DOES ROBUST ESTIMATION CHANGE CONCLUSIONS FROM ECONOMIC DATA?

by

Richard Andrew Simson

B.Comm. Honours (Economics) University of Natal

M.Comm. University of Natal

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
Economics

SIMON FRASER UNIVERSITY
July 1993

APPROVAL

Name: Richard A. Simson

Degree: Ph.D. (Economics)

Title of Thesis: Does Robust Estimation Change Conclusions From Economic Data?

Examining Committee

Chairman: Dr. Lawrence A. Boland

Dr. Peter E. Kennedy
Senior Supervisor

Dr. Stephen T. Easton
Supervisor

Dr. Mark Kamstra
Supervisor

Dr. R. A. Jones
Department of Economics
Internal/External Examiner

Dr. Lonnie J. Magee
Associate Professor
Department of Economics
McMaster University
External Examiner

Date Approved: JULY 12 1993

Abstract

Using ninety economic data sets from forty-four publications, robust estimates are compared with the original ordinary least squares' results. Robust methods, insensitive to large errors, have received a great deal of attention by statisticians and econometricians; economists have been exhorted to use them. Do these robust methods make a difference with economic data, in terms of the estimated coefficients, the results of hypothesis tests, or the quality of forecasts? The results of this thesis suggest that on the whole robust methods do make a difference, albeit more so on some criteria than on others. Forecasting differences are not great, for example, but hypothesis testing differences are large enough to be of special concern. This result led to an extension of the thesis, examining the quality of the standard error estimates produced by popular econometrics packages. This extension included the use of randomization tests and Monte Carlo work. Results indicate some methods in common use are markedly inferior to others and should be avoided.

Acknowledgements

This thesis would not have been possible without the support of Kimberley Simson.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Purpose of the Thesis

The purpose of this thesis is to confront robust estimators with economic data. These estimators are novel in the sense that while versions of them have been known for a long time, they do not enjoy such widespread use as ordinary least squares. Their description as robust is meant to convey the notion that these estimators can withstand a wide variety of unusual situations like very large errors or outlying points in the data. **Hogg and Craig** (1978) offer the practical definition of robust as:

> **An estimator that is fairly good (small variance, say) for a wide variety of distributions (not necessarily best for any one of them) is called a *robust* estimator. (p402)**

The idea is to select a number of these robust estimators, so as to avoid the possibility of results being dependent on a single estimator, and see if this set of estimators gives "different results" when used with economic data sets.

Many things can be meant by "different results" and here we approach the problem from a number of perspectives such as changes in the size of an interesting coefficient or the ability to provide better forecasts. Particular emphasis is placed on the hypothesis tested by the original researcher. We want to know if using robust estimators changes the results of these tests.

While this is a noble exercise, the immediate difficulty is where does one obtain many data sets? We obtain data sets from three sources. First, many articles provide data. Second, one journal sells data sets used in articles it has published. Finally, some econometric texts now provide data for replication of interesting studies in economics. Armed with ninety data sets (seveny-four time-series, sixteen cross-section) from forty-four articles we replicated the original study and applied the robust methods to the same data. With at least fourteen ways of determining what is a "different result" the robust estimates were compared to the original results as replicated.

1.2 A View of the Literature

Econometrics has seen an almost unprecedented expansion in the area of robust estimation. An impetus for this growth is the misgivings of some researchers, the most notable being **Mandelbrot** (1963a, 1963b and 1969), concerning the densities producing economic data. Economics must respond to Mandelbrot's work as he is attempting a critique of uses of the central limit theorem and the Gauss Markov theorem. Under a number of assumptions, least squares is the best linear unbiased estimator. A key assumption supporting the theorem is that the error density has a finite variance.

For Mandelbrot the requirement of a finite variance is the "Achilles heel" of econometrics. Mandelbrot, looking at cotton price changes, noticed there were too many outliers in the data for the density of the changes to be normal. When there are outliers of this type it becomes exceedingly difficult to estimate the population variance. As one draws a sample there is the possibility of obtaining some outliers and the exact values of these outliers can affect the sample variance estimate drastically. Mandelbrot then suggests the hypothesis that the population variance is infinite. Upon investigation, these extreme price changes and the possibility of an infinite variance seemed consistent with the probability density termed "Pareto's Law" or hyperbolic density. The term hyperbolic indicates some feature of the distribution is exaggerated and here that is the distribution has exaggerated tails. Mandelbrot argues the possibility of an infinite variance is best handled by a family of densities known as stable Lévy densities. Such densities have been called Lévy stable, Pareto-Lévy and stable

Paretian. This raises the possibility that economic series come from the non-normal siblings of the Lévy distribution. These non-normal densities have infinite variances. An infinite variance implies the sample estimate of the variance does not converge to a particular value as the sample size increases. Sadly the sample estimate can fluctuate wildly. Or as one wag put it: The standard deviation is highly variable.

It is not easy to formulate a response to the possibility of a density with an infinite variance if one is an econometrician. One might say "test" for such a possibility. There are three ways to do this. First plot the empirical cumulative density against the cumulative density produced by the normal. If the plot is a straight line one can reject the infinite variance hypothesis. A famous paper that adopts this method is **Fama** (1965) for first differences of the logarithm of stock prices. He uses data from 1957 to 1962 broken down into thirty samples with a range of observations from twelve hundred to seventeen hundred. A visual inspection of the plot supports the Mandelbrot hypothesis of an infinite variance.

The next method is to plot the sample variance as T (the sample size) changes. Evidence against the normal is a non-converging variance. The last and possibly the best test is the "log-tail" test. Plotting the number of errors greater than some large error against the logarithm of the errors one should obtain a straight line with slope equal to $-\alpha$ also known as the characteristic exponent of the density. If α is less than two in absolute value the density has an infinite variance. Based on these tests **Granger and Orr** (1972) state:

> **The evidence from the three tests we have described here is in some cases so strong that it should convince most research workers that infinite variance is a distinct possibility. Most certainly the longtailedness of some distributions arising in economics has been established. (p277)**

or as **Fama** (1963) concludes:

> **For commodity markets the most impressive single piece of evidence is a direct test of the infinite variance hypothesis for the case of cotton prices. Mandelbrot computed the sample second**

> **moments of the daily first differences of the logs of cotton prices for increasing samples .... He found as the sample size is increased the sample moment does not settle down to any limiting value but rather continues to vary in absolutely erratic fashion, precisely as would be expected under the stable Paretian hypothesis. (p428)**

Koenker and Basset (1978) in providing their regression quantile estimator, suggest there are other reasons for using robust estimators not related to the possibility of an infinite error variance. There is a view in econometrics that if the errors are not normally distributed, then least squares and statistical tests based on least squares are still valid. This view comes from the Gauss Markov theorem that, for errors of finite variance, posits least squares is the best linear unbiased estimator no matter the error distribution. Further, testing of hypotheses is possible if the researcher is willing to accept the idea the test can be supported on asymptotic grounds. **Koenker** (1982) points out least squares estimates can be changed substantially if the distribution of the errors only deviates slightly from the normal. Also, the asymptotic justifications for testing using the normal distribution have to do with the size of the test. Koenker shows how in an environment where there is some departure from the normal the power of least squares is much lower than that of robust tests. He states:

> ...**that classical tests need** 40% **more observations than the robust test to achieve the same power. (p236)**

As Monte Carlo studies have one drawback, the manufactured data may bear little resemblance to real data, **Stigler** (1977) maintains robust and non-robust estimators should be compared using *real* data. He does this using twenty data sets and eleven (ten robust) estimators. Stigler's data sets are measurements at least a century old of constants like the velocity of light for which there exist "modern" true values. The results indicate a small amount of trimming can produce gains. One would not say the result is a stunning endorsement of robust methods but at issue is the need to work with real data. This is a valid method to provide an environment in which estimators can be compared. The reason: Monte Carlo data does not guarantee to be like real data. **Rocke, Downs and Rocke** (1982) take up Stigler's idea to use real data; they compare twelve estimators using forty-seven analytical chemistry data sets. Unlike Stigler they recommend using a robust estimator. **Rocke, Downs and Rocke** also provide a concise statement of the methodology which is an alternative to the Monte Carlo technique as follows:

> **Given the variation in the quality and character of data that may exist across disciplines and over time, it seems probable that the utility of robust methods can be demonstrated only by applying the method described here to a substantial number of "typical" and current data sets chosen from specific fields. (p97)**

A cursory survey of the economics literature shows that very little work uses robust estimators. **Hogg** (1979) argues it is these estimators' computational complexity that causes econometricians and applied economists to shy away from them. This is changing but slowly. There are some very good reviews of robust estimators that are accessible to those with even the most rudimentary statistical skills, for example **Berk** (1990). Statistical software is available allowing easy computation of recommended robust estimators.

Another reason for the scarcity of robust estimation is that many rely on the usual asymptotic justifications for hypothesis testing with economic relationships. This notion is as old as the modern concern with robustness. When the term robust was first applied to statistics by **Box** (1953), it was evident even then statistical procedures could be robust regarding non-normality. This view could still have widespread support especially because of the acceptance of the Gauss Markov theorem.

A final reason for the less than widespread use of robust estimators is that the data to which the infinite variance hypothesis has application may be a small subset of all the data analysed by economists. The Mandelbrot and Fama empirical work, for example, concerns itself solely with the logarithm of price changes, specifically cotton and stock prices. In addition, the error environments that point to the efficiency of robust estimators have large variances, even infinite variances (See **Fama** (1963, p427)) and many economists could perceive these variances as being patently unrealistic.

The last two reasons above provide some impetus for a study, using economic data, attempting to find out if the use of robust estimators would have changed any conclusions based on least squares' associated hypothesis tests. Using forty-four published studies in applied economics we wish to replicate original results and then re-estimate and evaluate using robust methods. The following two quotes aptly summarize the problem. First **Taylor** (1974) states:

> **The implications of an infinite variance for conventional methods of estimation, least squares in particular, are rather grim. In a finite-variance world, we hardly need remind ourselves of the virtues of least squares: It provides the Gauss-Markov estimator and, in addition, the maximum likelihood estimator in the context of normality. Normality also opens up the door to the vast apparatus of classical and Bayesian inference. However in an infinite variance world, the Gauss-Markov theorem no longer applies, and least squares becomes another estimator. And a poor one at that. An infinite variance means thick tails and thick tails mean a lot of outliers. Least squares, as we know, gives outliers a lot of weight, and accordingly becomes extremely sample dependent. Thus, in this context, an estimator which gives relatively little weight to outliers is clearly to be preferred. (p170)**

Four years later **Koenker and Basset** (1978) note:

> **Indeed one somtimes encounters the view that infinite variance of the errors constitutes the only possible rationale for seeking robust alternatives to least squares in the linear model. This is emphatically false. While least squares is obviously abysmal for distributions having infinite variance (having zero efficiency for the Cauchy for example) its gross inferiority to a variety of nonlinear estimators is by no means confined to distributions with infinite variance. (p35)**

1.3 New Avenues of Research

Given the renewed interest in robust estimation (**Berk** (1990), for example) we feel a need for a study of robust estimators with real data. This thesis adds to and builds upon existing studies in some unique ways:

I All existing studies focus on a particular data set be it demand analysis or stock returns. This is obviously useful for those economists who specialize in these areas. This thesis attempts a new twist in that we want to know if general statements can be made concerning all economic data and move away from limiting ourselves to one data set.

II Each data set, of the many we use, has itself supported results in published studies. With this data set of "data sets" more can be gleaned about robust estimation than would be the case from applying least squares and a robust method to a new data set. We too could get data for a given specification and apply a robust method besides least squares but the choice of specification would be somewhat arbitrary. It makes sense to use specifications (and data) that have survived the refereeing process. Given the way applied economics is conducted, this means avoiding other problems, such as heteroskedasticity, resulting in fewer influences from these sources than if we used new data whether real or artificial.

III An important unaddressed question in the literature is: How often are results from published studies reversed with robust estimators? What we attempt to do here is to find out if robust methods change the conclusions others have reached with a simple testing procedure using economic data. Thus in a sense this thesis tests a "null" of robust estimators not making any difference. A failure to reject this "null" hypothesis could happen for two reasons:

1. Researchers may have employed both least squares and robust methods and only reported least squares results they *know* agree with robust methods;

2. Or economic data is such that robust methods give results close to those of least squares indicating economic data is free of the problems those robust methods avoid.

Thus we have the common problem of a joint "null" hypothesis. There is nothing that can be done to separate the two possibilities but taken together a finding in favour of the "null" is still a useful exercise. We obtain more information concerning testing and data in economics. One study, from which a data set is taken for this thesis, replaces a dependent variable observation with lower values whenever the author noted an outlying observation. This was done in a very *ad hoc* fashion; non of the usual robust procedures were performed to accomplish this. This study is the only one that explicitly deals with an outlier. To some extent, we do not know if others changed their data, not explicitly mentioning any form of robust analysis mitigates against the first part of the joint "null".

To address further the problem of a joint null, where our robust methods do show a difference for a particular data set we apply four of the diagnostic statistics in **Bollen and Jackman** (1990, page 268) that should reveal outlying errors (outliers proper) and leverage points or extreme values of the independent variables. Researchers have two choices, some use these tests to purge the data of the outliers before embarking on estimation: an example being **Granger, White and Kamstra** (1989). Other researchers use tests for outliers, and where the tests point to possible problems in the data, they may use robust methods. We want to address the objection that robust methods may not have been considered by the original researchers because available diagnostics failed to reveal a problem.

IV The study with the largest number of regressions comparing robust methods with least squares is twenty-one by **Connolly** (1989), for *one* data set divided into seven periods. We improve on this by looking at ninety regressions from forty-four published studies in many more economic contexts.

V **Hogg** (1979) argues for least squares to be accompanied by a robust method. Any sizeable difference between the two methods should be investigated. This thesis provides some "rules of thumb" to help one establish what is a sizeable difference.

VI Randomization tests, an alternative to the traditional testing methodology, do not require that errors be distributed normally and do not rely on asymptotic properties. Consequently, they offer an attractive alternative to robust estimation and testing. This thesis compares randomization tests, as well as traditional robust tests, to tests based on least squares' results. A discussion of randomization tests in the econometric context can be found in **Kennedy** (1993).

1.4 Why be Concerned with Robust Regression?

1.4.1 Introduction

We all at some time or another have used a regression package to estimate a linear equation from a favourite theory. It is usual to have the package perform its array of diagnostic tests and the temptation is always there to look at another model specification if the original equation fails one or more of these tests. This is not the place to discuss the merits of such a procedure but rather to point out there is one "diagnostic" that is not explicitly calculated by regression packages but is recommended by statisticians and econometricians. This "diagnostic" concerns the use of robust estimators. Robust has a particular meaning as **Stigler** (1973) has suggested:

In the eighteenth century, the word "robust" was used to refer to someone who was strong, yet boisterous, crude, and vulgar. By 1953 when Box first gave the word its statistical meaning,

> **the evolution of language had eliminated the negative connotation: robust meant simply strong, healthy, sufficiently tough to withstand life's adversities. (p872)**

Given that **Stigler** does refer to **Box** (1953) it is useful to review the latter's (p318) definition of robust:

> **The tests mentioned are derived on a number of assumptions, in particular, that the observations are normally distributed. Usually, however, since little is known of the populations from which the samples are drawn, these tests are used, of necessity, as if the assumption of normality could be ignored. So far as comparative tests on means are concerned it appears (perhaps rather surprisingly) that this practice is largely justifiable ...**
>
> **It would appear, however, that this remarkable property of "robustness" to non-normality which these tests for comparing means possess, and without which they would be much less appropriate to the needs of the experimenter, is not necessarily shared by other statistical tests, and in particular is not shared by the tests for equality of variances.**

1.4.2 Some Quotations Supporting Robust Methods

The following is a series of quotations reflecting the widespread call by statisticians and econometricians for the use of robust methods.

