Chapter 1

Introduction

Every scientific analysis builds upon a set of implicit or explicit assumptions. The validity of these assumptions can have important consequences for the validity of the final research results. Results are called *robust* if they are not affected much by (small) changes in the assumptions.

Robustness is an intriguing subject from both a theoretical and a practical point of view. From a theoretical perspective, the question of robustness stimulates researchers to determine the crucial assumptions underlying their results. The practical relevance of robustness is easily illustrated by considering, for example, the development of policy recommendations based on economic models. If the advice alters dramatically when the model or the assumptions underlying the model are changed only slightly, the policy maker might be just as well off as without any advice.

The above interpretation of the concept of robustness is generally applicable. In this thesis, however, I consider a more restricted notion of robustness. In order to explain the exact contents of this notion, I first restrict the general concept of robustness to that of statistical robustness. This is done in Section 1.1. In Section 1.2, I relate statistical robustness to the topic of outlier robustness, a theme that is underlying all remaining chapters of this thesis. In Section 1.3, I give some possible reasons for the present lack of interaction between robust statistics, especially outlier robust statistics, and econometrics. Finally, in Section 1.4, I present a brief summary of the remaining chapters of this thesis and give some notational conventions that are used throughout this book.

1.1 Statistical Robustness

Robustness concerns the sensitivity of one's results to deviations in the assumptions that are made to obtain those results. When defined like this, robustness covers the whole broad spectrum of scientific research. In this thesis I am mainly interested in the statistical aspects of certain econometric methods. Therefore, I restrict the general notion of robustness to that of statistical robustness.

Hampel et al. (1986, p. 6) provide the following definition of statistical robustness.

> In a broad informal sense, robust statistics is a body of knowledge, partly formalized into "theories of robustness," relating to deviations from idealized assumptions in statistics.

This definition is clearly covered by the more general definition presented earlier. Hampel et al. (1986) also give a second definition.

> Robust statistics, as a collection of related theories, is the statistics of approximate parametric models.

Although this definition captures most of the literature on robust statistics, it is much stricter than the first definition. It excludes, for example, most of the literature on semiparametric and nonparametric statistics. This is rather unsettling, because those branches of the statistical literature study procedures that are robust to, e.g., functional form and distributional misspecification. Therefore, I prefer the former definition to the latter.[1]

Several departures from idealized assumptions have already been dealt with in the literature. Much work has been done for the standard linear regression model. The effects of heteroskedasticity, autocorrelation, omitted variables, measurement error, and several other types of misspecification have been studied extensively (see, for example, standard econometric textbooks like Judge et al. (1988) and Davidson and MacKinnon (1993)). Procedures have been developed that give results (e.g., parameter estimates) that are robust to the mentioned departures. Some nice examples are the heteroskedasticity consistent covariance matrix estimator of White (1980) and Huber (1981) and its autocorrelation consistent extension by Newey and West (1987). These estimators can be used to construct inference procedures that are robust to certain types of heteroskedasticity and autocorrelation (see MacKinnon (1992) and the references cited therein).

Other types of robustness that receive much attention in the literature, nowadays, concern the robustness of estimation and inference procedures to distributional assumptions and functional form specification. Consider the following simple example. Assume that the model generating the data is

$$y_t = g(x_t^\top \beta) + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1.1}$$

where y_t and ε_t are scalars, x_t and β are k-dimensional column vectors, $^\top$ denotes transposition, $g(\cdot)$ is a smooth, real-valued function, and T denotes the number of observations. Assume that $\{\varepsilon_t\}_{t=1}^T$ is a set of independently and identically distributed (i.i.d.) random variables with zero mean and finite variance. If $g(\cdot)$ is the identity function, $g(x_t^\top \beta) = x_t^\top \beta$, then (1.1) is a standard linear regression model and the parameter β can be estimated by means

[1] Strictly speaking, the second definition also excludes the work in outlier robust nonparametric analysis as in, e.g., Boente and Fraiman (1990, 1991), Koenker and Ng (1992), and Wang and Scott (1994).

