PALGRAVE TEXTS IN ECONOMETRICS

# Bootstrap Tests for Regression Models

Leslie Godfrey

# Bootstrap Tests for Regression Models

*Palgrave Texts in Econometrics*

General Editors: Kerry Patterson

*Titles include:*

Simon P. Burke and John Hunter
MODELLING NON-STATIONARY TIME SERIES

Michael P. Clements
EVALUATING ECONOMETRIC FORECASTS OF ECONOMIC AND FINANCIAL VARIABLES

Leslie Godfrey
BOOTSTRAP TESTS FOR REGRESSION MODELS

Terence C. Mills
MODELLING TRENDS AND CYCLES IN ECONOMETRIC TIME SERIES

**Palgrave Texts in Econometrics**
**Series Standing Order ISBN 978–1–4039–0172–9 (hardcover)**
**Series Standing Order ISBN 978–1–4039–0173–6 (paperback)**
(*outside North America only*)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and the ISBN quoted above.

Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England

# Bootstrap Tests for Regression Models

Leslie Godfrey

palgrave
macmillan


Softcover reprint of the hardcover 1st edition 2009 978-0-230-20230-6



First published in 2009 by
PALGRAVE MACMILLAN

PALGRAVE MACMILLAN in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

PALGRAVE MACMILLAN in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

PALGRAVE MACMILLAN is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries.

ISBN-13: 978–0–230–20231–3 paperback
ISBN 978-0-230-20231-3 ISBN 978-0-230-23373-7 (eBook)
DOI 10.1057/9780230233737

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

A catalog record for this book is available from the Library of Congress.

10 9 8 7 6 5 4 3 2 1
18 17 16 15 14 13 12 11 10 09

*To Ben*

*This page intentionally left blank*

# Contents

# Preface

My wife is past-President of the Society for the Study of Addiction, but I suspect that even she finds it difficult to understand why I have not been able to free myself from an obsession with tests for econometric models in the last 30 years. My only defence is that I hoped that these tests would be useful to applied workers. Like many other researchers in the area, I had to make use of asymptotic theory when deriving tests. I now believe that the application of appropriate bootstrap techniques can greatly increase the usefulness of asymptotic test procedures at low cost and so I have a new obsession to combine with the old one.

Two types of problems associated with using asymptotic analysis to obtain tests are often mentioned. First, even when theory is tractable and leads to asymptotically valid critical values from standard distributions like Normal and Chi-Squared, which are convenient to use, there may be serious approximation errors in finite samples. In particular, the critical values implied by asymptotic theory may produce finite sample significance levels that are not close to the desired probabilities. Second, there are important test procedures for which asymptotic theory is intractable and does not provide a standard distribution from which critical values can be taken. The bootstrap has been used to tackle both types of problem. When a standard asymptotic test is available, the corresponding bootstrap test is often found to provide a better finite sample approximation and the improvement is sometimes remarkable. When no standard asymptotic test can be derived, the bootstrap can sometimes produce a test that is easy to carry out and has significance levels that are reasonably close to the desired values.

The bootstrap approach involves using computer programs to generate many samples from an artificial model that is intended to approximate the process assumed to generate the actual data. The values of test statistics calculated from these bootstrap samples can then be used to assess the statistical significance of the corresponding test statistic derived from the real observations. Given that many artificial samples are generated and each is subjected to the same statistical analysis as the genuine sample, there might be concerns about the computational costs of bootstrap tests. However, given the amazing increases in the power of personal computers, the real cost of the bootstrap approach is often very small in absolute terms, for example, the waiting time for results to appear. The

costs of bootstrapping are, therefore, often small and there is a great deal of evidence to suggest that the benefits can be very large.

The examples in this book that illustrate the value of the bootstrap and the dangers of relying upon asymptotically justified critical values are in the familiar framework of ordinary least squares (OLS) procedures for a linear regression model. The regression model is central to econometrics and its familiarity allows the reader to concentrate on the bootstrap techniques. The level of discussion is at an intermediate textbook standard and the aim has been to write a book that is useful to a fairly wide audience. However, references that cover more complicated models and more technical analyses of bootstrap procedures are provided.

Chapter 1 contains a discussion of regression models and OLS-based tests in order to summarize key results, to provide details of notation and to motivate going beyond conventional asymptotic theory as a basis for inference. The second chapter covers some basic ideas of simulation-based tests, with bootstrap procedures being given prominence but other approaches also being discussed. The application of simulation-based tests in regression models, under the assumption of independently and identically distributed (IID) errors, is examined in Chapters 3 and 4. The first of these two chapters covers test statistics that have standard asymptotic distributions, for example, Chi-Squared, when the null hypothesis is true. Chapter 4 is devoted to examples of situations of importance to empirical workers in which the bootstrap can be applied to statistics that, under the null hypothesis, have non-standard asymptotic distributions.

While the assumption that regression models have IID errors has often been made in the past when explaining results concerning the asymptotic properties of OLS estimators and test statistics, there has been a growing body of opinion that it is too restrictive. There are, of course, many ways in which data can be modelled using regression models with non-IID errors. The bootstrap world must mimic the process that is assumed to generate actual data under the null hypothesis. Consequently there is a need for bootstrap methods that allow for departures from the assumption of IID errors that are of interest to applied workers. Some of these methods are discussed in Chapter 5.

When the errors are not restricted to be IID, they can be assumed to be autocorrelated or heteroskedastic, according to precisely defined parametric models or in unspecified ways. The basic position taken in Chapter 5 is that there is rarely very clear guidance about the specification of parametric error models. There is, therefore, an emphasis on bootstrap methods that are designed to be asymptotically valid under unknown forms of autocorrelation and/or heteroskedasticity. Some examples of

the applications of these methods are examined in Chapter 6, which contains results on the finite sample behaviour of autocorrelation-robust and heteroskedasticity-robust bootstrap tests.

All of the tests discussed in Chapters 1 to 6 are based upon the assumption that the null-hypothesis model is a special case of the alternative-hypothesis model, that is, the former is nested in the latter. This assumption is required for much of the standard asymptotic theory of testing statistical hypotheses. However, competing specifications of linear regression models in applied econometric work are sometimes not nested and there is a considerable literature on tests for non-nested relationships. Chapter 7 contains a discussion of asymptotic and bootstrap tests for non-nested regression models. This discussion indicates how the bootstrap can help to overcome both of the above-mentioned general types of problem associated with reliance upon asymptotic theory when implementing tests of non-nested hypotheses. Finally, Chapter 8 contains an epilogue.