As regards normality as an assumption **Granger and Orr** (1972) suggest the following:

> **It is standard procedure in economic modeling and estimation to assume that random variables are normally distributed. In empirical work, confidence intervals and significance tests are widely used, and these usually hinge on the presumption of a**

> **normal population. Lately, there has been a growing awareness that some economic data display distributional characteristics that are flatly inconsistent with the hypothesis of normality. (p275)**

To justify the concern over normality the following is from the preface of **Robust Inference** by **Tan, Tiku and Balakrishnan** (1986):

> **Most classical statistical procedures are based on two assumptions, that the sample observations are independently and identically distributed, and that the underlying distribution is normal. While the first assumption may not be unrealistic in certain situations, it is the second assumption that is rather unrealistic from a practical point of view. To quote R. C. Geary (Biometrika, 1947): "Normality is a myth; there never was, and never will be, a normal distribution." This might be an overstatement, but the fact is that nonnormal distributions are more prevalent in practice, and to assume normality instead might lead to erroneous statistical inferences. (piii)**

One of the most enthusiastic proponents of robust methods is the statistician Robert Hogg who maintains in **Hogg** (1979):

> **The method of least squares and generalizations of it have served us well for many years. It is recognized, however, that outliers, which arise from heavy-tailed distributions or are simply bad data points due to errors, have an unusually large influence on the least squares estimators. That is, the outliers pull the least squares "fit" toward them too much, and a resulting examination of the residuals is misleading because then the residuals look more like normal ones. Accordingly, robust methods have been created to modify least squares procedures so that the outliers have much less influence on the final estimates. (p108)**

In the book **Robust Statistics** (1981) which attempts to deal with many aspects of these methods, **Huber** offers as a reason for wanting to consider robust methods the following:

> **During the past decades one has become increasingly aware that some of the most common statistical procedures (in particular, those optimized for an underlying normal distribution) are excessively sensitive to seemingly minor deviations from the assumptions, and a plethora of alternative "robust" procedures have been proposed. (p1)**

The same sentiments are echoed in the preface to a book entitled **Robust Regression** published in 1991 by **Tan, Tiku and Balakrishnan** (1986):

> **Statistical inference deals with the extraction of information from observations. Of equal importance to the empirical data are the assumptions underlying the analysis. While it is granted that these assumptions concerning randomness, independence, and so forth are not precisely true in the real setting, they are nonetheless invoked in order to provide theoretical foundations for the ensuing analysis.**
>
> **The implicit assumption made in many such situations is that small deviations from the assumed model will result in only small errors in the final results. Recent studies have indicated, however, that this is not always the case. As a result, the use of more robust statistical procedures has been an area of increased interest, both for the theoretician and the applied statistician. (piii)**

Another proponent of robust methods is **Andrews** (1974):

> **Least-squares is an optimal procedure in many senses when the errors in a regression model have a Gaussian distribution or when linear estimates are required (Gauss-Markov Theorem). Least**

> **squares is very far from optimal in many non-Gaussian situations with longer tails (see Andrews *et al.* 1972, Chapter 7 for further discussion). It is unlikely that the use of least squares is desirable in all instances. Some alternative to least squares is required. A recent study (Andrews *et al.* 1972) clearly demonstrates the inefficiency of least-squares relative to more robust estimates of location for a wide variety of distributions. (p523)**

In an article providing examples of uses of robust methods **Mallows** (1979) provides a summary of the **Hogg** (1979) suggestion for using robust estimators in empirical work:

> **A simple and useful strategy is to perform one's analysis both robustly and by standard methods and to compare the results. If the differences are minor, either set can be presented. If the differences are not, one must perforce consider why not, and the robust analysis is already at hand to guide the next steps. (p184)**

In the article concerning regression quantiles, a robust estimator, **Koenker and Basset** (1978) posit:

> **Unfortunately the extreme sensitivity of the least squares estimator to modest amounts of outlier contamination makes it a very poor estimator in many non-Gaussian, especially long-tailed, situations. (p34)**

Swartz and Welsh (1986) concerning themselves with problem independent variable values note:

> **The linear model is probably the most common statistical framework used in econometric analysis. Unfortunately the majority of applied work done with this model begins and ends with ordinary least squares (OLS) family of estimators. It is certainly true**

> **that OLS is an attractive estimator; it is very easy to compute, and, of course, it is efficient among linear unbiased estimators for problems that satisfy a certain set of assumptions. However, these assumptions seldom correspond to the situation facing us when we work with economic data. A more satisfactory approach to the linear model would still involve the OLS family of estimators, but only as a starting point around which diagnostic analysis can be used to identify assumptions that might be descriptive as well as theoretically useful. Given that understanding, the reliability of coefficient estimates and forecasts can then be enhanced by using estimators based on the assumptions we believe and robust to the breakdown of the assumptions we distrust. (p154)**

Given that we do use robust methods in conjunction with ordinary least squares **Hogg** (1979) does point to an additional advantage:

> **If a robust element is added to our present methods, we will detect many simple, and not so simple errors. These procedures have been used very successfully (e.g., Los Alamos Scientific Laboratory has the option of using them in all regression problems and this option is exercised frequently). Many interesting things have been discovered through of these procedures. My hope is that by 1980 almost all statistical investigation will include a robust aspect. (p114)**

The hope for the widespread use of robust methods in addition to ordinary least squares has yet to be realised. Contrasting least squares and robust methods could be termed the neglected "diagnostic" in econometrics. It is important to note that robust methods are not meant to substitute for conventional methods but rather act as a method complementary to least squares.

Koenker (1982) is an attempt to get more econometricians interested in robust methods. He maintains:

> **The classical statistical paradigm ... is gradually giving way to robust methods. This robustness revolution does not represent an attack on traditional statistical models, rather it reflects a heightened awareness of the potentially serious consequences of modest departures from classical hypotheses. Indeed the objective of robust methods is to extend the domain of validity of the classical models. (p246)**

Finally we have **Koenker** (1988) provide his view of robust estimation as it applies to the error distribution:

> **Robust estimation of the linear model as expounded for example, by Huber (1973, *M* estimators), Koenker and Basset (1978, *L* estimators), and Hettmansperger and McKean (1977, *R* estimators) has a very limited objective. Given a model specification, find an estimator of the regression parameter that achieves reasonable efficiency over some large neighborhood of error specifications. In practice this means finding estimation methods less sensitive than classical least squares to outliers in the y_i's. (p447)**

1.4.3 Summary

This section justifies the use of robust methods, especially robust regression by presenting some statements made by those who, either indirectly or directly, support their use. A noteworthy feature of these quotes is they point to the possibility of large absolute errors drawn with greater probability than the normal and the possibility of errors in the independent variables both resulting in problems with least squares. However if one uses a robust method in addition to ordinary least squares not only does one have another "diagnostic" but it is likely one also will have a greater knowledge of the data. This alone makes robust methods worth consideration.

1.5 Some Prior Work

1.5.1 Introduction

Although some have argued that the use of robust methods is likely to decline, **Baxter** (1990), our survey of the literature, did reveal recent applied studies that use robust estimators and economic data. **Eddy and Kadane** (1982) use robust regression as a check on least squares' results that give the cost of drilling an oil well in a particular region in the United States. **Lioukas** (1982) in a multinomial model uses robust methods to find the cost of business and non-business travel in Greece. To the extent marketing data is economic, it is necessary to mention **Mahajan, Sharma and Wind** (1982) who show how robust methods can help establish profitable areas of a company's activities. These last two endorse the use of robust methods. **Coursey and Nyquist** (1988) use robust regression to estimate price and income elasticities. They suggest taking the Mandelbrot and Fama findings seriously and push for robust estimation. In a study of the weekend effect, negative returns to stocks on a Monday, **Connolly** (1991) uses robust regression to show that this effect concerns assumptions about the error distribution, namely the distribution has "normal" tails, rather than economic forces. **Lichtenberg and Siegel** (1991) use outlier deletion to check whether least squares estimates are sensitive to large errors. They find their least squares results are not dramatically changed. Another study giving support to the use of least squares is **Swinton and King** (1991) but this may be due to the peculiar nature of their data as we explain in more detail below. **Geske and Torous** (1990) use robust methods to estimate the variability of stock returns and find robust methods can reduce price misspecifications. It seems as if robust estimation became more popular in the late eighties and nineties , although the number of studies is still small.

From these studies one gains the impression robust analysis is not used a lot and often it is used as something to try other than least squares. Not all the studies motivate the use of robust estimators. While the focus for this thesis is parametric robust regression, **Magee, Burbidge and Robb** (1991) use a nonparametric robust analysis to re-examine an economic hypothesis. The motivation for this thesis to use parmetric robust estimators is first, to determine if the choice of such estimators is warranted with economic data and second, we feel that although a nonparametric analysis to be equally, if not more, important than the usual parametric analysis, it is the parametric form that predominates in economics. About half of the studies discussed here find robust analysis does not change regression parameters, but it is not always made clear how this is determined by the authors of the studies. It has become an important question, given these results, whether robust estimation does give different results with economic data in general rather than in a few scattered data sets? At this stage one would be equivocal about the ability of robust regression to change the results from a technique like least squares. Further what it means to get a different result with a robust estimator needs to be covered in much greater detail than it has in the literature to date. We examine the studies in more detail in the next section. As an important aside, a feature of the nonparametric robust analysis is a "different" result is usually one that reverses a previously held economic hypothesis. While our focus here is parametric robust analysis we feel it is important enough to adopt as a criterion to determine if robust estimates are different from least squares' estimates.

1.5.2 Two important studies

We single out the following two studies as they both generalize from their respective data sets. This is not to say the studies in the next section are unimportant. We highlight **Swinton and King** (1991) and **Coursey and Nyquist** (1988) as they attempt what this thesis attempts, although they are concerned with particular economic data. We use many more data sets. Also we use techniques they develop for comparing estimates, robust or otherwise, and on this basis feel they must be given some prominence.

Where a researcher suspects an outlier, and can identify the same, the usual procedure, according to **Swinton and King** (1991) is to omit the outlier. This *ad hoc* procedure can be avoided if one uses a robust method as these estimators achieve the same purpose. Other approaches, see **Robb, Magee and Burbidge** (1992), transform the dependent variable, for instance. Returning to discuss **Swinton and King**, crop yield data has the peculiarity, when detrending, that an end of sample observation can influence the least squares' results. Despite this feature of such data, least squares is still better than robust methods argues **Swinton and King** (1991). For instance, corn yield data (used in agricultural economics) can have low values, at odds with what one expects from a normal distribution. Regressing corn yields on a variable year, with least squares, gives residuals skewed to the left. This might be an environment where robust estimates are better than the outlier removal method. If the disturbances are normal, least squares gives the maximum likelihood estimates. If they are not, least squares is still the best linear unbiased estimator and the variance estimator is unbiased and consistent. Also the maximum likelihood estimator is not least squares thus both the estimates of least squares will not be efficient or asymptotically efficient. **Swinton and King** attempt to see if robust estimates can produce estimates that are better than least squares. They do this in three ways.

1. By using a number of robust estimates, six of which we also use, they find the robust estimates lay within one least squares' standard deviation of the the least squares *estimate*. For them this meant least squares did not differ significantly from the robust estimates. Use is made of this idea in our own comparison of estimators. For them the *ad hoc* method would be better as long as you can identify the outlier. It is this problem which is addressed next.

2. **Swinton and King** obtain eight corn yield data series of varying length but ending in 1984. They then replaced the actual 1984 observation with an "artificial" observation. Thus they create eight data sets mostly of real observations but also having one outlier at the end of the series which they deliberately set below the actual value. Various robust estimators all available in SHAZAM (**White** (1978)) plus two popular measures of outliers are applied to these deliberately altered data sets. The idea is to determine if the robust estimators and the outlier measures are able to single out the created outlier in 1984. For samples sizes greater than five and less than twenty-five, trimmed least squares

performed the best with the benchmark of "within two standard deviations of the least squares estimate", based on the uncontaminated data, with the "artificial" observation replaced with the actual observation. Also, of the two widely accepted means of identifying outliers, one, DFBETAS, spotted the outlier. We employ this one, in addition to others, in this thesis.

3. Finally, least squares and a trimmed estimator are compared, using the Monte Carlo method with errors drawn from the beta and normal distribution and least squares gives "more accurate" estimates as the coefficient ranges and standard deviations are smaller.

On the basis of these three approaches **Swinton and King** (1991) argue, at least for crop yield data, least squares is preferred and DFBETAS can be used to identify outliers. In the first approach, use is made of the least squares' estimates to ask if a robust estimate lies beyond the usual range implied by those estimates. Robust estimates are deemed to be different if this range is exceeded. We employ the same idea with many data sets as one shortcoming of **Swinton and King** is they use one special data set.

Errors need not be , but sometimes are assumed to be normal. Without normality least squares is still best linear unbiased and this has led to the perception that least squares is impervious to aberrant errors. There are non-linear estimators that will do better than least squares. The economic literature is replete with evidence that economic data is generated by distributions having infinite variance and fat tails. The problem is what estimation techniques are required if this is the case? It is no secret that the whole opus of robust estimation arose to deal with such distributions. This provides a *raison d'être* for **Coursey and Nyquist** (1988) to examine price and income elasticities using Swedish and American data with five robust estimators. A feature of their study is to make explicit the distributional assumptions of the error distributions.

They have four regressions covering food, transportation, hotels and restaurants and housing for Sweden. For the United States there are five regressions as the category gas and oil is used in addition to the others already mentioned. Most income elasticities were significant. Results were not as good for the price elasticities. The novelty they use to determine if the robust estimates are different from least squares, is to determine if the largest percentage change of coefficients between least squares and the robust estimators is greater than thirty percent in the nine regressions. For three of the regressions the largest percentage change exceeded thirty percent for the income elasticities. For the price elasticities this is the case in four regressions. **Coursey and Nyquist** take these results to mean that the different distributional assumptions concerning the error term are "significant" using their term. With five robust estimators and different distributional assumptions concerning the error term **Coursey and Nyquist** obtain coefficient estimates that differ from least squares according to a thirty percent benchmark. For them these differences arise from the use of the robust estimators coupled with the error distribution assumptions they employ or, in other words, robust estimators do make a difference. An implicit implication of **Coursey and Nyquist**, for empirical work is the need to do preliminary tests to help determine the distribution of the errors in a regression. This thesis differs from their work in that we make the decision to remain ignorant about the error distribution. **Coursey and Nyquist** make particular distributional assumptions. We feel that ignorance describes most economic research and we can use their thirty percent benchmark to ascertain if robust estimators make a difference with many more data sets and, further, not just demand analysis. Even though we choose to remain ignorant of the error distribution, this does not preclude us from testing the errors of our ninety regressions for normality. This is done at the beginning of Chapter Four.

1.5.3 The Other Studies

Eddy and Kadane (1982) in a study of oil well drilling, argue the *logarithm* of the cost of drilling for an oil well has a distribution with outliers possible in both ends of the density. Instead of rejecting the outliers they adopt the so-called accommodation method where a robust estimator is used which does the same thing but according to a fixed rule. Rejecting outliers can be at the discretion of the researcher. The use of the word accommodation is confusing as the outlier rejection method is also a way of "accommodating" them. Accommodation means using a robust method to identify and deal with outliers. **Eddy and Kadane** are interested in estimating the cost of oil wells, given well characteristics and regression estimates of coefficients. As they are uninterested in the variability of their forecasts, they do not calculate the standard errors of the Huber estimates for twenty-nine American regions giving the same number of regressions. They also used least squares for comparison and conclude:

> **For this data set, the use of robust methods led to only a slight change in the estimates and predictions. Nonetheless, given that our task was to propose a method to be used on a data set as yet unseen, we feel the protection provided by robust methods is an important advantage and worth the premium paid in loss of efficiency.(p269)**

What is meant by **slight** is never explicitly defined by **Eddy and Kadane** as regards changes in estimates of coefficients. Comparing coefficient estimates of one of their least squares' regressions with the robust estimates, has no estimate changing by more than four percent. This is the lower limit of their unstated benchmark. It is not possible to determine the upper limit of their implicit benchmark. As **Eddy and Kadane** are no help in this regard we have used the percentage changes proposed by **Coursey and Nyquist** in this thesis. **Eddy and Kadane** also refer to predictions and use a single statistic, PRESS-the average forecast error from deleting one dependent variable at a time from the regression and forecasting (an analysis of PRESS can be found in **Magee and Veall** (1991)) the excluded value with least squares and the remaining observations-for a regression to determine whether robust or least squares provides the better fit. For their data, robust estimation provides the better fit. Again **Eddy and Kadane** provide no benchmark to determine what is a worrisome difference in their statistic relative to the residual mean squared error for the robust procedure and least squares. However, we do sympathize with the intention to use predictions and thus we also focus on forecasts of observations produced by least squares and robust techniques. Our addition is to make explicit the benchmark used to ascertain if the two methods produce different forecasts. Robust estimators have been used to good effect in forecasting macroeconomic time series observations. Two such studies are **Fair** (1974) and **Hallman and Kamstra** (1989) and it is these studies and the **Eddy and Kadane** study that provides the rational to compare estimators on their ability to forecast. Details are provided in Chapter Three.