of the ordinary least-squares (OLS) estimator. If $g(\cdot)$ is unknown, however, it becomes difficult to estimate β. Therefore, researchers in the area of semiparametric statistics have developed tools that allow one to simultaneously estimate β and $g(\cdot)$. Conditions can be found under which the estimators for β and $g(\cdot)$ are consistent. These estimation procedures are, therefore, robust to the exact specification of $g(\cdot)$ in (1.1). Related to the estimation of the link function $g(\cdot)$ is the estimation of the density of the errors ε_t. One can estimate this density along with the other unknown quantities and functions in the model. In this way one can try to exploit certain properties of the error distribution without having to postulate a particular parametric form for it (see Manski (1984)). In the nonparametric context, researchers proceed even further and replace (1.1) by $y_t = g(x_t) + \varepsilon_t$. The objective then becomes to estimate the function $g(\cdot)$, which is now a function from $\mathbb{R}^k$ to $\mathbb{R}$ instead of from $\mathbb{R}$ to $\mathbb{R}$. For an introduction into nonparametric statistics, I refer to Härdle (1990). For the semiparametric literature, one can start, for example, with the review article of Robinson (1988) or the lecture notes of Stoker (1991).

Another important departure from idealized assumptions in statistics is that of the occurrence of outlying observations. This topic is discussed in the next section.

1.2 Outlier Robustness

Outlier robustness plays an important role in this thesis. As implied by the term itself, outlier robustness is concerned with the construction of procedures that are robust to the occurrence of outliers in the data. The term outlier is mostly used rather informally. It often denotes observations that do not fit into a postulated model or that have some other 'surprising' characteristics. Davies and Gather (1993) attempt to define the outlier concept more formally. One of their main conclusions is that outliers are defined with respect to a model. As a consequence, observations might be outliers in one model, and, at the same time, be perfectly regular observations in another model.

Consider the following example. One has a sample of ten drawings from a mixture of normals. With probability $(1-p)$, the drawing comes from the standard normal distribution, and with probability p, the drawing comes from a normal with mean 100 and unit variance. For $p = 0.1$, the sample could look something like this:

$$\begin{array}{ccccc} 0.619 & -0.126 & -0.641 & -0.687 & 100.562 \\ 0.493 & -0.681 & -0.904 & -0.148 & -0.159. \end{array}$$

Most researchers would be inclined to regard the observation of 100.562 to be an outlier, because it is extremely large compared with the other observations. Indeed, if we assume that the above sample consists of i.i.d. drawings from a standard normal distribution, then the probability of drawing a number like 100.562 is extremely low. The observation is, thus, an outlier with respect to the model that the ε_t are i.i.d. standard normal for $t = 1, \ldots, 10$. If one relaxes the i.i.d. assumption, however, and considers the model that really generated the data, then all ten observations are perfectly regular.

The simple example above illustrates two points. First, outliers are always identified with respect to a specific *benchmark* or *null* model. Second, there exists a tradeoff between model simplicity and the number of outliers one is willing to tolerate. If one selects a simple model, e.g., that the sample in the example above consists of i.i.d. drawings from a normal distribution with mean μ and variance σ^2, then observations are more easily classified as outliers than in a larger model, e.g., the mixture of normals in the example.

The objective underlying the techniques presented in this thesis is to construct a model for the majority of the data. This formalizes the common practice in econometric model building (see also Section 2.1). Given this modeling objective, outlier robustness is an important topic.[2] When confronted with the possibility of outliers in the data, one should employ procedures that are not affected much by these anomalous observations. This brings us into the realm of outlier robust statistics.

One of the key techniques in outlier robust statistics is to reweight observations on the basis of their 'outlyingness' with respect to the postulated model. If an observation does not fit the model, then less weight is assigned to that particular observation. In the extreme case, the observation might be discarded completely. The use of such a weighting strategy implies that one tries to find a model that 'adequately' captures the features of the *bulk* of the data. One might object that this is an unsound modeling philosophy and that one should, instead, look for a model that fits all of the observed data. Three arguments can be raised against this critique.

First, there are many conceivable situations in which describing the bulk of the data is the best one can achieve. This can, for example, be due to the lack of additional information on the observations that are identified as outliers. Alternatively, if the data set is known to contain many recording errors, discarding these errors seems a reasonable strategy for most modeling purposes. Moreover, in the example above with the mixture of normals, extremely large samples are needed in order to estimate the parameters of the second component of the mixture reliably if p is close to zero. In such cases, it might also be useful to disregard the observations from the second mixture component and to focus on modeling the remaining observations.