In the discussions of the application of bootstrap methods to OLS-based tests in regression analysis, I have used some examples from articles that I have written with various coauthors. I owe many debts to Chris Orme, Hashem Pesaran, Joao Santos Silva, Andy Tremayne and Mike Veall. It was a pleasure to work with these fine researchers and Mike Veall deserves special acknowledgment because he introduced me to the bootstrap during his first visit to York. I am very much indebted to Kerry Patterson, editor of this series, for his careful and constructive comments on my drafts. I am also grateful to Taiba Batool, commissioning editor at Palgrave Macmillan, for her encouragement and help, and to Alina Spiru for her assistance with the indexes. Finally, my thanks go to Christine who probably never realized that marriage might lead to the burden of helping me to sort out my ideas about this book during our lunchtime walks around the York campus.

L. G. Godfrey

# 1
# Tests for Linear Regression Models

## 1.1. Introduction

The linear regression model is often used to study economic relationships and is familiar from standard intermediate and introductory courses at the level of, for example, Greene (2008), Gujarati (2003) and Wooldridge (2006). In such courses, considerable emphasis is usually placed on the important topic of testing hypotheses about the values of the parameters of the model. The text-book tests for regression models are developed using very strong auxiliary assumptions that simplify teaching but are of limited relevance in practical situations. As a consequence, applied workers often have to replace procedures that are exactly valid in finite samples under strong assumptions by tests that are based on weaker assumptions but are only asymptotically valid.

It is also often necessary to rely upon asymptotic, rather than finite sample, results when carrying out tests for misspecification of a regression model. It is now commonplace for the results of estimation to be accompanied by checks of the assumptions required to validate standard empirical analysis. Even under the restrictive assumptions of the classical textbook model, many of these checks have to be carried out using critical values that are only asymptotically valid. When these assumptions are relaxed, there is an even greater need to use asymptotic theory.

The problem for the empirical researcher is that asymptotic theory sometimes provides a poor approximation to the actual distribution of test statistics; so that the use of asymptotic critical values may lead to misleading inferences. Moreover, there is a second type of problem associated with the standard approach to deriving asymptotically valid tests. In some situations of importance, this approach is not capable of providing a usable tool for the applied worker. This failure can occur with some

tests when classical assumptions are relaxed, or when several separate large sample tests are being applied.

The purpose of this book is to explain how computers and appropriate software can be combined to tackle these problems. More precisely, the use of procedures involving the simulation of artificial sets of data is examined and some important cases are discussed in detail. The various computationally intensive simulation techniques, collectively known as *bootstrap methods*, provide:

1. ways to improve the finite-sample performance of well-known and widely-used large sample tests for regression models; and
2. new tests that can be employed when conventional asymptotic theory does not lead to a test statistic that can be compared with critical values from some standard distribution.

The reason for believing that it is worth providing a concise, but quite extensive, account of bootstrap tests in regression analysis is that, in recent years, personal computers have become so powerful and relatively cheap that it is now feasible to implement bootstrap procedures as part of routine econometric analysis. Also the linear multiple regression model provides a very useful framework for introducing ideas that can be used in more complicated models that are of interest to applied workers, students and others who carry out empirical econometric analyses.

The emphasis is on practical applications of bootstrap methods in regression models. There are many excellent treatments of theoretical issues associated with the validity and properties of bootstrap techniques in quite general settings. References to such technical material will be provided and key results will be summarized.

This chapter is intended to give an outline of the various frameworks for which results about regression model tests are available and widely used. The foundations required for the detailed treatments contained in later chapters are provided, along with notation. More thorough coverage of tests for regression models, including numerical examples, can be found in many text books, for example, J. Davidson (2000, chs 2 and 3) and Greene (2008, ch 5). The discussion in Davidson and MacKinnon (2004, ch 4) links the statistical underpinnings of tests with the use of simulation methods and so is especially useful for the purpose of this book.

The important problem of testing linear restrictions in the classical Normal linear regression model is covered in Section 1.2, which includes much of the required notation. Section 1.2 provides key results that are exactly valid under the very strong assumptions of the textbook classical

model. It is argued that, despite their value in simplifying the teaching of econometric tests, these assumptions should not be regarded as suitable for practical applications. Section 1.3 contains comments on carrying out tests under weaker assumptions about the error terms and explanatory variables of the regression model. However, the analysis of Section 1.3 is based upon the assumptions of independence and homoskedasticity. In Section 1.4, tests that are asymptotically valid in the presence of autocorrelation and/or heteroskedasticity are described. The tests of linear restrictions that are covered in Sections 1.3 and 1.4 are only asymptotically valid. Applied workers have to use data sets with a finite number of observations and may be concerned about relying on results that only hold as the sample size tends to infinity. Some examples are provided in Section 1.5 that illustrate the problems of inadequate approximations derived from asymptotic theory. Section 1.6 contains examples of situations in which it is not possible to derive an asymptotic test that permits reference to a standard distribution to assess statistical significance. A summary and some concluding remarks are given in Section 1.7.

## 1.2. Tests for the classical linear regression model

As in many texts, the starting point is the *classical linear regression model*

$$y_i = \sum_{j=1}^{k} x_{ij}\beta_j + u_i, \tag{1.1}$$

in which: $y_i$ is a typical observation on the dependent variable; the terms $x_{i1}, \ldots, x_{ik}$ are the nonrandom values of a typical observation on the $k$ regressors; the unknown regression coefficients to be estimated are $\beta_1, \ldots, \beta_k$; and the unobservable errors, with typical term $u_i$, are independently and identically distributed (IID), each having the Normal distribution with zero mean and variance $\sigma^2$. The classical assumptions concerning the error term will sometimes be written using the notation NID$(0, \sigma^2)$, with NID standing for "Normally and independently distributed."

Suppose that there are $n > k$ observations for statistical analysis. It follows from (1.1) that the random variables $y_1, \ldots, y_n$ are independently distributed, with individual distributions being given by

$$y_i \sim N\left(\sum_{j=1}^{k} x_{ij}\beta_j, \sigma^2\right), i = 1, \ldots, n. \tag{1.2}$$

The system of $n$ equations with typical member (1.1) can be written in matrix-vector notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{1.3}$$

in which: $\mathbf{y}$ and $\mathbf{u}$ are the $n$-dimensional vectors with typical elements $y_i$ and $u_i$, respectively; $\mathbf{X}$ is the $n$ by $k$ matrix with typical element $x_{ij}$, which is assumed to have rank equal to $k$, that is, there is no perfect multicollinearity; and $\boldsymbol{\beta}$ is the $k$-dimensional vector with typical element $\beta_j$.