Another study using robust methods is **Lioukas** (1982). In order to evaluate transport investment projects, the value of time needs to be estimated. Typically what is done in developing countries is to use estimates of the opportunity cost of time from developed countries which is not feasible for Greece which some consider a developing country. Also as Greek market wage rates are distorted these cannot be used. Using a microeconomic justification **Lioukas** shows how the the choice of travel is influenced by the relative price of travel and other characteristics, such as income. After estimating the preferred equation it is discovered that there are many outliers and thus robust estimates are made using estimators from the class called M-Estimators. These estimators reduce the weight that least squares gives to large errors. Although the robust standard errors are reported no mention is made of how they were calculated and the study does not do any significance tests. **Lioukas** ends by reporting:

> **The robust maximum-likelihood method of estimation appears somewhat superior to others for the analysis of residuals.(p174)**

Lioukas never provides any details of the analysis of the residuals so it is difficult to determine how the superiority of one method over another is established . The problem of not making explicit criteria used to choose amongst estimators is a general feature of studies using robust methods. This thesis makes explicit the criteria used to evaluate whether robust estimators are different from least squares.

Many are uncomfortable with the practice of deleting outliers and subsequent use of the reduced data set. **Mahajan, Sharma and Wind** (1984) fall into this category. Their primary aim is to identify salespeople and departments within firms that are different from each other usually according to performance criteria. In this context to discard outliers may be inappropriate as they could be the most valuable observation points in that they contain useful information on performance. You can identify outliers with robust methods. They recommend robust analysis as it is a procedure that can identify outliers that may not be identified by least squares or other *ad hoc* procedures. They also recommend the use of robust procedures as the impact of outliers on estimates is reduced. This may be at odds with their dislike of outlier rejection as you may want the outlier to affect the estimates. The outlier may be telling you something important about sales performance. However, they do exonerate themselves with the following statement:

> ...**perform the usual OLS analysis along with a robust procedure. If the resulting estimates are in essential agreement, report OLS estimates and relevant statistics. If substantial differences are found, however, carefully examine the observations with large residuals and check to determine whether they contain errors of any type or if they represent significant situations in which the postulated regression model is not appropriate.(p276)**

This study does provide a formula for calculating the variance covariance matrix although the focus is not on hypothesis testing. The standard deviations are reported but not used in the analysis of sales and department performance. This is also a study that does not make explicit what is a **substantial difference** between estimates.

A study which uses a robust estimator is **Lichtenberg and Siegel** (1991). They use the Least Absolute Errors estimator to calculate rates of return to various research activities. This estimator gives rates of return as *high* as thirty percent *lower* than obtained with weighted (to correct for heteroskedasticity) least squares. No hypothesis tests were performed using this estimator as **Lichtenberg and Siegel** maintain the "standard errors for the parameter estimates are unknown" for Least Absolute Errors. As these changes are cause for some concern, hypothesis tests of interest were conducted using the parameter estimates, after deleting influential outliers. The hypothesis tests were unchanged and for **Lichtenberg and Siegel** this meant there is little to worry about outliers influencing their results. It is recognized by many, such as **Hogg** (1989), that there are a number of ways to estimate the variance covariance matrix for robust estimators. We found that while the standard errors are unknown they can be estimated and thus play an important role in determining whether hypothesis tests change. We did estimate standard errors in a number of ways in this thesis and found there is cause for concern with these estimates but we do offer some solutions to the problem in Chapter Five.

Another study falling into this vein is **Small** (1986) who uses hypothesis tests as a means to compare robust results with least squares and:

> **...conclude that although the robust estimations weaken the results somewhat, there is still tentative evidence for an energy-scarcity effect.(p379)**

As we are to use standard deviations it makes sense to compare robust estimators with least squares, at the level of hypothesis testing, following **Small**. For each data set in this thesis there is an associated hypothesis test performed by the original researchers. We redo these tests, be they t or F-tests, to see if results are altered using the robust estimates.

In the financial economics literature there is some evidence that the return of a stock market index is significantly negative on Mondays. **Connolly** (1989) investigates this issue but from the perspective that financial data is not normally distributed and thus inferences as usually made with coefficients and standard deviations are not valid, if there are outliers. **Connolly** worries that inferences, especially those made about the negative return on Mondays may be reversed if inference is performed using robust estimators.

Using data on three return measures and covering twenty years **Connolly** finds a significant and negative return on Mondays. The residuals from the regressions are tested for normality and it is found "there is substantial evidence of nonnormality". The suggestion is the original inferences on the negative return on Mondays is cast in doubt. Using Least Absolute Errors, Trimmed Least Squares and the Huber M-Estimator, the general result, for inferences based on the robust estimates, is the Monday effect on returns disappears in the middle of the seventies. This study estimates the variance covariance matrix of the estimates with a different formula than that used by **Mahajan, Sharma and Wind** (1984). A study similar to **Connolly's** is **Geske and Torous** (1990) who examine the prices of a specialized financial instrument and find robust estimation changes conclusions based on least squares results.

1.5.4 Least Median Squares

An increasingly popular robust estimator is one introduced by **Rousseeuw** (1984) called the least median squares (LMS) estimator. This estimator chooses parameter estimates to minimize the magnitude of the squared **median** residual. In contrast, OLS minimizes the **sum** of the squared residuals. Finding the parameter estimates which produce the lowest median squared residual is not as easy as estimating the least squares coefficient, but there is an algorithm to do this and is available in the statistical programme S-Plus. **Rousseeuw** (1984) points to a shortcoming of LMS:

> **A disadvantage of the LMS method is its lack of efficiency** ...
> (p876)

In recognition of this, the statistics programme S-Plus does not even provide estimated standard errors for LMS. On this basis alone we do not use LMS, given that one focus of this thesis is hypothesis testing. There is, however, another reason to exclude this estimator from the present study. LMS, as pointed out by **Hetmansperger and Sheather** (1992), is not itself robust to slight changes in the data, but unlike least squares it is not data points a long way from the mean that are the problem. It turns out, LMS is very sensitive to changes in data points close to the mean and **Hetmansperger and Sheather** provide some examples which show this is the case. Given these **two** problems we decided not to include LMS as one of the estimators used in this thesis.

1.5.5 Summary

It appears not much robust work has been done to date with economic data and nor does it seem that the robust estimators emerge as winners in a race against more conventional techniques. Clearly what is needed is an analysis of many economic data sets using robust techniques to help form an initial conclusion as to their general impact. We also found a weakness of existing work to be the lack of any explicit benchmarks that helps determine whether robust analysis does indeed give different results.

1.5.6 What is to be done?

From reviewing these studies, using robust estimators, it is clear that there is some reticence to use the robust standard errors. **Lichtenberg and Siegel** (1991) is an example. Others like **Connolly** (1989) grip the nettle and use them to reëxamine well-established inferences. One study, **Swinton and King** (1991) uses the robust estimators available in the statistical package SHAZAM (**White** (1978)). A hitherto unexamined question is whether methods of calculating the robust standard errors, such as SHAZAM (**White** (1978)) provides, are satisfactory? We attempt to answer this question in Chapter Five. Also an obvious lacuna in the empirical robust literature is an extension to *many* data sets in a particular field like economics. There are even not many for particular data-sets. This paucity provides a springboard from which this thesis can begin to add to our understanding of robust estimation.

Chapter 2

Robust Estimators

2.1 Introduction

Classical statistical tests in the regression context require the errors to be normally, independently and identically distributed. We are concerned with the sampling properties of estimators for β and test statistics when considering $y = \mathbf{X}\beta + \epsilon$. If ϵ is normally distributed it is well known what are the finite sampling distributions of estimators. However this knowledge of normality is tenuous at best as pointed out by **Koenker** (1978, 1982) including many others. To overcome this problem econometricians have had recourse to large sample theory, also known as asymptotic theory, to make inferences about β. The result of this theory enables one, when the finite distribution of an estimator is not known to be normal, to presume the large sample approximation distribution is normal. Thus one can still employ the Classical t-test to reject or accept hypotheses.

Our knowledge of the extent of outliers is also limited. This limitation can be reduced by comparing robust estimates with ordinary least squares as suggested by **Hogg** (1979). Thus this chapter has a twofold purpose: first there is a need to review the robust estimators and cover their interesting properties. Second we need to determine how tests of hypotheses can be performed using robust methods.

2.2 Empirical Work In Economics

This section gives a brief outline of empirical economics. Central to empiricism is hypothesis testing, thus some discussion is devoted to this topic. Further there is a need to bring together hypothesis testing and robust estimators in the context in which a comparison is made with ordinary least squares.

Darnell and Evans (1990) give the procedure of empirical work as:

1. A hypothesis is deduced from economic theory that is feasible to refute;

2. This hypothesis is recast as a linear regression model and is estimated using ordinary least squares;

3. Residual analysis is undertaken to ascertain if the actual errors satisfy a number of assumptions-such as zero expected value, uniform variance and are uncorrelated -allowing the researcher to test the hypothesis;

4. Once three has been completed the appropriate statistical test can be performed and the hypothesis accepted or rejected. It is likely that all the studies used here have gone through this process.

What is not certain is whether, in going through the above, much, if any, attention is paid to the possibility of outliers in the errors. It could be, applied economists take the robust literature seriously and use robust methods in conjunction with least squares and find little evidence of outliers. In the next chapter we consider what it means for a robust estimator to give the same results as least squares.

2.3 Properties of Ordinary Least Squares

To provide a basis for considering the properties of robust estimators we pause to review ordinary least squares. This and the next section draws from **Judge, Hill, Griffiths, Lutkepohl and Lee** (1988).

Linear regression uses the following equation

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e} \tag{1}$$

where $\mathbf{X}$ is fixed with rank k ($\lim_{T\to\infty} T^{-1}\mathbf{X}^T\mathbf{X}$ is a nonsingular and finite matrix) and it is assumed the residuals have the following properties: $E(\mathbf{e}) = \mathbf{0}$ and $E(\mathbf{e}\mathbf{e}')= \sigma^2\mathbf{I}$

with σ^2 finite. If $\mathbf{e}$ is not drawn from the normal distribution and if it is the case $E(\mathbf{e}) = \mathbf{0}$ and $E(\mathbf{ee}')= \sigma^2\mathbf{I}$ with σ^2 finite then:

• The estimator $\mathbf{b}$ is unbiased and has the minimum variance from the class of linear unbiased estimators and is consistent;

• $\mathbf{b}$ is not efficient or asymptotically efficient, especially for those distributions with fatter tails than the normal. If the exact nature of the distribution is unknown, as is likely to be the usual case, a robust technique may be better than the classical normal linear estimator $\mathbf{b}$;

• Hypothesis tests are no longer valid unless one is willing to accept an asymptotic justification for their use. In a sense we are forced to accept this, despite the large sample requirement never being an empirical possibility. It is the best a researcher can possibly accomplish.

Assuming $\mathbf{e}$ is normally distributed brings the following results:

• The ordinary least squares estimator $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ is asymptotically efficient;

• The distribution for $\mathbf{b}$ is Normal, in small samples;

• F-tests of linear restrictions of the form $\mathbf{R}\beta = \mathbf{r}$, and t-tests on individual coefficients are possible in small samples.

Thus even if non-normality is true, the estimator $\mathbf{b}$ is still appropriate if one wishes to use a linear unbiased estimator and hypothesis tests are possible asymptotically. Obviously it is not at all a settled issue in economic testing whether one should choose a linear estimator. The adjective linear, in linear estimator, refers to estimator being a linear function of the dependent variable. In the formula $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ notice the estimator is a matrix times the $\mathbf{y}$'s. Robust estimators cannot be written in this fashion. The problem with using linear estimators and an asymptotic justification for hypothesis testing, as is pointed out by **Koenker** (1982), is first, to restrict oneself to linear estimators may cause one to ignore many non-linear estimators and second, the power of the asymptotic hypotheses' tests, or the ability of the tests to reject the null when the null is false, is sensitive to the assumed distribution of the errors and the power is reduced when, for example, the error distribution has "fat" tails.

We saw in Chapter One that outlying errors affect the ordinary least squares regression line. Robust estimators attempt to reduce the effect of these errors. Thus it seems reasonable to exploit this difference in responsiveness between ordinary least squares and robust estimates to these errors in the context of hypothesis testing. In principle there is no difference in the method of hypothesis testing with either a robust estimator or ordinary least squares. One needs a coefficient estimate and an estimate of the coefficient's standard error. These can be calculated. To be frank, the calculation of the variance-covariance matrix with robust estimators, which provides estimates of the standard errors, is not as easy nor does it it leave one the same degree of comfort as would be the case for ordinary least squares. The reason is the formulæ associated with the robust estimators are complicated and econometric packages do not use the recommended ones in the literature but have their own formulæ for estimating standard errors. However as long as one is aware of the possible pitfalls that might arise with robust estimators and take steps to guard against those situations that could present problems, enough of the difference in responsiveness remains to be able to compare estimators at the level of hypothesis testing. The one step we take in this thesis is to perform a Monte Carlo study for one robust estimator using three ways to estimate the standard deviations. If any of these three ways gives different results compared to the "true" standard error, the hypotheses tests are redone using the alternate standard error estimates. This is done in Chapter Five.

2.4 Some Robust Estimators

The following sections gives details of robust estimators. We know robust estimators are recommended where the error density is infinite or for those densities for which the sample median is a more efficient estimator of location. From examining the small number of studies in economics using robust estimators it appears three classes of robust estimators are popular, namely M-Estimators, L-Estimators and Trimmed Least Squares. The first class is so named because of its relationship to *m*aximum likelihood techniques. The second class is named because its estimators are derived from *l*inear combinations of the quantiles. In the third class some observations are discarded, or *trimmed*, from the data.

2.5 M-Estimators

2.5.1 M-Estimators

Least Squares and Maximum Likelihood

Least squares finds a β to minimize $\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\beta)^2$ and this β must satisfy the normal equations

$$\sum_{t=1}^{T} \mathbf{x}_t(y_t - \mathbf{x}_t'\beta) = 0 \tag{2}$$

Using a maximum likelihood approach we need to find a β that minimizes

$$-\sum ln f(y_t - \mathbf{x}_t'\beta)$$

where the density function is $f(y_t - \mathbf{x}_t'\beta)$. This procedure entails finding a β that must satisfy

$$\sum_{t=1}^{T} \mathbf{x}_t f'/f = 0 \tag{3}$$

It is well known that if $f(y_t - \mathbf{x}_t'\beta)$ is the normal distribution then f'/f is $(y_t - \mathbf{x}_t'\beta)$ and thus least squares is the maximum likelihood estimator for this distribution.

The Link To M-Estimators

Notice the height and slope (see equation three above) of the density, crudely speaking, are important for the maximum likelihood approach as one would expect when fitting a density to data both play a role in finding the best "fit" of the density to that data, which is at a simple or conceptual level, the procedure of maximum likelihood. The ratio of the slope and height, a function of the errors, is the $\psi(\epsilon)$ function and in equation form is

$$\sum_{i=1}^{T} \mathbf{x}_t \psi(y_t - \mathbf{x}_t'\beta) = 0 \tag{4}$$

and it is at this point M-Estimators come into their own by suggesting different ψ functions. The above is the usual way to characterize M-Estimators. We now look at an alternative way that is easier to grasp.

A Better Formulation

A more intuitive method is to examine the original minimization problem $\sum_{t=1}^{T}(y_t - \mathbf{x}_t'\beta)^2$ which can be re-written as $\sum_{t=1}^{T} \mid y_t - \mathbf{x}_t'\beta \mid\mid y_t - \mathbf{x}_t'\beta \mid$ and we can use this in a unique manner. Notice we could have have obtained this last equation by minimizing $\sum_{t=1}^{T} \mid y_t - \mathbf{x}_t'\beta \mid \psi$ and replacing ψ equal to $\mid y_t - \mathbf{x}_t'\beta \mid$. Now ordinary least squares is an M-Estimator where the absolute error is weighted by the absolute error. Any other M-estimator is derived, rather easily, by changing the weight used in the function to be minimized but keeping in mind the weights should have the feature of not becoming large if indeed the absolute value of the error is large. Going back to the usual formulation, it is the function $\psi(\epsilon_t)$ which is the central part of M-Estimators for they capture the intuitive notion of how the errors are weighted. To be consistent with their description in the literature we return to the usual formulation of M-Estimators.