Second, finding a model for the bulk of the data can also be viewed as an intermediate goal in the process of model specification. Robust techniques sometimes provide stronger signals of model misspecification and more clearly indicate directions for model respecification than methods that accommodate all observations (see, e.g., Machado (1993) and Ronchetti and Staudte (1994)).

[2] Of course, also if one is reluctant to reweight or discard a part of the sample, robust methods are still useful tools for data analysis. For example, diagnostic measures based on robust estimators are good indicators of certain types of model failure (see also Subsection 1.2.2).

Third, also when trying to find a model for all available observations, one sometimes implicitly uses weighting procedures. For example, consider the simple location model

$$y_t = \mu + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{1.2}$$

where μ is the unknown location parameter, and the ε_t's are i.i.d. standard Cauchy. If one uses the maximum likelihood estimator in order to estimate μ, one is implicitly calculating the arithmetic mean of the reweighted sample $\{\tilde{y}_t\}_{t=1}^T$, where

$$\tilde{y}_t = y_t(1+\varepsilon_t^2)^{-1}/(T^{-1}\textstyle\sum_{i=1}^T(1+\varepsilon_i^2)^{-1}). \tag{1.3}$$

Thus, the asymptotically optimal estimator for the model that generated the data can in this case be reinterpreted as a simpler estimate, namely the arithmetic mean,[3] computed on a reweighted sample.

I conclude the discussion of outlier robustness by describing its relation to three other branches of the statistical literature. In Subsection 1.2.1, I briefly relate outlier robustness to semiparametric and nonparametric statistics. In Subsection 1.2.2, the relation to the diagnostics school is discussed. Finally, in Subsection 1.2.3, some remarks can be found on the link between outlier robustness and the common practice of including dummy variables in order to correct for anomalous observations.

1.2.1 Outlier Robustness and Semiparametrics or Nonparametrics

The goals of outlier robust and semiparametric statistics (as in, e.g., Robinson (1988) and Stoker (1991)) or nonparametric statistics (as in, e.g., Härdle (1990)) partially overlap. In all three research areas methods are studied that are robust to misspecification of the model. There are, however, two main differences. First, in the semiparametric and nonparametric literature almost no distributional assumptions are made. Using only some regularity conditions on the existence of certain moments of the distribution and on the smoothness of the functions that are studied, one is usually able to obtain quite strong results (see Manski (1984)). In the literature on outlier robustness, however, quite precise distributional assumptions are often made. Given these assumptions, one then tries to develop procedures that give satisfactory results even if the assumptions are only *approximately* satisfied. Second, semiparametric and nonparametric analyses are, in general, concerned with modeling all observations without discarding the possibly aberrant ones.[4] This contrasts with the

[3]Note that the mean of the $\tilde{y}_t$'s cannot be computed directly, as the ε_t's in (1.3) are not observed. Therefore, the maximum likelihood estimator for μ has to be obtained by means of nonlinear optimization. At the maximum likelihood estimate $\hat{\mu}$, however, $\hat{\mu}$ can be interpreted as the mean of the $\tilde{y}_t$ sample, with ε_t in (1.3) replaced by $y_t - \hat{\mu}$. For more details, see the discussion on M estimators in Chapter 2.

[4]Although some researchers claim that nonparametric procedures automatically provide protection to outliers, other researchers stress the need to correct for anomalous observations

philosophy underlying the robustness literature, where one assigns less weight to those observations that do not seem to fit into the postulated model.

An advantage of the semiparametric estimators is that they produce consistent estimates under more general conditions than outlier robust estimators. Robust estimators are, however, designed in such a way that the asymptotic bias is small. Moreover, it can be argued that consistency is only one out of many possible criteria for evaluating the performance of estimators. If one also takes account of the asymptotic variance of the estimator, robust estimators can easily dominate semiparametric estimators in a neighborhood of the null model (see, Peracchi (1990, reply)).

It is also known that semiparametric and nonparametric techniques usually require many observations in order to work reasonably well (compare the simulation results in Section 3.3). Outlier robust statistical techniques, in contrast, mostly make use of parametric models and, therefore, require much less observations to produce useful results.