The classical assumptions about the errors imply that their joint distribution can be written in the form

$$\mathbf{u} \sim N(\mathbf{0}_n, \sigma^2\mathbf{I}_n), \tag{1.4}$$

in which: $N(\boldsymbol{\mu}, \boldsymbol{\Omega})$ denotes the multivariate Normal distribution with *mean vector* $\boldsymbol{\mu}$ and *covariance matrix* $\boldsymbol{\Omega}$; $\mathbf{0}_n$ is the $n$-dimensional column vector with every element equal to zero; and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. These assumptions, combined with those made about the regressor terms, also imply that the joint distribution of the elements of $\mathbf{y}$ is given by

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n). \tag{1.5}$$

The parameters to be estimated are, therefore, the elements of $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \sigma^2)$.

Under classical assumptions, there are strong incentives to use the *ordinary least squares* (OLS) estimator for $\boldsymbol{\beta}$ because it is best unbiassed and also the maximum likelihood estimator (MLE). The OLS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{1.6}$$

and so (1.5) implies that

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta},\sigma^2(\mathbf{X}'\mathbf{X})^{-1}), \tag{1.7}$$

with $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)'$. The implied vector of OLS predicted values is denoted by

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{1.8}$$

using (1.6).

In (1.8), pre-multiplication of $\mathbf{y}$ by $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ produces $\hat{\mathbf{y}}$, which is read as "y-hat." The $n$ by $n$ matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is sometimes referred to as the *hat-matrix* and is denoted by $\mathbf{H}$, that is,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{1.9}$$

The diagonal elements of $\mathbf{H}$, denoted by $h_{ii}, i = 1, \ldots, n$, are called the *leverage values* in the literature on diagnostics for regression models. By combining (1.7) and (1.8), it can be seen that, in the classical framework,

$$\hat{\mathbf{y}} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{H}), \tag{1.10}$$

in which $\mathbf{H}$ is a matrix that is symmetric and idempotent, having rank equal to $k$.

It remains to estimate the error variance $\sigma^2$. The errors are not observed but their variability can be estimated by using the OLS residuals as proxies. The OLS residuals are the elements of

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I}_n - \mathbf{H})\mathbf{y} = \mathbf{M}\mathbf{y} = \mathbf{M}\mathbf{u}, \tag{1.11}$$

in which $\mathbf{M} = \mathbf{I}_n - \mathbf{H} = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ has rank equal to $(n-k)$. Like $\mathbf{H}$, $\mathbf{M}$ is a symmetric, idempotent matrix; so that (1.4) implies

$$\hat{\mathbf{u}} \sim N(\mathbf{0}_n, \sigma^2(\mathbf{I}_n - \mathbf{H})), \tag{1.12}$$

with a typical OLS residual having a Normal distribution according to

$$\hat{u}_i \sim N(0, \sigma^2(1 - h_{ii})). \tag{1.13}$$

The residual sum of squares (*RSS*) from OLS estimation is

$$RSS = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}. \tag{1.14}$$

Under the assumptions of the classical regression model, the distribution of *RSS* is given by

$$RSS \sim \sigma^2\chi^2(n-k), \tag{1.15}$$

in which $n-k$ is the number of degrees of freedom associated with the estimation of (1.3). It follows from properties of the $\chi^2$ distribution that if $s^2$ is defined by

$$s^2 = \frac{RSS}{(n-k)}, \tag{1.16}$$

then $s^2$ is unbiassed and consistent for $\sigma^2$. The MLE of $\sigma^2$ is given by

$$\hat{\sigma}^2 = \frac{RSS}{n} = \frac{(n-k)}{n} \cdot \frac{RSS}{(n-k)}, \tag{1.17}$$

and so is consistent, but not unbiassed.

It will be assumed that $\hat{\boldsymbol{\beta}}$ of (1.6) and $s^2$ of (1.16) are to be used for the estimation of $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \sigma^2)$, whether or not the restrictive assumptions of nonrandom regressors and Normally distributed errors are made. In addition to the unrestricted estimation of the elements of $\boldsymbol{\theta}$, there is often interest in testing restrictions that reduce the number of elements of $\boldsymbol{\beta}$ that require estimation. Such restrictions can take many forms. If the restrictions are linear, that is, they specify the values of known linear combinations of the regression coefficients, the assumptions of the classical model permit the application of tests that are exactly valid. In such a case, let the restrictions to be tested be written as the null hypothesis

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}, \tag{1.18}$$

in which $\mathbf{R}$ is a known $q$ by $k$, $q \leq k$, matrix with rank equal to $q$ and $\mathbf{r}$ is a known $q$-dimensional vector.

The alternative hypothesis is assumed to be

$$H_1 : \beta_1, \ldots, \beta_k \text{ are unrestricted.}$$

The OLS estimator $\hat{\boldsymbol{\beta}}$ of (1.6) minimizes the residual sum of squares

$$Q(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

under $H_1$, and will be called the *unrestricted estimator*. The elements of $\hat{\mathbf{u}}$ will be referred to as the *unrestricted residuals*. It is convenient to add to the notation by using $RSS(H_1)$ to stand for the unrestricted residual sum of squares, that is, the quantity defined by (1.14) and to denote the number of degrees of freedom for the unrestricted model by $df(H_1)$.

The estimator that minimizes $Q(\boldsymbol{\beta})$ subject to $H_0$, that is, subject to the restrictions of $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, will be called the *restricted estimator* and is denoted by $\tilde{\boldsymbol{\beta}}$. The *restricted residuals* are defined by

$$\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}, \tag{1.19}$$

and the restricted residual sum of squares is written as

$$RSS(H_0) = \tilde{\mathbf{u}}'\tilde{\mathbf{u}}.$$

The standard $F$-statistic for testing $H_0$ against $H_1$ can then be calculated as

$$F = \frac{RSS(H_0) - RSS(H_1)}{RSS(H_1)} \cdot \frac{df(H_1)}{q}, \tag{1.20}$$

where, in this case, $df(H_1) = n-k$. When the null hypothesis is true and the classical assumptions are satisfied, $F$ of (1.20) has the F distribution with $q$ and $df(H_1)$ degrees of freedom. This result, which is exactly valid, is written as

$$F \sim F(q, df(H_1)),$$

under $H_0$. Large values of the test statistic in (1.20) indicate that there is strong evidence against $H_0$, so that a one-sided test should be conducted. If the required significance level is $\alpha$, the decision rule can be written as:

