** which completes the proof.■

Now we replace values of A_{it} ,(i=0,1,2,3) to get biases of all HCCMEs in an explicit form. The following subsections explain these results.

5.3: Bias of Eicker-White (HC0) Estimator

In this section, we provide results of finite as well as second order asymptotic bias of HC0.

5.3.1: Finite Sample Bias of HC0

Theorem 5-2: Finite sample bias of HC0 for slope parameter is given by:

EQ: 5-8 $$B_{22}(A_0)=\frac{1}{T^2\text{Var}(Z)}\left\{EV+2\left(EW^3\right)EWV+\left(EW^4-2\right)EW^2V-2EW^4V\right\}$$

Remark: Finite sample bias of HC0 for slope parameter can be written in summation form as follows:

EQ: 5-9 $$B_{22}\left(A_0\right)=\frac{1}{T^3\text{Var}(Z)}\sum_{t=1}^{T}p_0\left(w_t\right)\sigma_t^2$$

Where,

EQ: 5-10 $$p_0\left(w_t\right)=1+2\left(EW^3\right)w_t+\left(EW^4-2\right)w_t^2-2w_t^4$$

Proof: Replacing, $i=0$, i.e. $A_0=I$, $A_{0t}=1,\forall t$ in ***EQ: 5-5*** above, and making use of the fact that $EW^2=1$, we get;

$$B_{22}\left(A_0\right)=\frac{1}{T^3Var(Z)}\sum_{t=1}^{T}\left[1+2\left(EW^3\right)w_t+\left(EW^4-2\right)w_t^2-2w_t^4\right]\sigma_t^2 .$$

Writing in terms of polynomial, $B_{22}(A_0) = \frac{1}{T^3 \mathrm{Var}(Z)} \sum_{t=1}^{T} p_0(w_t)\sigma_t^2$

Where, $p_0(w_t) = 1 + 2(EW^3)w_t + (EW^4 - 2)w_t^2 - 2w_t^4$

Changing summation expression to expectations, we get:

$$B_{22}(A_0) = \frac{1}{T^2 \mathrm{Var}(Z)} \left\{ EV + 2(EW^3)EWV + (EW^4 - 2)EW^2V - 2EW^4V \right\}$$

This proves theorem. ■

5.3.2: Second Order Asymptotic Bias of HC0

In order to derive second order asymptotic bias, we need to make some assumptions about the asymptotic behavior of the regressors and variances. In particular, we assume the following hmits EXIST:

EQ: 5-11
$$\begin{cases} \sigma_t^2 \le U, \forall t, \text{ U is some upper bound on variances} \\ \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sigma_t^2 = \lim_{T\to\infty} EV_T = EV_\infty \\ \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} w_t^\alpha = \lim_{T\to\infty} EW_T^\alpha = EW_\infty^\alpha, \qquad \alpha \le 11 \\ \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} w_t^\alpha \sigma_t^2 = \lim_{T\to\infty} EW_T^\alpha V_T = EW_\infty^\alpha V_\infty, \qquad \alpha \le 11 \\ \lim_{T\to\infty} Var(Z_T) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} z_t^2 = \lim_{T\to\infty} EZ_T^2 = EZ_\infty^2 \end{cases}$$

Remark: The above limits in ***EQ: 5-11*** exist under weaker condition. Specifically, powers of regressors and the sequence of variances and certain cross products are required to be Cesàro summable. One important special case where the limits exist is when the distribution of the random variables V_T, W_T and Z_T converge to some limiting distribution. In this case, we can use V_∞, W_∞ and Z_∞ to denote a random variable with this limiting distribution, and the notation for limit above is accurate. It is important to note that the limits assumed above will exist under much weaker conditions. It is convenient to continue to use the same notation V_∞, W_∞ and Z_∞ even when the limiting distribution of random variables do not exist, to indicate the required limits in the expressions above.

From the above assumptions, it follows that the absolute moments $\lim_{T\to\infty} E\left|W_T^\alpha\right|$ also exist for all $\alpha \leq 10$ [See Chung (2001), Theorem 6.4.1, page 166].

Now we are in a position to present the second order asymptotic bias of HC0 and the following theorem provides the same.

Here and when necessary, we will use the notation, $ASOB(HCi)$, $i = 0,1,2,3$ to denote the second order asymptotic bias of corresponding HCCMEs (namely HC0, HC1, HC2 and HC3). i.e. $\lim_{T\to\infty}\left\{T^2 B_{22}\left(A_i\right)\right\} = ASOB(HCi)$, $i = 0,1,2,3$

Theorem 5-3: Under the assumptions stated in ***EQ: 5-11*** the second order asymptotic bias of HC0 is given by:

EQ: 5-12 $$ASOB(HC0)=\frac{1}{EZ_{\infty}^{2}}\left\{EV_{\infty}+2\left(EW_{\infty}^{3}\right)EW_{\infty}V_{\infty}+\left(EW_{\infty}^{4}-2\right)EW_{\infty}^{2}V_{\infty}-2EW_{\infty}^{4}V_{\infty}\right\}$$

Proof: Taking limit as 'T' approaches infinity after multiplying both sides of **EQ: 5-8** by T^2 and noting that limit exists by the assumptions in **EQ: 5-11.** This leads to required result.■

5.4: Bias of Hinkley (HC1) Estimator

Here we present results regarding finite samples as well as second order asymptotic bias of HC1.

5.4.1: Finite Sample Bias of HC1

Theorem 5-4: Finite sample bias of HC1 for slope parameter is given by:

EQ: 5-13

$$B_{22}(A_1)=\frac{1}{T^2 Var(Z)}\left(\frac{T}{T-2}\right)\left(EV+2\left(EW^3\right)EWV+\left(EW^4\right)EW^2V-2EW^4V\right)$$

Remark: Finite sample bias of HC1 for slope parameter can be written in summation form as:

EQ: 5-14 $$B_{22}(A_1)=\frac{1}{T^3\text{Var}(Z)}\frac{T}{T-2}\sum_{t=1}^{T}p_1(w_t)\sigma_t^2$$

Where,

EQ: 5-15 $$p_1(w_t)=1+2\left(EW^3\right)w_t+\left(EW^4\right)w_t^2-2w_t^4$$

Proof: Replacing, $i=1$, $A_1=\left(\frac{T}{T-2}\right)I$, $A_{1t}=\left(\frac{T}{T-2}\right), \forall t$ in ***EQ: 5-5*** above, and making use of $EW^2=1$, leads to:

$$B_{22}(A_1)=\left(\frac{T}{T-2}\right)\frac{1}{T^3Var(Z)}\sum_{t=1}^{T}\left[1+2\left(EW^3\right)w_t+\left(EW^4\right)w_t^2-2w_t^4\right]\sigma_t^2$$

Writing in terms of polynomial, $B_{22}(A_1)=\left(\frac{T}{T-2}\right)\frac{1}{T^3\text{Var}(Z)}\sum_{t=1}^{T}p_1(w_t)\sigma_t^2$

Where, $p_1(w_t)=1+2\left(EW^3\right)w_t+\left(EW^4\right)w_t^2-2w_t^4$

Changing summation expression to expectations, we get:

$$B_{22}(A_1)=\left(\frac{T}{T-2}\right)\frac{1}{T^2\text{Var}(Z)}\left\{EV+2\left(EW^3\right)EWV+\left(EW^4\right)EW^2V-2EW^4V\right\}$$

This proves theorem. ∎

5.4.2: Second Order Asymptotic Bias of HC1

As before, for second order asymptotic bias, we need to make some assumptions about the asymptotic behavior of the regressors and variances. Under the same assumptions as in case of HC0 above, (see ***EQ: 5-11***), the asymptotic bias of HC1 is given in the following theorem:

Theorem 5-5: Under the assumptions stated in ***EQ: 5-11***, the second order asymptotic bias of HC1 is given below:

EQ: 5-16 $$ASOB(HC1)=\frac{1}{EZ_\infty^2}\left\{EV_\infty+2\left(EW_\infty^3\right)EW_\infty V_\infty+\left(EW_\infty^4\right)EW_\infty^2V_\infty-2EW_\infty^4V_\infty\right\}$$

Proof: Taking limit as 'T' approaches infinity on both sides of **EQ: 5-13** after multiplying both sides by T^2 and noting that $\left(\frac{T}{T-2}\right)$ approaches one as 'T' approaches infinity. In addition, assumptions in ***EQ: 5-11*** guarantees that limit exists. This proves the theorem.■

5.5: Bias of Horn, Horn and Duncan (HC2) Estimator

The results of finite sample bias of HC2 along with second order asymptotic bias of the same are provided in this section. The following two sub-sections provide their details.

5.5.1: Finite Sample Bias of HC2

Theorem 5-6: Finite sample bias of HC2 for slope parameter is given by:

EQ: 5-17

$$B_{22}(A_2)=\frac{1}{T^2Var(Z_T)}\left\{EV_T+2\left(EW_T^3\right)EW_TV_T+\left(EW_T^4-1\right)EW_T^2V_T-EW_T^4V_T+R_{1T}\right\}$$

Where, R_{1T} is the remainder term given in **EQ: 5-18** below.

Remark: We will show later [See **Theorem 5-7** below] that the remainder (R_{1T}) is small for large 'T'.

Proof: To get the bias of HC2, take, $i=2$, $A_{2t}=\frac{1}{1-h_{tt}}, \forall t$ in ***EQ: 5-5*** and simplifying, we get:

$$\left\{T^2Var(Z_T)\right\}B_{22}(A_2)=\frac{1}{T}\left(\sum_{t=1}^{T}\frac{w_t^2}{(1-h_{tt})}\right)EV_T+\frac{2}{T}\left(\sum_{t=1}^{T}\frac{w_t^3}{(1-h_{tt})}\right)EW_TV_T+\sum_{t=1}^{T}\frac{w_t^2\sigma_t^2}{(1-h_{tt})}$$
$$-T\,EW_T^2V_T+\frac{1}{T}\left(\sum_{t=1}^{T}\frac{w_t^4}{(1-h_{tt})}\right)EW_T^2V_T-\frac{2}{T}\sum_{t=1}^{T}\frac{w_t^2\sigma_t^2}{(1-h_{tt})}-\frac{2}{T}\sum_{t=1}^{T}\frac{w_t^4\sigma_t^2}{(1-h_{tt})}$$

Note that, we can write,

$$\frac{1}{1-h_{tt}}=1+h_{tt}+h_{tt}^2+h_{tt}^3+.....=1+h_{tt}+h_{tt}^2\left(\frac{1}{1-h_{tt}}\right)$$

Replacing this in above expression, we get:

$$\left\{T^2Var(Z_T)\right\}B_{22}(A_2)=\frac{1}{T}\left(\sum_{t=1}^{T}w_t^2\right)EV_T+\frac{2}{T}\left(\sum_{t=1}^{T}w_t^3\right)EW_TV_T-TEW_T^2V_T$$
$$+\sum_{t=1}^{T}w_t^2\sigma_t^2+\frac{1}{T}\left(\sum_{t=1}^{T}w_t^4\right)EW_T^2V_T-\frac{2}{T}\sum_{t=1}^{T}w_t^2\sigma_t^2-\frac{2}{T}\sum_{t=1}^{T}w_t^4\sigma_t^2+\frac{1}{T}\left(\sum_{t=1}^{T}h_{tt}w_t^2\right)EV_T$$
$$+\frac{2}{T}\left(\sum_{t=1}^{T}h_{tt}w_t^3\right)EW_TV_T+\sum_{t=1}^{T}h_{tt}w_t^2\sigma_t^2+\frac{1}{T}\left(\sum_{t=1}^{T}h_{tt}w_t^4\right)EW_T^2V_T-\frac{2}{T}\sum_{t=1}^{T}h_{tt}w_t^2\sigma_t^2$$
$$-\frac{2}{T}\sum_{t=1}^{T}h_{tt}w_t^4\sigma_t^2+\frac{1}{T}\left(\sum_{t=1}^{T}\frac{h_{tt}^2w_t^2}{1-h_{tt}}\right)EV_T+\frac{2}{T}\left(\sum_{t=1}^{T}\frac{h_{tt}^2w_t^3}{1-h_{tt}}\right)EW_TV_T+\sum_{t=1}^{T}\frac{h_{tt}^2w_t^2\sigma_t^2}{1-h_{tt}}$$
$$+\frac{1}{T}\left(\sum_{t=1}^{T}\frac{h_{tt}^2w_t^4}{1-h_{tt}}\right)EW_T^2V_T-\frac{2}{T}\sum_{t=1}^{T}\frac{h_{tt}^2w_t^2\sigma_t^2}{1-h_{tt}}-\frac{2}{T}\sum_{t=1}^{T}\frac{h_{tt}^2w_t^4\sigma_t^2}{1-h_{tt}}$$

Using $h_{tt}=\frac{1}{T}\left(1+w_t^2\right)$ and $h_{tt}^2=\frac{1}{T^2}\left(1+w_t^2\right)^2=\frac{1}{T^2}\left(1+2w_t^2+w_t^4\right)$, and simplifying, we get:

$$B_{22}(A_2)=\frac{1}{T^2}\left[\frac{1}{EZ_T^2}\left\{EV_T+2\left(EW_T^3\right)EW_TV_T+\left(EW_T^4-1\right)EW_T^2V_T-EW_T^4V_T\right\}+R_{1T}\right]$$

Where,

EQ: 5-18

$$R_{1T}=\frac{1}{T\,EZ_T^2}\left[\left(1+EW_T^4\right)EV_T+2\left(EW_T^3+EW_T^5\right)EW_TV_T+\left(EW_T^4+EW_T^6-2\right)EW_T^2V_T\right]$$

$$-\frac{1}{T\,EZ_T^2}\left[4EW_T^4V_T+2\left(EW_T^6V_T\right)\right]+\frac{1}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2+2w_t^4+w_t^6\right)}{1-h_{tt}}\right)EV_T$$

$$+\frac{2}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^3+2w_t^5+w_t^7\right)}{1-h_{tt}}\right)EW_TV_T+\frac{1}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2+2w_t^4\sigma_t^2+w_t^6\sigma_t^2\right)}{1-h_{tt}}\right)$$

$$+\frac{1}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4+2w_t^6+w_t^8\right)}{1-h_{tt}}\right)EW_T^2V_T-\frac{2}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2+2w_t^4\sigma_t^2+w_t^6\sigma_t^2\right)}{1-h_{tt}}\right)$$

$$-\frac{2}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4\sigma_t^2+2w_t^6\sigma_t^2+w_t^8\sigma_t^2\right)}{1-h_{tt}}\right)$$

This completes the proof. ■

5.5.2: Second Order Asymptotic Bias of HC2

To get an explicit expression for the second order asymptotic bias of HC2, as before, we have to make some assumptions about the asymptotic behavior of the regressors and variances. The assumptions are stated in ***EQ: 5-11*** above.

We also assume throughout that hat matrix, $H=X\left(X'X\right)^{-1}X'$ is asymptotically balanced, i.e.

EQ: 5-19 $\lim_{T\to\infty} \max_{1\le t\le T} \left(h_{tt}\right) = 0$

This assumption assumes that all regressors are of similar order of magnitude and that there are no extreme outliers.

Theorem 5-7: Under the assumptions stated in ***EQ: 5-11*** and ***EQ: 5-19***, the second order asymptotic bias of HC2 is given below:

EQ: 5-20

$$ASOB(HC2) = \frac{1}{EZ_\infty^2}\left\{EV_\infty + 2\left(EW_\infty^3\right)EW_\infty V_\infty + \left(EW_\infty^4 - 1\right)EW_\infty^2 V_\infty - EW_\infty^4 V_\infty\right\}$$

Proof: We have established that

$$\mathrm{T}^2\mathrm{B}_{22}\left(A_2\right) = \frac{1}{EZ_T^2}\left\{EV_T + 2\left(EW_T^3\right)EW_T V_T + \left(EW_T^4 - 1\right)EW_T^2 V_T - EW_T^4 V_T\right\} + R_{1T}$$

Where, R_{1T} is defined in **EQ: 5-18** above. To prove the theorem, it is sufficient to show that $\lim_{T\to\infty}\left(R_{1T}\right) = 0$. The limit of the remaining part exists and equals the expression in the theorem by assumptions in ***EQ: 5-11***.