1.2.2 Outlier Robustness and Diagnostics

The outlier robustness school and the diagnostics school (see, e.g., Belsley et al. (1980), Cook and Weisberg (1982), and Cook (1986)) consider the outlier problem from a different perspective. Perhaps this has been the reason for the fact that, until recently, not much interaction was taking place between both fields of research. The diagnostics school tries to develop procedures for detecting outlying and influential observations. After these observations have been detected, one can correct for them, either by assigning less weight to them or by discarding them altogether. 'Corrected' versions of standard estimates or tests result in conclusions that are more robust. The chief goal of the diagnostics school, however, is to *detect*, i.e., to identify outliers and influential observations. Numerous difficulties can arise during the outlier identification stage. The most notorious one is the masking effect (see, e.g., Rousseeuw and van Zomeren (1990)). If there are several outliers grouped close together in a region of the sample space far away from the bulk of the data, most nonrobust outlier detection methods fail to identify these observations as outliers. In other words, the outliers mask one another. The robust statistics school, in contrast, tries to develop procedures that are insensitive to anomalous observations. Its chief goal is to *protect* (see Huber (1991)). Robust statistical procedures automatically include the identification of outliers into the estimation stage. The discordant observations can often be identified in a second stage as a byproduct of the first analysis. This facilitates the task for the researcher. Moreover, robust statistical procedures are designed such that they can cope with the masking effect described above.

The approaches mentioned above need not be mutually exclusive.[5] First,

---

in the nonparametric context as well. This leads to a branch of the literature called '(outlier) robust nonparametrics' (see the references mentioned in Footnote 1).

[5]Not everyone agrees on this statement, see, for example, Fieller (1993, comment, last

outlier robust methods can provide excellent tools for identifying outliers and influential observations (see Rousseeuw and van Zomeren (1990) and Fung (1993)). Second, every statistician, including those of the robust statistics school, should be interested in whether there are any observations that do not fit into the model and whether there is any particular reason for this lack of fit. This, in general, leads to a better understanding of the phenomenon that is studied. It is, therefore, not remarkable that some reconciliation has taken place. The two approaches can be regarded as being complementary (see, e.g., Huber (1991) and Davies and Gather (1993, rejoinder)).

1.2.3 Outlier Robustness and Dummy Variables

A final point worth mentioning is the relation between robust statistics and the use of dummy variables[6] (see, e.g., Box and Tiao (1975)). Robust statistical procedures are designed to be insensitive to outliers. Procedures that use dummy variables also possess this property in some situations, e.g., when dummies are used to correct for individual aberrant observations. Therefore, procedures that use dummy variables can be considered as informal robust procedures. There are, however, three important differences between the two approaches.

First, when using the dummy approach, the outlying observations have to be identified in advance. This indicates that there is a link between the dummy approach and the diagnostics school that was described in the previous subsection. As mentioned in Subsection 1.2.2, the identification of outliers can be very difficult if one uses nonrobust techniques.

Second, because the robustness school incorporates the identification of outliers formally into the rest of the analysis, it also allows for a more formal treatment of the effects of discarding or reweighting observations. This is especially relevant for the construction of confidence intervals.

Third, most formal robust procedures use a more gradual form of weighting than the dummy approach. The latter classifies each observation as either fully trustworthy or untrustworthy. Put differently, an observation receives either weight one or weight zero. Smoother weight functions can easily be thought of and are commonly used in robust statistics. As a consequence, observations can be classified as dubious, i.e., neither completely trustworthy nor totally untrustworthy. This can lead to substantial efficiency gains, as is illustrated

---

sentence), who calls “robust outlier identification” an oxymoron.

[6] Dummies are used for both time series and cross-sectional data. A related technique that can be applied in the cross-section context is that of prescanning/screening the data. Using simple techniques, one tries to identify anomalous observations and data inconsistencies beforehand. All discordant observations are removed from the data set. The main analysis is then performed using the remaining observations. As the focus in this thesis is on time series models, I restrict my attention in the remaining discussion to the use of dummy variables only. It should be kept in mind, however, that the points raised against the use of dummy variables apply equally well to the use prescanning/screening techniques for dealing with outliers.

in, e.g., Relles and Rogers (1977).

More on the distinction between formal outlier robust statistics and the dummy approach can be found in Huber (1981, p. 4,5) and Hampel et al. (1986, Section 1.4).

1.3 Outlier Robust Statistics and Econometrics

At present, there is not much interaction between outlier robust statistics and econometrics. There are at least five possible explanations for this.