$$\textit{reject } H_0 \textit{ if } F \geq f(\alpha; q, df(H_1)), \tag{1.21}$$

in which the *critical value* $f(.)$ is determined by

$$\Pr(F(q, df(H_1)) \leq f(\alpha; q, df(H_1))) = 1 - \alpha.$$

If there is a single linear restriction to be tested, there is an alternative to calculating the $F$-statistic of (1.20). Suppose that the null hypothesis has the form $H_0 : \mathbf{R}\boldsymbol{\beta} = r_1$, where $\mathbf{R}$ is the row vector $(R_{11}, \ldots, R_{1k})$ and $r_1$ is a specified scalar, and the alternative hypothesis is $H_1 : \mathbf{R}\boldsymbol{\beta} \neq r_1$. With this combination of a single restriction in $H_0$ and a two-sided alternative, the reference distribution for the $F$-test is $F(1, df(H_1))$. A random variable with the same distribution is the square of a random variable that has the Student $t$ distribution with $df(H_1)$ degrees of freedom. This relationship is denoted by

$$F(1, df(H_1)) = [t(df(H_1))]^2.$$

It follows that a test of a single restriction against a two-sided alternative can be based upon the $t$-ratio defined by

$$t = \frac{\mathbf{R}\hat{\boldsymbol{\beta}} - r_1}{SE(\mathbf{R}\hat{\boldsymbol{\beta}} - r_1)}, \tag{1.22}$$

in which $SE(.)$ denotes the estimated standard error, that is,

$$SE(\mathbf{R}\hat{\boldsymbol{\beta}} - r_1) = \sqrt{s^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'}.$$

Under the assumptions of the classical regression model, a two-sided $t$-test with significance level $\alpha$ can be based upon the decision rule

$$\textit{reject } H_0 \textit{ if } |t| \geq t(\alpha/2; df(H_1)), \tag{1.23}$$

in which $\Pr(t(df(H_1)) \leq t(\alpha/2; df(H_1))) = 1-\alpha/2$. This two-sided $t$-test is equivalent to the $F$-test, with the sample values of test statistics obeying $t^2 = F$.

If there is a priori (non-sample) information about the sign of $\mathbf{R}\boldsymbol{\beta} - r_1$ when $H_0$ is false, a one-sided $t$-test can be applied in the usual way. With $H_1^+ : \mathbf{R}\boldsymbol{\beta} > r_1$, the decision rule is

$$\textit{reject } H_0 \textit{ if } t \geq t(\alpha; df(H_1)), \tag{1.24}$$

and with $H_1^- : \mathbf{R}\boldsymbol{\beta} < r_1$, it is

$$\textit{reject } H_0 \textit{ if } -t \geq t(\alpha; df(H_1)), \tag{1.25}$$

where $\Pr(t(df(H_1)) \leq t(\alpha; df(H_1))) = 1 - \alpha$.

Rules (1.21), (1.23), (1.24) and (1.25) have all been written so that the rejection region is in the right-hand tail of the relevant reference distribution. It is convenient, for the subsequent discussions, to assume that all tests are set up in this form. Some diagnostic checks, for example, the widely-used *test for heteroskedasticity* proposed in Breusch and Pagan (1979), involve the use of criteria that are asymptotically distributed as $\chi^2$ under the null hypothesis. The rejection region for such tests are, as with those given above, in the right-hand tail.

It is worth noting that, as an alternative to a $\chi^2$-form, many diagnostic checks can be computed as seemingly conventional tests of the significance of artificial (constructed) variables that are added to the regressors of (1.1). For example, tests for autocorrelation, structural change, errors-in-variables etcetera can be computed using standard formulae for $F$ or $t$ statistics, which are applied to an appropriate artificial regression model; see Davidson and MacKinnon (2004, section 15.2) for a general discussion. In such cases, (1.1) is viewed as the *restricted (null) model*. The nature of the *unrestricted (alternative) model*, which contains the restricted model (1.1) as a special case, has important implications for the properties of the test of the latter against the former. The unrestricted model required for the convenient calculation of a diagnostic check is often such that

the $F$ and $t$ tests are not exactly valid even when the classical Normal regression model (1.1) is the correct specification.

The problems associated with appealing to classical finite sample theory in the context of testing for misspecification can be illustrated by considering the well-known *Breusch-Godfrey Lagrange Multiplier* (LM) test for autocorrelation; see Breusch (1978) and Godfrey (1978). Suppose that quarterly data are being used and that the researcher believes that it is useful to test the assumption that the errors are independent against the fourth-order alternative

$$u_i = \phi_1 u_{i-1} + \cdots + \phi_4 u_{i-4} + \epsilon_i,$$

with the variates $\epsilon_i$ being NID$(0, \sigma_\epsilon^2)$. The required Breusch-Godfrey test can be implemented by applying the $F$-test of the four linear restrictions $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 0$ in the augmented version of (1.1) given by

$$y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \sum_{j=1}^{4} \phi_j \hat{u}_{i-j} + u_i, \tag{1.26}$$

where terms $\hat{u}_{i-j}$ with $i \leq j$ are set equal to zero. Even under the restrictive assumption that the errors $u_i$ are NID$(0, \sigma^2)$, the $F$-test of (1.1) against (1.26) is not exactly valid, but does have a significance level that tends to the required level $\alpha$ as $n \longrightarrow \infty$, that is, it is asymptotically valid.

The failure of standard finite sample theory to apply to the $F$-test of (1.1) against (1.26) might be anticipated on the grounds that the regressors of the latter, which serves as the alternative or unrestricted model, include random variables, namely, the lagged residuals. However, there are cases of diagnostic checks in which $F$-tests are exact even though the regressors of the alternative model include random variables. An important example is the *RESET test* proposed in Ramsey (1969).

The RESET test provides a check of the specification of the mean function of (1.1), with the OLS predicted values from estimation of this model being employed to obtain the additional regressors required for the alternative model. More precisely, in the formula for the RESET $F$-statistic with $q$ test variables, $RSS(H_0)$ is derived from OLS estimation of (1.1), that is, it is given by (1.14), and $RSS(H_1)$ is obtained after estimation of the artificial model

$$y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \sum_{j=1}^{q} \hat{y}_i^{j+1}\delta_j + u_i, i = 1, \ldots, n, \tag{1.27}$$

with $df(H_1) = n - k - q$. The $F$-statistic for testing $\delta_1 = \cdots = \delta_q = 0$ in (1.27) is denoted by $F_R$ and

$$F_R \sim F(q, n - k - q),$$

under the null hypothesis, when the assumptions of the classical model concerning $\mathbf{X}$ and $\mathbf{u}$ hold. Consequently, under these assumptions, it is possible to have perfect control of finite sample significance levels of the RESET test. This result follows from a general property of tests involving functions of $\hat{\mathbf{y}}$; see Milliken and Graybill (1970).