Note that, we can write R_{1T} as, $R_{1T} = \frac{1}{T}\left(F_T\right)$, where F_T is given by:

EQ: 5-21

$$F_T = \frac{1}{EZ_T^2}\left(1+EW_T^4\right)EV_T + \frac{2}{EZ_T^2}\left(EW_T^3+EW_T^5\right)EW_TV_T + \frac{1}{EZ_T^2}\left(EW_T^4+EW_T^6-2\right)EW_T^2V_T$$

$$-\frac{4}{EZ_T^2}EW_T^4V_T - \frac{2}{EZ_T^2}\left(EW_T^6V_T\right) + \frac{1}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2+2w_t^4+w_t^6\right)}{1-h_{tt}}\right)EV_T$$

$$+\frac{2}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^3+2w_t^5+w_t^7\right)}{1-h_{tt}}\right)EW_TV_T + \frac{1}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4+2w_t^6+w_t^8\right)}{1-h_{tt}}\right)EW_T^2V_T$$

$$+\frac{1}{EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2+2w_t^4\sigma_t^2+w_t^6\sigma_t^2\right)}{1-h_{tt}}\right) - \frac{2}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2+2w_t^4\sigma_t^2+w_t^6\sigma_t^2\right)}{1-h_{tt}}\right)$$

$$-\frac{2}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4\sigma_t^2+2w_t^6\sigma_t^2+w_t^8\sigma_t^2\right)}{1-h_{tt}}\right)$$

Under the assumption that the design matrix is balanced (See ***EQ: 5-19***),

i.e. $\lim_{T\to\infty}\max_{1\le t\le T}\left(h_{tt}\right)=0$,

So for 'T' large enough, $h_{tt}\le\frac{1}{2},\forall t$ and hence, $\frac{1}{1-h_{tt}}<2$.

Note that we can write $\frac{1}{T}\sum_{t=1}^{T}w_t^\alpha \le \frac{1}{T}\sum_{t=1}^{T}\left|w_t\right|^\alpha$

and, $\left|\frac{1}{T}\sum_{t=1}^{T}\frac{w_t^\alpha}{1-h_{tt}}\right| \le \frac{1}{T}\sum_{t=1}^{T}\left|\frac{w_t^\alpha}{1-h_{tt}}\right| \le 2\frac{1}{T}\sum_{t=1}^{T}\left|w_t^\alpha\right|$

Since variances are bounded, i.e. $\sigma_t^2\le U,\forall t$

So, we can write; $\left|\frac{1}{T}\sum_{t=1}^{T}\frac{w_t^\alpha\sigma_t^2}{1-h_{tt}}\right| \le \frac{1}{T}\sum_{t=1}^{T}\left|\frac{w_t^\alpha\sigma_t^2}{1-h_{tt}}\right| \le 2U\frac{1}{T}\sum_{t=1}^{T}\left|w_t^\alpha\right| = 2U\,E\left|W_T^\alpha\right|$.

This means that each of the terms in the F_T above is separately bounded and is of order $\mathcal{O}(1)$. Hence the sum of all terms (i.e. F_T) is also bounded and is of order $\mathcal{O}(1)$, so $R_{1T} = \frac{1}{T}(F_T)$ goes to zero as 'T' approaches infinity, as claimed. ■

5.6: *Bias of Mackinnon and White (HC3) Estimator*

This section presents results of finite sample as well as second order asymptotic bias of HC3. The details are provided in the following two sub-sections.

5.6.1: Finite Sample Bias of HC3

Theorem 5-8: Finite sample bias of HC3 for slope parameter is given by:

EQ: 5-22
$$B_{22}(A_3) = \frac{1}{T^2 Var(Z_T)} \left\{ EV_T + 2\left(EW_T^3\right) EW_T V_T + \left(EW_T^4\right) EW_T^2 V_T + R_{2T} \right\}$$

Where, R_{2T} is the remainder term given in **EQ: 5-23** below.

Remark: We will show later [See **Theorem 5-9** below] that the remainder (R_{2T}) is small for large 'T'.

Proof: To get the bias of HC2, replace $i=3$, i.e. $A_{3t}=\frac{1}{(1-h_{tt})^2}, \forall t$ in **EQ: 5-5** above, and simplifying, we get:

$$\{T^2Var(Z_T)\}\mathrm{B}_{22}(A_3)=\frac{1}{T}\left(\sum_{t=1}^{T}\frac{w_t^2}{(1-h_{tt})^2}\right)EV_T+\frac{2}{T}\left(\sum_{t=1}^{T}\frac{w_t^3}{(1-h_{tt})^2}\right)EW_TV_T+\sum_{t=1}^{T}\frac{w_t^2\sigma_t^2}{(1-h_{tt})^2}$$

$$-T\,EW_T^2V_T+\frac{1}{T}\left(\sum_{t=1}^{T}\frac{w_t^4}{(1-h_{tt})^2}\right)EW_T^2V_T-\frac{2}{T}\sum_{t=1}^{T}\frac{w_t^2\sigma_t^2}{(1-h_{tt})^2}-\frac{2}{T}\sum_{t=1}^{T}\frac{w_t^4\sigma_t^2}{(1-h_{tt})^2}$$

Note that,

$$\frac{1}{(1-h_{tt})^2}=1+2h_{tt}+3h_{tt}^2+4h_{tt}^3+5h_{tt}^4+....$$

$$=1+2h_{tt}+3h_{tt}^2\left(\frac{1}{1-h_{tt}}\right)+h_{tt}^3\left(\frac{1}{(1-h_{tt})^2}\right)$$

Replacing it in bias expression given above, we get:

The bias of this estimator is $B_{22}^{\cdot}(\alpha) = E\hat{\Omega}_{22}(\alpha) - \Omega_{22}$. This bias depends on the unknown variance under heteroskedasticity. Assuming that the variances are bounded, i.e. $\forall t : \sigma_t^2 \leq U$, there is a maximum possible bias which obtains when the heteroskedasticity has the worst possible configuration. For each value of 'α', we can calculate this worst possible bias: $MB(\Omega(\alpha)) = \max(B^+, -B^-)$

Define the Minimax Value of 'α' to be the one which minimizes this maximum bias. This is the best possible value of 'α' in a certain sense.

In terms of these notations, we can state our conjecture.

CONJECTURE: The minimax value of 'a' and hence 'α' (because $\alpha = 1 + \frac{a}{T}$) does not depend on the exact sequence of regressors but only on the sample size 'T', and the kurtosis of the regressors, 'K'.

Heuristic Proof: This was done by simulations. For each sample size 'T', and fixed value of kurtosis 'K', we generated random sequences of 'T' regressors having kurtosis 'K', and numerically computed the minimax value of 'a'. A heuristic proof of the conjecture is obtained by showing that this minimax value always comes out the same. Some of the simulations which support this conjecture have already been reported earlier (See **Section 4.4**: **above**). The details of simulation results is provided in the following table, where we calculated maximum bias and the optimal value of 'a', where both positive and negative bias are equal, for six different samples with matching kurtosis and skewness measures. In particular, we fixed kurtosis at '3' and took two values of skewness, first fixing it to zero and then at one and calculated maximum bias and also calculated optimal value of 'a' for all '12' samples (6 samples taking skewness zero and for remaining six samples skewness is taken as one). The second half of table shows results with kurtosis measure fixed at '4' while skewness measure taking values zero and one respectively.

Table 4-1: Maximum Bias with varying skewness with fixed kurtosis at Sample size 100 for different set of regressors

	K=3 & S=0		K=3 & S=1		K=4 & S=0		K=4 & S=1	
Samples of Regressors	**a***	**MB**	**a***	**MB**	**a***	**MB**	**a***	**MB**
1	4.167	2.080	4.167	2.070	5.263	3.046	5.263	2.885
2	4.167	2.012	4.167	2.131	5.263	4.166	5.263	4.095
3	4.167	2.394	4.167	1.833	5.263	4.162	5.263	3.750
4	4.167	2.771	4.167	1.933	5.263	3.535	5.263	2.743
5	4.167	2.137	4.167	1.837	5.263	4.166	5.263	3.358
6	4.167	2.256	4.167	2.070	5.263	3.391	5.263	3.358

Note: K and S are Kurtosis and Skewness measures respectively and MB represents the maximum bias.

We can see from the above table that, for all six samples with randomly chosen regressors, whether skewness is zero or one but with same kurtosis measure, the optimal value of 'a' is same.

Similar results were obtained for sample sizes 25 and 50 which are not reported to save space.

4.4.2: The Minimax Value of 'a'

Once we assume that the invariance conjecture is valid, it becomes possible to compute analytically the minimax value of 'a'. This is because we now compute the minimax value for a particular sequence of regressors for which easy analytic calculation is possible. On the basis of the invariance conjecture, this calculation should be valid simultaneously for all sequences of regressors with matching Kurtosis and equal sample size. We use this method to compute analytically (instead of numerically) the optimal minimax value of 'a' in this section.

This is what we do to find the minimax value of 'a' in **Theorem 4-4** below. Even though the chosen sequence of regressors is very different from the normal regressors case solved earlier, identical asymptotic minimax values of 'a' emerge, giving further support to our invariance conjecture.

Theorem 4-4: Assume that the variances are bounded so that $\forall t : \sigma_t^2 \le U$. For each α define the maximum bias $MB(\alpha) = \max_{\sigma_t^2 \le U} B_{22}(\alpha)$ -- this is maximum possible bias obtainable by setting the heteroskedastic sequence of variances to the least favorable configuration. Let K $\doteq$ EW^4 be the kurtosis of the sequence of regressors. Define $\alpha^* = 1 + \frac{a^*}{T}$, where,

EQ: 4-16 $a^* = \frac{K+1}{1-\frac{1}{T}(K+1)}$

Then $MB(\alpha^*) \le MB(\alpha)$ for all α.

Proof: Assuming that the exact sequence of regressors does not matter, we pick a particular sequence for which the computations are easy. We can easily calculate the positive and negative bias functions for the simple regressor sequence described below.

We choose sample size *T* and a large constant M [*M>1*], such that $k = T/(2M^2)$ is an integer smaller than T/2.Now consider the sequence of regressors $x_1, \cdots, x_T$ such that $x_1 = x_2 = \cdots = x_k = -M$, $x_{k+1} = \cdots = x_{T-k} = 0$, and $x_{T-k+1} = \cdots = x_T = +M$. As before, letting Z be the random variable such that Z = x_t with probability 1/T, we can easily check that EZ=0, $EZ^2=1$, and $EZ^4=M^2$. That is Z is centered and standardized and has kurtosis $K = EZ^4 = M^2$. Note that we assume M>*1*, this means that we are constraining the kurtosis to be bounded below by 1; equivalently, we assume Excess Kurtosis (EK) to be greater than -2: EK=K-3>-2. Because Z is standardized, W=Z/Std(Z)=Z and the standardized regressors are just $w_t = x_t$. Noting that $EW^3=0$ and that the kurtosis is $EW^4=M^2$, so, we can write the polynomial p(w) in ***EQ: 4-7*** as:

$$p(w) = \alpha + \left(T(\alpha-1) + \alpha\left(M^2-2\right)\right)w^2 - 2\alpha w^4$$

Note that $p(0)=\alpha=1+\frac{a}{T}>0$ for all positive values of 'a'. Also p(+M)=p(-M) is a polynomial in M^2 with positive root:

$$r^{+}=\frac{\left(a-2\left(1+\frac{a}{T}\right)\right)+\sqrt{\left(a-2\left(1+\frac{a}{T}\right)\right)^{2}+4\left(1+\frac{a}{T}\right)^{2}}}{2\left(1+\frac{a}{T}\right)}.$$

Note that $p(\pm M)<0, \quad \forall M^2>r^{+}\geq 1$ and $p(\pm M)>0, \quad \forall 1<M^2<r^{+}$

First we consider the case where the value of 'a' is below $M^2+2-\frac{1}{M^2}$, and T is large. In this case, it is easily checked from above calculations that $p(\pm M)<0$. The maximum positive and negative bias functions can be calculated as follows:

$$\begin{aligned}
B^{+}&=\frac{1}{T^3\text{Var}(Z)}\sum_{t=k+1}^{t=T-k}p(w_t)U=\frac{1}{T^2\text{Var}(Z)}\frac{T-2k}{T}p(0)U\\
&=\frac{1}{T^2\text{Var}(Z)}\left(1-\frac{1}{M^2}\right)\left(1+\frac{a}{T}\right)U
\end{aligned}$$

$$\begin{aligned}
B^{-}&=\frac{1}{T^3\text{Var}(Z)}\left\{\sum_{t=1}^{k}p(w_t)U+\sum_{t=T-k+1}^{T}p(w_t)U\right\}\\
&=\frac{1}{T^2\text{Var}(Z)}\frac{2k}{T}p(M)U=\frac{1}{T^2\text{Var}(Z)}\frac{p(M)U}{M^2}\\
&=\frac{1}{T^2\text{Var}(Z)}\frac{U}{M^2}\left\{\left(1+\frac{a}{T}\right)+\left(a+(1+a/T)\left(M^2-2\right)\right)M^2-2\left(1+\frac{a}{T}\right)M^4\right\}
\end{aligned}$$

Simplifying, the above expressions, we get:

EQ: 4-17 $$M^2\left\{\mathrm{T}^2\mathrm{Var}(Z)\right\}\mathcal{B}^+/U=\left(M^2-1\right)+\frac{\mathrm{a}}{T}\left(M^2-1\right)$$

EQ: 4-18 $$-M^2\left\{\mathrm{T}^2\mathrm{Var}(Z)\right\}\mathcal{B}^-/U=M^4+(2-\mathrm{a})M^2-1+\frac{\mathrm{a}}{T}\left(M^4+2M^2-1\right)$$

Solving $M^2\left\{\mathrm{T}^2\mathrm{Var}(Z)\right\}\mathcal{B}^+/U=-M^2\left\{\mathrm{T}^2\mathrm{Var}(Z)\right\}\mathcal{B}^-/U$ yields the minimax (optimal) value of 'a' in case of finite samples:

EQ: 4-19 $$\mathrm{a}^*=\frac{M^2+1}{1-\frac{1}{T}\left(M^2+1\right)}=\frac{K+1}{1-\frac{1}{T}(K+1)}$$

This proves that a* minimizes the maximum bias for the range of values of 'a' less than $M^2+2-\frac{1}{M^2}$. Note that this range includes a=0 and a=2, so that the minimax estimator a* dominates Eicker-White and Hinkley in terms of maximum risk. For larger values of 'a', i.e. when $\mathrm{a}>M^2+2-\frac{1}{M^2}$, the polynomial $p(w)$ becomes positive at +M and –M, i.e. $p(\pm M)>0$ and so the above calculations do not apply. For this case, maximum negative bias becomes zero and we have only maximum positive bias which can be calculated as follows:

$$\mathcal{B}^+=\frac{1}{T^3\mathrm{Var}(Z)}\left\{\sum_{t=1}^{k}p(w_t)U+\sum_{t=k+1}^{t=T-k}p(w_t)U+\sum_{t=T-k+1}^{T}p(w_t)U\right\}$$
$$=\frac{1}{T^2\mathrm{Var}(Z)}\frac{T-2k}{T}p(0)U+\frac{1}{T^2\mathrm{Var}(Z)}\frac{2k}{T}p(M)U$$

This can be simplified as follows:

$$\frac{\{T^2\mathrm{Var}(Z)\}\mathcal{B}^+}{U} = \left(1-\frac{1}{M^2}\right)\left(1+\frac{\mathrm{a}}{T}\right)$$
$$+\frac{1}{M^2}\left\{\left(1+\frac{\mathrm{a}}{T}\right)+\left(\mathrm{a}+(1+\mathrm{a}/T)(M^2-2)\right)M^2-2\left(1+\frac{\mathrm{a}}{T}\right)M^4\right\}$$
$$=\left((\mathrm{a}-1)-M^2\right)-\frac{\mathrm{a}}{T}\left(1+M^2\right)$$

Minimizing this maximum positive bias over range of values of 'a' larger than $M^2+2-\frac{1}{M^2}$, we can get minimax bias for the case when $p(\pm M)>0$. Further note that, these values of 'a', cannot lead to reductions in maximum risk. It follows that the value of a* in **EQ: 4-19** for the case when $p(\pm M)<0$ minimizes the maximum risk over all possible non-negative values of 'a'. This proves the Theorem.■

Taking the limit as T goes to infinity, we can get the asymptotic minimax value of 'a', which is given below:

EQ: 4-20 $\quad \mathrm{a}^* = M^2+1 = K+1.$

Note that when K=3, matching the kurtosis of normal regressors, we get the same asymptotic minimax value of a=4.

4.4.3: Evaluation of Relative Performance

Using the results obtained, we can analytically compare the relative performance of the Eicker-White and the Minimax Estimator in terms of asymptotic bias. Note that both the maximum bias functions $\mathcal{B}^+$ and $\mathcal{B}^-$ are proportional to *U*, the upper bound on the variances. To get a reasonable performance measure which is invariant to this arbitrary upper bound, it seems reasonable to divide by this factor.

The asymptotic maximum positive bias is then $\lim_{T\to\infty}\{\mathrm{T}^2\mathrm{Var}(Z)\}\mathcal{B}^+/U = 1-1/M^2$, which does not depend on 'a', and is bounded above by 1.