M-Estimators Again

M-Estimators use different forms of the $\psi(\epsilon_t)$. There is a natural way one might choose the form of the function. Where very large absolute errors are possible we would like the $\psi(\epsilon_t)$ to be smaller as the error gets larger and larger in absolute value. Thus one would choose $\psi(\epsilon_t)$ ensuring the resulting estimator has acceptable properties if the distribution is normal. It must also have excellent properties if the errors come from a distribution where outliers are possible with greater probability than the normal. Given that one, to begin with, is ignorant of the true distribution one trys a weighting scheme that does well if the distribution is not normal but also one that does not result in too great a loss if it turns out the distribution is normal. In the following section we present the weighting schemes of some M-Estimators.

M-Estimators thus exploit the ideas of maximum likelihood but in a strict sense they are not maximum likelihood as one does not specify any particular distribution. As maximum likelihood estimation uses the logarithm of the density, any attempt at optimization entails the derivative of a logarithm. This can be written as f'/f. To the extent that M-Estimators use substitutes for this derivative, not from any particular distribution, proponents regard them as a close cousin of the method of maximum likelihood. Many describe them as "maximum likelihood like". Typically, looking for estimators like these involves changing the particular form of f'/f that will do well if there are outliers in samples. The estimators must be efficient, but not perfectly so, if the distribution turns out to be normal. The estimator must be extremely efficient if the distribution produces outliers. Less-than-perfect efficiency with normal errors is a premium paid to insure against distributions other than the normal.

2.5.2 Three M-Estimators

In the regression context, we want to use M-Estimators to estimate the vector β $(k \times 1)$ in

$$y_t = \mathbf{x}_t' \beta + e_t \tag{5}$$

with e_t independent and identically distributed. The e_t have a distribution function $F(e_t)$ and density $f(e_t)$ symmetrical about zero. If the distribution is not symmetric, **Carroll** (1979) has shown in the regression context, the slope coefficients are hardly affected.

We cover the ψ functions of Huber, Tukey and Hampel. In the following exposition all errors are divided by a robust measure of the spead of the errors. Different studies use different robust estimates and common measures are the so-called median of the absolute values of the Least Absolute Errors' residuals,

$$median|residual_i - median(residual_i)|$$

known in the literature as the MAD, or the interquartile range of Least Absolute Errors' residuals is used instead.

Huber

For the Huber M-Estimator the $\psi(e_t)$ function is

$$\psi(e_t) = \begin{cases} e_t & if \mid e_t \mid \leq a \\ asign(e_t) & otherwise \end{cases}$$

Residuals less than a in absolute value are treated in the usual manner. Any other residual has a ψ weight of either plus or minus a. This $\psi(e_t)$ function looks like a "linear" hill passing through the origin. The hill starts at sea level $(-a)$ with a rise until a plateau is reached at (a). The new set of residuals are the "winsorized" (defined in Section **2.5.3**) residuals. **Hogg** (1974) recommends using one and a half as the value for a.

Tukey (Also called the Biweight or Bisquare)

$$\psi(e_t) = \begin{cases} e_t\left(1 - \frac{e_t^2}{a^2}\right)^2 & if \mid e_t \mid \leq a \\ 0 & otherwise \end{cases}$$

This estimator has a ψ function that is wave shaped. For a value of $a = 6$ as recommended by **Hogg** (1974), the wave rises out of the origin reaches a peak just beyond two then falls to zero at six. Thus until unity this estimator is much like least squares but after that it is very different until it reaches its peak where an error just above two has a ψ weight just below two. After the peak all the error weights are replaced by lower and lower values until the point is reached, say six, above which all error weights are zero. M-Estimators that have ψ functions eventually reaching zero are called "hard redescenders". The Tukey is one.

Hampel

$$\psi(e_t) = \begin{cases} sign(e_t)\mid e_t \mid & if \mid e_t \mid \leq a \\ sign(e_t)a & if a < \mid e_t \mid \leq b \\ sign(e_t)a((c - \mid e_t \mid)/(c-b)) & if b < \mid e_t \mid \leq c \\ 0 & otherwise \end{cases}$$

This estimator looks complicated but is really like a linear version of the Tukey estimator. It requires one to set three "tuning" constants. Up to the first constant all error weights are replaced by themselves, the least squares solution. After that until the next constant is reached all error weights are the second constant. This is like the plateau of the Huber or the peak of the Tukey. Between the second and the third constant each error weight within that range is replaced by lower and lower values and once the error is larger than the third constant its weight is zero. This M-Estimator is a "hard redescender".

The M-Estimators are weighted least squares estimators. For instance, BMDP

Statistical Software (**Dixon** (1990)) uses an iterative procedure to minimize

$$\sum_{i=1}^{T} w_t(y_t - \mathbf{x}_t'\beta)^2 \tag{6}$$

with

$$w_t = \frac{\psi(e_t)}{e_t} \tag{7}$$

All N of the w_t are formed into a diagonal matrix W and the estimator becomes, where the data is defined for a Gauss-Newton procedure,

$$(X'WX)^{-1}X'W\mathbf{y} \tag{8}$$

The robust estimators provide the appropriate $\psi(e_t)$ to use for determining the weights. The econometrician used to classical least squares can see M-Estimators in the same way except as least squares on transformed data where the transformations are provided by the ever active ψ functions. The effect of this transformation is the resulting residuals, once all the iterations have been completed, are like "normal" ones. The logic behind this is we may begin with errors that are not normal. As we apply the robust method it "ignores" outliers and on each iteration the resulting residuals will contain fewer and fewer outliers. The residuals thus will begin to look like they came from a normal distribution. The process gives residuals that are more "normal" looking. On the basis of this, **Koenker** (1982) maintains hypothesis testing is possible, on an asymptotic basis, using the transformed data but points out this does not work for an L-Estimator like least absolute errors.

2.5.3 The Estimation Process

One problem with M-Estimators is they are not always scale invariant. If we multiplied all the errors by a constant the estimate based on the transformed errors would not be that constant times the estimate based on the raw errors. To overcome this problem one divides the errors in the ψ function by a robust measure of scale, such as the interquartile range of Least Absolute Errors' residuals. A better approach is to estimate both β and σ together. Before we begin one defintion from **Dixon and Tukey** (1968) is needed:

If $y_1 \leq y_2 \leq \ldots \leq y_n$ are the ordered observations of a sample, the g-and-g times Winsorized observations are defined by:

$$z_1 = z_2 = \ldots = z_g = y_{g+1}$$

$$z_{g+i} = y_{g+i}, for 1 < i < h = n - 2g \ (1)$$

$$z_n = z_{n-1} = \ldots = z_{n-g+1} = y_{n-g}$$

and their mean

$$\tilde{z} = \frac{z_1 + z_2 + \ldots + z_n}{n} = y_w \ (2)$$

is the (g-and-g times) Winsorized mean of the original sample. (p83)

This is to be distinguised from a trimmed mean

$$\tilde{z} = \frac{y_{g+1} + y_{g+2} + \ldots + y_{n-g}}{n - 2g} = y_T$$

As an example, some raw data and the Winsorized values are shown in Table 2.1 taken from **Dixon and Tukey** (1968). Armed with this definition of "winsorized" we can examine one of the many algorithms to do this, usually variants of the following structure:

1. Start with a least absolute errors estimate of β and a robust estimate of σ (**Harvey**, 1977);

2. For the Huber ψ function replace those errors with absolute values greater than a with $-a$ or $+a$. These are the "winsorized" residuals. The term a is the tuning constant associated with the ψ function. Subsection **2.5.2** gives values for the tuning constants. For the Huber estimator this might be one and a half. Other estimators would "winsorize" using their ψ functions;

3. Use the "winsorized" errors to determine the weight each error receives;

4. Find the weighted least squares estimate using the weights calculated using the "winsorized" errors, the so-called updated estimate of β;

Raw Values	Once Winsorized	Twice Winsorized
63	24	19
24	24	19
19	19	19
12	12	12
5	5	5
-1	-1	-1
-3	-3	-3
-3	-3	-3
-17	-17	-17
-24	-24	-24
-30	-30	-24
-36	-30	-24

Table 2.1: An Example of Raw and Winsorized Data

5. Use this updated estimate of β to obtain the new or updated errors to make a new robust estimate of σ using, for example, the interquartile range of the new residuals (**Street, Carroll and Ruppert**, 1988);

6. Now steps one to five can be repeated until the final estimates are those for which certain convergence requirements have been satisfied. These requirements usually entail the estimates not changing by a specified very small (10^{-4}) amount from one iteration to the next or the residual sum of squares from the regression not changing by a specified very small (10^{-5}) amount from one iteration to the next.

There are also a number of algorithms to calculate the asymptotic variance and co-variance matrix for the estimators and thus hypothesis testing is possible with M-Estimators. More details are provided in Subsection **2.5.4**. The software package BMDP (**Dixon** (1990)) is used to generate M-Estimator in this thesis.

2.5.4 Hypothesis Testing

Hypothesis testing with M-Estimators, speaking loosely, takes the usual form of coefficients divided by standard errors. There are complications. These concern the estimated variance covariance matrix. While there exist formulæ for the estimated variance covariance matrix of each M-Estimator, see equation (11) below for the Huber M-Estimator, but statistical software such as BMDP (**Dixon** (1990)) does not use them. Rather they have another way of estimating the standard errors. This problem is recognized by **Hogg, Ruberg and Yuh** (1990) who point out programmes use other estimates of the variance covariance matrix. They also mention there is no consensus on which estimate is the best one to use. We adopt the BMDP (**Dixon** (1990)) approach here as it is likely that economists would have calculated M-Estimators using BMDP (**Dixon** (1990)) as it is widely available and easy to use. Thus our results are likely to have been conducted in the same manner as that adopted by other researchers. In addition, some other results concerning hypothesis testing are discussed below.

Asymptotic Distributions of M-Estimators

It is possible for the asymptotic distribution of the estimator to be normal no matter the density of the errors. This means t tests of hypotheses are feasible with M-Estimators.

Yohai and Maronna (1979) show that

$$\sqrt{T}(\hat{\beta}-\beta)\xrightarrow{d}N(0,\sigma^2(\psi,F)(T^{-1}X'X)^{-1}) \tag{9}$$

and where ψ' is the first derivative of ψ

$$\sigma^2(\psi,F)=E_F\psi(e_t)^2/[E_F\psi'(e_t)]^2 \tag{10}$$

For the Huber estimator the above asymptotic variance covariance matrix is estimated with

$$(1/T-k)(((1+(k/T))((1-m)/m))/m)^2e^{*'}e^*(X'X)^{-1} \tag{11}$$

where e^* are the "winsorized" residuals and m is the proportion of errors such that,

$$-a < e_t < a \tag{12}$$

This estimate of the variance-covariance matrix has done well in a Monte Carlo study by **Huber** (1973) although one should exercise caution for T/k lower than eight. If one looks at the expression for the variance covariance matrix one sees familiar terms except for an odd looking squared adjustment in the middle. The reason for this adjustment is an attempt to correct for not knowing the true density of the errors. The Tukey and the Hampel have their own correction factors. The programme we are to use to calculate our set of M-Estimators is BMDP (**Dixon** (1990)) as we feel others would have opted for this easy to use programme when employing the **Hogg** (1978) call to do both least squares and use a robust method. BMDP (**Dixon** (1990)) uses a Gauss-Newton algorithm to find estimates of β. This process is iterative and ends when certain criteritia are met. One criterion is the weighted residual sum of squares from one iteration to the next change by a small pre-set amount. During the iterative process a matrix that BMDP (**Dixon** (1990)) calls the **G** matrix is inverted. The asymptotic variance covariance matrix of the estimated coefficients is calculated as a combination of the elements of the inverted **G** matrix, after the process has converged and the final weighted residual sum of squares adjusted by the number of observations and the number of estimated coefficients including a constant term as outlined by **Dixon** (1990, p1299).

Iterative Weighted Least Squares

Given a robust estimate, termed $\tilde{\beta}^r$ by **Bickel** (Discussion appended to **Bickel** (1976), page 167) creates the pseudo observations

$$\tilde{y}_t = x_t\tilde{\beta}^r + \tilde{\lambda}^{-1}\psi \tag{13}$$

where

$$\tilde{\lambda} = n^{-1}\sum \psi' \tag{14}$$

The least squares estimate of β using the pseudo observations is the robust estimate using the original data. **Bickel** argues testing of hypotheses is possible with the pseudo observations using least squares techniques. **Koenker** (1982, p235) also makes the same point. When the robust regression takes the form of estimating β and σ simultaneously **Huber** (1977, p38) maintains ordinary weighted least squares is one possible way to generate estimates. Obviously the ψ function plays a major role in determining the weights. One way to view weighted least squares is as least squares applied to the transformed data. Now applying the same idea suggesting the pseudo observations could be used in the classical manner to the iterative reweighted method. With the latter method the robust estimate based on the original data is the same as least squares applied to the reweighted or tranformed data. The residuals, calculated using the original data, the robust estimate of β and an estimate of σ, after all iterations and reweighting should be close to "normal" ones, or as if they were generated from the normal distribution. The idea is to exploit this and conduct hypothesis tests in the usual way based on the estimated variance covariance matrix calculated as it would be for least squares but with the residuals-calculated using the original data, the robust estimate of β and an estimate of σ-after all the iterations are completed. These estimates are inconsistent. **Schrader and Hettmansperger** (1980) put it this way:

> **The most natural tests derive from the iteratively reweighted least squares algorithm for computing $\hat{\beta}$, ... If the final configuration of weights is treated as fixed and given *a priori*, a least squares weighted analysis of variance could be done. This may be a reasonable procedure with small sample sizes; we have no evidence to the contrary. Asymptotic theory does not support it, however. (p96)**

For the Tukey estimator this inconsistency produces lower standard errors than would be the case had the correct method been used. This would mean the t value would be higher than it should. To overcome this problem one can adjust the critical values from tables upwards (**Magee** (1991) points out it is often easier to adjust the test statistic) to produce the correct decision when testing hypotheses. Using the Tukey with $a = 6$ and a sample size of twenty (**Gross**, 1977) one uses at the five percent level 1.36 times the critical t (from tables) with degrees of freedom

$$\frac{(3T-19)}{4} \tag{15}$$

where T is the sample size. This procedure will produce critical t values about twenty percent higher.

Other Methods

Usually in economics the researcher evaluates the Student t test when testing hypotheses. If we entertain the possibility the e_t come from a distribution other than the normal, the Student t test will result in an acceptance of the null with greater frequency. The reason: the estimate of the variance can be come very large if the errors come from a distribution with fatter tails than the normal. A solution to this is to substitute robust estimates of parameters and scale into the test statistic. It is not obvious what degrees of freedom one must use to obtain the critical values from tables for purposes of comparison.

One way to avoid this is to construct critical t values using the M-Estimators and Monte Carlo methods. For the Tukey (biweight or bisquare) estimator critical values have been worked out for the location model and shown in the box below the five percent level and a value of $a = 9$. A problem with these methods is they are not comprehensive and are limited to given levels of significance and particular sample sizes.

CRITICAL VALUES*		
Tukey	T	t
a=9	10	2.57
	20	2.18
	30	2.09
	40	2.05
	50	2.03
	100	1.99

* **Iglewicz** (1983)

Where the critical values are calculated in the above manner the t value is adjusted to ensure the robust test has the same level of significance of the most conservative test, usually that assuming a normal distribution.

Another approach for location is to take the t values from tables and increase them by a constant where the constant is determined by Monte Carlo methods. For a Hampel (with $a = 2.25, b = 3.75, c = 15.0$) one would use $0.6(T-1)$ degrees of freedom. For the Tukey with $a = 9$ this corresponds to using $0.75(T-1)$ degrees of freedom for sample size of ten. For larger sample sizes the usual degrees of freedom suffice. **Hogg** (1979) recommends one use the Tukey with $a = 6$. Also, notice in the table the t values are close to two. Even though the critical values are for the case where $a = 9$ we are to do hypothesis testing in the usual manner with the Tukey M-Estimator as it is unlikely that economic researchers made these refinements. Also the problem is compounded by the fact that the proponents of robust estimators like **Hogg** (1979) recommend an estimator yet the refinements apply to a different estimator not recommended. We opted for the recommended estimator under the supposition that is what others would have done.