Notwithstanding the interest to theorists of results such as those in Milliken and Graybill (1970) and also in Stewart (1997), there is a need to weaken the assumptions of the classical model and to see what can be established about the properties of tests under more general conditions.

## 1.3. Tests for linear regression models under weaker assumptions: random regressors and non-Normal IID errors

From the viewpoint of the applied econometrician, the results concerning the exact validity of the $F$ and $t$ tests in the classical linear regression model are of doubtful relevance. The assumption that the regressors are non-random and would be fixed if repeated sampling were possible may well be appropriate for the analysis of data obtained, for example, from experiments in a laboratory. However, in econometric models, the regressors will usually include economic variables that are properly regarded as random. Thus, in general, the regressor set will include both random and non-random terms. The applicability of the results of the previous section is now open to question.

Suppose first that the following two conditions hold: the regressors are such that any random term $x_{ij}$ is independent of $u_m$ for $i, m = 1, \ldots, n$; and the errors $u_m$ are NID$(0, \sigma^2)$ for $m = 1, \ldots, n$. When the first of these conditions is satisfied, the regressors are said to be *strictly exogenous* or, less precisely, *exogenous*. The complete independence of errors and regressors implies that conditioning on regressor values has no impact on the distribution of the errors. Consequently, given the two conditions, we can write the conditional error distribution as

$$\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}_n, \sigma^2 \mathbf{I}_n), \tag{1.28}$$

and for the conditional distribution of the dependent variable, given the regressor values, we have

$$\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n). \tag{1.29}$$

Comparison of (1.4) and (1.5) with (1.28) and (1.29), respectively, indicates how results given in the previous section for the former pair of equations will now apply in a conditional sense under the latter pair. In particular, when testing restrictions of the form (1.18), the $F$ statistic of (1.20) will have the conditional distribution

$$F|\mathbf{X} \sim F(q, df(H_1)), \tag{1.30}$$

under the null hypothesis. This conditional distribution is completely characterized by the values of $q$ and $df(H_1)$, but neither of these items depends upon $\mathbf{X}$. Hence, when the null hypothesis is true, the unconditional distribution is the same as the conditional distribution in (1.30), which is the same as the reference distribution appropriate for the classical model. The $F$-test is, therefore, exactly valid. Similar arguments apply to the $t$-test.

However, the conditions that underpin this argument are very restrictive. The assumption that all regressors are strictly exogenous is inconsistent with the common practice of including lagged values of the dependent variable as explanatory variables when estimating regression models using time series data. For example, the standard partial adjustment model leads to the inclusion of $y_{i-1}$ as a regressor and this regressor cannot be independent of all past errors (obviously $E(y_{i-1}u_{i-1}) \neq 0$). Moreover, there is rarely precise information available about the shape of the error distribution and, in particular, there seems little reason to believe that the errors are Normally distributed, even if they are assumed to be IID.

If the assumption of Normally distributed errors is relaxed, tests involving the use of critical values from standard distributions must, in general, be based upon asymptotic theory. Appeal has to be made to versions of a Law of Large Numbers and a Central Limit Theorem (CLT). Discussions of these topics and their application to tests for regression models can be rather technical and readers are referred to Davidson (1994), McCabe and Tremayne (1993), and White (1984) for detailed treatments. For the purpose of providing an outline of the relevant arguments of asymptotic theory, it is useful to introduce the ideas of *orders of magnitude* for random variables due to Mann and Wald (1943).

Given a sequence of real random variables, denoted by $\{S_{(n)}\}$, and some real number $a$, we say that $S_{(n)}$ is of order of probability $n^a$ if for any $\varepsilon > 0$ there exists $b_\varepsilon > 0$ such that

$$\Pr(-b_\varepsilon \leq n^{-a} S_{(n)} \leq b_\varepsilon) \geq 1 - \varepsilon,$$

for all $n$. The standard notation for such a variable is to write $S_{(n)} = O_p(n^a)$. If, for some real number $c$, $p\lim n^{-c} S_{(n)} = 0$, we say that $S_{(n)}$ is of smaller order of probability than $n^c$ and write $S_{(n)} = o_p(n^c)$.

For example, assume that the observations $y_1, \ldots, y_n$ are NID$(\mu, \sigma^2)$ and $S_{(n)} = y_1 + \cdots + y_n$. Since, in this case, $S_{(n)}$ is $N(n\mu, n\sigma^2)$, it follows that: (i) $S_{(n)} = O_p(n)$ with $plim\ n^{-1} S_{(n)} = \mu$; (ii) $S_{(n)} - n\mu$ is $O_p(n^{1/2})$ with $n^{-1/2}(S_{(n)} - n\mu)$ being $N(0, \sigma^2)$; and (iii) $S_{(n)} - n\mu$ is $o_p(n)$ with $n^{-1}(S_{(n)} - n\mu)$ being $N(0, n^{-1}\sigma^2)$ so that $plim\ n^{-1}(S_{(n)} - n\mu) = 0$.

In standard textbook discussions of linear regression models, assumptions are made that imply that $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + O_p(n^{-1/2})$, with $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ being asymptotically Normally distributed with zero mean vector and finite, positive-definite covariance matrix. Strictly speaking, the notation used in the discussion of asymptotic theory for regression models should reflect the dependence of estimators and test statistics on the sample size $n$, for example, $\hat{\boldsymbol{\beta}}_{(n)}$ rather than $\hat{\boldsymbol{\beta}}$. However, no confusion should be caused by adopting the less cluttered style employed above and the key results can be summarized as follows. First, when using $F$ of (1.20) to test the null hypothesis of (1.18), asymptotic theory predicts that, when the null is true, $F$ is $O_p(1)$ with

$$F \sim_a \frac{\chi^2(q)}{q},$$

in which $\sim_a$ is used as a shorthand for "is asymptotically distributed as". Second, if $q = 1$, the $t$-ratio of (1.22) is $O_p(1)$ and is asymptotically distributed as $N(0, 1)$ when the null hypothesis is true.