On the other hand, the asymptotic maximum negative bias is:

$$\lim_{T\to\infty}\left\{-\{\mathrm{T}^2\mathrm{Var}(Z)\}\mathcal{B}^-/U\right\} = M^2 + (2-\mathrm{a}) - 1/M^2$$

This increase with the kurtosis $K = M^2$ of regressors and is unbounded.

Let maximum of both biases (maximum positive and minus the maximum negative bias) is represented by $\mathcal{B} = \max(\mathcal{B}^+, -\mathcal{B}^-)$, where $\mathcal{B}^+$ and $\mathcal{B}^-$ respectively are positive and negative biases.

For the Eicker-White estimator with a=0, this maximum bias is somewhat larger than the kurtosis:

$$\mathcal{B}_0 = \max(\mathcal{B}_0^+, -\mathcal{B}_0^-) = K + 2 - \frac{1}{K}$$

Hinkley's bias correction has a=2, which yields the maximum bias of: $\mathcal{B}_1 = \max(\mathcal{B}_1^+, -\mathcal{B}_1^-) = K - \frac{1}{K}$. Note that Hinkley bias correction is too timid – it knocks out the middle term, but does nothing to the dominant bias term K= M^2.

The minimax value sets $\mathrm{a} = M^2 + 1$, which results in:

$$\lim_{T\to\infty}\left[-\{\mathrm{T}^2\mathrm{Var}(Z)\}\mathcal{B}^-/U\right] = 1 - 1/M^2 = \lim_{T\to\infty}\left[\{\mathrm{T}^2\mathrm{Var}(Z)\}\mathcal{B}^+/U\right]$$

Thus maximum bias of the minimax estimate is:

$$\mathcal{B}_{\text{Minimax}} = \max\left(\mathcal{B}^{+}_{\text{Minimax}}, -\mathcal{B}^{-}_{\text{Minimax}}\right) = 1 - \frac{1}{K}$$

By knocking out the leading bias term, this results in maximum bias bounded above by 1. When Kurtosis (K) is large, the minimax estimator is substantially superior to both Eicker-White and Hinkley in terms of maximum possible bias.

These results above are for the asymptotic bias. Next we consider the finite sample case by taking into account the terms of order $\mathcal{O}(1/T)$ which have been ignored in the above calculations. These $\mathcal{O}(1/T)$ terms further enhance the superiority of the minimax bias estimator over the Eicker-White. The leading $\mathcal{O}(1/T)$ term in **EQ: 4-18** is M^4 which dominates others for large values of M. The $\mathcal{O}(1/T)$ term in the minimax estimator knocks out this term and substantially reduces maximum possible finite sample bias over Eicker-White.

$$\left\{T^2 Var(Z_T)\right\} \mathrm{B}_{22}(A_3) = \frac{1}{T}\left(\sum_{t=1}^{T} w_t^2\right) EV_T + \frac{2}{T}\left(\sum_{t=1}^{T} w_t^3\right) EW_T V_T + \sum_{t=1}^{T} w_t^2 \sigma_t^2 - T\, EW_T^2 V_T$$

$$+\frac{1}{T}\left(\sum_{t=1}^{T} w_t^4\right) EW_T^2 V_T - \frac{2}{T}\sum_{t=1}^{T} w_t^2 \sigma_t^2 - \frac{2}{T}\sum_{t=1}^{T} w_t^4 \sigma_t^2 + \frac{2}{T}\left(\sum_{t=1}^{T} h_{tt} w_t^2\right) EV_T$$

$$+\frac{4}{T}\left(\sum_{t=1}^{T} h_{tt} w_t^3\right) EW_T V_T + 2\sum_{t=1}^{T} h_{tt} w_t^2 \sigma_t^2 + \frac{2}{T}\left(\sum_{t=1}^{T} h_{tt} w_t^4\right) EW_T^2 V_T - \frac{4}{T}\sum_{t=1}^{T} h_{tt} w_t^2 \sigma_t^2$$

$$-\frac{4}{T}\sum_{t=1}^{T} h_{tt} w_t^4 \sigma_t^2 + \frac{3}{T}\left(\sum_{t=1}^{T} \frac{h_{tt}^2 w_t^2}{1-h_{tt}}\right) EV_T + \frac{6}{T}\left(\sum_{t=1}^{T} \frac{h_{tt}^2 w_t^3}{1-h_{tt}}\right) EW_T V_T + 3\sum_{t=1}^{T} \frac{h_{tt}^2 w_t^2 \sigma_t^2}{1-h_{tt}}$$

$$+\frac{3}{T}\left(\sum_{t=1}^{T} \frac{h_{tt}^2 w_t^4}{1-h_{tt}}\right) EW_T^2 V_T - \frac{6}{T}\sum_{t=1}^{T} \frac{h_{tt}^2 w_t^2 \sigma_t^2}{1-h_{tt}} - \frac{6}{T}\sum_{t=1}^{T} \frac{h_{tt}^2 w_t^4 \sigma_t^2}{1-h_{tt}} + \frac{1}{T}\left(\sum_{t=1}^{T} \frac{h_{tt}^3 w_t^2}{(1-h_{tt})^2}\right) EV_T$$

$$+\frac{2}{T}\left(\sum_{t=1}^{T} \frac{h_{tt}^3 w_t^3}{(1-h_{tt})^2}\right) EW_T V_T + \sum_{t=1}^{T} \frac{h_{tt}^3 w_t^2 \sigma_t^2}{(1-h_{tt})^2} + \frac{1}{T}\left(\sum_{t=1}^{T} \frac{h_{tt}^3 w_t^4}{(1-h_{tt})^2}\right) EW_T^2 V_T$$

$$-\frac{2}{T}\sum_{t=1}^{T} \frac{h_{tt}^3 w_t^2 \sigma_t^2}{(1-h_{tt})^2} - \frac{2}{T}\sum_{t=1}^{T} \frac{h_{tt}^3 w_t^4 \sigma_t^2}{(1-h_{tt})^2}$$

Using, $h_{tt} = \frac{1}{T}\left(1+w_t^2\right)$, $h_{tt}^2 = \frac{1}{T^2}\left(1+w_t^2\right)^2 = \frac{1}{T^2}\left(1+2w_t^2+w_t^4\right)$,

$h_{tt}^2 = \frac{1}{T^3}\left(1+w_t^2\right)^3 = \frac{1}{T^3}\left(1+w_t^6+3w_t^2+3w_t^4\right)$, and simplifying, we get:

$$\mathrm{B}_{22}(A_3) = \frac{1}{T^2}\left[\frac{1}{EZ_T^2}\left\{EV_T + 2\left(EW_T^3\right)EW_T V_T + \left(EW_T^4\right)EW_T^2 V_T\right\} + R_{2T}\right]$$

Where,

EQ: 5-23

$$R_{2T} = \frac{2}{T\,EZ_T^2}\left[EV_T + \left(EW_T^4\right)EV_T + 2\left(EW_T^3 + EW_T^5\right)EW_T V_T\right]$$

$$+\frac{2}{T\,EZ_T^2}\left[\left(EW_T^4 + EW_T^6 - 2\right)EW_T^2 V_T - 4EW_T^4 V_T - 2EW_T^6 V_T\right]$$

$$+\frac{3}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2 + 2w_t^4 + w_t^6\right)}{1-h_{tt}}\right)EV_T + \frac{6}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^3 + 2w_t^5 + w_t^7\right)}{1-h_{tt}}\right)EW_T V_T$$

$$+\frac{3}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + 2w_t^4\sigma_t^2 + w_t^6\sigma_t^2\right)}{1-h_{tt}}\right) + \frac{3}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4 + 2w_t^6 + w_t^8\right)}{1-h_{tt}}\right)EW_T^2 V_T$$

$$-\frac{6}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + 2w_t^4\sigma_t^2 + w_t^6\sigma_t^2\right)}{1-h_{tt}}\right) - \frac{6}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4\sigma_t^2 + 2w_t^6\sigma_t^2 + w_t^8\sigma_t^2\right)}{1-h_{tt}}\right)$$

$$+\frac{1}{T^3\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2 + w_t^8 + 3w_t^4 + 3w_t^6\right)}{\left(1-h_{tt}\right)^2}\right)EV_T$$

$$+\frac{2}{T^3\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^3 + w_t^9 + 3w_t^5 + 3w_t^7\right)}{\left(1-h_{tt}\right)^2}\right)EW_T V_T$$

$$+\frac{1}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + w_t^8\sigma_t^2 + 3w_t^4\sigma_t^2 + 3w_t^6\sigma_t^2\right)}{\left(1-h_{tt}\right)^2}\right)$$

$$+\frac{1}{T^3\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4 + w_t^{10} + 3w_t^6 + 3w_t^8\right)}{\left(1-h_{tt}\right)^2}\right)EW_T^2 V_T$$

$$-\frac{2}{T^3\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + w_t^8\sigma_t^2 + 3w_t^4\sigma_t^2 + 3w_t^6\sigma_t^2\right)}{\left(1-h_{tt}\right)^2}\right)$$

$$-\frac{2}{T^3\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4\sigma_t^2 + w_t^{10}\sigma_t^2 + 3w_t^6\sigma_t^2 + 3w_t^8\sigma_t^2\right)}{\left(1-h_{tt}\right)^2}\right)$$

This completes the proof. ■

5.6.2: Second Order Asymptotic Bias of HC3

Under the same assumptions about the asymptotic behavior of the regressors and variances as in case of HC2, (See ***EQ: 5-11*** and ***EQ: 5-19***), we can get an explicit expression for the second order asymptotic bias of HC3.

Theorem 5-9: Under assumptions (See ***EQ: 5-11*** and ***EQ: 5-19***), the second order asymptotic bias of HC3 is given below:

EQ: 5-24 $$ASOB(HC3)=\frac{1}{EZ_\infty^2}\left\{EV_\infty+2\left(EW_\infty^3\right)EW_\infty V_\infty+\left(EW_\infty^4\right)EW_\infty^2V_\infty\right\}$$

Proof: Note, we have established that

$$T^2B_{22}(A_3)=\frac{1}{EZ_T^2}\left\{EV_T+2\left(EW_T^3\right)EW_TV_T+\left(EW_T^4\right)EW_T^2V_T\right\}+R_{2T}$$

Where, R_{2T} is defined in **EQ: 5-23** above.

To prove the theorem, it is sufficient to show that $\lim_{T\to\infty}(R_{2T})=0$.

The limit of the remaining part exists and equals the expression in the theorem by assumptions in ***EQ: 5-11***.

Note that, we can write R_{2T} as, $R_{2T}=\frac{1}{T}(G_T)$, where G_T is given by:

EQ: 5-25

$$G_T = \frac{2}{EZ_T^2}\left[EV_T + \left(EW_T^4\right)EV_T + 2\left(EW_T^3 + EW_T^5\right)EW_TV_T\right]$$

$$+\frac{2}{EZ_T^2}\left[\left(EW_T^4 + EW_T^6 - 2\right)EW_T^2V_T - 4EW_T^4V_T - 2EW_T^6V_T\right]$$

$$+\frac{3}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2 + 2w_t^4 + w_t^6\right)}{1-h_{tt}}\right)EV_T + \frac{6}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^3 + 2w_t^5 + w_t^7\right)}{1-h_{tt}}\right)EW_TV_T$$

$$+\frac{3}{EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + 2w_t^4\sigma_t^2 + w_t^6\sigma_t^2\right)}{1-h_{tt}}\right) + \frac{3}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4 + 2w_t^6 + w_t^8\right)}{1-h_{tt}}\right)EW_T^2V_T$$

$$-\frac{6}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + 2w_t^4\sigma_t^2 + w_t^6\sigma_t^2\right)}{1-h_{tt}}\right) - \frac{6}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4\sigma_t^2 + 2w_t^6\sigma_t^2 + w_t^8\sigma_t^2\right)}{1-h_{tt}}\right)$$

$$+\frac{1}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2 + w_t^8 + 3w_t^4 + 3w_t^6\right)}{\left(1-h_{tt}\right)^2}\right)EV_T$$

$$+\frac{2}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^3 + w_t^9 + 3w_t^5 + 3w_t^7\right)}{\left(1-h_{tt}\right)^2}\right)EW_TV_T$$

$$+\frac{1}{T\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + w_t^8\sigma_t^2 + 3w_t^4\sigma_t^2 + 3w_t^6\sigma_t^2\right)}{\left(1-h_{tt}\right)^2}\right)$$

$$+\frac{1}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4 + w_t^{10} + 3w_t^6 + 3w_t^8\right)}{\left(1-h_{tt}\right)^2}\right)EW_T^2V_T$$

$$-\frac{2}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^2\sigma_t^2 + w_t^8\sigma_t^2 + 3w_t^4\sigma_t^2 + 3w_t^6\sigma_t^2\right)}{\left(1-h_{tt}\right)^2}\right)$$

$$-\frac{2}{T^2\,EZ_T^2}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\left(w_t^4\sigma_t^2 + w_t^{10}\sigma_t^2 + 3w_t^6\sigma_t^2 + 3w_t^8\sigma_t^2\right)}{\left(1-h_{tt}\right)^2}\right)$$

Under the assumption that the design matrix is balanced (See ***EQ: 5-19***), i.e. $\lim_{T\to\infty}\max_{1\le t\le T}(h_{tt})=0$.

So for 'T' large enough $h_{tt}\le\frac{1}{2},\forall t$ and $\frac{1}{(1-h_{tt})^2}<4$.

Note that we can write, $\left|\frac{1}{T}\sum_{t=1}^{T}\frac{w_t^\alpha}{(1-h_{tt})^2}\right|\le\frac{1}{T}\sum_{t=1}^{T}\left|\frac{w_t^\alpha}{(1-h_{tt})^2}\right|\le 4\frac{1}{T}\sum_{t=1}^{T}\left|w_t^\alpha\right|$

Since variances are bounded, i.e. $\sigma_t^2\le U,\ \forall t$

So, we can write

$$\left|\frac{1}{T}\sum_{t=1}^{T}\frac{w_t^\alpha\sigma_t^2}{(1-h_{tt})^2}\right|\le\frac{1}{T}\sum_{t=1}^{T}\left|\frac{w_t^\alpha\sigma_t^2}{(1-h_{tt})^2}\right|\le 4U\frac{1}{T}\sum_{t=1}^{T}\left|w_t^\alpha\right|=4U\,E\left|W_T^\alpha\right|.$$

This means that each of the terms in the G_T above is separately bounded and is of order $\mathcal{O}(1)$. Hence the sum of all terms (i.e. G_T) is also bounded and is of order $\mathcal{O}(1)$, so $R_{2T}=\frac{1}{T}(G_T)$ goes to zero as 'T' approaches infinity, as claimed. ■

Chapter 6: ASYMPTOTIC MAXIMUM BIASES OF HCCMEs

6.1: Introduction

This chapter provides results regarding second order asymptotic maximum biases of all HCCMEs. Since we have the analytical expressions for the biases in previous chapter, this allows us to calculate the configuration of variances which leads to the maximum bias. We characterize this least favorable form of heteroskedasticity and the associated maximum bias.

To state the results of this chapter in a self-contained manner, let's recapitulate our basic model and related definitions.

We start out with a heteroskedastic regression model with centered regressors, without loss of generality: $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$

Where, $E(\varepsilon) = 0$, $Cov(\varepsilon) = \Sigma$, where $\Sigma = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_T^2)$. As before, (V_T, Z_T) represents artificial random variable which takes one of 'T' possible values

$\left(\sigma_t^2, x_t\right)$ for $t = 1,2,\cdots,T$ with equal probabilities $\frac{1}{T}$. Let W_T be the standardized regressors, $W_T = \frac{x_t - \bar{x}}{\sqrt{Var\left(Z_T\right)}}$, where, $\bar{x} = (1/T)\sum x_t$, and $\text{Var}\left(Z_T\right) = (1/T)\sum (x_t - \bar{x})^2$ as usual. Note that $EW_T = 0$ and $Var\left(W_T\right) = EW_T^2 = 1$.

The following subsections provide results for maximum asymptotic bias of all HCCMEs.

6.2: Asymptotic Maximum Bias

In this section, our aim is to find the *worst possible configuration of heteroskedasticity*. For this we provide the results of asymptotic maximum bias of all HCCMEs including (HC0, HC1, HC2 and HC3).

Note that, if the variances are unconstrained, then the maximum bias can be arbitrarily large, so we assume some upper bound on the variances: $\forall t : \sigma_t^2 \leq U$. Under this assumption, we proceed to derive the largest possible second order bias for all HCCMEs under consideration.