A first possible reason is that some econometricians tend to consider outlier robust statistics as a black-box procedure in which the observations are manipulated or weighted in such a way that the postulated model almost invariably fits the data.[7] This view is too simplistic. It ignores the fact that every serious econometrician inspects the data and often introduces dummy variables for the observations that seem to be outlying. As mentioned in Subsection 1.2.3, this corresponds to the introduction of a zero-one weighting scheme for the observations. More gradual weighting schemes are, therefore, not that uncomfortable to live with. Moreover, outlier robust procedures are conceptually more appealing, as they incorporate the outlier identification stage into the estimation procedure. The fact that outlier robust procedures assign weights to observations is not suspicious, but should rather be regarded as an additional source of information. The weights are available after the estimation stage and can be subjected to further analysis (see Franses and Lucas (1995)). In this way one can obtain insight into which observations one is actually describing with the estimated model. Also note that weighting can arise naturally if one considers non-Gaussian regression errors (see Section 1.2). Finally, after computing the outlier robust estimates, adapted versions of the usual diagnostic measures are available for assessing model adequacy and for detecting the outliers (see, e.g., Machado (1993), Markatou and He (1994), Ronchetti and Staudte (1994), and Rousseeuw and van Zomeren (1990)). When the discordant observations have been identified, the researcher can decide whether to leave these observations out in a second stage, whether to assign less weight to them, or whether to incorporate them in a standard analysis after all. It is also possible that a particular pattern emerges from the anomalous observations, such that the researcher is motivated to reformulate the initial model. Note that the identified outliers need not be bad observations in some absolute sense. It might well be the case that the observations that are not captured by the model are the really interesting ones. For example, when predicting asset returns, extreme above-average returns can be very interesting for investors. Concluding, when applied with care, outlier robust methods provide useful additional information to the information obtained with the standard econometric toolkit.

[7] This impression is based on an oral discussion with professor S. van Wijnbergen and professor A. Kapteijn during the AIO presentation day held in Amsterdam, March 1994.

A second explanation for the lack of interest among econometricians in robust methods is the availability of semiparametric and nonparametric techniques. These techniques are very flexible, which makes them attractive for econometricians. As mentioned in Subsection 1.2.1, however, these methods generally require much more observations to work reasonably well. Moreover, also for these flexible modeling techniques, robustness considerations remain important (see Stahel (1991) and the references mentioned in Footnote 1).

A third reason for the low level of interaction between econometrics and outlier robustness is that outlier robust procedures are not standardized enough. They leave the modal user with too many degrees of freedom. For example, in the case of M estimators, the user can choose between several different functional forms of the pseudo likelihood. Moreover, for a specific chosen functional form, he or she has to set several tuning constants. Often a method for estimating certain nuisance parameters, like scale, has to be provided as well. Part of these choices can be automated, e.g., if users are willing to specify the efficiency loss they are willing to tolerate at a certain benchmark model. The lack of attention paid to these problems, however, has probably prevented the incorporation of advanced robust techniques into large scale standard statistical software packages, which is one of the main conditions for the widespread use of a statistical technique. Some software packages for performing an outlier robust analysis are the S-plus language, which has several macros for outlier robust estimation and testing, PROCOVIEW, a program for computing least median of squares and minimum volume ellipsoid estimators (see Rousseeuw and Leroy (1987)), and RIPE, an interactive estimation program that incorporates low-breakdown and high-breakdown linear regression estimators.[8]

A fourth, important reason for the lack of interest among econometricians in robust methods is the underdevelopment of diagnostic tests and model evaluation procedures for outlier robust fitting techniques. The development of appropriate testing procedures is something that requires much attention in the near future.

A fifth and final possible reason for the lack of interest in outlier robust procedures among economists and econometricians, is that the computational effort required to compute robust estimators is higher than for traditional methods, like OLS. This argument does not apply to all robust estimators. Moreover, the argument becomes less important as parallel computing becomes more popular and the speed of computers increases. It will remain an important topic for at least some time, however, for most high breakdown methods like the least median of squares (Rousseeuw (1984)) and the minimum volume ellipsoid estimator (Rousseeuw (1985)). Some progress in reducing the computing time of these estimator has recently been made (see Atkinson (1994) and Woodruff and Rocke (1994)).