For the Huber Estimator of location and values of a between unity and two, one possible interpretation of **Boos** (1980) is the usual t statistic will be adequate for sample sizes of twenty or more.

Summary

If we calculate the M-Estimators in a weighted least squares manner, using alternative critical values by adjusting t values can be avoided. To do this requires consistent estimates of the standard errors. Also if one wanted to use adjusted t values it is necessary to have a consistent estimate of the standard errors. None of these options are possible with software we chose to use, namely BMDP. We chose to use BMDP as it is widely available and easy to use. We found with just a few commands, one can produce estimates of coefficients and standard errors without spending hours in manuals and learning a complicated language. Though BMDP does not use the estimate of the variance covariance matrix we would like to see, we feel that those doing economic research in economics would have used BMDP to check their least squares' results. There are adjustments that can be used with the M-Estimators when doing hypothesis tests, although they are not used by those doing empirical work a great deal. These we feel are not widely known and nor do they apply to recommended M-Estimators. For these reasons we use the BMDP estimates to do hypothesis testing.

2.5.5 Monte Carlo

The evaluation of M-Estimators with Monte Carlo methods has focussed on an estimate of the average often called location. These studies, such as **Gross** (1976), do not, except a column of ones for the independent variable, cover the linear regression case. One exception is **Holland and Welsch** (1977) although it is not comprehensive. They consider eight M-Estimators including the Tukey and the Huber. Using the normal distribution these estimators lose nine and six percent in efficiency for a sample size of twenty. With a long tailed distribution, the Tukey achieves the lowest variance for sample sizes ten, twenty and forty.

The general impression from Monte Carlo studies of M-Estimators in the regression context is they tend to outperform least squares in non-normal situations. The more difficult problem is to decide whether these studies can point to which M-estimator to use as an alternative to least squares. Monte Carlo studies point to the "hard redescenders" being better than other options. Examples of "hard redescenders" are the Tukey and the Hampel (See Subsection **2.5.3** for details of these M-Estimators). The Monte Carlo studies also show the sacrifice of efficiency at the normal with the "hard redescenders" can be high. One M-Estimator that is not a "hard redescender" is the Huber estimator and does not give up as much efficiency at the normal. It does do well when the distribution has "fat-tails" but not as well as the "hard redescenders". Looking at the Monte Carlo results it seems wise to use a "hard redescender" and the Huber. This provides good efficiency at the normal and gives excellent protection for more elongated distributions. A problem we have is the need to perform hypothesis testing and the "hard redescender" we use for this purpose is the Tukey.

2.5.6 So Which M-Estimator?

There is a clear message from the work on M-Estimators for econometricians. The normal distribution is not the only distribution to consider. For instance some economic data has errors associated with fatter tailed distributions. Some economists recognize this and may employ some form of observation deletion when doing empirical work. An example is **Coate and Uri** (1988). Although, one might do much better with an M-Estimator. **Relles and Rogers** (1977) asked statisticians to make estimates of a location parameter using outlier rejection as a means for making those estimates. M-Estimators were also used to make estimates. The estimates by the statisticians and the M-Estimators were compared using variances of estimates in a Monte Carlo study. Three M-Estimators of location prevail over the statisticians. We now look at this study in more detail.

Outlier Rejection and Robust Estimates

Relles and Rogers (1977) are interested in the performance of the outlier rejecting statistician relative to robust estimators. Five statisticians were shown fifty data "configurations", drawn from the Student distribution and they had to provide their own estimate of location based on the "configurations". Also, robust estimates are made using 250 "configurations" from the Student distribution. To show the statisticians all the "configurations" would have been too time consuming. You can show the statisticians less than 250, but this requires a means of obtaining the distribution of their location estimates, had they been able to look at all 250 "configurations". This is done with a numerical procedure amounting to a Bayesian calculation and, using crude terms but which capture the essence of the procedure, the posterior distribution, of the statisticians' estimates, is calculated. Or as they put it:

> **they are exactly the computations one would perform to obtain ...Bayesian posterior means, (p109)**

Relles and Rogers has been cited by **Simonoff** (1984), *inter alia*, and to quote from the latter is constructive:

> **One comparison of robust and outlier-detection method estimates is a study by Relles and Rogers (1977). Rather than use the "objective" outlier detection procedures in the literature, however, they used the subjective opinions of several statisticians to trim off outliers. They found that this outlier detection procedure worked fairly well, although not as well as the robust procedures.(p815)**

While this result may not carry over to the regression case it is nonetheless reassuring, and we have also seen that M-estimators are nothing more than weighted least squares estimators. They are more familiar to us than we realize. Statistical software has not caught up with theoretical developments as **Hogg, Ruberg and Yuh** (1990) do point out that available programmes provide their own *estimate* of the variance covariance matrix, and thus hypothesis testing must be done with caution, although as we point out others are likely to have made use of this software. In Monte Carlo studies, a Tukey estimator with a "tuning" constant of 6 is popular as it has performed well (and some say the best in the location case) with many distributions, and it has done well in forecasting (**Fair** (1974)) and as **Hogg** does suggest 6 we adopt it in this thesis. **Hogg, Ruberg and Yuh** (1990) do point out that the weighted least squares method does provide an *estimate* of the variance covariance matrix. As long as we realize it is an estimate it can still be used. Thus, in addition, we are to estimate the Huber with a "tuning" constant of 1.5. We are careful to use robust starting values, from Least Absolute Errors, in the BMDP procedures.

2.6 L-Estimators

This section has the following structure:

1. The formula for the θth regression quantile an ingredient of L-Estimators, is given;
2. The asymptotic distribution results are shown for the quantile estimate β^*;
3. For expository purposes, attention is paid to the least absolute error estimator and how it is used to conduct hypothesis testing.

The process underlying hypothesis testing with a robust estimator such as the least absolute error estimator is in principle no different from testing with ordinary least squares. We need an estimate, and an estimate of the standard error of the estimate. Usually with these a t-test would be constructed and the significance of the estimate checked with a critical value from the t-distribution. Some adjustments to this procedure are presented.

2.6.1 Regression Quantiles

The θth regression quantile $(0 < \theta < 1)$, results from minimizing the following with respect to β

$$\left[\sum_{(t|y_t \geq \mathbf{x}_t'\beta)} \theta(|y_t - \mathbf{x}_t'\beta|) + \sum_{(t|y_t < \mathbf{x}_t'\beta)} (1-\theta)(|y_t - \mathbf{x}_t'\beta|)\right]. \tag{16}$$

If $\theta = 1/2$, (16) is the same as minimizing the sum of the absolute values of the residuals or the least absolute errors estimator. Although this looks formidable to calculate, there are algorithms (**Narula and Wellington** (1986) and **Rech, Schmidbauer and Eng** (1989)) that exploit the linear programming nature of the problem and can find a solution efficiently.

Koenker and Basset {Theorem 4.2, page 43}(1978) is reproduced exactly below.

Let $\{\beta_T^*(\theta_1), \beta_T^*(\theta_2), \ldots \beta_T^*(\theta_M)\}$ with $0 < \theta_1 < \theta_2 < \ldots < \theta_M < 1$ denote a sequence of unique regression quantiles from model (16). Let $\xi(\theta) = F^{-1}(\theta)$, $\boldsymbol{\xi}(\theta) = (\xi(\theta), 0, \ldots, 0) \in \Re^K$, and $\boldsymbol{\xi}_T^*(\theta) = \beta_T^*(\theta) - \beta$. Assume:

(i) F is continous and has continous and positive density, f, at $\xi(\theta_i)$, $i = 1, 2, \ldots$M, and

(ii) $\mathbf{x}_{1t} = 1 : t = 1, 2, \ldots$ and $\lim_{T\to\infty} T^{-1}X'X = Q$, a positive definite matrix.

Then,

$$\sqrt{T}[\boldsymbol{\xi}_T^*(\theta_1) - \boldsymbol{\xi}(\theta_1), \ldots, \boldsymbol{\xi}_T^*(\theta_M) - \boldsymbol{\xi}(\theta_M)]$$

converges in distribution to an (MK)-variate Gaussian random vector with mean 0 and covariance matrix $\Omega(\theta_1, \ldots, \theta_M; F) \otimes Q^{-1}$ where Ω is the covariance matrix of the corresponding M ordinary sample quantiles from random samples from distribution F.

In the theorem Ω is the variance-covariance matrix of the M quantiles' with elements,

$$w_{ij} = \frac{min(\theta_i \theta_j) - \theta_i \theta_j}{f(\xi_{\theta i}) f(\xi_{\theta j})} \tag{17}$$

Let us try and make sense of this theorem. First T represents the sample size and the * indicates an estimate. Second $\xi(\theta)$ has many zero elements. In fact only the intercept has a non-zero element. This non-zero element is an adjustment equal to the θth quantile. So if we look at the expression

$$\sqrt{T}[\boldsymbol{\xi}_T^*(\theta_1) - \boldsymbol{\xi}(\theta_1), \ldots, \boldsymbol{\xi}_T^*(\theta_M) - \boldsymbol{\xi}(\theta_M)]$$

we note that the estimated value is subtracted from the true value except for the intercept where there is an adjustment made for each θ equal to the θth quantile. The quantile estimates of coefficients can be compared to their true values in the usual way but not for any constant term you need to estimate. The notation of the proof, especially concerning $\xi(\theta)$ attempts to capture this notion concerning the intercept term. Notice the data $\mathbf{x}_{1t} = 1 : t = 1, 2, \ldots$ includes a column of ones for the intercept. Missing is the fact that there must be T, t's. This theorem is paramount as it is the most important result of a procedure that allows hypothesis testing with L-Estimators. The key elements of the theorem are a set of quantile estimates for various values of θ, an adjustment to the intercept, a given distribution with a continuous non-zero density and the data must have some benign features. In this environment the limiting distribution of the the difference between an estimate and its true value has a normal distribution.

A special case of the regression quantile estimator is the least absolute error estimator and is obtained when $\theta = 1/2$. Further the estimate is denoted as $\beta^*(1/2)$. Without limiting ourselves to any distribution of the errors we can use the limiting distribution of the quantile estimator to obtain the elements of the variance-covariance matrix for this estimator. If we consider the median error it is going to be the median of the errors, zero: the value dividing the distribution into two equal halves, or $F(0) = 1/2$. Recall $\xi_\theta = F^{-1}(\theta)$ is the θth quantile. If this is the case $\xi_{1/2} = 0$, or those values of the errors for which $P(y - x'\beta < 0) = 1/2$. The Ω matrix which is an element of the limiting distribution of the quantile estimator has $\xi_{1/2}$ thus we have $\omega_{ij} = \omega_{ii}$ and i equal to the number of quantiles which is one, thus $\omega_{ij} = \omega_{ii}$ and using (17) obtain

$$w_{ij} = [2f(0)]^{-2} \tag{18}$$

giving the limiting distribution of $\sqrt{T}(\hat{\beta}^*(1/2) - \beta)$ as $N(0, [2f(0)]^{-2}Q^{-1})$. As we are not making any assumptions about f, we need to estimate f. A way to do this has been suggested by **Cox and Hinkley** (1974) and has

$$\hat{f}(0) = \frac{2d}{T(\hat{e}_{m+d} - \hat{e}_{m-d})} \tag{19}$$

and $\hat{e}_{(1)}, \hat{e}_{(2)}, \ldots, \hat{e}_{(T)}$ are the ordered $\hat{\beta}^*(1/2)$ residuals. The residual $\hat{e}_{(m)}$ is the one closest to $T/2$ that is zero. The main difficulty with this estimate is "d". As can be seen from the formula for $\hat{f}(0)$, "d" is used to select which ordered residuals one uses to estimate the height of the density. A well-known (**Rech, Schmidbauer and Eng** (1989)) result in Least Absolute Errors estimation is at least k of the residuals are zero and thus it is possible if two zero residuals are selected and subtracted from one another the denominator in equation 10 will be zero and thus $[2\hat{f}(0)]^{-2}$ is very small. Thus the estimated variance covariance matrix is sensitive to the choice of "d". The statistical package SHAZAM (**White** (1978)) without much justification calculates "d" as $\frac{T-k}{6}$, where T allows. As the asymptotic distribution of the least absolute error estimator is normal and the variance covariance matrix is capable of being estimated, it is possible to perform t-tests of hypotheses concerning β. But "d" is going to be a problem when k/T is large. We should guard against this possibility by checking to see if the estimated variances produced for our data sets are too low. This we do in

Chapter Four. It is possible the t-test is sensitive to the chosen value of "d" and we now consider this issue.

The least absolute error estimator is the 0.5th quantile estimator, so the asymptotic distribution results of the quantile estimator can be used to aid in the construction of the "P-value" with this estimator. Recognizing that with the least absolute error estimator, θ equals one half produces a particular expression for the variance-covariance matrix of the estimator. The constituent parts of this expression are either calculated from the data or estimated. The one item that is estimated is the height of the density function of the errors for the least absolute error that is zero. The econometrics package uses equation 19 to do this. We could rewrite this equation as

$$\hat{f}(0) = \frac{h(d)}{Tg(d)} \tag{20}$$

notice $h'(d)$ is positive and if the residuals are ordered from highest to lowest then it must be $g'(d)$ is also positive. This can help us examine the question:-what happens to $\hat{f}(0)$ when "d" changes? This estimate could fall or rise depending on the sign of $g(d) - dg'(d)$, obtained from using equation 20 and the quotient rule and selecting the possibly non-positive parts of the resulting expression. If this latter expression is positive, a higher value of "d" raises the "P-Value". For every data set, at least for the least absolute value estimator it is possible to examine how the "P-value" changes when changing "d" around the value used by the econometrics package SHAZAM (**White** (1978)) . This suggests a Monte Carlo study is needed to see if the sensitivity is so great that it makes inference difficult. Such a study is performed in Chapter Five.

As regards the quantile estimator usually what is done is to estimate the coefficients based on a number of quantiles. These estimates are a function of θ and denoted $\beta^*(\theta)$. Having done this, these estimates are then combined with a weighting scheme. The weights denoted by π, themselves functions of the chosen θ's. Possible weighting schemes that have been suggested are the following:

1. The five-quantile with $\theta = 0.1$, 0.25, 0.5, 0.75, and 0.9 and $\pi(\theta) = 0.5$, 0.25, 0.4, 0.25, 0.05;

2. The Gastwirth with $\theta = 0.33$, 0.5, 0.67 and $\pi(\theta) = 0.3$, 0.4, 0,3;
3. The Tukey trimean with $\theta = 0.25$, 0.5, 0.75 and $\pi(\theta) = 0.25$, 0.5, 0.25.

The symbol π captures the symmetrical weighting scheme and the quantile estimators are linear functions of the regression quantiles. Generally these estimators can be written as

$$\beta^*(\pi) = \sum_{i=1}^{m} \pi_i(\theta_i)\beta^*(\theta_i) \tag{21}$$

where $\pi = [(\pi(\theta_1), \pi(\theta_2), \ldots, \pi(\theta_m))]^T$. We need to have some way of differentiating between the five-quantile, the Gastwirth and Tukey: one scheme is β_f^*, β_g^* and β_t^*. As $\pi(\theta_i)$ and ϵ_i are symmetric $\beta(\pi)$ is an estimator of β and the limiting distribution of $\sqrt{T}\beta(\pi) - \beta$ is $N(O, \pi^T\Omega\pi Q^{-1})$. This is theorem 4.3 of **Koenker and Basset** (1978, page 46). The limiting distribution provides the avenue for statistical testing as long as we are able to obtain a consistent estimator for Ω. However the elements of Ω, ω_{ij} require an estimate of $f(\xi_{\theta i}), \hat{f}(\xi_{\theta i})$. As in the case of the least absolute errors estimator we were able to estimate densities for the regression quantiles as

$$\hat{f}(\xi_\theta) = \frac{2d}{T(\hat{e}_r - \hat{e}_s)} \tag{22}$$

where $r = [T\theta] + d$ and $s = [T\theta]$-d. Also note the estimate is sensitive to the value of "d". Considering the least absolute value estimator as a special case of estimation using quantiles, so far we have four alternatives to ordinary least squares. Further we can perform hypothesis testing using these four procedures although there are problems associated with their calculation. However this does not stop us using them as long as we realize the limitations of such an exercise.