In addition, we note that these all estimators have second order asymptotic bias and these biases are given below:

EQ: 6-1 $\mathcal{ASOB}(HC0)=\frac{1}{EZ_{\infty}^{2}}\left\{EV_{\infty}+2\left(EW_{\infty}^{3}\right)EW_{\infty}V_{\infty}+\left(EW_{\infty}^{4}-2\right)EW_{\infty}^{2}V_{\infty}-2EW_{\infty}^{4}V_{\infty}\right\}$

EQ: 6-2 $\mathcal{ASOB}(HC1)=\frac{1}{EZ_{\infty}^{2}}\left\{EV_{\infty}+2\left(EW_{\infty}^{3}\right)EW_{\infty}V_{\infty}+\left(EW_{\infty}^{4}\right)EW_{\infty}^{2}V_{\infty}-2EW_{\infty}^{4}V_{\infty}\right\}$

EQ: 6-3 $\mathcal{ASOB}(HC2)=\frac{1}{EZ_{\infty}^{2}}\left\{EV_{\infty}+2\left(EW_{\infty}^{3}\right)EW_{\infty}V_{\infty}+\left(EW_{\infty}^{4}-1\right)EW_{\infty}^{2}V_{\infty}-EW_{\infty}^{4}V_{\infty}\right\}$

EQ: 6-4 $\mathcal{ASOB}(HC3)=\frac{1}{EZ_{\infty}^{2}}\left\{EV_{\infty}+2\left(EW_{\infty}^{3}\right)EW_{\infty}V_{\infty}+\left(EW_{\infty}^{4}\right)EW_{\infty}^{2}V_{\infty}\right\}$

In addition, note that throughout in the current chapter, asymptotic maximum bias means the second order asymptotic maximum bias.

The maximum possible positive and negative biases occur at different least favorable sequences of variances. We give the expression for both in our preliminary result below.

Theorem 6-1: Let $\mathcal{B}_i^+$ and $\mathcal{B}_i^-$, i=0,1,2,3 be the maximum possible second order asymptotic positive and negative biases of HCCMEs which are given by:

EQ: 6-5

$$\mathcal{B}_i^+ = \max_{\sigma_i^2 \le U}\left[\mathcal{ASOB}(HCi)\right] = \lim_{T\to\infty}\left\{\frac{1}{TVar(Z_T)}\sum_{t=1}^{T}\max\left(p_i(w_t),0\right)U\right\}, i=0,1,2,3.$$

EQ: 6-6

$$\mathcal{B}_i^- = \min_{\sigma_i^2 \le U}\left[\mathcal{ASOB}(HCi)\right] = \lim_{T\to\infty}\left\{\frac{1}{TVar(Z_T)}\sum_{t=1}^{T}\min\left(p_i(w_t),0\right)U\right\}, i=0,1,2,3.$$

Where polynomials p_i (i=0,1,2,3) are given below:

EQ: 6-7 $$p_0(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4 - 2\right)w_t^2 - 2w_t^4$$

EQ: 6-8 $$p_1(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4\right)w_t^2 - 2w_t^4$$

EQ: 6-9 $$p_2(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4 - 1\right)w_t^2 - w_t^4$$

EQ: 6-10 $$p_3(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4\right)w_t^2$$

Proof: Similar to **Theorem 4-2** above.■

We now try to obtain more explicit characterizations of these maxima and minima. Note that in **Chapter 4:**, we derived a class of estimators which considers HC0, HC1 as special cases and we derived minimax estimator and that particular minimax value of Minimax estimator 'a' was found to be independent of exact sequence of regressors and it only depends on the sample size, T, and the kurtosis of regressor (K). Here in this chapter, we have formulae of two estimators under study (HC2 and HC3) which do not fall into that particular class of estimators. Further, we do not have such minimax value which exists in this class of estimators, so the bias of all HCCMEs proposed in **Chapter 5:** may depend on the exact sequence of regressors as well as the sample size. We are presenting the maximum bias of all HCCMEs for a particular sequence of regressors for which computations are easy. In particular, we consider two cases, one with symmetric regressors and the other one as asymmetric case. The issue of finding a minimax estimator which covers all four estimators (HC0 to HC3) requires more work and we leave it as an open problem for future research.

First we consider the case where the regressors are symmetric. The case of asymmetric regressors will follow this case.

6.2.1: Asymptotic Maximum Bias with Symmetric Regressors

Under the assumption that the regressors are symmetric, we derive analytical formulae for the approximate large sample maximum biases $\mathcal{B}_i = \max\left(\mathcal{B}_i^+, -\mathcal{B}_i^-\right), i = 0,1,2,3$ for all HCCMEs. In this case the average value of regressors (EW_T) and the skewness (EW_T^3) will be zero, while the kurtosis (EW_T^4) will vary for different samples. This simplifies the polynomials to:

$$p_0(w_t) = 1 + (EW_T^4 - 2)w_t^2 - 2w_t^4$$

$$p_1(w_t) = 1 + (EW_T^4)w_t^2 - 2w_t^4$$

$$p_2(w_t) = 1 + (EW_T^4 - 1)w_t^2 - w_t^4$$

$$p_3(w_t) = 1 + (EW_T^4)w_t^2$$

We choose sample size T and a constant M [$M>1$], such that $k = T/(2M^2)$ is an integer, and consider the sequence of regressors $x_1, \cdots, x_T$ such that $x_1 = x_2 = \cdots = x_k = -M$, $x_{k+1} = \cdots = x_{T-k} = 0$, and $x_{T-k+1} = \cdots = x_T = +M$. As before, letting Z_T be the random variable such that $Z_T = x_t$ with probability 1/T, we can easily check that $EZ_T = 0$; $EZ_T^2 = 1$, and $EZ_T^4 = M^2$. That is Z_T is centered and standardized and has kurtosis $K = EZ_T^4 = M^2$. Note that we assume M>1, this means that we are constraining the kurtosis to be bounded below by '1'. Because Z_T is standardized, $W_T = Z_T/\text{Std}(Z_T) = Z_T$ and the standardized regressors are just $w_t = x_t$. Noting that $EW_T^3 = 0$ and that the kurtosis is $EW_T^4 = M^2$.

Under this setup, the analytical expressions for the maximum asymptotic positive and negative bias functions and the maximum bias, $\mathcal{B}_i = \max(\mathcal{B}_i^+, -\mathcal{B}_i^-), i = 0,1,2,3$ are provided in the following subsections.

Note that, the results of asymptotic maximum bias of HC0 and HC1 in case of symmetric regressors are exactly the same as we have derived earlier in **Chapter 4:**. But in **Chapter 4:**, we were interested in finding the maximum bias of a class of estimators and hence the minimax estimator. Since HC0 and HC1 fall into that particular class of estimators. So their maximum biases were also calculated. But here in this chapter, we are presenting the results of both HC0 and HC1 along with HC2 and HC3 which do not fall into that particular class of estimators and hence there is no minimax value of 'a' which remains same of all regressor sequences. So for completeness, the results for the asymptotic maximum bias of HC0 and HC1 are also presented here.

In addition, we provide the results of Minimax estimator as well to compare it with HC2 and HC3.

6.2.1.1: Asymptotic Maximum Bias of Eicker-White (HC0) Estimator

This subsection provides the asymptotic maximum bias of Eicker-White (HC0) estimator.

Theorem 6-2: Second order asymptotic maximum bias of HC0 estimator, when the regressors are symmetric, is given by:

EQ: 6-11 $$B_0 = 2 + M^2 - \frac{1}{M^2}$$

Proof: When regressors are symmetric, polynomial corresponding to HC0 is:

$$p_0(w_t) = 1 + \left(EW_T^4 - 2\right)w_t^2 - 2w_t^4$$

In order to find maximum positive and negative bias functions, we have to evaluate $\max\left(p_0(w_t), 0\right)$ which involve finding signs of polynomial $p_0(w_t)$ at w_t=+M, 0 and –M.

Note that, $p_0(0) = 1 > 0$ and $p_0(\pm M) = 1 - 2M^2 - M^4 < 0 \qquad \left(\forall\ M^2 > 1\right)$

So we can write the asymptotic maximum positive and negative bias functions for HC0 as follows:

$$\mathcal{B}_0^+ = \max_{\sigma_t^2 \le U}\left\{\lim_{T\to\infty}\left\{T^2 Var(Z_T) B_{22}(A_0)\right\}\right\} = \left\{p_0(0)\left(1 - \frac{1}{M^2}\right)\right\}U = \left(1 - \frac{1}{M^2}\right)U$$

$$\mathcal{B}_0^- = \min_{\sigma_t^2 \le U}\left\{\lim_{T\to\infty}\left\{T^2 Var(Z_T) B_{22}(A_0)\right\}\right\} = \left\{p_0(M)\left(\frac{1}{M^2}\right)\right\}U = \left\{\left(1 - 2M^2 - M^4\right)\left(\frac{1}{M^2}\right)\right\}U$$

Note that, both the maximum bias functions $\mathcal{B}_0^+$ and $\mathcal{B}_0^-$ are proportional to U, the upper bound on the variances. To get a reasonable performance measure which is invariant to this arbitrary upper bound, it seems reasonable to divide by this factor. Reversing the sign of the negative bias to get the magnitude, we obtain:

$$\mathcal{B}_0^+ = \max_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2 Var(Z_T) B_{22}(A_0)}{U}\right\}\right] = 1 - \frac{1}{M^2}$$

$$-\mathcal{B}_0^- = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2 Var(Z_T) B_{22}(A_0)}{U}\right\}\right] = \left(2 + M^2\right) - \frac{1}{M^2} = \left(1 - \frac{1}{M^2}\right) + \left(1 + M^2\right)$$

Taking maximum of the two biases leads to the required result. ■

6.2.1.2: Asymptotic Maximum Bias of Hinkley (HC1) Estimator

Here we provide the results of asymptotic maximum bias of Hinkley (HC1) estimator.

Theorem 6-3: Second order asymptotic maximum bias of HC1 estimator when the regressors are symmetric, is given by:

EQ: 6-12 $$\mathcal{B}_1 = M^2 - \frac{1}{M^2}$$

Proof: Polynomial corresponding to HC1, for symmetric regressors, is given by:

$$p_1(w_t) = 1 + (M^2)w_t^2 - 2w_t^4$$

Working in the same lines as in the proof of the asymptotic maximum bias of HC0, we have to evaluate the above polynomial at different values of W_T's, i.e. at $W_T = 0$, +M and –M:

$$p_1(0) = 1 > 0 \text{ and } p_1(\pm M) = 1 - M^4 < 0 \qquad (\forall M^2 > 1)$$

This leads us to write the asymptotic maximum positive and negative bias functions as follows:

$$\mathcal{B}_1^+ = \max_{\sigma_t^2 \le U}\left\{\lim_{T\to\infty}\left\{T^2 Var(Z_T)\, B_{22}(A_1)\right\}\right\} = \left\{p_1(0)\left(1 - \frac{1}{M^2}\right)\right\}U = \left(1 - \frac{1}{M^2}\right)U$$

$$\mathcal{B}_1^- = \min_{\sigma_i^2 \le U}\left\{\lim_{T\to\infty}\left\{T^2 Var(Z_T)B_{22}(A_1)\right\}\right\} = \left\{p_1(M)\left(\frac{1}{M^2}\right)\right\}U = \left\{\left(1-M^4\right)\left(\frac{1}{M^2}\right)\right\}U$$

Dividing both the bias functions with U and reversing the sign of negative bias function, we have:

$$\mathcal{B}_1^+ = \max_{\sigma_i^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)B_{22}(A_1)}{U}\right\}\right] = 1-\frac{1}{M^2}$$

$$-\mathcal{B}_1^- = \min_{\sigma_i^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)B_{22}(A_1)}{U}\right\}\right] = M^2-\frac{1}{M^2} = \left(1-\frac{1}{M^2}\right)+\left(M^2-1\right)$$

Taking the maximum of two bias functions, leads to required result.■

6.2.1.3: Asymptotic Maximum Bias of Horn, Horn and Duncan (HC2) Estimator

In this subsection, we collect results regarding asymptotic maximum bias of Horn, Horn and Duncan (HC2) estimator.

Theorem 6-4: Second order asymptotic maximum bias of HC2 estimator, when the regressors are symmetric, is given by:

EQ: 6-13 $$\mathcal{B}_2 = 1-\frac{1}{M^2}$$

Proof: When regressors are symmetric, polynomial corresponding to HC2 is:

$$p_2(w_t) = 1 + (M^2 - 1)w_t^2 - w_t^4$$

As before, $p_2(0) = 1 > 0$ and $p_2(\pm M) = 1 - M^2 < 1 \qquad (\forall M^2 > 1)$

Again value of polynomial at '0' is '1', which is always positive, while, value of polynomial at +M and –M is negative for all M^2 (kurtosis) greater than 1. So; asymptotic maximum positive and negative bias functions for HC2 can be written in a compact form as follows:

$$\mathcal{B}_2^+ = \max_{\sigma_t^2 \le U}\left\{\lim_{T\to\infty}\left\{T^2 Var(Z_T)\, B_{22}(A_2)\right\}\right\} = \left\{p_2(0)\left(1-\frac{1}{M^2}\right)\right\}U = \left(1-\frac{1}{M^2}\right)U$$

$$\mathcal{B}_2^- = \min_{\sigma_t^2 \le U}\left\{\lim_{T\to\infty}\left\{T^2 Var(Z_T)\, B_{22}(A_2)\right\}\right\} = \left\{p_2(M)\left(\frac{1}{M^2}\right)\right\}U = \left\{(1-M^2)\left(\frac{1}{M^2}\right)\right\}U$$

Dividing by the U (the upper bound to variances), and reversing the sign of the negative bias function, we have:

$$\mathcal{B}_2^+ = \max_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)\, B_{22}(A_2)}{U}\right\}\right] = 1 - \frac{1}{M^2}$$

$$-\mathcal{B}_2^- = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)\, B_{22}(A_2)}{U}\right\}\right] = 1 - \frac{1}{M^2}$$

Note that, both are exactly the same, so maximum bias is the bias of any one of them. This leads to desired result. ■

6.2.1.4: Asymptotic Maximum Bias of Mackinnon and White (HC3) Estimator

In this portion, we present results of the asymptotic maximum bias of Mackinnon and White (HC3) estimator.

Theorem 6-5: Second order asymptotic maximum bias of HC3 estimator, when the regressors are symmetric, is given by:

EQ: 6-14 $$\mathcal{B}_3 = M^2 + \frac{1}{M^2}$$

Proof: When regressors are symmetric, polynomial corresponding to HC3 is:

$$p_3(w_t) = 1 + (M^2) w_t^2$$

As before, we can see that, $p_3(0) = 1 > 0$, $p_3(\pm M) = 1 + M^4 > 0$ $\quad (\forall M^2 > 1)$

Now, we can write the asymptotic maximum positive and negative bias functions for HC3 as follows:

$$\mathcal{B}_3^+ = \max_{\sigma_t^2 \le U} \left\{ \lim_{T\to\infty} \left\{ T^2 Var(Z_T) B_{22}(A_3) \right\} \right\} = \left\{ p_3(0)\left(1 - \frac{1}{M^2}\right) + p_3(M)\left(\frac{1}{M^2}\right) \right\} U$$

$$= \left\{ \left(1 - \frac{1}{M^2}\right) + \left(1 + M^4\right)\left(\frac{1}{M^2}\right) \right\} U$$

$$\mathcal{B}_3^- = \min_{\sigma_t^2 \le U}\left\{\lim_{T\to\infty}\left\{T^2 Var(Z_T) B_{22}(A_3)\right\}\right\} = 0$$

Re-scaling the bias functions by U and reversing the sign of negative bias function, we have:

$$\mathcal{B}_3^+ = \max_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T) B_{22}(A_3)}{U}\right\}\right] = \frac{1}{M^2} + M^2$$

$$-\mathcal{B}_3^- = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T) B_{22}(A_3)}{U}\right\}\right] = 0$$

The maximum of two is $\mathcal{B}_3 = \max\left(\mathcal{B}_3^+, -\mathcal{B}_3^-\right) = M^2 + \frac{1}{M^2}$, which is the required result.■

6.2.1.5: Asymptotic Maximum Bias of Minimax Estimator

In this portion, we present results of the asymptotic maximum bias of Minimax estimator.

Theorem 6-6: Second order asymptotic maximum bias of Minimax estimator, when the regressors are symmetric, is given by:

EQ: 6-15 $$\mathcal{B}_{\text{Minimax}} = 1 - \frac{1}{M^2}$$

Proof: The proof is same as that of **Theorem 4-4** in **Chapter 4:** above.■

6.2.1.6: Asymptotic Comparison of HCCMEs (Symmetric Regressors Case)

Using the results obtained, we can analytically compare the relative performance of Eicker-White, Hinkley, Horn, Horn & Duncan, Mackinnon and White and Minimax estimators in terms of asymptotic bias when regressors follow symmetric distribution. Below we present the formulae derived in previous sub-sections in a comparable manner.