Although the interaction between econometrics and robust statistics has

[8]The S-plus language is a commercial product. PROCOVIEW and RIPE, however, are public domain programs and can be obtained from the authors, namely Peter Rousseeuw for PROCOVIEW and André Lucas for RIPE, respectively.

not been overwhelming over the last decades, some applications of robust techniques to econometric modeling can be found in the literature. For review articles, one can look at the papers by Koenker (1982), Krasker et al. (1983), and Peracchi (1990, 1991b). For the linear regression model, some references on estimators with a bounded influence function are Krasker (1980) and Krasker and Welsch (1982). In the field of simultaneous equations and outlier robustness, work has been done by, e.g., Krasker and Welsch (1985) and Furno (1988). Bierens (1981) discusses the asymptotic theory of robust procedures in nonlinear econometric models. Furthermore, Koenker and Portnoy (1990) and Peracchi (1991a) discuss the use of outlier robust estimators for seemingly unrelated regressions. Also the papers of Koenker and Basset (1978, 1982) and Basset and Koenker (1978, 1982, 1986) on quantile regression, which includes least-absolute-deviations estimation, demonstrate that some interaction has taken place between outlier robustness and econometrics. The same holds if one also includes applications of the Student t distribution, as in, e.g., Prucha and Kelejian (1984). These references are only a few out of the many articles on outlier robustness that are of possible interest to econometricians. Especially if one also considers articles from the statistical literature that deal with problems that are of potential interest to econometricians, the number of references becomes much too large to treat satisfactorily in this thesis.

1.4 Summary of the Contents

This thesis quite naturally falls into three parts. The first part consists of Chapters 1, 2 and 3, and forms an introduction to robust statistics. This part is discussed in Subsection 1.4.1. The second part, which consists of Chapters 4 through 6, deals with the application of robust statistics to the univariate unit root testing problem. This part is described in Subsection 1.4.2. The final part, consisting of Chapters 7 and 8, discusses the usefulness of robust procedures for the multivariate unit root or cointegration testing problem. This part is dealt with in Subsection 1.4.3. Finally, Subsection 1.4.4 presents some notational conventions that are adopted throughout this thesis.

1.4.1 An Introduction into Robust Techniques

The first part of this thesis comprises Chapters 1 through 3. The present chapter, Chapter 1, contains a general discussion on the notion of robustness. It also presents the notion of robustness that is the focus of this thesis, namely outlier robustness. Moreover, it discusses the relation between outlier robustness and several other statistical techniques that are often applied in the econometric literature.

Chapter 2 provides a more technical and more detailed introduction into robust statistics. It discusses several of the main concepts that can be used for assessing the robustness of statistical procedures. Moreover, it presents several well known outlier robust estimators. The material in this chapter is not new and mainly serves as an introduction for econometricians who are not familiar with the outlier robust strand of the statistical literature.

Chapter 3 tries to bridge the gap between the introductory Chapters 1 and 2 and the more technical Chapters 4 through 8. It presents some results on the maximum likelihood estimator based on the Student t distribution. This estimator is sometimes used in the econometric literature as a first alternative to the Gaussian maximum likelihood estimator. Using the concepts introduced in Chapter 2, it is shown that the Student t based maximum likelihood estimator qualifies as a nonrobust estimator, at least if the degrees of freedom parameter is estimated from the data. This result contrasts with the regular view of the Student t based estimator as being outlier robust. A simulation experiment is provided, comparing the performance of the Student t based maximum likelihood estimator to that of other estimators. It turns out that even though the Student t based maximum likelihood estimator is nonrobust, its performance in finite samples is good under a variety of circumstances. This demonstrates that statements about the robustness or nonrobustness of a procedure based on the concepts put forward in Chapter 2, should always be supplemented with evidence, e.g., simulations, that illustrate the relevance of the robustness (or nonrobustness) in situations that are of practical interest.

1.4.2 Univariate Unit Root Testing

The second part of this thesis discusses the autoregressive unit root testing problem for univariate time series. It comprises the Chapters 4 through 6.

The literature on testing for autoregressive unit roots has exploded over the last decade. Most research in the classical context has concentrated on the least-squares estimator and the associated t-test statistic (see, e.g., Fuller (1976), Said and Dickey (1984), Phillips (1987), Phillips and Perron (1988), Diebold and Nerlove (1990), Campbell and Perron (1991), Stock (1994), and the references cited in these articles). The aim of this second part of my thesis is to propose some outlier robust unit root tests and to evaluate their performance using both simulated and empirical data.