So Which L-Estimator?

One advantage to L-Estimators is they can be used to generate regression counterparts to the location estimator of Gastwirth and Tukey's trimean, see items 2. and 3. above. With the regression quantiles, we get three estimators to choose from, four counting least absolute errors. Monte Carlo work on these estimators is sparse partly because it is well known that they will outperform least squares for a number of error densities where the median is preferred to the mean on efficiency grounds. These include densities leading to Mandelbrot's concerns with economic data. It also includes densities with variances that are defined such as the double exponential. We covered, in Chapter One the applied study that has used all four estimators, **Swinton and King** (1991) and with the particular data of interest not much is gained using these robust estimates. We propose to do the same. **Koenker and Basset** (1982) do point out the loss of efficiency of the least absolute errors estimator can be "quite high". This estimator can be better in non-normal situations. It is the maximum-likelihood estimator for the double exponential or Laplace distribution.

In the context of tests of location, we noted earlier that for M-Estimators any critical t value may have to be adjusted or the degrees of freedom need to be changed. Similar adjustments are recommended for estimates of location using the L-Estimators of Tukey and Gastwirth by **Patel, Mudholkar and Fernando** (1988). Even though these adjustments exist, few employ them in empirical work. Thus we do not employ them here as we want to use the methods that the original researchers used to test hypotheses. Also these adjustments are for the location case which we do not have.

2.7 Trimmed Least Squares

Another estimator that has been used by (**Hallman and Kamstra** (1989)), for example, to deal with outliers is the trimmed least squares estimator. This estimator, like the M-Estimators, has an easy characterization. One begins by selecting a trimming proportion α between zero and one half. Applied work uses five, ten and twenty percent and we use the same. Using a quantile estimator determine $\hat{\beta}^*(\alpha)$ and $\hat{\beta}^*(1-\alpha)$. Drop all observations where

$$y_t - \mathbf{x}_t'\hat{\beta}^*(\alpha) \leq 0 \tag{23}$$

or

$$y_t - \mathbf{x}_t'\hat{\beta}^*(1-\alpha) \geq 0 \tag{24}$$

and use least squares on the remaining observations. This is the trimmed least squares estimator and its limiting distribution is normal and the variance covariance matrix is easy to calculate. Monte Carlo work, **Ruppert and Carrol** (1980) and **de Jongh and de Wet** (1985) show this to be a good alternative to least squares where large errors are possible. Empirical studies have used this estimator, for example **Connolly** (1989). Even though **Patel, Mudholkar and Fernando** (1988) recommended using adjusted degrees of freedom we do not use them as others have not done so. We do not wish to complicate matters but are to use the methods that others have followed when doing empirical work.

2.8 Summary and Conclusion

Chapter One proposes a novel idea. Given that economists use data to substantiate various economic ideas, obtain data used in empirical work and determine if robust estimators give different results from least squares when using robust estimators. Before showing exactly how to implement this idea, which is done in the next chapter, Chapter Three, details of hypothesis testing and some important features of the robust estimators are provided here. On the basis of available software, hypothesis testing, Monte Carlo studies and the robust estimators used to date in economics we are to use the Tukey and Huber (M-Estimators) and four quantile estimators (L-Estimators). One of the quantile estimators is also called the Tukey but should not be confused with the M-Estimator of the same name. The least absolute error estimator can be both an L-estimator and an M-Estimator but we place it in the former category for this thesis. We also propose to use the trimmed least squares estimator, using two trimming proportions. Thus, in total, we use eight robust estimators in this thesis. Although there exist adjustments that can be made to critical values for hypothesis tests we do not make them as they are not made in the empirical literature.

Chapter 3

Criteria to Compare Robust to Least Squares

3.1 Criteria for Comparing Estimates

3.1.1 Introduction

In the previous chapter, we pointed out that statisticians and econometricians have been suggesting that for diagnostic purposes researchers should compare ordinary least squares and robust methods. Casual observation indicates this suggestion is rarely taken seriously. A possible reason for this is that data typically used in economic research does not result in robust methods making any difference.

The purpose of this thesis is to investigate this issue. What does it mean to say that robust methods produce "different" results? Fourteen metrics are provided below to address this question, falling into five general categories. Four criteria fall into Category One, called Coefficient Estimates' Differences:

- Criteria I and II which are measures comparing the magnitudes of particular estimates;
- Also in Category One are Criteria III and IV which are magnitudes relative to standard deviations.

Making up Category Two: Tests of Differences are:

- Criteria V and VI being tests of the joint "null" mentioned in Chapter One: either outliers are not a problem in economic data or applied economists use the **Hogg** (1978) technique and only publish least squares results if they agree with a robust estimate.

Variance Differences comprise the Third Category:

- Criteria VII and VIII concern themselves with the variances of estimates.

The fourth category is called Forecasting Differences and:

- Criteria IX and X use forecasting methods to establish whether robust estimates are different from least squares.

The final and fifth category is called the Hypothesis Testing Differences' Category made up of:

- Criteria XI through XIV which use hypothesis testing as a basis for comparison.

For each criterion is provided a subjective opinion of how different on that criterion the robust estimators need to be, to be thought "different" from least squares to a "significant" degree. The idea is to provide target levels for the various criteria, nearly all the ones mentioned in this chapter. If in the next chapter where the actual outcomes are given, the calculated number exceeds the target then least squares and robust methods are viewed as different. For example, to have the robust techniques be viewed as different on the first criterion the average absolute percentage change must be above fifteen percent, where this number, fifteen, reflects my personal judgement. In addition, each criterion is provided a subjective "grade" on a scale of one to ten to indicate its importance. Here a "grade" of ten indicates extremely important and lower grades indicates less importance for the purposes at hand.

3.1.2 The Fourteen Criteria

Coefficient Estimate Differences-Category One

Criterion–I *Percentage of all robust estimates differing from least squares by more than thirty percent. Also by more than fifty percent.*

This criterion builds the distribution of the percentage change in the estimated slope coefficients relative to a "cutoff" (such as used by **Coursey and Nyquist** (1988, p607)) and a decision made as to whether robust analysis is different from least squares. The average is calculated from the many coefficient estimates: both robust and least squares. In the next chapter where the results are given we provide the sample sizes for each calculation. At the end of this chapter is a summary of the sample sizes.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above fifteen for changes more than thirty and above ten for changes more than fifty.
- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve a seven. Usually in economics less attention is given the size of a coefficient and more weight put on the estimated standard errors. If one objects to this, the size has a high "grade" relative to other criteria.

Criterion–II *Percentage of the regressions in which at least one estimated slope coefficient changes by more than thirty percent. Also, changes by more than fifty percent are noted.*

In one of the few studies that attempts to compare least squares with robust estimators **Coursey and Nyquist** (1988, p606) adopt this thirty percent criterion for the largest percentage change in a coefficient and intercept in nine regressions, all with the same number of coefficients. While the thirty and fifty percent are arbitrary we adopt them here as **Coursey and Nyquist** seem to feel it was appropriate. In a footnote **Coursey and Nyquist** (1988, p607) report the number of cases where the largest difference is fifty percent for another data set.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above fifteen for changes more than thirty and above ten for changes more than fifty.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve a seven. Usually in economics less attention is given the size of a coefficient and more weight put on the estimated standard errors. If one objects to this, the size has a high "grade" relative to other criteria.

Criterion–III *Percentage of the regressions in which at least one estimated slope coefficient changes by more than one ordinary least squares' standard deviation. Also changes by more than two standard deviations are calculated.*

Swinton and King (1991, pp446–47) adopt this criterion for determining whether robust methods provide coefficient estimates that differ significantly from those of least squares. Economists often work with an estimate and using the idea of repeated samples will build a confidence interval for an estimate that is based on the point estimate plus or minus some multiple of the standard error of that estimate. The reason is, for the normal distribution ninety-five percent of the estimates from repeated samples should lie with two standard deviations of the estimate. A robust estimate, if it falls outside of that confidence interval of the least squares' estimate, is regarded as different from the least squares' estimate. In a sense, the robust estimate is producing an unusual value, relative to the confidence interval that is expected in repeated samples and thus is regarded as different. A problem that we face is the interval produced with the least squares' estimate and the estimated standard error of the estimate, is itself an *estimate* of this confidence interval. However it has been used to compare estimates and we use it in the same manner.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is more than eighty-two for one standard deviation and more than twenty-eight for two standard deviations.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an ten. This high grade reflects the importance attached, either rightly or wrongly, to the estimated variances in economic hypothesis testing.

This is a critical criterion and the hurdle has been set at eighty-five percent for one standard deviation. It is twenty-three percent for two standard deviations, for reasons explained below. If we suppose that the coefficient differences are distributed normally with mean zero (the null hypothesis) then for any equation the biggest coefficient difference, which is the measure relevant for this criterion, will not be distributed normally. A rough calculation was performed to find (on the null) the percentile corresponding to the biggest coefficient being more than one (or two) OLS standard deviation(s). These percentiles were estimated to be eighty-five and twenty-three respectively. This calculation is described below.

- We have ninety regressions and on average there are five parameters per regression.

- Assume that (the null hypothesis) the estimated coefficient differences are are distributed normally with mean zero.

- These assumptions imply for the average equation, the probability that at least exceeds one parameter exceeds two least squares' standard deviations is:

$$1 - (0.95)^5$$

  and the probability that at least exceeds one parameter exceeds one least squares' standard deviations is

$$1 - (0.68)^5$$

Criterion-IV *Percentage of robust estimates differing from least squares' by more than one standard deviation. Also more than two standard deviations.*

This builds a distribution of the extent to which slope estimates differ relative to a benchmark. The benchmark is the least squares' standard deviation.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above thirty-two for one standard deviation, and above five for two standard deviations.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an ten. This high grade reflects the importance attached, either rightly or wrongly, to the estimated variances in economic hypothesis testing.

Tests of Differences-Category Two

Criterion–V *Test of whether the intercept and slope are equal to zero and one in a regression of the least squares' slope estimates on the robust slope estimates.*

The rationale behind this is to see whether there is some systematic way in which ordinary least squares differs from robust regression. **Hogg** (1979) counsels researchers to perform least squares and robust methods. Further, if these provide different estimates they should both be reported. Causal empiricism indicates very few robust results are reported. This may be because there are no differences when using economic data. The issue of whether or not there are differences has not been studied before and thus we do it here. We make an adjustment for heteroskedasticity by dividing the observations by the square root of the estimated variance for the ordinary least squares' estimate.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above five percent.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve a six. This "grade" is given as the measurement error problem may not be overcome.

There is one other problem. The independent variable is measured with error in the regression of interest which is

$$OLS = \alpha + \beta E(ROBUST) + \epsilon_{ols}$$

where $E(ROBUST)$ is the "true" variable. But we only observe

$$ROBUST = E(ROBUST) + \epsilon_{robust}$$

making the obvious substitutions gives

$$OLS = \alpha + \beta ROBUST + \epsilon_{ols} - \beta\epsilon_{robust}$$

and the disturbance $\epsilon_{ols} - \beta\epsilon_{robust}$ is correlated with the the variable $ROBUST$ which violates an assumption of least squares. Let us examine this estimator in more detail

$$b = \frac{(1/T)\sum ROBUST \times OLS}{(1/T)\sum ROBUST^2}$$

seting $\alpha = 0$ for ease of exposition

$$plimb = \beta + \frac{plim(1/T)\sum ROBUST(\epsilon_{ols} - \beta\epsilon_{robust})}{plim(1/T)\sum ROBUST^2}$$

Assuming $Cov(\epsilon_{robust}, E(ROBUST))$ and $Cov(\epsilon_{ols}, E(ROBUST))$ are zero the bias term reduces to

$$\frac{-\beta\sigma^2_{\epsilon_{robust}} + Cov(\epsilon_{robust}, \epsilon_{ols})}{P + \sigma^2_{\epsilon_{robust}}}$$

The usual negative bias in b is offset to some degree if $Cov(\epsilon_{robust}, \epsilon_{ols})$ is positive. This is likely to be the case since both least squares and robust estimates are based on the same data . If least squares gives an overestimate of β, we expect the robust estimators to "make a mistake" in the same direction. When least squares is "out", it is likely the robust results are awry in the same direction. Even though the usual downward bias will not be completely removed, for want of a better surrogate, this measurement error is ignored.

Criterion–VI χ^2 *and Normal test of whether the sets of coefficients-least squares or robust-are different from one another.*

This criterion uses a χ^2 and non-parametric tests to determine if the sets of coefficients-least squares or robust-are different from one another. Assume:

1. Each coefficient estimate can be considered independent and as this may not be the case for more than one coefficient taken from the same regression, the test is also performed on a sample that includes one coefficient taken at random from each of the regressions;

2. Each coefficient estimate is normally distributed, at least asymptotically.

If the above conditions are met,

$$\sum \left(\frac{b^{ols} - b^{robust}}{se(b^{ols})} \right)^2 = testsix$$

could be seen as a χ^2 with the degrees of freedom equal to the number of coefficients.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above five percent.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an seven but the non-parametric tests are given a nine. The χ^2 test has a higher grade than the test in Criterion V as we select random coefficients here. Non-parametric tests get the highest grade as they require no distributional assumptions.

An adjustment

$$\sqrt{Varb^{ols} + Varb^{robust} - 2Covb^{ols}b^{robust}}$$

could be made to the denominator of "testsix" reflecting the stochastic nature of b^{robust} and capturing the possible covariance between the least squares and robust estimates.

However to obtain a "good" estimate of the covariance could prove to be a difficult task and thus this correction is one of the items to include in the future research section. For this reason this test must at this stage rank lower than the Vth criterion. One problem we face here is that the degrees of freedom are going to be large. Thus when we cannot interpolate because available tables have at most degrees of freedom, v, equal to 200 we use the approximation (**Murdoch and Barnes** (1976)) that the critical value for degrees of freedom (v) can be calculated from:

$$v\left[1 - \frac{2}{9v} \pm \frac{x}{\sigma}\sqrt{\frac{2}{9v}}\right]^3$$

and $\frac{x}{\sigma}$ is the standard normal "ordinate" associated with the area in the tail of the normal distribution such as, but not limited to, 1.96.

This χ^2 test is supplemented with two non-parametric tests of whether the least squares' sample is different from the robust sample of coefficients when considering both the size of the estimated coefficients divided by the the estimated standard errors of those coefficients. One way to bring these separate elements together is to use a non-parametric test to determine if the two samples come from populations that differ as to location. We can also test whether the samples come from populations that differ in variability. To borrow terminology we try to ascertain with these two tests whether there is any "treatment" effect both in location and variability from using robust methods rather than least squares. The first of these two non-parametric tests is the Wilcoxon Rank-Sum test. The least squares' coefficient estimates divided by their standard deviations are the control group and the robust coefficient estimates divided by their standard deviations become the "treatment" group. One coefficient and associated standard error is selected from each regression. This test is selected as the it is not the case that observations for least squares' and robust are independent of one another. Indeed they are *paired* with each other. The original research used least squares' and we have used robust to estimate the same coefficient and standard error. This test is also appropriate as it relies not only on whether least squares' or robust is larger, but also considers the size of the difference. Under the null hypothesis of no effect from robust estimation we would expect the direction (plus or minus) and size difference to be random over the sample. There should be as many pluses as minuses and large changes with small changes. Also the assignment of pluses and minuses to large and small changes should be random. We consider the alternative that there is a difference between the least squares' estimated coefficient divided by the standard deviation and the robust estimate of the coefficient divided by its standard deviation. One proceeds as follows:

1. Pick a coefficient at random from each regression and divide by its standard deviation, call these y_i;

2. Make a robust estimate of the same coefficient and divide each by its standard error, call these x_i;

3. Calculate T which is the sum of the differences of ranks with the least frequent sign. If there are less negative differences between y_i and x_i, then sum these ranks, otherwise sum the ranks associated with the positive differences.

4. As our sample size is greater than twenty then

$$z = \frac{(T - \mu)}{\sigma}$$

where

$$\mu = \frac{N(N+1)}{4}$$

$$\sigma = \sqrt{\frac{N(N+1)(2N+1)}{24}}$$

has a standard Normal distribution and N is the number of matched pairs of estimates.