Asymptotic theory can also be used as a source of approximations to the behaviour of test statistics when the null hypothesis is false. Consider the case of testing a single restriction, which is written as $H_0 : \mathbf{R}\boldsymbol{\beta} = r_1$, as above. The relevant $t$-statistic can be written as

$$\frac{\mathbf{R}\hat{\boldsymbol{\beta}} - r_1}{SE(\mathbf{R}\hat{\boldsymbol{\beta}} - r_1)} = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})}{SE(\mathbf{R}\hat{\boldsymbol{\beta}} - r_1)} + \frac{(\mathbf{R}\boldsymbol{\beta} - r_1)}{SE(\mathbf{R}\hat{\boldsymbol{\beta}} - r_1)}, \tag{1.31}$$

in which the first term on the right-hand side of (1.31) tends to $N(0, 1)$, whether or not $H_0$ is true, but the asymptotic behaviour of the second

term does depend upon the value of $\mathbf{R}\boldsymbol{\beta} - r_1$. If $H_0$ is true, $\mathbf{R}\boldsymbol{\beta} - r_1 = 0$ and the second term vanishes. If $\mathbf{R}\boldsymbol{\beta} - r_1$ is a fixed nonzero number (so that $H_0$ is untrue), the second term on the right-hand side of (1.31) is $O_p(n^{1/2})$, under the standard assumptions of asymptotic theory for regression models. (The standard errors of OLS estimators are $O_p(n^{-1/2})$, given these assumptions.) Hence, as $n \longrightarrow \infty$, the $t$-statistic goes to $\pm\infty$, according to the sign of the nonzero constant $\mathbf{R}\boldsymbol{\beta} - r_1$. Thus, with *fixed alternatives* $H_1 : \mathbf{R}\boldsymbol{\beta} - r_1 \neq 0$, asymptotic theory cannot lead to the limit of a proper distribution with finite mean and variance as a basis for approximating the behaviour of the $t$-statistic. A device known as a *sequence of local alternatives*, or as *Pitman drift*, does allow asymptotic theory to provide such an approximation for the study of power; see, for example, Godfrey and Tremayne (1988).

The device is to introduce the sequence of alternatives

$$H_{1n} : \mathbf{R}\boldsymbol{\beta} - r_1 = \frac{\lambda}{\sqrt{n}}, |\lambda| < \infty, \tag{1.32}$$

which clearly tends to the null hypothesis as $n$ increases. The second term of (1.31) is, under (1.32), given by

$$\frac{\lambda}{\sqrt{n}SE(\mathbf{R}\hat{\boldsymbol{\beta}} - r_1)},$$

which tends to a finite constant, say $\mu_\lambda$. Consequently, under the local alternatives assumption, the asymptotic distribution of the $t$-ratio can be written as

$$\frac{\mathbf{R}\hat{\boldsymbol{\beta}} - r_1}{SE(\mathbf{R}\hat{\boldsymbol{\beta}} - r_1)} \sim_a N(\mu_\lambda, 1),$$

and this distribution satisfies the requirements to have finite mean and variance. Local alternatives are often used when researchers seek to choose between two or more asymptotically valid tests on the basis of their sensitivity to departures from the null hypothesis. A similar result is available when the $F$-test is used to check several restrictions.

Several researchers, while acknowledging a reliance on asymptotic theory, prefer to use the conventional $F(q, df(H_1))$ and $t(df(H_1))$ distributions for critical values, rather than the corresponding limiting forms of $\chi^2(q)/q$ and $N(0, 1)$. There may be reason to be concerned about the relevance of asymptotic theory if $df(H_1)$ is not large enough for the choice between, for example, $t(df(H_1))$ and $N(0, 1)$ to be unimportant. Indeed, from a practical point of view, a question of real interest is how large does

the sample size have to be before a CLT will give a useful approximation for controlling the significance level when testing the null hypothesis. Unfortunately there is no generally valid answer.

The robustness of the standard regression $F$-test to *non-Normality* of the errors is investigated in Ali and Sharma (1996). In addition to the sample size, degrees of freedom and the actual distribution of the errors, important determinants of the robustness to non-Normality are the non-Normality of the regressors and the presence of observations with relatively high leverage values. The relevance of such characteristics of the regressor set is not surprising, given the dependence of the test statistic on OLS residuals and the form of (1.11). In view of the uncertain quality of the approximation provided by asymptotic theory in the case of a linear regression model with IID, but non-Normal, errors and the evidence that the approximation is sometimes poor, it is natural to look for an alternative approach to testing. Chapter 2 contains a discussion of a simulation-based approach that can be applied in the context of linear regression models with IID errors. However, like the standard $t$ and $F$ tests, these simulation techniques may produce misleading inferences when the errors of (1.1) are not IID, that is, the data are generated by a *generalized regression model*.

## 1.4. Tests for generalized linear regression models

The generalized regression model with exogenous regressors is derived by combining the model in (1.3), that is,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u},$$

with

$$E(\mathbf{u}|\mathbf{X}) = \mathbf{0},$$

and

$$E(\mathbf{u}\mathbf{u}'|\mathbf{X}) = \sigma^2\boldsymbol{\Omega}, \tag{1.33}$$

in which $\boldsymbol{\Omega}$ is an $n$ by $n$ matrix that is symmetric and positive definite. If the errors are independent but heteroskedastic, the elements of $\boldsymbol{\Omega}$ are such that: $\omega_{ij} = 0$ if $i \neq j$; and $\omega_{ii} \neq \omega_{jj}$ for some $i \neq j$. If the errors are correlated but homoskedastic, the elements of $\boldsymbol{\Omega}$ are such that: $\omega_{ii} = \omega_{jj} = 1$, say, for all $i$ and $j$; and $\omega_{ij} \neq 0$ for some $i \neq j$. In the latter case, it is assumed that there are time series data and the errors are autocorrelated.

(Tests can be developed for models with spatial correlation; see Anselin, 2006.) It will be assumed that autocorrelated errors are generated by (weakly) stationary processes so that $\omega_{ij}$ depends upon $|i-j|$, rather than on $i$ and $j$ separately. For example, if the errors were generated by a stationary first-order autoregression

$$u_i = \phi u_{i-1} + \epsilon_i, |\phi| < 1, \epsilon_i \quad NID(0, \sigma_\epsilon^2),$$

a typical element of $\mathbf{\Omega}$ in (1.33) would be $\omega_{ij} = \phi^{|i-j|}$.