Chapter 4 discusses unit root tests with a low breakdown point, i.e., tests that can stand up to only a small number of outliers. The chapter is based on Hoek, Lucas, and van Dijk (1995). The proposed test is a simple Wald test based upon pseudo maximum likelihood estimates of the parameters in the model. The pseudo likelihood that is chosen is the Student t. It is proved that the Student t based pseudo likelihood estimator has a bounded influence function in the autoregressive time series context. Moreover, the unit root test based on this estimator also has some robustness properties. The performance of these tests relative to the tests that are used in Chapter 5 is, however, somewhat disappointing. Due to their computational ease, the low breakdown tests still deserve some attention.

Another part of Chapter 4 is devoted to the robustness of the unit root testing problem in the Bayesian setting. It is argued that Bayesian posterior unit root inference based on flat priors is extremely sensitive to outliers if a Gaussian likelihood is used to describe the data. Again, it is investigated whether this problem is (partially) solved if the Gaussian likelihood is replaced by the Student t likelihood. It turns out that this is the case.

Chapter 5 discusses the use of high breakdown methods for the unit root testing problem. The chapter is based on Lucas (1995a). A problem with high breakdown methods is that they are very computer intensive. The results reveal, however, that the additional computation time can be worthwhile. In several settings the robust unit root test performs better than both the standard Dickey-Fuller test (see Fuller (1976)) and its extension by Phillips and Perron (1988). The high breakdown methods are also used in an empirical example, using fourteen economic time series for the U.S. For some of the series, the robust and nonrobust methods give different results, both in terms of unit root inference and in terms of model selection. The outliers that cause the discrepancy between the two methods can, in some cases, be visualized using partial regression techniques. This is also explained in Chapter 5.

Chapter 6 focusses on the asymptotic distribution of maximum likelihood type (M) estimators in the context of random walk models. It builds upon the material presented in Lucas (1995b). The asymptotic distribution of unit root test statistics is, in general, nonstandard under the hypothesis of a unit root. For the OLS estimator, this has been acknowledged quite some time ago (see, e.g., White (1958, 1959)). For other estimators than OLS, some recent references are Cox and Llatas (1991), Knight (1989, 1991), and Herce (1993). Chapter 6 contributes to this literature in several ways. First, in contrast to Knight (1989, 1991), it treats the case of finite variance errors. Second, in contrast to Herce (1993), it treats the case of smooth M estimators. Third, in contrast to Cox and Llatas (1991) and Herce (1993), it allows for deviations from the i.i.d. assumption. Fourth, the asymptotic distribution of the unit root t-test is derived for regression models that may contain quite general deterministic functions of time as additional regressors. One of the findings of Chapter 6 is that particular linear combinations of unit root tests based upon the OLS estimator and upon an M estimator, are asymptotically normally distributed. This finding was simultaneously established for the LAD estimator by Herce (1993). The particular linear combination depends upon the (long term) correlation between the error term and the pseudo score that defines the M estimator. This fact can be exploited for constructing unit root tests that have standard asymptotic distributions. Again the performance of several unit root tests is evaluated by means of simulation for several data generating processes. A major conclusion from the simulation experiments is that part of the nonrobustness of the OLS based unit root test can be removed by using heteroskedasticity consistent standard errors, as in White (1980).

1.4.3 Multivariate Unit Root Testing

The third part of this thesis consists of Chapters 7 and 8. It deals with the problem of cointegration or multivariate unit root testing (see, e.g., Engle and Granger (1987), Johansen (1988, 1991), and Stock and Watson (1988)). Both the theoretical and empirical literature in this area has been expanding rapidly. Again, most attention is paid to Gaussian (pseudo) maximum likelihood estimation of the model parameters. Using similar arguments as in the univariate case, one can expect that the use of outlier robust estimators for constructing multivariate unit root tests helps to improve the properties of these tests in terms of, e.g., size or power if there are outliers.

The research on robust statistical procedures for multivariate observations has mostly concentrated on the joint estimation of location and scatter (see, e.g., Hampel et al. (1986, Chapter 5), Lopuhaä (1990), Maronna (1976), and Tyler (1983, 1991)). Applications to multivariate regression models are scarce (see, e.g., Koenker and Portnoy (1990), Peracchi (1991a)). The procedures presented in this part of the thesis combine the research areas of robust statistics, multivariate time series analysis, and unit root econometrics.