Second order asymptotic maximum bias of HC0 is:

$$\mathcal{B}_0 = M^2 + 2 - \frac{1}{M^2} = \left(1 - \frac{1}{M^2}\right) + \left(1 + M^2\right)$$

Second order asymptotic maximum bias of HC1 is:

$$\mathcal{B}_1 = M^2 - \frac{1}{M^2} = \left(1 - \frac{1}{M^2}\right) + \left(M^2 - 1\right)$$

Second order asymptotic maximum bias of HC2 is:

$$\mathcal{B}_2 = 1 - \frac{1}{M^2}$$

Second order asymptotic maximum bias of HC3 is:

$$\mathcal{B}_3 = M^2 + \frac{1}{M^2} = -\left(1 - \frac{1}{M^2}\right) + \left(1 + M^2\right)$$

Second order asymptotic maximum bias of Minimax is:

$$\mathcal{B}_{\text{Minimax}} \doteq 1-\frac{1}{M^2}$$

From the bias functions provided above, we can see that the second order asymptotic maximum bias of HC0, HC1 and HC3 increases with the kurtosis (M^2) and differ marginally. Second order asymptotic maximum bias of Minimax and HC2 is exactly same and goes to 1 with an increase in kurtosis (M^2), i.e. $\left(\mathcal{B}_{\text{Minimax}} = \mathcal{B}_2 = 1-\frac{1}{M^2}\right) \to 1, \ \left(As\, M^2 \to \infty\right)$. This is because the $1/M^2$ term goes to zero as M^2 goes to infinity. So we conclude that bias of HC2 is bounded above by '1'. The bias of HC0 is largest of all making it the least favorable estimator. The bias of HC3 is the 2nd least favorable while HC1 is 2nd favorable estimator. Minimax along with HC2 is the clear winner among the five.

6.2.2: Asymptotic Maximum Bias with Asymmetric Regressors

In this subsection, we consider the case of asymmetric regressors and derive analytical formulae for the approximate large sample second order maximum biases $\mathcal{B}_i = \max\left(\mathcal{B}_i^+, -\mathcal{B}_i^-\right), i = 0,1,2,3$. As in case of symmetric regressors, the average value of regressors is zero by construction but due to asymmetry of regressors, the skewness will be different from zero and both skewness and kurtosis will vary for different samples. Under this setup, the polynomials simplifies to:

$$p_0(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4 - 2\right)w_t^2 - 2w_t^4$$

$$p_1(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4\right)w_t^2 - 2w_t^4$$

$$p_2(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4 - 1\right)w_t^2 - w_t^4$$

$$p_3(w_t) = 1 + 2\left(EW_T^3\right)w_t + \left(EW_T^4\right)w_t^2$$

As before, evaluation of the expressions for finding maximum positive and negative bias functions, we have to evaluate $\max\left(p_i(w_t), 0\right)$, $i = 0,1,2,3$ which involve finding signs of polynomials $p_i(w_t)$ at w_t=+M, 0 and –M.

Following the same lines as in case of symmetric regressor, here again we pick a particular sequence of regressor for which the computations are easy.

We choose a sample of size 'T' such that $k = \dfrac{T}{MN\left(M+N\right)^2}$, [M and N are some positive scalars] is an integer, and consider the sequence of regressors $x_1, \cdots, x_T$ such that $x_1 = x_2 = \cdots = x_k = M$, $x_{k+1} = \cdots = x_{T-k} = 0$, and $x_{T-k+1} = \cdots = x_T = -N$.

As before, letting W_T be the standardized regressors with $EW_T=0$, $EW_T^2 = 1$, $EW_T^3 = M - N$ and $EW_T^4 = M^2 + N^2 - MN$.

Using these values in above polynomials, we have;

$$
\begin{aligned}
p_0\left(w_t\right) &= 1 + 2\left(M-N\right)w_t + \left\{\left(M^2+N^2-MN\right)-2\right\}w_t^2 - 2w_t^4 \\
p_1\left(w_t\right) &= 1 + 2\left(M-N\right)w_t + \left(M^2+N^2-MN\right)w_t^2 - 2w_t^4 \\
p_2\left(w_t\right) &= 1 + 2\left(M-N\right)w_t + \left\{\left(M^2+N^2-MN\right)-1\right\}w_t^2 - w_t^4 \\
p_3\left(w_t\right) &= 1 + 2\left(M-N\right)w_t + \left(M^2+N^2-MN\right)w_t^2
\end{aligned}
$$

As before, in order to find the second order asymptotic maximum positive and negative bias functions, we have to evaluate the corresponding polynomials at $W_T=0$, +M and –N.

The value of polynomials at $w_t=0$ is 1, i.e; $p_i\left(0\right) = 1$, $\quad i = 0,1,2,3$.

Note that; value of polynomial at '0' is 1, which is always positive for any combination of M and –N.

For finding the exact signs of polynomials at +M and –N, Let, $\frac{N}{M} = \rho$, $\quad or, \quad N = \rho M \quad (0 < N < M \& 0 < \rho < 1)$

With this notation, skewness and kurtosis measures become:

$EW_T^3 = M(1-\rho)$ and $EW_T^4 = (\rho^2 - \rho + 1)M^2$.

Putting value of N in above polynomials, we get:

$$p_0(+M) = 1 - 2\rho M^2 + (\rho^2 - \rho - 1)M^4$$

$$p_0(-N) = 1 - 2\rho M^2 + \rho^2(1 - \rho - \rho^2)M^4$$

$$p_1(+M) = 1 + 2(1-\rho)M^2 + (\rho^2 - \rho - 1)M^4$$

$$p_1(-N) = 1 - 2\rho(1-\rho)M^2 + \rho^2(1 - \rho - \rho^2)M^4$$

$$p_2(+M) = 1 + (1-2\rho)M^2 + \rho(\rho - 1)M^4$$

$$p_2(-N) = 1 - \rho(2-\rho)M^2 + \rho^2(1-\rho)M^4$$

$$p_3(+M) = 1 + 2(1-\rho)M^2 + (1 - \rho + \rho^2)M^4$$

$$p_3(-N) = 1 - 2\rho(1-\rho)M^2 + \rho^2(1 - \rho + \rho^2)M^4$$

Note that, these polynomials are quadratic in M^2, and so we can easily find their signs analytically. This will allow us to find the second order asymptotic maximum positive and negative bias functions of all HCCMEs, these are provided in the following subsections.

In addition, we present here the second order asymptotic maximum bias of Minimax estimator as well to show its performance against the HC2 and HC3 in case of asymmetric regressors.

Note here that the second order asymptotic maximum bias of Minimax estimator depends on the exact sequence of regressors etc. but the optimal value of 'a' is independent of exact regressors sequence but only depends on the kurtosis of regressors and the sample size 'T'. So here it is important to see how second order asymptotic maximum bias of Minimax estimator against HC2, HC3 as well.

6.2.2.1: Asymptotic Maximum Bias of Eicker-White (HC0) Estimator

This subsection provides the second order asymptotic maximum bias of Eicker-White (HC0) estimator.

Theorem 6-7: Second order asymptotic maximum bias of HC0 estimator, when the regressors are asymmetric, is given by:

EQ: 6-16 $$\mathcal{B}_0 = 2 - \frac{1}{M^2(1+\rho)} - \frac{M^2}{1+\rho} - \frac{2}{1+\rho} + 2M^2 - \rho M^2, \quad 0 < \rho < \frac{(\sqrt{5}-1)}{2}$$

and

EQ: 6-17 $$\mathcal{B}_0 = (\rho^2 - \rho + 1)M^2 + 2 - \frac{1}{\rho M^2}, \quad \frac{(\sqrt{5}-1)}{2} < \rho < 1$$

Proof: When regressors are asymmetric, polynomial corresponding to HC0 is:

$$p_0(w_t) = 1 + 2(EW_T^3)w_t + (EW_T^4 - 2)w_t^2 - 2w_t^4 \quad .$$

Using the values of $EW_T^3 = M - N$ and $EW_T^4 = M^2 + N^2 - MN$ in above polynomial, we have;

$$p_0(w_t) = 1 + 2(M - N)w_t + \{(M^2 + N^2 - MN) - 2\}w_t^2 - 2w_t^4$$

Evaluating the polynomial at w=0, +M and –N, we have:

$$p_0(1) = 1$$

$$p_0(+M) = 1 + 2(M - N)M + \{(M^2 + N^2 - MN) - 2\}M^2 - 2M^4$$

$$p_0(-N) = 1 - 2(M-N)N + \{(M^2 + N^2 - MN) - 2\}N^2 - 2N^4$$

Since $N = \rho M \quad (0 < \rho < 1)$, so we have:

$$p_0(+M) = 1 - 2\alpha M^2 + (\alpha^2 - \alpha - 1)M^4$$

$$p_0(-N) = 1 - 2\alpha M^2 + \alpha^2(1 - \alpha - \alpha^2)M^4$$

Note that, these polynomials are quadratic in M^2, so, we can find the signs of the polynomials by evaluating the roots of the same.

Note that, $p_0(0) = 1 > 0$, when $0 < \rho < 1$

When, $0 < \rho < 1$, $p_0(+M) < 0, \quad when \quad M^2 \geq r_{0,M}$

Where, $r_{0,M}$ is the positive root of the polynomial $p_0(+M)$ and is given by:

$$r_{0,M} = \frac{\alpha - \sqrt{\alpha+1}}{\alpha^2 - \alpha - 1}$$

When, $0 < \rho < \frac{(\sqrt{5}-1)}{2}$, $p_0(-N) > 0, \quad when \quad M^2 > r_{0,-N}$

When, $\frac{(\sqrt{5}-1)}{2} < \rho < 1$, $p_0(-N) < 0, \quad when \quad M^2 > r_{0,-N}$

Where, $r_{0,-N}$ is the positive root of the polynomial $p_0(-N)$ and is given by:

$$r_{0,-N} = \frac{1 - \sqrt{\rho(1+\rho)}}{\rho(1-\rho-\rho^2)}$$

Consider two cases,

Case 1: When $0 < \rho < \frac{\left(\sqrt{5}-1\right)}{2}$

The second order asymptotic maximum positive bias of HC0 is:

$$\mathcal{B}_0^+ = \max_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\,Var\left(Z_T\right)B_{22}\left(A_0\right)}{U}\right\}\right] = p_0\left(1\right)\left(1-\frac{1}{\rho M^2}\right)+p_0\left(-N\right)\left(\frac{1}{\rho M^2\left(1+\rho\right)}\right)$$

Replacing the values of polynomials and simplifying, we get:

$$\mathcal{B}_0^+ = 1-\frac{1}{M^2\left(1+\rho\right)}-\frac{M^2}{1+\rho}-\frac{2}{1+\rho}+M^2-\rho^2 M^2$$

The second order asymptotic maximum negative bias of HC0

$$\mathcal{B}_0^- = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\,Var\left(Z_T\right)B_{22}\left(A_0\right)}{U}\right\}\right] = p_0\left(M\right)\left(\frac{1}{M^2\left(1+\rho\right)}\right)$$

Replacing the value of polynomial and simplifying, we get:

$$\mathcal{B}_0^- = -2+\frac{1}{M^2\left(1+\rho\right)}+\frac{M^2}{1+\rho}+\frac{2}{1+\rho}-2M^2+\rho M^2$$

Reversing the sign of second order asymptotic maximum negative bias function, we get:

$$-\mathcal{B}_0^- = 2-\frac{1}{M^2\left(1+\rho\right)}-\frac{M^2}{1+\rho}-\frac{2}{1+\rho}+2M^2-\rho M^2$$

Case 2: When $\frac{(\sqrt{5}-1)}{2} < \rho < 1$

The second order asymptotic maximum positive bias of HC0 is given by:

$$\mathcal{B}_0^+ = \max_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)\, B_{22}(A_0)}{U}\right\}\right] = p_0(1)\left(1-\frac{1}{\rho M^2}\right) = 1-\frac{1}{\rho M^2}$$

The second order asymptotic maximum negative bias of HC0 is given by:

$$\mathcal{B}_0^- = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)\, B_{22}(A_0)}{U}\right\}\right] = p_0(M)\left(\frac{1}{M^2(1+\rho)}\right) + p_0(-N)\left(\frac{1}{\rho M^2(1+\rho)}\right)$$

Replacing the values of polynomials and simplifying, we get:

$$\mathcal{B}_0^- = \frac{1}{\rho M^2} - \left(\rho^2-\rho+1\right)M^2 - 2$$

Reversing the sign of second order asymptotic maximum negative bias function, we get:

$$-\mathcal{B}_0^- = \left(\rho^2-\rho+1\right)M^2 + 2 - \frac{1}{\rho M^2}$$

Overall second order asymptotic maximum bias of HC0 is given by:

When, $0 < \rho < \frac{(\sqrt{5}-1)}{2}$

$$\mathcal{B}_0 = \max\left(\mathcal{B}_0^+, -\mathcal{B}_0^-\right) = -\mathcal{B}_0^-$$

Where,

$$\mathcal{B}_0^+ = 1 - \frac{1}{M^2(1+\rho)} - \frac{M^2}{1+\rho} - \frac{2}{1+\rho} + M^2 - \rho^2 M^2$$

$$-\mathcal{B}_0^- = 2 - \frac{1}{M^2(1+\rho)} - \frac{M^2}{1+\rho} - \frac{2}{1+\rho} + 2M^2 - \rho M^2$$

When $\frac{(\sqrt{5}-1)}{2} < \rho < 1$

$$\mathcal{B}_0 = \max\left(\mathcal{B}_0^+, -\mathcal{B}_0^-\right) = -\mathcal{B}_0^-$$

Where, $$\mathcal{B}_0^+ = p_0(1)\left(1 - \frac{1}{\rho M^2}\right) = 1 - \frac{1}{\rho M^2}$$

$$-\mathcal{B}_0^- = \left(\rho^2 - \rho + 1\right)M^2 + 2 - \frac{1}{\rho M^2}$$

This completes the proof. ■

6.2.2.2: Asymptotic Maximum Bias of Hinkley (HC1) Estimator

This subsection provides the second order asymptotic maximum bias of Hinkley (HC1) estimator.

Theorem 6-8: Second order asymptotic maximum bias of HC1 estimator when the regressors are asymmetric, is given by:

EQ: 6-18 $$\mathcal{B}_1 = 2 - \frac{1}{M^2(1+\rho)} - \frac{4}{1+\rho} - \frac{M^2}{1+\rho} - \rho M^2 + 2M^2, \qquad 0 < \rho < \frac{(\sqrt{5}-1)}{2}$$

and

EQ: 6-19 $$\mathcal{B}_1 = (\rho^2 - \rho + 1)M^2 - \frac{1}{\rho M^2}, \quad \frac{(\sqrt{5}-1)}{2} < \rho < 1$$

Proof: When regressors are asymmetric, polynomial corresponding to HC1 is:

$$p_1(w_t) = 1 + 2(EW_T^3)w_t + (EW_T^4)w_t^2 - 2w_t^4$$

Using the values of EW_T^3 and EW_T^4 in above polynomial, we have;

$$p_1(w_t) = 1 + 2(M-N)w_t + (M^2 + N^2 - MN)w_t^2 - 2w_t^4$$

Evaluating the polynomial at w=0, +M and –N, we have:

$$p_1(1) = 1$$

$$p_1(+M) = 1 + 2(M-N)M + (M^2 + N^2 - MN)M^2 - 2M^4$$

$$p_1(-N) = 1 - 2(M-N)N + \left(M^2 + N^2 - MN\right)N^2 - 2N^4$$

Since $N = \rho M \qquad (0 < \rho < 1)$, so we have:

$$p_1(+M) = 1 + 2(1-\rho)M^2 + \left(\rho^2 - \rho - 1\right)M^4$$

$$p_1(-N) = 1 - 2\rho(1-\rho)M^2 + \rho^2\left(1 - \rho - \rho^2\right)M^4$$

Note that, these polynomials are quadratic in M^2, so, we can find the signs of the polynomials by evaluating the roots of the same.