The Wilcoxon Rank-Sum test tests for differences in *location.* Another question we could ask is whether the least squares' estimate is more dispersed around the "true" value than is a robust method? The **Siegel-Tukey** (1960) test answers this question. The test orders both samples, mixed together but keeping track of which is least squares and which is robust, from smallest to largest and then assigns ranks such that the smallest receives a one, two and three are given to the two largest, four and five to the second and third lowest, six and seven to the third and fourth highest and so on: an alternating method. Under the null the mean rank, assigned in this way. of least squares should be equal to the mean rank of the robust sample. If the alternative is correct, more observations from the more variable population, are going to be at the ends of the the ordered mixed samples and be given a lower rank with the alternating method.

1. Pick a coefficient at random from each regression and divide by its standard deviation, call these y_i;
2. Make a robust estimate of the same coefficient and divide each by its standard error, call these x_i;
3. Combine y_i and x_i in a single series and order them from highest to lowest;
4. Assign ranks according to the alternating method;
5. Calculate R_1 which is the sum of the ranks for the one of the estimators.
6. As our sample size is greater than twenty then

$$z = \frac{2R_1 - N(2N+1) \pm 1}{\sqrt{\frac{N(2N+1)}{\frac{N}{3}}}}$$

   has a standard Normal distribution and the $\pm$ in the numerator must be chosen so as to make z smaller.

Both these tests measure different aspects. The Wilcoxon Rank-Sum test concerns itself with location. The **Siegel-Tukey** focuses on variability. The observations we have are the estimated coefficients divided by estimated standard errors.

Variance Differences-Category Three

Criterion–VII *Percentage of the estimated slope coefficients for which the robust estimated variance is lower than the estimated least squares variance.*

Assuming a correct specification, least squares and robust estimators are both unbiased and thus there is a need to examine other features such as variances. Although a lower estimated variance does not necessarily correspond to a lower actual variance it should be of interest to compare estimated variances.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is greater than or equal to fifty percent.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an four. This criterion receives a low "grade" as we suspect the estimated variances may be understated.

Criterion–VIII *Average percentage change in the estimated variance for those instances out of the many estimated slope coefficients for which the estimated robust variance is smaller than the least squares variance* **versus** *the average percentage change in the estimated variance for those instances out of the many cases in which the ordinary least squares estimated variance is smaller. The median percentage change is also calculated.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above fifteen and above ten percent for the two possible outcomes (each of robust or least squares smaller) for each of the average and the median.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an four. This criterion receives a low "grade" as we suspect the estimated variances may be understated.

Forecast Differences-Category Four

Criterion–IX *Percentage of the forecasting equations for which forecasts using robust estimates beat forecasts using least squares' estimates on a mean absolute percentage error criterion. Also calculated for the median absolute percentage error.*

A similar criterion is employed by **Hallman and Kamstra** (1989) and **Fair** (1974). For each of the forecasting equations ten percent of the observations are removed from the end of time series data and randomly from cross-section data, estimation is undertaken using both least squares and a robust method, and the omitted observations are forecast.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above seventy and below thirty percent. If the percentage is in some sense an intermediate one one might be equivocal in recommending either technique. We have set the zone of indifference at forty thus robust is different for percentages above seventy and below thirty.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an eight. This "grade" is given as forecasting has an important role in econometrics and should be ranked higher than say the criterion concerned with the size of coefficients.

Average Mean Absolute Percentage Error Comparison

Criterion–X *Average improvement in the mean absolute percentage error for those instances out of the forecasting equations for which the robust forecast is superior* **versus** *the improvement in the average mean absolute percentage error for those instances out of the forecasting equations for which the ordinary least squares forecast is superior. This is also done for the improvement in the median absolute percentage error.*

The reason for this calculation is to determine in what type of environment does each estimator do well on forecasting considerations.

- In my subjective judgement, on this criterion the robust estimators are thought "different" from least squares to a "significant" degree if this improvement is higher for the robust estimators.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an eight. This "grade" is given as forecasting has an important role in econometrics and should be ranked higher than say the criterion concerned with the size of coefficients.

Hypothesis Testing Differences-Category Four

Criterion–XI *Percentage of the tested coefficients for which robust estimation changes the hypothesis that the author deemed to be of interest.*

For each of the many tested coefficients the hypothesis test is the one the author of the study saw as important. We are concerned with those instances where the robust regression changes the results of an important hypothesis test.

This is the most important criterion and the hurdle we have set is five percent. It is this low as hypothesis testing is so fundamental to empirical work and anyone using the results of empirical work would be worried if five percent of the results did not stand up to scrutiny. In addition to the above, four popular measures of aberrant observations are used, with each data set, to determine if there are outliers.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above five percent. It is this low as hypothesis testing is so fundamental to empirical work and anyone using the results of empirical work would be worried if five percent of the results did not stand up to scrutiny in much the same spirit in which a five percent level is chosen for a Type I error.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an ten, it is the most important.

In addition to the usual t and F tests, randomization tests of hypotheses, a testing methodology alleged to be robust, are also used to test the same hypotheses. As described in **Kennedy** (1993), a randomization test recalculates the original test statistic after shuffling the data. Repeating the process builds a "distribution" of test statistics allowing the researcher to decide whether the original test statistic is very different from the "distribution" of test statistics.

In the regression context we randomly shuffle the relevant independent variable, redo the test and calculate the percentage of times the t values from many shuffles exceeds the original t value; this estimates the test's P-Value. Armed with this P-Value it is possible to make a comparison with the original P-Value from least squares and determine if the hypothesis test outcome is changed. An alternative way of conducting the randomization test is also employed. The generic procedure remains the same however one uses the residuals from a regression of the dependent variable on the data bar the independent variable being subject to the hypothesis test, known as residualized "y". The variable being used in the hypothesis test is residualized as well. **Kennedy** (1993) indicates that it is not clear which method is superior.

Criterion–XII *Percentage of times the P-value gets bigger using robust regression.*

The P-Value is the probability under the null hypothesis of obtaining a test statistic value greater than or equal to the value actually obtained. As pointed out by **Goldberger** (1991 pp 238–240) among many others, since it is the readers of research who must make a decision on the outcome of the hypothesis test, the test statistic's P-value should be reported rather than simply declaring coefficients that are statistically significant at an arbitrary significance level.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is less than or equal to ten, and equal to, or above ninety percent. For each coefficient there is a P-Value produced by least squares and also by the robust estimators. If many of the P-values are smaller with least squares then many will be larger with the robust estimator. Based on this we have to decide if robust is "different" from least squares. So, if only thirty percent of the P-values are larger with a robust estimator this means seventy percent are smaller with least squares and it is the "seventy" which gains attention. The problem is how do you set these bounds? This is not an easy question to answer but on a subjective basis they have been set at less than or equal to ten, or greater than, or equal to ninety.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve seven as the results have been covered in the tests of hypotheses.

Criterion–XIII *Average percentage change in the P-value when it gets larger. Also the average percentage point change. Also calculated are the medians of these changes. Further, all of this is calculated for the P-value becoming smaller.*

- In my subjective judgement, on this criterion the robust estimators need to are thought "different" from least squares to a "significant" degree if this percentage is above eighty percent for the means and above sixty percent for the medians.
- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an seven.

P-values are reported as decimal fractions of one, for example, 0.049 and this indicates a marginal significance level. Because of the link to significance levels we also show changes in these decimal fractions of the P-value and refer to these changes as percentage point changes.

Criterion–XIV *Percentage of times the P-value changes by less than five percent or five percentage points. Also ten and twenty percent or ten and twenty percentage points.*

- In my subjective judgement, on this criterion the robust estimators are thought "different" from least squares to a "significant" degree if this percentage is above five percentage points for the means, and above two percentage points for the medians.
- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an seven.

3.1.3 Sample Sizes

In the following chapter regression estimates, both robust and least squares are made of many slope coefficients from ninety data sets typical in economics. The criteria of comparison, in this chapter, make use of the estimated coefficients in addition to their standard deviations. It must be pointed out the *sample size* applicable to the empirical appraisal of the data sets on the basis of these criteria differs from criterion to criterion. For instance, not all of the slope coefficients were subjected to conventional hypothesis tests as the original research may have included many variables in a regression but only chose to test a subset of them. This subset sometimes only comprises one coefficient. Further some of the robust estimation methods cannot be employed with small sample sizes, so the number of "observations" for the robust estimators varies from technique to technique for the same reason. The number of estimated coefficients varies from a high of two hundred and twenty-nine for the Least Absolute Errors Estimator to a low of forty-nine for the Five-Percent Trimmed Estimator. Table 3.1 presents the number of coefficients. Where a description of a criterion refers to a coefficient it is the numbers in the table that are used. As the L-Estimators use linear programming methods to obtain estimates there is the possibility some solutions are going to be non-unique. But to do hypothesis tests one needs unique solutions. As the important focus of this study is hypothesis tests, the regressions, where there are no unique solutions, are not used. Where this effect is most evident is for the Five Quantile estimator. Also the trimmed estimators do not function well if the number of observations is small. So the trimmed estimators also lost data sets. Another reason is the trimmed estimators rely on initial quantile estimates to identify which observations to discard. The particular quantiles depend on the trimming proportion. If the trimming proportion is low with a small number of observations the necessary quantile cannot be calculated. The number of estimated regressions corresponding to the number of coefficients of Table 3.1 is given in Table 3.2.

Criteria IX and X which use forecasts, refer to forecasting equations and Table 3.3 gives the number of equations for which forecasting is possible with each estimator.

Estimator	Sample Size-Coefficients
Least Absolute Errors	229
Five Quantile	57
Gastwirth	222
Tukey L-Estimator	221
Ten Percent Trimmed	136
Five Percent Trimmed	49
Huber M-Estimator	214
Tukey M-Estimator	210

Table 3.1: Sample Size-Coefficients

Estimator	Sample Size-Regressions
Least Absolute Errors	73
Five Quantile	26
Gastwirth	71
Tukey L-Estimator	70
Ten Percent Trimmed	47
Five Percent Trimmed	21
Huber M-Estimator	71
Tukey M-Estimator	70

Table 3.2: Sample Size-Regressions

Estimator	Equations Forecast
Least Absolute Errors	64
Ten Percent Trimmed	31
Five Percent Trimmed	11
Huber M-Estimator	66
Tukey M-Estimator	65

Table 3.3: Equations Forecast

Chapter 4

Results

4.1 Introduction

This chapter presents the results for our data and the criteria outlined in the previous chapter. Recall the previous chapter outlined fourteen ways to compare robust estimates with estimates from least squares. This chapter gives the outcomes of those criteria for comparison. The results are provided in tables and care has been taken to ensure each table is explicitly linked to a method. Commentary is provided for the tables. The aim is to see if all the criteria point towards a particular conclusion on the basis of five categories. It is as if we have a "null hypothesis" of robust estimators being no different from least squares and this chapter determines how that "null" is altered as we consider the five categories which include measures that allow one to see if robust estimates are different. We are interested to see if robust methods would have resulted in a widespread reversal of the original results or caused those original researchers to say "the robust methods are trying to tell us something", prompting further investigation. Based on the studies in Chapter One our prior belief is that robust estimators make no difference. It appears, however, that on the basis of some of the criteria examined here that robust estimators do indeed make some difference. For the variance criterion the robust estimators made so much of a difference that the variance results are treated with caution. This caution led to a further examination of the robust variances in the next chapter.

4.2 The Data

To make the comparisons, described in the previous chapter, a collection of data sets is required. Also, one has to be fairly sure the **data sets** are those the the original researcher used and that the results of the researchers do not differ markedly from those we obtained upon replication. We began with about eighty studies which were reduced ultimately to forty-four published **studies** in economics which could be used here. Exact replication is not always possible in economics, as **Dewald, Thursby and Anderson** (1986) have found. The process of selecting data sets, in this thesis, uses a weaker definition of replication, in anticipation of not being able to replicate exactly each study. For twenty-six studies exact replication was possible.

For eleven studies although exact replication was not possible, the coefficients were sufficiently similar that we judged them to be usable for our purposes. In two of these the coefficient estimates are matched to two decimal places. To give the reader a sense of what subjective judgements were made here, the worst nine are reported in Table 4.1. The first four of these have the biggest differences. In these studies we have ignored differences in the intercept (on the grounds that interest is seldom shown in such estimates) and have subjectively judged the replication to be close enough for the studies to be employed in our sample. In all cases the test statistics used to test the hypothesis of interest was little affected, and and continued to produce the same test result. These examples show what judgements were made. What is important here is our goal: we want to be confident that we are working with the same data set that the original authors used. Few would argue that, for the nine really problematic data sets, we are working with a completely different data set than the one originally used. In all cases here this is the data set that the authors said they used. It is interesting that most of the nine come from the *Journal of Money, Credit and Banking.* For some of those data sets the observations were captured electronically by hand from typewritten pages provided by the journal. The pages of data had been provided to the journal by the original authors. Other data sets were transferred from microcomputer floppy diskettes. We are confident that we are working with the original data.

A data set is also eligible for inclusion if the original inference is replicated by

Rounding		
Study	Original	Replicated
Boyer and Adams	0.0015	0.00144
	0.502	0.421
Allen	-14.01	-13.619
	1.38	1.3286
	-0.268	-0.3084
	-0.071	-0.0798
	0.302	0.353
	0.050	0.054
Edwards	-0.610	-1.4168
	0.360	0.357
	0.430	0.433
	-0.171	-0.180
	0.324	0.326
Kim	-0.699	-0.707
	0.952	0.877
	0.057	0.057
	-0.557	-0.566
Rassekh and Wilbratte	-0.0002	-0.0002
	0.08	0.0849
	-0.05	-0.0511
	0.02	0.0255
	0.7	0.673
	-0.13	-0.14
Gerlach	0.946	0.955
	-0.000136	-0.000136
Ladenson and Bombara	3.36	3.12
	0.04	0.039
	0.91	0.93
	-0.88	-0.89
	0.53	0.54
	-0.54	-0.53
	-0.34	-0.339
Nerlove	0.721	0.7206
Walsh	-0.79	-0.798
	0.33	0.38
	-0.38	-0.378
	-0.09	-0.078

Table 4.1: Rounding

rounding the test statistic to two decimal places (but not so as to reverse the original conclusion). There are two studies in this category. **Berndt** (1991) includes exercises which use economic data sets, and describes replications as "close", and in some instances gives the least squares results for a data set or updates an older data set. These seven data sets are included. This is the process by which **studies** are selected. The final number of studies (total **forty-four**) in each category is given in Table 4.2 and note most of the studies fall into the exact replication category. The extent of exact replication is much better than obtained in other replication work, for example, **Dewald, Thursby and Anderson** who obtained a success rate of one point three percent as pointed out by **Mirowski and Sklivas** (1991). Our success rate, on their terms for least squares is just over fifty percent but in fairness about thirty data sets, in addition to the four described above, were obtained for which replication proved difficult and thus were discarded. But we feel the above process of selection is a fair one and representative of economic data. Given the problem with replication in economics we have done well to produce a data set of data sets.

In Chapter Two we saw how changing assumptions about the distributional form of the errors can have an impact on the regression, robust or otherwise, we might use. With this in mind, the residuals from the ninety regressions are tested for normality using a χ^2 goodness of fit test, testing whether the actual distribution of the residuals in groups follows a normal distribution. The test has been applied to residuals by **Klein** (1974). Of the ninety regressions, the null that the errors follow a normal distribution is rejected for 45.6 percent of the regressions. This provides some comfort that the data is not overly representative of only normal residuals or of only non-normal residuals. In fact, based on the percent failing the normality test the odds are skewed somewhat in favour of least squares.

The data, for this thesis, making it through the above selection process comes from three sources. The book by **Berndt** serves as a source for thirteen of the **studies**. The *Journal of Money, Credit and Banking* requires authors of articles to submit the data on which their results are based. The journal makes these data sets available to others for a nominal fee. Most data sets from the 1980's were purchased. The process of replication, described above, produced fourteen **studies** for which the data could

STUDIES	
Type of Replication	Number
Exact	26
Replicated Inference Two Decimal Places	2
With Rounding Nearest Integer	9
Berndt	7

Table 4.2: Types of Replication

be used. Finally the econometrics book by **Lott and Ray** (1992) provides seventeen usable **studies**. Overall this gave forty-four **studies** and as some studies estimate more than one equation the result: ninety data sets. The computer package used is SHAZAM (**White** (1978)) primarily because it is the only econometrics package providing an easy to use command for robust estimation that allows one to properly estimate (since Version 7.0) the standard errors of coefficients. Given SHAZAM's robust estimates can be obtained at low cost we feel other researchers are likely to have opted for the same easy to use robust command, and thus we also use the command and its many options. During the replication procedure where exact replication proved difficult the student version of TSP (**Hall and Lilien** (1990)) was used to ensure different results were not just as a result of using one econometrics package. Both packages gave the same result for the same data when using least squares. To obtain M-Estimators, the programme 3R in BMDP (**Dixon** (1990)) is used.