Chapter 7 discusses the use of pseudo maximum likelihood estimators for constructing cointegration tests. The asymptotic theory of these tests is derived for the case of i.i.d. innovations. Only for the Gaussian pseudo maximum likelihood estimator, the asymptotic distribution of the (pseudo) likelihood ratio test is shown not to depend on nuisance parameters, asymptotically. For all other choices of the pseudo likelihood, nuisance parameters enter the asymptotic distribution. In contrast to the derivations in Chapter 6, Chapter 7 also considers the asymptotic distribution of the test statistics under the alternative hypothesis of cointegration (or stationarity). It is proved that for non-Gaussian innovations, power can be gained by exploiting the nonnormality of the innovations in the estimation procedure. This finding is supported by simulation evidence.

A disadvantage of proposing new unit root tests is that for every test new critical values are required. These are mostly obtained by means of simulation. Setting up a simulation experiment for every (nonlinear) estimator in a multivariate time series context is very time consuming. Therefore, a simple Bartlett type correction for the cointegration tests is proposed in Chapter 7, such that only one table with critical values is required. The procedure appears to work reasonably well in some simulation experiments (Gaussian innovations), but not in others (e.g., truncated Cauchy innovations).

Chapter 8 investigates a second possibility for constructing outlier robust cointegration tests. Instead of starting from the likelihood ratio test of Johansen (1988, 1991), the point of departure is the Wald cointegration test of Kleibergen and van Dijk (1994). In contrast to Chapter 7, Chapter 8 devotes much more attention to the presence of deterministic drift terms in either the data generating process or the regression model.

Compared to the likelihood ratio based approach proposed in Chapter 7, an advantage of the Wald test is that less nuisance parameters enter the limiting distribution of the test. The main problem with the approach of Kleibergen and van Dijk is that the variables have to be ordered according to their long run causality relationships. Chapter 8 shows that it is, as yet, not possible to check for the validity of the ordering by means of a pretesting procedure. Moreover, the effect of an incorrect ordering of the variables can have dramatic effects for the size of the test. Some possibilities to resolve the ordering problem are briefly discussed in Chapter 8. The main conclusion that emerges from this discussion is that the Lagrange Multiplier test seems the most suitable candidate for outlier robust cointegration testing. It has less nuisance parameters than the likelihood ratio based tests proposed in Chapter 7, and it is not sensitive to the ordering of the variables.

Chapter 9 concludes this thesis. It summarizes the main findings of the different chapters and presents interesting directions for future research.

1.4.4 Notational Conventions used in this Thesis

The following notation is used throughout this thesis. The abbreviation i.i.d. stands for independently and identically distributed. Furthermore, I make quite extensive use of the notion of weak convergence as described in Billingsley (1968). If X_T weakly converges to X for $T \to \infty$, this is denoted as $X_T \Rightarrow X$.

Most of the test statistics used in this thesis weakly converge to random variables that can be represented as functionals of stochastic processes. Let $B(s)$ denote a stochastic process defined on the unit interval, e.g., a Brownian motion or an Ornstein-Uhlenbeck process. Then integrals with respect to Lebesgue measure are denoted in the following way: $\int B$ denotes $\int_0^1 B(s)ds$. A similar convention is adopted for stochastic integrals, with $\int BdB$ denoting $\int_0^1 B(s)dB(s)$. For vector stochastic processes, the definitions are extended quite easily, with $\int BdB^\top = \int_0^1 B(s)(dB(s))^\top$ and $\int BB^\top = \int_0^1 B(s)(B(s))^\top ds$, where $^\top$ denotes transposition.

Especially in Chapters 7 and 8 of this thesis, I use the orthogonal complement of a matrix. Let M denote a $k \times r$ matrix of full column rank r, with $k > r$. Then $M_\perp$ denotes the orthogonal complement of M, i.e., a $k \times (k - r)$ matrix of full column rank, such that $M_\perp^\top M = 0$. If k is equal to r, then $M_\perp$ is left undefined.

Finally, the end of an example is marked by a $\triangle$, while the end of a proof is marked by a $\Box$.