Note that, $p_1(0) = 1 > 0$, when $0 < \rho < 1$

When, $0 < \rho < 1$, $p_1(+M) < 0, \quad when \quad M^2 \geq r_{1,M}$

Where, $r_{1,M}$ is the positive root of the polynomial $p_1(+M)$ and is given by:

$$r_{1,M} = \frac{\rho - 1 - \sqrt{2-\rho}}{\rho^2 - \rho - 1}$$

When, $0 < \rho < \frac{\left(\sqrt{5}-1\right)}{2}$, $p_1(-N) > 0, \quad when \quad M^2 > r_{1,-N}$

When, $\frac{\left(\sqrt{5}-1\right)}{2} < \rho < 1$, $p_1(-N) < 0, \quad when \quad M^2 > r_{1,-N}$

Where, $r_{1,-N}$ is the positive root of the polynomial $p_1(-N)$ and is given by:

$$r_{1,-N} = \frac{1 - \rho - \sqrt{2\rho^2 - \rho}}{\rho\left(1 - \rho - \rho^2\right)}$$

Again, consider two cases arise,

Case 1: When $0 < \rho < \frac{(\sqrt{5}-1)}{2}$

The second order asymptotic maximum positive bias of HC1 is:

$$\mathcal{B}_1^+ = \max_{\sigma_i^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\,Var(Z_T)\,B_{22}(A_1)}{U}\right\}\right] = p_1(1)\left(1-\frac{1}{\rho M^2}\right)+p_1(-N)\left(\frac{1}{\rho M^2(1+\rho)}\right)$$

Replacing the values of polynomials and simplifying, we get:

$$\mathcal{B}_1^+ = 3-\frac{1}{M^2(1+\rho)}-\frac{4}{1+\rho}-\frac{M^2}{1+\rho}-\rho^2M^2+M^2$$

The second order asymptotic maximum negative bias of HC1

$$\mathcal{B}_1^- = \min_{\sigma_i^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\,Var(Z_T)\,B_{22}(A_1)}{U}\right\}\right] = p_1(M)\left(\frac{1}{M^2(1+\rho)}\right)$$

Replacing the value of polynomial and simplifying, we get:

$$\mathcal{B}_1^- = \frac{1}{M^2(1+\rho)}-2+\frac{4}{1+\rho}+\rho M^2-2M^2+\frac{M^2}{1+\rho}$$

Reversing the sign of second order asymptotic maximum negative bias function, we get:

$$-\mathcal{B}_1^- = 2-\frac{1}{M^2(1+\rho)}-\frac{4}{1+\rho}-\frac{M^2}{1+\rho}-\rho M^2+2M^2$$

Case 2: When $\frac{(\sqrt{5}-1)}{2} < \rho < 1$

The second order asymptotic maximum positive bias of HC1 is given by:

$$\mathcal{B}_1^+ = \max_{\sigma_i^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)\, B_{22}(A_1)}{U}\right\}\right] = p_1(1)\left(1-\frac{1}{\rho M^2}\right) = 1-\frac{1}{\rho M^2}$$

The second order asymptotic maximum negative bias of HC1 is given by:

$$\mathcal{B}_1^- = \min_{\sigma_i^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\; Var(Z_T)\, B_{22}(A_1)}{U}\right\}\right] = p_1(M)\left(\frac{1}{M^2(1+\rho)}\right) + p_1(-N)\left(\frac{1}{\rho M^2(1+\rho)}\right)$$

Replacing the values of polynomials and simplifying, we get:

$$\mathcal{B}_1^- = \frac{1}{\rho M^2} - (\rho^2-\rho+1)M^2$$

Reversing the sign of second order asymptotic maximum negative bias function, we get:

$$-\mathcal{B}_1^- = (\rho^2-\rho+1)M^2 - \frac{1}{\rho M^2}$$

Overall second order asymptotic maximum bias of HC1 is given by:

When $0 < \rho < \frac{(\sqrt{5}-1)}{2}$

$$\mathcal{B}_1 = \max\left(\mathcal{B}_1^+, -\mathcal{B}_1^-\right) = -\mathcal{B}_1^- = 2 - \frac{1}{M^2(1+\rho)} - \frac{4}{1+\rho} - \frac{M^2}{1+\rho} - \rho M^2 + 2M^2$$

Where,

$$\mathcal{B}_1^+ = 3 - \frac{1}{M^2(1+\rho)} - \frac{4}{1+\rho} - \frac{M^2}{1+\rho} - \rho^2 M^2 + M^2$$

$$-\mathcal{B}_1^- = 2 - \frac{1}{M^2(1+\rho)} - \frac{4}{1+\rho} - \frac{M^2}{1+\rho} - \rho M^2 + 2M^2$$

When $\frac{(\sqrt{5}-1)}{2} < \rho < 1$

$$\mathcal{B}_1 = \max\left(\mathcal{B}_1^+, -\mathcal{B}_1^-\right) = -\mathcal{B}_1^- = \left(\rho^2 - \rho + 1\right)M^2 - \frac{1}{\rho M^2}$$

Where,

$$\mathcal{B}_1^+ = p_1(1)\left(1 - \frac{1}{\rho M^2}\right) = 1 - \frac{1}{\rho M^2}$$

$$-\mathcal{B}_1^- = \left(\rho^2 - \rho + 1\right)M^2 - \frac{1}{\rho M^2}$$

This completes the proof. ■

6.2.2.3: Asymptotic Maximum Bias of Horn, Horn and Duncan (HC2) Estimator

This subsection provides the second order asymptotic maximum bias of Horn, Horn and Duncan (HC2) estimator.

Theorem 6-9: Second order asymptotic maximum bias of HC2 estimator, when the regressors are asymmetric, is given by:

EQ: 6-20 $$\mathcal{B}_2 = 2 - \frac{1}{M^2(1+\rho)} - \frac{3}{1+\rho} - \frac{2M^2}{1+\rho} + 2M^2 - \rho M^2, \quad 0 < \rho < 1$$

Proof: When regressors are asymmetric, polynomial corresponding to HC2 is:

$$p_2(w_t) = 1 + 2(EW_T^3)w_t + (EW_T^4 - 1)w_t^2 - w_t^4$$

Using the values of EW_T^3 and EW_T^4 in above polynomial, we have;

$$p_2(w_t) = 1 + 2(M-N)w_t + \{(M^2 + N^2 - MN) - 1\}w_t^2 - w_t^4$$

Evaluating the polynomial at w=0, +M and –N, we have:

$$p_2(1) = 1$$

$$p_2(+M) = 1 + 2(M-N)M + \{(M^2 + N^2 - MN) - 1\}M^2 - M^4$$

$$p_2(-N) = 1 - 2(M-N)N + \{(M^2 + N^2 - MN) - 1\}N^2 - N^4$$

Since $N = \rho M \qquad (0 < \rho < 1)$, so we have:

$$p_2(+M) = 1 + (1-2\rho)M^2 + \rho(\rho-1)M^4$$

$$p_2(-N) = 1 - \rho(2-\rho)M^2 + \rho^2(1-\rho)M^4$$

Note that, these polynomials are quadratic in M^2, so, we can find the signs of the polynomials by evaluating the roots of the same.

Note that, When, $0 < \rho < 1$,

$p_2(0) = 1 > 0$, $p_2(+M) < 0$, *when* $M^2 \geq r_{2,M}$ and $p_2(-N) > 0$, *when* $M^2 > r_{2,-N}$

Where, $r_{2,M}$ and $r_{1,-N}$ are the positive roots of the polynomials $p_2(+M)$ and $p_1(-N)$ respectively, and are given by: $r_{2,M} = r_{2,-N} = \dfrac{1}{\rho}$

The second order asymptotic maximum positive bias of HC2 is:

$$\mathcal{B}_2^+ = \max_{\sigma_t^2 \leq U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)B_{22}(A_2)}{U}\right\}\right] = p_2(1)\left(1 - \frac{1}{\rho M^2}\right) + p_2(-N)\left(\frac{1}{\rho M^2(1+\rho)}\right)$$

Replacing the values of polynomials and simplifying, we get:

$$\mathcal{B}_2^+ = 2 + \frac{1}{M^2(1+\rho)} - \frac{3}{1+\rho} - \frac{2M^2}{1+\rho} + 2M^2 - \rho M^2$$

The second order asymptotic maximum negative bias of HC2

$$\mathcal{B}_2^- = \min_{\sigma_i^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T)B_{22}(A_2)}{U}\right\}\right] = p_2(M)\left(\frac{1}{M^2(1+\rho)}\right)$$

Replacing the value of polynomial and simplifying, we get:

$$\mathcal{B}_2^- = \frac{1}{M^2(1+\rho)} + \frac{1-2\rho}{1+\rho} + \frac{\rho(\rho-1)}{1+\rho}M^2$$

Reversing the sign of second order asymptotic maximum negative bias function, we get:

$$-\mathcal{B}_2^- = 2 - \frac{1}{M^2(1+\rho)} - \frac{3}{1+\rho} - \frac{2M^2}{1+\rho} + 2M^2 - \rho M^2$$

Overall second order asymptotic maximum bias of HC2 is given by:

$$\mathcal{B}_2 = \max\left(\mathcal{B}_2^+, -\mathcal{B}_2^-\right) = 2 - \frac{1}{M^2(1+\rho)} - \frac{3}{1+\rho} - \frac{2M^2}{1+\rho} + 2M^2 - \rho M^2$$

Where,

$$\mathcal{B}_2^+ = 2 - \frac{1}{M^2(1+\rho)} - \frac{3}{1+\rho} - \frac{2M^2}{1+\rho} + 2M^2 - \rho M^2$$

$$-\mathcal{B}_2^- = 2 - \frac{1}{M^2(1+\rho)} - \frac{3}{1+\rho} - \frac{2M^2}{1+\rho} + 2M^2 - \rho M^2$$

This completes the proof. ■

6.2.2.4: Asymptotic Maximum Bias of Mackinnon and White (HC3) Estimator

This subsection provides the second order asymptotic maximum bias of Mackinnon and White (HC3) estimator for asymmetric regressors.

Theorem 6-10: Second order asymptotic maximum bias of HC3 estimator, when the regressors are asymmetric, is given by:

EQ: 6-21 $$\mathcal{B}_3 = 1 + \left(1 - \rho + \rho^2\right) M^2, \qquad 0 < \rho < 1$$

Proof: When regressors are asymmetric, polynomial corresponding to HC3 is:

$$p_3(w_t) = 1 + 2\left(EW_T^3\right) w_t + \left(EW_T^4\right) w_t^2$$

Using the values of EW_T^3 and EW_T^4 in above polynomials, we have;

$$p_3(w_t) = 1 + 2(M - N) w_t + \left(M^2 + N^2 - MN\right) w_t^2$$

Evaluating the polynomial at w=0, +M and −N, we have:

$$p_3(1) = 1$$

$$p_3(+M) = 1 + 2(M - N) M + \left(M^2 + N^2 - MN\right) M^2$$

$$p_3(-N) = 1 - 2(M - N) N + \left(M^2 + N^2 - MN\right) N^2$$

Since $N = \rho M \qquad (0 < \rho < 1)$, so we have:

$$p_3(+M) = 1 + 2(1-\rho)M^2 + (1-\rho+\rho^2)M^4$$

$$p_3(-N) = 1 - 2\rho(1-\rho)M^2 + \rho^2(1-\rho+\rho^2)M^4$$

Note that, these polynomials are quadratic in M^2, so, we can find the signs of the polynomials by evaluating the roots of the same.

Note that, roots the two polynomials at +M and –N are imaginary, so polynomials do not cut the x-axis and will remain above it or we can say both will always be positive for all values of ρ $(0 < \rho < 1)$.

So, For $0 < \rho < 1$, $p_3(0) = 1 > 0$, $p_3(+M) > 0$ and $p_3(-N) > 0$

The second order asymptotic maximum positive bias of HC3 is:

$$B_3^+ = \max_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var(Z_T) B_{22}(A_3)}{U}\right\}\right] = p_3(1)\left(1 - \frac{1}{\rho M^2}\right) + p_3(+M)\left(\frac{1}{M^2(1+\rho)}\right) + p_3(-N)\left(\frac{1}{\alpha M^2(1+\rho)}\right)$$

Replacing the values of polynomials and simplifying, we get:

$$B_3^+ = 1 + (1-\rho+\rho^2)M^2$$

The second order asymptotic maximum negative bias of HC3

$$\mathcal{B}_3^- = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\ Var(Z_T)B_{22}(A_3)}{U}\right\}\right] = 0$$

Reversing the sign of second order asymptotic maximum negative bias function, we get:

$$-\mathcal{B}_3^- = 0$$

Overall second order asymptotic maximum bias of HC3 is given by:

$$\mathcal{B}_3 = \max\left(\mathcal{B}_3^+, -\mathcal{B}_3^-\right) = \mathcal{B}_3^+ = 1 + \left(1-\rho+\rho^2\right)M^2$$

Where,

$$\mathcal{B}_3^+ = 1 + \left(1-\rho+\rho^2\right)M^2$$

$$-\mathcal{B}_3^- = 0$$

This completes the proof. ■

6.2.2.5: Asymptotic Maximum Bias of Minimax Estimator

This subsection provides the second order asymptotic maximum bias of Minimax estimator for asymmetric regressors.

Theorem 6-11: Second order asymptotic maximum bias of Minimax estimator, when the regressors are asymmetric, is given by:

EQ: 6-22 $$\mathcal{B}_{\text{Minimax}} = 2 - \frac{1}{M^2(1+\rho)} - \frac{3}{1+\rho} - \frac{4M^2}{1+\rho} + 4M^2 - 2\rho M^2, \qquad 0 < \rho < 1$$

Proof: When regressors are asymmetric, polynomial corresponding to Minimax estimator is:

$$p_{\text{Minimax}}(w_t) = \alpha + \left(2\alpha EW_T^3\right) w_t + \left(a + \alpha\left(EW_T^4 - 2\right)\right) w_t^2 - 2\alpha w_t^4$$

Note that, in large samples $\alpha = 1$, and also note that, $a = K + 1 = EW_T^4 + 1$

Using these values in the polynomial and simplifying, we get:

$$p_{MAAZ}(w_t) = 1 + \left(2EW_T^3\right) w_t + \left(2EW_T^4 - 1\right) w_t^2 - 2w_t^4$$

Using the values of EW_T^3 and EW_T^4 in above polynomials, we have;

$$p_{\text{Minimax}}(w_t) = 1 + \left(2\left(M - N\right)\right) w_t + \left\{2\left(M^2 + N^2 - MN\right) - 1\right\} w_t^2 - 2w_t^4$$

Evaluating the polynomial at w=0, +M and –N, we have:

$$p_{\text{Minimax}}(1) = 1$$

$$p_{\text{Minimax}}(+M) = 1 + \left(2\left(M - N\right)\right)M + \left\{2\left(M^2 + N^2 - MN\right) - 1\right\}M^2 - 2M^4$$

$$p_{\text{Minimax}}(-N) = 1 - \left(2\left(M - N\right)\right)N + \left\{2\left(M^2 + N^2 - MN\right) - 1\right\}N^2 - 2N^4$$

Since $N = \rho M \qquad \left(0 < \rho < 1\right)$, so we have:

$$p_{\text{Minimax}}\left(+M\right) = 1 + \left(1 - 2\rho\right)M^2 + 2\rho\left(\rho - 1\right)M^4$$

$$p_{\text{Minimax}}\left(-N\right) = 1 + \rho\left(\rho - 2\right)M^2 + 2\rho^2\left(1 - \rho\right)M^4$$

Note that, for large M, the leading term is the one involving M^4, and, since $0 < \rho < 1$, so we can see that, $p_{\text{Minimax}}\left(0\right) = 1 > 0$, $p_{MAAZ}\left(+M\right) < 0$ and $p_{\text{Minimax}}\left(-N\right) > 0$.