The OLS estimator of $\boldsymbol{\beta}$, under the assumptions of the generalized regression model, has conditional mean vector

$$E(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta},$$

and conditional covariance matrix given by

$$\mathbf{V}_G(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{1.34}$$

In general, the matrix of (1.34) is not equal to the one that appears in (1.7) and so the tests described above cannot be expected to be asymptotically valid.

In some special models, the elements of $\mathbf{\Omega}$ are known constants. For example, if each element of $\mathbf{u}$ is the sum of a known number of basic IID disturbances, $\mathbf{\Omega}$ can be calculated very simply; see Rowley and Wilton (1973) for an example based upon the "four-quarter overlapping-change" model in wage analysis. When $\mathbf{\Omega}$ is known, the OLS estimator can be replaced by the more efficient Generalized Least Squares (GLS) estimator

$$\check{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}, \tag{1.35}$$

which has conditional covariance matrix equal to $\sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}$. The estimator of $\sigma^2$ is no longer given by $s^2$ of (1.16) but is now defined by

$$\check{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\check{\boldsymbol{\beta}})'\mathbf{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\check{\boldsymbol{\beta}})}{(n-k)}.$$

Given $\check{\boldsymbol{\beta}}$ and $\check{\sigma}^2$, an asymptotically valid test of $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ in (1.18) can be based upon the result that, when $H_0$ is true, the *Wald statistic*

$$W_{GLS} = (\mathbf{R}\check{\boldsymbol{\beta}} - \mathbf{r})' \left[\check{\sigma}^2\mathbf{R}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}'\right]^{-1} (\mathbf{R}\check{\boldsymbol{\beta}} - \mathbf{r}), \tag{1.36}$$

is asymptotically distributed as $\chi^2(q)$; see, for example, Greene (2008, ch. 8, section 8.3.1) for the corresponding asymptotically valid $F$-statistic.

Significantly large values of $W_{GLS}$ indicate that the restrictions of $H_0$ : $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ are not consistent with the sample data.

Unfortunately, the test statistic of (1.36) is rarely available in practical situations because, in general, $\boldsymbol{\Omega}$ is unknown and it is not feasible to calculate the GLS estimator $\breve{\boldsymbol{\beta}}$.

When the elements of $\boldsymbol{\Omega}$ are continuous functions of the elements of an unknown parameter vector $\boldsymbol{\psi}$, estimates of the parameters of the generalized regression model can be obtained by minimizing the *Nonlinear Least Squares* (NLS) criterion

$$Q_{NLS}(\boldsymbol{\beta}, \boldsymbol{\psi}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'[\boldsymbol{\Omega}(\boldsymbol{\psi})]^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$. Alternatively, if some consistent estimator of $\boldsymbol{\psi}$, denoted by $\hat{\boldsymbol{\psi}}$, is available and necessary regularity conditions are satisfied, $\boldsymbol{\beta}$ can be estimated by minimizing the *Feasible Generalized Least Squares* (FGLS) function

$$Q_{FGLS}(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'[\boldsymbol{\Omega}(\hat{\boldsymbol{\psi}})]^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

However, both of these estimation methods are based upon the assumptions that: (i) there is a parametric model that determines the structure of $\boldsymbol{\Omega}$; and (ii) the general form of $\boldsymbol{\Omega}(\boldsymbol{\psi})$ is known, with only its finite-dimensional parameter vector $\boldsymbol{\psi}$ being unknown. While economics might be a source of useful information about the mean function of $\mathbf{y}$, there is little reason to suppose that applied workers will know the form of, for example, heteroskedasticity. Thus it will often be difficult to have confidence in an assumed error model.

Misspecification of the model for autocorrelation and/or heteroskedasticity will, in general, lead to an inconsistent estimator of the covariance matrix of the minimizers of $Q_{NLS}(\boldsymbol{\beta}, \boldsymbol{\psi})$ and $Q_{FGLS}(\boldsymbol{\beta})$. Hence errors made in modelling $\boldsymbol{\Omega}$ may imply misleading outcomes of tests of hypotheses such as (1.18), because such tests use the estimated covariance matrix to assess the significance of sample outcomes. An investigation of the effects of misspecifying the model for heteroskedasticity is reported in Belsley (2002). It is found that effects can be serious and Belsley concludes that

> Correction for heteroskedasticity clearly does best when both the proper arguments and the proper form of the skedasticity function are known. But this is an empty conclusion since misspecification is probably the rule. (Belsley, 2002, p. 1396)

Moreover, it has been argued that, even with correct specification of the model underlying $\boldsymbol{\Omega}$, it is not clear that FGLS is superior to OLS in finite samples because of the extra variability associated with the estimation of $\boldsymbol{\psi}$; see, for example, Greene (2008, p. 158).

In view of these findings, it is not surprising that there has been an interest in deriving tests using the uncorrected OLS estimator $\hat{\boldsymbol{\beta}}$ and an appropriate estimator of its covariance matrix, which is no longer given by the *IID-valid* formula $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ used in (1.7). If the errors are assumed to be independent and heteroskedastic, a *Heteroskedasticity-Consistent Covariance Matrix Estimator* (usually denoted by *HCCME*) is required. If the errors are heteroskedastic and autocorrelated, a *Heteroskedasticity and Autocorrelation Consistent* (usually denoted by *HAC*) estimator is needed. The former provides standard errors that are *heteroskedasticity-robust*. The latter provides standard errors that are *heteroskedasticity and autocorrelation robust*.

Many computer programs offer users the chance to use robust standard errors from either some HCCME or some HAC estimate, rather than relying on the traditional IID-valid standard errors given by the matrix $s^2(\mathbf{X}'\mathbf{X})^{-1}$. However, the traditional standard errors are often provided as the default and this approach has been criticized. Stock and Watson remark that

> In econometric applications, there is rarely a reason to believe that the errors are homoskedastic and normally distributed. Because sample sizes are typically large, however, inference can proceed . . . by first computing the heteroskedasticity-robust standard errors. (Stock and Watson, 2007, p. 171)

Similarly, it is argued in Hansen (1999) that a modern approach should involve the use of test statistics that are valid under heteroskedasticity and do not require the assumption of Normality. (It is also suggested in Hansen (1999) that applied workers should think about using the bootstrap for inference, rather than relying on asymptotic theory. Much of what follows in this book is concerned with presenting evidence to support this suggestion and to help empirical researchers to select the appropriate form of the bootstrap.)