This permits us to write the second order asymptotic maximum positive bias of Minimax estimator as:

$$\begin{aligned} \mathcal{B}^{+}_{\text{Minimax}} &= \max_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var\left(Z_T\right)B_{22}\left(A_{\text{Minimax}}\right)}{U}\right\}\right] \\ &= p_{\text{Minimax}}\left(1\right)\left(1 - \frac{1}{\rho M^2}\right) + p_{\text{Minimax}}\left(-N\right)\left(\frac{1}{\rho M^2\left(1+\rho\right)}\right) \end{aligned}$$

Replacing the values of polynomials and simplifying, we get:

$$\mathcal{B}^{+}_{\text{Minimax}} = 2 - \frac{1}{M^2\left(1+\rho\right)} - \frac{3}{1+\rho} - \frac{4M^2}{1+\rho} + 4M^2 - 2\rho M^2, \qquad 0 < \rho < 1$$

The second order asymptotic maximum negative bias of Minimax estimator is:

$$\mathcal{B}^{-}_{\text{Minimax}} = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var\left(Z_T\right) B_{22}\left(A_{\text{Minimax}}\right)}{U}\right\}\right] = p_{\text{Minimax}}\left(+M\right)\left(\frac{1}{M^2\left(1+\rho\right)}\right)$$

Replacing the value of polynomial and simplifying, we get:

$$\mathcal{B}^{-}_{\text{Minimax}} = \min_{\sigma_t^2 \le U}\left[\lim_{T\to\infty}\left\{\frac{T^2\, Var\left(Z_T\right) B_{22}\left(A_{\text{Minimax}}\right)}{U}\right\}\right]$$
$$= 2 - \frac{1}{M^2\left(1+\rho\right)} - \frac{3}{1+\rho} - \frac{4M^2}{1+\rho} + 4M^2 - 2\rho M^2$$

Note that, second order asymptotic maximum positive and negative biases are exactly the same, so we can write the overall second order asymptotic maximum bias of Minimax estimator as follows:

$$\mathcal{B}_{\text{Minimax}} = 2 - \frac{1}{M^2\left(1+\rho\right)} - \frac{3}{1+\rho} - \frac{4M^2}{1+\rho} + 4M^2 - 2\rho M^2, \qquad 0 < \rho < 1$$

This completes the proof. ■

6.2.2.6: Asymptotic Comparison of HCCMEs (Asymmetric Regressors Case)

In this section we compare all four HCCMEs on the basis of overall second order asymptotic maximum bias. In order to provide a suitable comparison, we first present the formulae of the second order asymptotic maximum biases of all four HCCMEs in a simplified form below:

Second order asymptotic maximum bias of HC0 is:

$$B_0 = \frac{2\rho}{1+\rho} + \left(\frac{1+\rho-\rho^2}{1+\rho}\right)M^2 - \frac{1}{M^2(1+\rho)}, \quad 0 < \rho < \frac{\left(\sqrt{5}-1\right)}{2}$$

$$B_0 = 2 + \left(\rho^2 - \rho + 1\right)M^2 - \frac{1}{\rho M^2}, \quad \frac{\left(\sqrt{5}-1\right)}{2} < \rho < 1$$

Second order asymptotic maximum bias of HC1 is:

$$B_1 = \frac{2\rho-2}{1+\rho} + \left(\frac{1+\rho-\rho^2}{1+\rho}\right)M^2 - \frac{1}{M^2(1+\rho)}, \quad 0 < \rho < \frac{\left(\sqrt{5}-1\right)}{2}$$

$$B_1 = \left(\rho^2 - \rho + 1\right)M^2 - \frac{1}{\rho M^2}, \quad \frac{\left(\sqrt{5}-1\right)}{2} < \rho < 1$$

Second order asymptotic maximum bias of HC2 is:

$$B_2 = \frac{2\rho-1}{1+\rho} + \left(\frac{\rho-\rho^2}{1+\rho}\right)M^2 - \frac{1}{M^2(1+\rho)}, \quad 0 < \rho < 1$$

Second order asymptotic maximum bias of HC3 is:

$$B_3 = 1 + \left(1 - \rho + \rho^2\right) M^2, \qquad 0 < \rho < 1$$

Second order asymptotic maximum bias of Minimax is:

$$B_{\text{Minimax}} = \frac{2\rho - 1}{1+\rho} + 2\left(\frac{\rho - \rho^2}{1+\rho}\right) M^2 - \frac{1}{M^2(1+\rho)}, \qquad 0 < \rho < 1$$

When M is very large, then the terms containing ($1/M^2$) will go to zero also the constant terms are very small. So the only term that matters for the comparison is the one involving M^2. Note that the coefficient of M^2 in B_2 is $\frac{\rho - \rho^2}{1+\rho}$ for all $0 < \rho < 1$, yielding the smallest value as compared to coefficients of M^2 corresponding to all other HCCMEs. So we can say that second order asymptotic maximum bias of HC2 is the smallest, declaring it the clear winner against its rivals. Similar examination leads to the conclusion that Minimax estimator is the second most preferable.

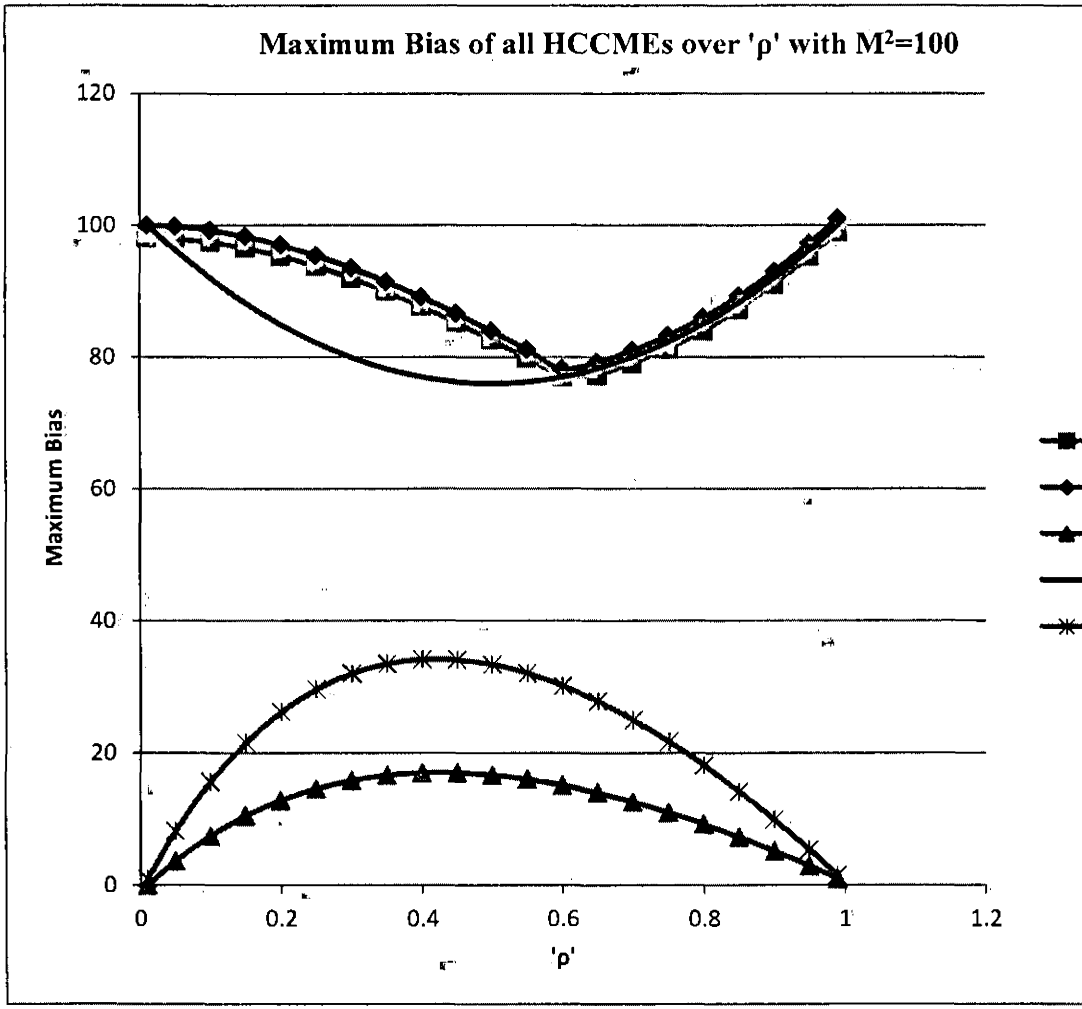
Also note that when the value of ρ (which is a measure of skewness) is close to zero or one, and M^2 is relatively a small number, then second order asymptotic maximum bias of HC2 and Minimax are comparatively close to each other but when ρ is around 0.5 then the second order asymptotic maximum bias of Minimax estimator is almost double in magnitude than that of HC2.

In addition, note that, when $0 < \rho < \frac{\left(\sqrt{5} - 1\right)}{2}$, the second order asymptotic maximum bias of HC3 is larger than both Minimax and HC2 but smaller than both HC1 and HC0 and when $\frac{\left(\sqrt{5}-1\right)}{2} < \rho < 1$, the second order asymptotic maximum bias of HC3 is larger in magnitude than HC1, HC2 and Minimax estimator while it is about the same as that of HC0.

Overall our results are in favor of HC2.

To make the comparison more clear, we plotted the second order asymptotic maximum bias of all HCCMEs over all possible values of ρ while keeping M^2 fixed at 100. The following figure represents the comparison.

Figure 6-1: Comparison of second order asymptotic maximum Bias of all HCCMEs over 'ρ'



From the above figure, we can see that the second order asymptotic maximum bias of HC2 is the smallest, making HC2 as the best estimator against its rivals including Minimax estimator, however, the Minimax estimator came out to be the second best.

Further note that, similar results were obtained with higher values of M^2.

In the end we provide a comment on different findings of existing studies. As explained earlier, existing studies are based on simulation by taking a particular set of regressors and a skedastic function. But since the performance of HCCMEs is different for different set of regressors as well as skedastic function, conflicting results are obtained. Different estimators are superior for different configurations of parameters. The minimax bias criterion provides a global comparison of the HCCMEs utilizing the worst case configuration of heteroskedasticity. This evaluation shows that HC2 is by far the best among all the estimators considered.

6.3: EXTENSION TO MULTIPLE REGRESSORS CASE

A major limitation on our results above is the restriction to the single regressor model. As we show in this section, it is possible to extend these results to multiple regressor models under certain conditions. This can be done by considering a sequence of increasingly complex cases.

6.3.1: CASE 1: Orthogonal Regressors

Consider a standard regression model with regressors $x_0, x_1, \ldots, x_{k-1}$, where the first regressor is a constant, and all regressors are orthogonal, i.e.

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_{k-1} x_{k-1\,t} + \varepsilon_t$$

Let β be the OLS estimate of β and consider the alternative model $y^* = \beta_0 + \beta_1 x_1 + u$, where $y^* = y - \hat{\beta}_2 X_2 - \cdots - \hat{\beta}_{k-1} X_{k-1}$.

Thus we can now use our results to assess the significance of regressor x_1 exactly as before. The minimax variance estimate can be applied to obtain a minimally biased estimator for the variance of β_1.

6.3.2: CASE 2: Prioritized Regressors

Next, consider a situation where the sequence of priorities of the regressor is known in advance. That is, we are testing for significance of a regressor x_i within a nested sequence of models where the j-th model consists of all regressors x_1 to x_j. This situation arises in polynomial regression, or in ARDL models, where we would like to choose the simplest models, with minimal order polynomial or lag. In this situation, we can use the Gram-Schmidt procedure to orthogonalize the regressors. Once we have orthogonal regressors, we can use the procedure of Case 1 to evaluate the significance of any regressor. Note that different priority orderings will lead to different calculations for the variances. Intuitively, the question we ask of the data is the following:

Given the x_1 to x_{i-1} are explanatory variables, does x_i add sufficient explanatory power to the model to be worth including? This is answered by purging x_i of the influence of the preceding regressors prior to testing for significance. This differs from the conventional t-statistic which treats x_i on par with the other variables.

6.3.3: CASE 3: Categorized Variables with Unique Ranking

In the general case, to evaluate the significance of a regressor x_j, we must categorize the relevant regressors into three categories. The first category is $x_1, \ldots, x_i$ where $i<j$. These are the regressors which have higher priority than x_j – they must be included in the model. This would be the case if, for example, theory dictates their inclusion. The second category is those variables which are of equal priority; these would

be variables $x_{i+1},\ldots,x_k$, where $k>j$. The third category is variables of lower priority; these are $x_{k+1}, \ldots, x_K$. These variables are to be included only if they add explanatory power AFTER x_j has already been included in the model.

First consider the case where $i=j-1$, and $k=j+1$, so that there are no variables of equal priority. In this case, a Gram-Schmidt procedure will convert the model to orthogonal regressors, and applying the procedures of the first case will yield the desired results. The variable x_j will be judged significant if and only if it adds significant explanatory power after the inclusion of all variables x_1 to x_{j-1}.

6.3.4: CASE 4: General Case

Now suppose that there are other variable of equal priority. This includes the possibility that $i=1$ and $j=K$ so that all variables are of equal priority. This is typically assumed in regression models. This is similar to the case of multicollinearity. If two variables are close substitutes, then it can happen that both have insignificant t-statistics. This means that one of the two is sufficient; neither variable adds explanatory power in presence of the other. In such cases, it is impossible to decide which of the two is significant on purely statistical grounds. Many applied cases can be cited where multicollinearity leads to wrong decisions on significance on statistical grounds. The suggested decision procedure for this situation is as follows.

First, position x_j as the first in the group of equivalent variables, then apply Gram-Schmidt to orthogonalize the regressors, and follow the procedure of Case 1 to determine significance. If x_i is not significant in the first position, then it is not significant. In any later position, it cannot acquire significance after orthogonalization.

Next, position x_i as the last in the group of equivalent variables. If it is significant in the last position after a Gram-Schmidt orthogonalization, then it will always be significant in earlier positions.

The only remaining possibility is that the variable is significant in the first position, but insignificant in the last position. In this case, the data does not lead to a firm conclusion. The variable may or may not be significant, depending on whether or not other variables of equal priority are included. In absence of statistical evidence, decisions about significance must be made on a priori or theoretical grounds.

6.4: A REAL WORLD APPLICATION OF HCCMEs

In this section, we provide application of these results using some real world data. The details are provided below:

Consider a two-covariate linear regression model:

[6.4.1] $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t, t = 1,2,...,T$.

The dataset used consists of data on per capita expenditure on public schools and per capita income by state in the U.S. in 1979, and is taken from Greene (1997, p.541). The dependent variable y is per capita expenditure on public schools and covariates x_2 and x_3 are per capita income (scaled by 10^{-4}) and the squarer of per capita income respectively, totaling T=50 observations. We considered only 47 observations omitting three high leverage observations (Alaska, Mississipi, Washington D.C.), so our sample size is T=47. Since this is a multiple regressor model so we first orthogonalized the regressors by using Gram-Schmidt procedure and then estimated the above model using OLS. The multiple linear regression model is transformed to simple linear regression model using the procedure outlined in section 6.3 above. i.e. we estimated the following regression model:

[6.4.2] $y^* = \beta_1 + \beta_2 x_2 + u$, where $y^* = y - \hat{\beta}_3 X_3$, and, β's are OLS estimate of β's.

We calculated the maximum positive and maximum negative bias of all HCCMEs, HC0, HC1, HC2, HC3 and Minimax Estimator using the analytical formulae developed. The overall maximum bias is also calculated for all HCCMEs including Minimax estimator. The results are provided in the following table.

Table 6-1: Overall Maximum Bias of HCCMEs

MB of HC0	3.63
MB of HC1	2.25
MB of HC2	0.96
MB of HC3	3.22
MB of Minimax	1.64

Note: MB stands for maximum bias

From the above table, we can see that overall maximum bias of HC2 is smallest of all while minimax estimator has second lowest maximum bias. The bias of HC0 is found to be largest of all.

We experimented with other data sets as well and obtained the similar results favoring HC2 though the maximum bias changed due to change in the design matrix.

In the end we strongly recommend the practitioners to use HC2 estimator while performing heteroskedasticity corrections.

Chapter 7: CONCLUSIONS AND RECOMMENDATIONS

This thesis deals with the issues of comparing the most famous HCCMEs namely HC0 proposed by White, HC1 proposed by Hinkley, HC2 suggested by Horn, Horn and Duncan and HC3 by Mackinnon and White. Much of the existing literature is concerned to compare their performances using Monte Carlo simulations. Some of the studies provided the bias expression for different HCCMEs but their analytics is too complex to compare them in detail. So, still no clear cut winner has emerged, though, a number of studies suggested HC2 and HC3 using different criteria, e.g. size distortion etc. Since performance of HCCMEs depends on the design matrix of regressors as well as the skedastic function of variances, so simulations are not the right choice, since simulation takes one particular set of regressors and skedastic function. So the only solution is to compare HCCMEs using analytical comparison.

In this thesis, we gave exact analytical expressions for the biases of HCCMEs. Due to complexity of analytics, we consider one regressor model and provided the comparison of HCCMEs by comparing their asymptotical worst case biases.

We have obtained elementary explicit analytical formulae for the bias of variance estimates in a single regressor model with heteroskedasticity. This allows us to calculate the pattern of the least favorable heteroskedastic sequence, and to compute worst case bias. In the past, simulation studies chose different patterns of heteroskedasticity in an ad-hoc fashion. This ad-hoc choice does not allow for accurate evaluation of strengths and weaknesses of different classes of estimators. Our methodology permits an analytical assessment and comparison of estimators on the basis of their worst case bias. In some cases, this minimax assessment can be too pessimistic. Our formulae also permit alternative methods of evaluation, which may be explored in future research.

One very important payoff from our research is an explicit formulae for a minimax estimator which has substantially lower maximum bias than conventional estimators, HC0, HC1, HC3 [in case of symmetric as well as asymmetric regressors] and HC2 [in case of symmetric regressors only]. The proof of minimaxity is not analytic but heuristically based on simulations. We prove minimaxity for a restricted class of regressor sequences, and numerically showed that particulars of the regressor sequence do not matter. It is possible that the estimator obtained is minimax among the class of all estimators – not just the special one parameter family analyzed in this thesis. Even if this is not so, numerical calculations show that it cannot be improved upon by much. This solves the problem raised in the introduction: how to choose a specific HCCME from among a broad class with competing claims to superiority and widely different small sample properties.

An unsolved puzzle is the invariance conjecture. The maximum bias functions $\mathcal{B}^{+}(a)$ and $\mathcal{B}^{-}(a)$ depend directly on the sequence of regressors. Why the value of 'a' at their intersection depends only the kurtosis is a mystery we leave for future researchers to resolve. Further note that we proved that invariance conjecture is valid in case of a special class of estimators which takes into account HC0 and HC1 as special cases, however, for the other two HCCMEs (HC2 and HC3), this conjecture may holds as well. Further research is required to explore this issue.