Since the use of procedures based upon HCCME and HAC estimates offers the chance to derive tests that are asymptotically valid in the presence of unspecified forms of departure from the assumption of IID errors, such robust tests are of real interest in practical applications. Moreover, the availability of suitable software means that there is no important

obstacle to hinder the use of robust tests. Given the potential importance of these alternatives to the conventional IID-based asymptotic $t$ and $F$ tests, each will discussed.

### 1.4.1. HCCME-based tests

Suppose first that the errors $u_i$ are independently distributed with zero means and variances $\sigma_i^2, i = 1, \ldots, n$, with all variances being finite and positive. It is not assumed that there is any precise information available to support the specification of a parametric model of the heteroskedasticity. The tests that are to be discussed are asymptotically robust to heteroskedasticity of unspecified form and are also asymptotically valid under the classical assumption of homoskedasticity. The key results for HCCME-based inference in linear regression models will now be discussed.

If the regressors were not random and the errors were Normally distributed, the OLS estimator would, in the presence of unspecified forms of heteroskedasticity, have the following distribution

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}),$$

or equivalently,

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N(\mathbf{0}_k, n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}), \tag{1.37}$$

in which $\boldsymbol{\Sigma}$ is the $n$ by $n$ diagonal matrix with the variances $\sigma_i^2, i = 1, \ldots, n$, as the nonzero elements on its leading diagonal. The random vector $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is $O_p(1)$, with the covariance matrix that appears in (1.37) being assumed to tend to a finite positive-definite matrix as $n \longrightarrow \infty$. This property of the covariance matrix is more easily seen when it is noted that

$$n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}}{n}\right)\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}.$$

The covariance matrix that appears in (1.37) is sometimes referred to as a *sandwich covariance matrix*; the term depending on error variances, that is, $\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}$, being sandwiched between the two terms equal to $(\mathbf{X}'\mathbf{X})^{-1}$. The problem of finding useful estimates for the sandwich form in order to develop methods for feasible inference was studied in the statistics literature, for example, Eicker (1967). However, interest and applications in econometrics were stimulated by an important paper by White who relaxed the assumptions of fixed regressors and Normally distributed errors; see White (1980).

White showed that, under suitable regularity conditions, the OLS estimator $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$, with $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ being $O_p(n^{-1/2})$ and $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ having a limiting distribution ( as $n \longrightarrow \infty$) given by

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim_a N(\mathbf{0}_k, plim \quad n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}). \tag{1.38}$$

It is (1.38) that provides the basis for asymptotically valid heteroskedasticity-robust tests. If the null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ is to be tested, we can use the result that (1.38) implies that

$$n^{1/2}\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim_a N(\mathbf{0}_q, plim \quad n\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'),$$

and so, if the null hypothesis is true,

$$n^{1/2}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim_a N(\mathbf{0}_q, plim \quad n\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}').$$

Consequently, if the restrictions of $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ are valid, standard asymptotic theory implies that

$$n(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'[plim \quad n\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim_a \chi^2(q).$$

However, this result does not yield a feasible test procedure because it concerns a random variable that depends upon the probability limit of a matrix that is, in part, determined by the unknown matrix $\boldsymbol{\Sigma}$.

White provided a very simple and convenient solution to the problem of deriving a feasible large sample test. In White (1980), it is shown that, under certain regularity conditions that place mild restrictions on the behaviour of errors and random regressors,

$$plim \quad n^{-1}\mathbf{X}'\dot{\boldsymbol{\Sigma}}\mathbf{X} = plim \quad n^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}, \tag{1.39}$$

in which $\dot{\boldsymbol{\Sigma}}$ is obtained from $\boldsymbol{\Sigma}$ by replacing the unknown variance $\sigma_i^2$ by the calculable squared OLS residual $\hat{u}_i^2, i = 1, \ldots, n$. Consequently feasible and asymptotically robust tests can be derived by using the heteroskedasticity-consistent estimator

$$HC0 = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\dot{\boldsymbol{\Sigma}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \tag{1.40}$$

for the covariance matrix that appears in (1.38). A heteroskedasticity-robust test of $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ can then be based upon the statistic

$$W_{HC0} = n(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'\left[n\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\dot{\boldsymbol{\Sigma}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim_a \chi^2(q),$$

with significantly large values of $W_{HC0}$ indicating the data inconsistency of the linear restrictions of $H_0$.

There is a large literature on the construction and analysis of heteroskedasticity-robust tests for regression models and a summary will be given in Chapter 6. However, it is worth noting that statistics that are asymptotically equivalent to $W_{HC0}$, that is, differ from it by terms that are $o_p(1)$, can be obtained by modifying $HC0$ of (1.40). Three modifications are often discussed. First, a simple degrees-of-freedom adjustment is employed, which leads to

$$HC1 = (n-k)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\dot{\mathbf{\Sigma}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{1.41}$$

The second and third standard modifications both involve taking the leverage values $h_{ii}$ (see (1.9) above) into account, with the estimators being defined by

$$HC2 = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\ddot{\mathbf{\Sigma}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \tag{1.42}$$

and

$$HC3 = n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\dddot{\mathbf{\Sigma}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \tag{1.43}$$

in which $\ddot{\mathbf{\Sigma}}$ and $\dddot{\mathbf{\Sigma}}$ are derived from $\dot{\mathbf{\Sigma}}$ by replacing the terms $\hat{u}_i^2$ by $(1-h_{ii})^{-1}\hat{u}_i^2$ and $(1-h_{ii})^{-2}\hat{u}_i^2, i = 1, \ldots, n$, respectively. Clearly $HC0$ and $HC1$ have the same probability limit, with

$$HC1 = \frac{(n-k)}{n} \cdot HC0 = HC0 + O_p(n^{-1}),$$

so that $(HC1 - HC0)$ is asymptotically negligible relative to $HC0$. Similarly the differences $(HC2 - HC0)$ and $(HC3 - HC0)$ are also asymptotically negligible since each term $h_{ii}$ is $O_p(n^{-1})$, with $h_{11} + \cdots + h_{nn} = k$ for all $n \geq k$. An examination of these variants is provided in, for example, Long and Ervin (2000) and MacKinnon and White (1985).

Many textbooks point out that heteroskedasticity could be present when regression models are estimated using cross-section data. It is, therefore, not surprising that the assumption that the regressors are independently distributed over the observations is made in White (1980). However, while this assumption concerning the behaviour of regressors may often be appropriate for cross-section applications, it is too restrictive when time series regressions are estimated and heteroskedasticity certainly cannot be ruled out in such cases. Fortunately, it is possible to extend White's results by establishing that a HCCME can be obtained