We can summarize our main findings as follows:

1. Minimax has lowest second order asymptotic maximum bias in the class of all estimators including HC0 and HC1 as special cases. Hence Minimax estimator is a clear winner against all estimators which fall in this specific class of estimators.

2. HC2 and HC3 do not fall into that particular class, so, we compared them for a particular sequence of regressors (symmetric as well as asymmetric). In case of symmetric regressors, the second order asymptotic maximum bias of HC2 is exactly same as that of Minimax estimator and both are lowest as compared to all rival estimators (HC0, HC1, HC3). But in case of asymmetric regressors, the second order asymptotic maximum bias of HC2 is best among all estimators including Minimax estimator.

3. Overall we can say that HC2 is the real minimax estimator whether regressors are symmetric or asymmetric but for a restricted class of estimator Minimax estimator is best.

The analysis can be extended to cover high-leveraged estimators (HC4 and HC5) in a future research and one can devise a minimax estimator covering these two estimators as well.

REFERENCES

1) Adèr, H. J., Mellenbergh G. J., & Hand, D. J. "Advising on research methods: A consultant's companion." Huizen, The Netherlands: Johannes van Kessel Publishing, 2008.

2) Akahira Masafumi, and Takeuchi Kei. Asymptotic efficiency of statistical estimators: concepts and higher order asymptotic efficiency. Springer-Verlag, 1981.

3) Bera A.K., T. Suprayitno, G. Premaratne. "On some heteroskedasticity-robust estimators of variance-covariance matrix of the least-squares estimators." Journal of Statistical Planning and Inference 108 (2002): 121–136.

4) Berkowitz, J. and Kilian, L. "Recent developments in bootstrapping time series." Econometric Reviews 19 (2000): 1-48.

5) Chernick, M. R. Bootstrap Methods: A Practicioner's Guide. New York: Wiley, 1999.

6) Chesher A., and I. Jewitt. "The Bias of Heteroskedasticity Consistent Covariance Matrix Estimators." Econometrica 55.5 (1987): 1217-1222.

7) Chung, Kai Lai. A Course in Probability Theory: Third Edition. London: Academic Press, 2001.

8) Cragg J. G. "More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form." Econometrica 51.3 (1983): 751-763.

9) Cribari-Neto F, Zarkos S. "Leverage-adjusted heteroskedastic bootstrap methods." Journal of Statistical Computation and Simulation 74.3 (2004): 215 - 232.

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

11) Cribari-Neto F. Ferrari, S. L. P. Cordeiro G. M. "Improved Heteroskedasticity-Consistent Covariance Matrix Estimator." Biometrica 87.4 (2000): 907-918.

12) Cribari-Neto F., Nila M. S. Galvao. "A Class of Improved Heteroskedasticity-Consistent." Communications in Statistics-Theory and Methods 32.10 (2003): 1951 - 1980.

13) Cribari-Neto F., T. C. Souza, Vasconcellos K. L. P. "Inference under Heteroskedasticity and Leveraged Data." Communications in Statistics - Theory and Methods 36.10 (2007): 1877 - 1888.

14) Cribari-Neto F., Zarkos S. "Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing." Econometric Reviews 18.2 (1999): 211 – 228.

15) —. "Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap." Journal of Statistical Computing and Simulation 68 (2001): 391-411.

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

17) Efron B. "The jackknife, the bootstrap, and other resampling plans." Society of Industrial and Applied Mathematics 38 (1982).

18) Efron B. "Bootstrap methods: Another look at the jackknife." Annals of Statistics 7 (1979): 1-26.

19) Eicker F. "Asymptotic Normality and Consistency of the Least squares Estimators for Families of Linear Regressions." Annals of Mathematical Statistics 34 (1963): 447-456.

20) —. "Limit Theorems for Regression with Unequal and Dependent Errors." Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (1967): 59-82.

21) Furno M. "A Robust Heteroskedasticity Consistent Covariance Matrix Estimator." Statistics 30.3 (1997): 201 - 219.

22) Greene W. H. Econometric Analysis. Prentice Hall, 2003.

23) Griffiths W. E., Hill R.C., Judge G. G. Learning and Practicing Econometrics. New York: John Wiley & Sons, 1993.

24) Gujarati D. Basic Econometrics. 4th . The McGraw–Hill Company, 2004.

25) Hendry David F. Dynamic econometrics. Oxford University Press, 1995.

26) Hinkley, V. D. "Jackknifing in Unbalanced Situation." Technometrics 19 (1977): 285-292.

27) Horn, S. D., Horn, R. A., Duncan, D. B. "Estimating heteroskedastic variances in linear models." Journal of the American Statistical Association 70 (1975): 380–385.

28) Horowitz, J. L. Bootstrap methods in econometrics: Theory and numerical performance. Vol. 3. New York: Cambridge University Press, 1997.

29) Kennedy, Peter. A guide to econometrics. MIT Press, 2003.

30) Li, H. and Maddala, G. S. "Bootstrapping time series models." Econometric Reviews 15 (1996): 115–195.

31) Long, J. S. and Ervin, L. H. "Using heteroskedasticity consistent standard errors in the linear regression model." The American Statistician 54 (2000): 217–224.

32) Mackinnon, J. and H. White. "Some Heteroskedasticity-Consistent Covariance Matrix Estimator with Improved Finite Sample Properties." Journal of Econometrics 29 (1985): 305-325.

33) Newey, Whitney K and Kenneth D West. "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." Econometrica 53.3 (1987): 703–708.

34) Orhan, Mehmet. Robust Regression, HCCM Estimators, and an Empirical Bayes Application. PhD Thesis. Bilkent University, Turkey, 2000.

35) Pearce J.M, and Hall. "A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli." Psychological Review 87 (1980): 532-552.

36) Prais, Sig J., and Hendrik S. Houthakker. The Analysis of Family Budgets. Cambridge University Press, 1955.

37) Ramsey, J.B. "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis." Journal of the Royal Statistical Society 31.2 (1969): 350–371.

38) Shepp, L.A. "Distinguishing a Sequence of Random Variables from a Translate of Itself." The Annals of Mathematical Statistics 36.4 (1965): 1107-1112.

39) Stock, J. H. and Watson, M. W. Introduction to Econometrics. Boston: Addison-Wesley, 2003.

40) White, H. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." Econometrica 48.4 (1980): 817-838.

41) Wooldridge, J. M. Introductory Econometrics: A Modern Approach. Cincinnati: South-Western College Publishing, 2000.

42) Wu, C.F.J. "Jackknife, bootstrap and other resampling methods in Regression analysis." Annals of Statistics 14 (1986): 1261–1295.

43) Zietz, Joachim. "Heteroskedasticity and Neglected Parameter Heterogeneity." Oxford Bulletin of Economics and Statistics 63.2 (2001): 263–273.

APPENDICES

APPENDIX A: ESTIMATORS NOT COVERED IN THE STUDY

This appendix includes the estimators which have not been considered in our study but are presented here for completeness.

A.1: BOOTSTRAPPED BASED HCCMEs

One of the alternatives to HCCMEs is to use bootstrap methods to estimate the covariance matrix of OLS estimator. Efron (1979) proposed this method the first time, called naïve bootstrap. The basic idea is to resample the available information in data to get the information about the unknown statistic of interest. Bootstrap based methods are useful, according to Ader et al. (2008), the bootstrap method should be adopted in the following situations:

i. When theoretical distribution of a statistic of interest is complicated or unknown.
ii. When the sample size is insufficient for straightforward statistical inference.
iii. When power calculations have to be performed, and a small pilot sample is available.

[For more detailed survey on bootstrap methods, See Li and Maddala (1996), Horowitz (1997) and Berkowitz and Kilian (2000) etc.]

In the following sub-sections, we provide brief review of bootstrap methods developed so far.

A.1.1: NAÏVE BOOTSTRAP ESTIMATOR

Efron (1979) proposed the first bootstrap estimator known as naïve bootstrap. The bootstrap scheme is as follows:-

a. Draw a random sample, $e_i^*, i = 1,2,\dots,T$, with replacement form OLS residual, $e_i, i = 1,2,\dots,T$.

b. Use $y_i^* = X_i\beta^{\wedge} + e_i^*$ to get a bootstrap sample.

c. Compute OLS estimate $\beta^* = (X'X)^{-1}X'y^*$.

d. Repeat first three steps a large number of times, (say N times) to get N vectors of OLS estimates (β^*).

e. Calculate covariance matrix of N vectors of (β^*).

A.1.2: JACKKNIFE ESTIMATOR (JA)

In 1982, Efron proposed an estimator known as Jackknife estimator, the idea is to drop one observation each time the regression model is estimated and parameters are estimated. At the end, the variance of the estimated parameters gives an estimate of the variance of true parameter. The jackknife estimator is given by:

$$\Omega_{Ja} = \frac{T-1}{T}(X'X)^{-1}\left[X'\hat{\Sigma}_{ja}X - \frac{1}{T}X'ee'X\right](X'X)^{-1}$$

Where, $\hat{\Sigma}_{ja} = diag\left(\frac{e_t^2}{(1-h_{tt})^2}\right), \qquad t = 1,2,...,T$

Here e^2's are OLS squared residuals and h_{tt} is the t-th entry of the Hat matrix (H), $H = X(X'X)^{-1}X'$.

A.1.3: WEIGHTED BOOTSTRAP ESTIMATOR

The estimator proposed by Wu (1986) can work evenly well in situations where the data, in population, is not IID as opposed to Efron (1979) which works well when data, in population, is IID and gives inconsistent estimates. Wu's (1986) proposed a scheme based on resampling the residuals in such a way that can yield HCCME. This estimator is known as weighted bootstrap estimator.

Later Chernick (1999) suggested to resample the actual data (Y, X) instead of resampling the OLS residuals when the model is mis-specified or there is heteroskedasticity.

A.1.4: LEVERAGED ADJUSTED BOOTSTRAP ESTIMATORS

In 2004, Cribari-Neto & Zarkos proposed three alternative bootstrapped estimators, namely, *adjusted weighted bootstrap estimator*, *linearly adjusted weighted bootstrap* and *inversely adjusted weighted bootstrap* to take into account the effect the high-leveraged observations. Readers are referred to Cribari-Neto & Zarkos (2004) for more detailed discussion of these methods.

A.2: ESTIMATORS BASED ON GMM

Cragg (1983) proposed an estimator based on generalized method of moments (GMM) which is proved to be more efficient that OLS based estimator.

Cragg estimator is given below:

$$\beta_{Cragg} = \left(X'W\left(W'\Sigma W\right)^{-1} W'X \right)^{-1} X'W\left(W'\Sigma W\right)^{-1} W'y$$

Where, W is the matrix of instruments, which includes, regressors X, their cross-products and successive positive powers.

When W=X, Cragg estimator reduces to Eicker-White estimator. By adding additional instruments, a gain in efficiency can be obtained. Small samples performance of this estimator is very poor, due to large size distortions of tests based on it. So this approach could not get much popularity.

A.3: ROBUST HCCMEs

Furno (1997) advocated to use robust HCCME and proposed the robust versions of HC0, HC1, HC2 and HC3.According to her, her approach has three main advantages. First, one need not to specify the form of heteroskedasticity, second, the sample bias of HCCMEs can be reduced and third one need not to do any preliminary analysis for the detection of outliers and thus this saves us from losing the additional information which the outliers contain and can be lost if we delete them.

The robust version of Eicker-White (HC0) estimator is:

$$HC0_R = (X'WX)^{-1} X'W\Sigma_{0R}WX(X'WX)^{-1}$$

Where, W is an TxT diagonal matrix with $w_t = \min(1, c/h_{tt})$ and $c = 1.5k/T$. Here, $\Sigma_{0R} = diag\left\{\tilde{e}_t^2\right\}, t = 1,2,...,T$ where, $\tilde{e}_t^2$ is the t-th weighted least squares (WLS) residual obtained from WLS regression of Y on X with least squares regression parameter $\beta_R = (X'WX)^{-1} X'Wy$.

The robust version of Hinkley (HC0) estimator is:

$$HC1_R = (X'WX)^{-1} X'W\Sigma_{1R}WX(X'WX)^{-1}$$

Where, $\Sigma_{1R} = diag\left\{\frac{\tilde{e}_t^2}{1-k/T}\right\}, t = 1, 2, ..., T$

The robust version of Horn, Horn and Duncan (HC2) estimator is:

$$HC2_R = (X'WX)^{-1} X'W\Sigma_{2R}WX(X'WX)^{-1}$$

Where, $\Sigma_{2R} = diag\left\{\frac{\tilde{e}_t^2}{1-h_{tt}^*}\right\}, t = 1, 2, ..., T$

The robust version of Mackinnon & White (HC3) estimator is:

$$HC3_R = (X'WX)^{-1} X'W\Sigma_{3R}WX(X'WX)^{-1}$$

Where, $\Sigma_{3R} = diag\left\{\left(\frac{\tilde{e}_t}{1-h_{tt}^*}\right)^2\right\}, t = 1, 2, ..., T$

With h_{tt}^* is the t-th diagonal element of $\sqrt{W}X(X'WX)^{-1}X'\sqrt{W}$

Furno compared the performance of these four robust heteroskedasticity consistent covariance matrix estimators (RHCCMEs) using Monte Carlo Simulations. Her results showed that in case of heteroskedasticity, the RHCCMEs reduce the biases and thus are more efficient.

A.4: BIAS CORRECTED HCCMEs

There is another approach that seems sensible, which first finds the bias and then simply subtract this estimated bias from the estimator to get a bias corrected estimator.

The first set of bias corrected HCCMEs is proposed by Cribari-Neto et al. (2000) who analytically calculated the bias of Eicker-White (HC0) and then defined bias corrected estimators recursively, that is, they calculated bias of HC0 and subtracted it from the actual HC0 estimator recursively. They also provided the expressions for the variances of these estimators.

Later, Cribari-Neto and Galvao (2003) generalized the results of Cribari-Neto et al (2000) and analytically calculated the bias corrected versions of HC1, HC2 and HC3 along with HC0 and also provided the expressions for the variances of these estimators.

A.5: HIGH-LEVERAGED BASED ESTIMATORS

Two estimators developed by Cribari-Neto (2004) and Cribari-Neto et al. (2007) for high-leveraged observations in the design matrix. These are discussed briefly in the following sub-sections.

A.5.1: CRIBARI-NETO (HC4) ESTIMATOR

This estimator was proposed by Cribari-Neto (2004) and it takes into account the effect of high-leveraged observations in the regression design. He adjusted HC3 by taking the exponent of the discounting term as minimum of the 3 and the ratio between each leverage measure and the mean leverage. This estimator was named as HC4 and is given as follows:

$$\Omega_{HC4} = (X'X)^{-1}(X'\hat{\Sigma}_{HC4}X)(X'X)^{-1}$$

Where, $$\hat{\Sigma}_{HC4} = diag\left(\frac{u_t^2}{(1-h_{tt})^{g_t}}\right), \qquad t = 1,2,...,T$$

With, $$g_t = \min\left\{\frac{h_{tt}}{T^{-1}\sum_{t=1}^{T} h_{tt}}, 3\right\} = \min\left\{\frac{T * h_{tt}}{K}, 3\right\}, \qquad \left(\therefore \sum_{t=1}^{T} h_{tt} = K\right)$$

Here the discount factor is the ratio between the leverage measure of each observation and the average of all leverage measures. Cribari-Neto (2004) showed using simulation suggested that HC4 should be used when regression design contains influential observations using size distortion as the deciding criteria.

A.5.2: CRIBARI-NETO et al. (HC5) ESTIMATOR

This estimator known as HC5 was proposed by Cribari-Neto et al. (2007) and is given by:

$$\Omega_{HC5} = (X'X)^{-1}(X'\hat{\Sigma}_{HC5}X)(X'X)^{-1}$$

Where, $$\hat{\Sigma}_{HC5} = diag\left(\frac{u_t^2}{(1-h_{tt})^{\alpha_t}}\right), \qquad t = 1,2,...,T$$

With, $$\alpha_t = \min\left\{\frac{h_{tt}}{T^{-1}\sum_{t=1}^{T} h_{tt}}, \max\left(4, \frac{kh_{max}}{T^{-1}\sum_{t=1}^{T} h_{tt}}\right)\right\} = \min\left\{\frac{T\,h_{tt}}{p}; \max\left(4, \frac{Tkh_{max}}{K}\right)\right\}$$

Where $h_{max} = \max\{h_1, h_2, ..., h_T\}$ is the maximal leverage.

The constant α_i determines how much the i^{th} squared residual should be inflated in order to account for the i^{th} observation leverage. According to Cribari-Neto et al. (2007), this estimator is useful when the regression design contains extreme high-leveraged observations. This estimator (HC5) reduces to HC4 when the regression design contains no extreme influential observations.