4.3 Results for the thirteen criteria

Coefficient Estimate Differences-Category One

4.4 Criterion I

Criterion–I *Percentage of all robust estimates differing from least squares by*

Percentage Changes More Than Stated			
Criterion–I	Percentage changes more than thirty	Sample Size	Percentage changes more than fifty
Least Absolute Errors	48.04	229	34.93
Five Quantile	19.3	57	14.04
Gastwirth	39.64	222	27.48
Tukey L-Estimator	41.17	221	27.6
Ten Percent Trimmed	27.21	136	15.44
Five Percent Trimmed	32.65	49	18.37
Huber M-Estimator	22.43	214	12.62
Tukey M-Estimator	32.38	210	25.71

Table 4.3: Percentage Changes of Robust Slope Coefficients Changing By More than Stated

more than thirty percent. Also by more than fifty percent.

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above fifteen for changes more than thirty and above ten for changes more than fifty.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve a seven. Usually in economics less attention is given the size of a coefficient and more weight put on the estimated standard errors. If one objects to this, the size has a high "grade" relative to other criteria.

In Table 4.3 (the item in the third column is the sample size) is reported the percentage of robust estimated coefficients differing from least squares by more than thirty and fifty percent. Given the benchmarks of **Coursey and Nyquist** (1988), and the target percentages, all the robust estimates are different. Thus we reject the crudely constructed "null" that robust estimators are the same as least squares.

4.5 Criterion II

Criterion–II *Percentage of the regressions in which at least one estimated slope coefficient changes by more than thirty percent. Also, changes by more than fifty percent are noted.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above fifteen for changes more than thirty and above ten for changes more than fifty.
- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve a seven. Usually in economics less attention is given the size of a coefficient and more weight put on the estimated standard errors. If one objects to this, the size has a high "grade" relative to other criteria.

In Table 4.4 are the results from the second criterion where we are concerned with the percentage of the regressions in which at least one estimated slope coefficient changes by more than thirty percent when estimated using robust methods. Overall, what is of note is all estimators have more than fifteen percent of the regressions in which at least one estimated slope coefficient changes by more than thirty percent. In Table 4.4 we consider the percentage of the regressions in which at least one estimated slope coefficient changes by more than fifty percent using robust methods. All estimators exceed the cut-off of ten percent posited in the previous chapter as a critical percentage for the robust estimators to be considered having produced different results.

4.6 Criterion III

Criterion–III *Percentage of the regressions in which at least one estimated slope coefficient changes by more than one ordinary least squares' standard deviation. Also changes by more than two standard deviations are calculated.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree this percentage is more

Average Percentage Change of Slope Coefficients			
Criterion–II	Percentage changes more than thirty	Sample Size	Percentage changes more than fifty
Least Absolute Errors	58.3	73	51.4
Five Quantile	34.6	26	15.4
Gastwirth	50	71	41.4
Tukey L-Estimator	52.2	70	40.6
Ten Percent Trimmed	45.7	47	37
Five Percent Trimmed	38.1	21	38.1
Huber M-Estimator	38.6	71	24.3
Tukey M-Estimator	47.8	70	36.2

Table 4.4: Percentage Changes More Than Thirty and Fifty

than eighty-five for one standard deviation and more than twenty-three for two standard deviations.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an ten. This high grade reflects the importance attached, either rightly or wrongly, to the estimated variances in economic hypothesis testing.

No estimators (Table 4.5) exceed the crucial cut-off for one standard deviation of eight-five percent and here we would conclude robust estimators are not making a marginal difference on the basis of this standard. Changes more than two standard deviations are calculated in Table 4.5 and the percentages presented for each estimator. The reason for choosing two standard deviations has some basis in hypothesis testing. The largest is the twenty-three percent for the Tukey M-Estimator. The order changes rather markedly from the one standard deviation case, suggesting the Tukey M-Estimator has some large percentage changes as it placed fourth when considering changes more than one standard deviation. For this criterion all estimators have percentages below the target, with the exception of the Tukey M-Estimator. Robust estimators are not different at two standard deviations.

Regressions			
Criterion–III	Changes more than one standard deviation	Sample Size	Changes more than two standard deviations
Least Absolute Errors	55.6	73	18.1
Five Quantile	23.1	26	15.4
Gastwirth	35.7	71	8.57
Tukey L-Estimator	27.5	70	5.8
Ten Percent Trimmed	34.8	47	8.7
Five Percent Trimmed	33.3	21	9.5
Huber M-Estimator	18.6	71	8.57
Tukey M-Estimator	33.3	70	23.2

Table 4.5: Changes More Than OLS Standard Deviation

4.7 Criterion IV

Criterion–IV *Percentage of robust estimates differing from least squares' by more than one standard deviation. Also more than two standard deviations.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree this percentage is above thirty-two for one standard deviation, and above five for two standard deviations.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an ten. This high grade reflects the importance attached, either rightly or wrongly, to the estimated variances in economic hypothesis testing.

Table 4.6 shows the percentage changes of the robust estimates differing from least squares by more than one and two least squares' standard deviation. Only two exceed the cutoff for one least squares' standard deviation and thus we would accept the "null" that robust and least squares are the same on this basis. For two least squares'

Coefficients			
Criterion–IV	Changes more than one standard deviation	Sample Size	Changes more than two standard deviations
Least Absolute Errors	32.62	229	8.30
Five Quantile	14.04	57	7.02
Gastwirth	19.36	222	3.15
Tukey L-Estimator	18.1	221	1.81
Ten Percent Trimmed	23.52	136	5.15
Five Percent Trimmed	38.61	49	4.08
Huber M-Estimator	7.94	214	2.34
Tukey M-Estimator	22.86	210	10.0

Table 4.6: Changes more than OLS standard deviations

standard deviations, only four exceed the cutoff and here we might also want to accept our "null".

Summary-Category One

At this stage what can be made of the results? Looking at coefficients we see some changes that suggest robust estimators are making a difference, especially for Criterion II. For changes relative to one least squares' standard deviation the results are not in the robust estimators' favour and when we consider changes relative to two standard deviations we also see the robust estimators are not different. The direction of the previous result is also confirmed when looking at the total number of coefficients. Further these are important criterion as has been indicated, although more importance is attached to the standard deviation criteria. On the latter basis, we might say we are beginning, speaking loosely, to lean towards the "null" that robust estimators do not make a difference. These results, especially those related to standard deviations, are in agreement with **Swinton and King** (1991).

For the criteria in this category each estimator is examined to see if there is anything unusual about any of them. This is ultimately done for all the categories but the process is outlined here.

1. For each table in a category the estimators were ranked according to whether they produced a small or large change for the criterion in question.

2. The two robust estimators associated with the extreme rankings were noted.

3. In deciding which estimators were unusual or extreme, the two estimators with low sample sizes, namely the Five-Quantile and Five Percent Trimmed , are not considered.

4. This process is then repeated for the next criterion or where possible the next table as some criteria generated more than a single table of information.

For this category it is the case that two estimators stand out. The first is the Huber M-Estimator which has the feature that it usually has the lowest percentage for the differences of the criteria. Least Absolute Errors is distinctive as it usually has the largest percentage. If one were to recommend which robust estimators to use, these would thus be obvious candidates. The reason two are suggested is both seem to behave differently. Thus if you were to use only one, you would never know what the other estimator could tell you. On the basis of this and other categories these two estimators are on separate sides of a spectrum so to speak. Given their different behavior there is some reluctance to endorse only one as one aspect of robust estimation requires looking for differences and attempting to explain them away or deal with them. Using only one of two apparently different estimators may mean possible differences from the other are not considered. And it is quite possible these missed differences could provide additional insights.

4.8 Criterion V

Tests of Differences-Category Two

Criterion–V *Test of whether the intercept and slope are equal to zero and one in a regression of the least squares' slope estimates on the robust slope estimates.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above five percent.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve a six. This "grade" is given as the measurement error problem may not be overcome.

If estimates are not too far apart then a regression of the least squares' estimates on the robust estimates should pass through the origin and have slope of unity. On the basis of the regression test it is the case that the coefficients are not in agreement as Table 4.7 shows. The intercept and slope estimates are shown in Table 4.7. The P-Values, sometimes called marginal significance values, are all below the five percent cutoff and we should reject the null. To obtain some confirmation of these results (for the slope estimates are "close" to unity) the slope standard deviations are provided. For example, for the Least Absolute Errors case we reject the null that the slope is unity and if one adds twice the slope standard deviation (0.0156) to the slope estimate one obtains 0.97926 which is below unity. An analysis of the constituent parts of the estimated variance of the slope estimate for the test, reveals the observations are matched over the range minus ten to plus ten. This means the elements of $(X'X)^{-1}$ are very small: this is the source of the low estimated variances of the slope estimates of the test.

These results were robust in the sense that instead of using least squares (not to be confused with the least squares' observations) to perform the test of this criterion, the regression of the test is estimated using the Least Absolute Errors estimator. With these Least Absolute Errors' results the P-Values of the test are recalculated. For all but the Gastwirth Estimator are the results, based on these new P-Values, unchanged. Looking at the observations (least squares' and Gastwirth estimates of many coefficients) for this estimator a number of standardized residuals were found to be outliers. For the Gastwirth the P-Value becomes 0.1043 suggesting this estimator is no different from least squares. For the seven other estimators this robust analysis does not change the original test results. On doing this robust analysis it became

Intercept and Slope Test					
Criterion–V	Intercept	Slope	P-Value	Slope SD	Sample Size
Least Absolute Errors	0.0000008	0.96366	0.0003	0.0078	229
Five Quantile	0.0004	0.97472	0.0021	0.0068	57
Gastwirth	0.0000014	0.97472	0.00093	0.0067	222
Tukey L-Estimator	0.0000008	0.97904	0.00012	0.0048	221
Ten Percent Trimmed	0.000007	0.97833	0.0005	0.0054	136
Five Percent Trimmed	0.000068	0.97965	0.02792	0.0073	49
Huber M-Estimator	0.000016	0.97877	0.00000	0.004	214
Tukey M-Estimator	0.00001	0.9629	0.0001	0.007	210

Table 4.7: Standard Deviation

clear that using the **Hogg** suggestion of doing your analysis with least squares and a robust estimator is more difficult than first appears. Assume you have solved the problem of what it means for a robust estimator to make a difference. Here we used changes in P-Values and compared 0.0067 and 0.1043 to five percent to say the robust estimators produced a different result. The difficulty is what to do now. We identified some outliers but there is no apparent reason why these are outliers. But what do you do? There is no reason to discard them. It is probably this uncertainty that has kept robust methods from taking hold in economics.

For the tests of this criterion all variables were adjusted by dividing by the standard error of the least squares estimate as this is a known form of heteroskedasticity. It is also the case that the robust estimate is measured with error but as if any least squares' estimate is an overestimate say, it is believed the robust estimate will be in error in the same direction. This should mitigate against the usual downward bias in the estimated coefficients of the regression for this criterion. Despite this optimism, the slope estimates all lie below one and it may be the measurement error effect is stronger than anticipated in the previous chapter. This may be what causes the rejection of our "null". Because of this we might want to place less emphasis on this particular test and thus the "grade" given is a six. The same test is performed,as already mentioned, using Least Absolute Errors and for the Tukey L-Estimator the

Results of χ^2 test				
Estimator	testsix	DOF	Critical	Null
Least Absolute Errors	468.6	229	272.8	R
Five Quantile	79.734	57	79.1	R
Gastwirth	317.4	222	265.2	R
Tukey L-Estimator	169.2	221	264.1	A
Ten Percent Trimmed	116.3	136	170.2	A
Five Percent Trimmed	65.5	49	66.3	A
Huber M-Estimator	101.5	214	236.4	A
Tukey M-Estimator	397.3	210	252.0	R

Table 4.8: χ^2 Test Results

P-Value becomes 0.1043 suggesting this estimator is no different from least squares. For the seven other estimators this robust analysis does not change the original test results.

4.9 Criterion VI

Criterion–VI χ^2 *and non-parametric tests of whether the sets of coefficients-least squares or robust-are the same.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above five percent.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an seven but nine for the non-parametric tests. The χ^2 test has a higher grade than the test in Criterion V as we select random coefficients here. Non-parametric tests get the highest grade as they require no distributional assumptions.

Results of χ^2 test: random coef				
Estimator	testsix	DOF	Critical	Null
Least Absolute Errors	78.3	73	93.9	A
Five Quantile	8.5	26	38.9	A
Gastwirth	49.7	71	91.7	A
Tukey L-Estimator	35.5	70	90.5	A
Ten Percent Trimmed	20.6	47	63.0	A
Five Percent Trimmed	8.2	21	32.7	A
Huber M-Estimator	21.02	71	91.7	A
Tukey M-Estimator	72.3	70	90.5	A

Table 4.9: χ^2 Test Results: random coef

The "testsix" as described in Chapter Three for each estimator is shown in Table 4.8 and with the critical values also shown in Table 4.8 we have to reject the null hypothesis for four of the robust estimators. The problem here is the degrees of freedom (v) are large and above even comprehensive "critical" value tables. For these large degrees of freedom, to obtain critical values, one can use the approximation f rom (**Murdoch and Barnes** (1976)) which gives the critical values for degrees of freedom (v):

$$v\left[1 - \frac{2}{9v} \pm \frac{x}{\sigma}\sqrt{\frac{2}{9v}}\right]^3$$

where $\frac{x}{\sigma}$ is the standard normal "ordinate" associated with the area in the tail of the normal distribution such as, but not limited to, 1.96.

The test is repeated but selecting one estimated coefficient at random from each regression and this is reported in Table 4.9. The results are the reverse four of the "Reject" results in Table 4.8 and given the random selection of coefficients one might be more inclined to accept them. Based on the latter random test it is the case that in all the cases we accept the null and thus this test is in not in agreement with the F-test. But in order to be sure two non-parametric tests are also performed.

To complement Category Two, two non-parametric tests are used. The first is the Wilcoxon Rank-Sum test which is available in BMDP (**Dixon** (1990)). The least squares' coefficient estimates divided by their standard deviations are the control group and the robust coefficient estimates divided by their standard deviations become the "treatment" group. One coefficient and associated standard error is selected at random from each regression. For all but the Ten Percent Trimmed Estimator, it is the case that the null hypothesis is rejected (Table 4.10) and the robust estimates of the coefficients divided by standard deviations are larger than the least squares' estimates divided by their standard deviations. The Ten Percent Trimmed observations are larger than least squares, but not significantly larger. The **Siegel-Tukey** (1960) tests the *null* hypothesis with the alternative that the populations have different variances about the median.

The results of the **Siegel-Tukey** non-parametric test show (Table 4.11) the Huber M-Estimator, Tukey M-Estimator (there is no relation between the two other than they are named after Tukey to credit his having pointed them out), Least Absolute Errors and the Ten Percent Trimmed have the same *variability* as least squares. It must be pointed out that this test will not correctly reject as often as it should if the locations of the two samples are not the same, or in other words, the power is reduced. This problem manifests itself for the Least Absolute Errors Estimator. The Wilcoxon test shows the location to be very different from least squares but the P-Value for the second test is 0.238 yet there is a large difference in the standard deviations.

It appears that the populations from which these samples are drawn have very different medians (of estimates divided by their standard deviations) although the variability about those medians is the same, for half of the estimators, with the exceptions noted above. The results of the Wilcoxon Rank-Sum test suggest that robust estimators make a difference for Category Two. One problem here is the estimated variances associated with the robust estimators are underestimated and this could be causing us to conclude the medians (of estimates divided by their standard deviations) are different.

Summary-Category Two

This is the third most important category as indicated by the evaluation given to the tests here when compared with other criterion. The F-test indicates that the results

Wilcoxon	
Estimator	P-Value
Least Absolute errors	0.0022
Five Quantile	0.0001
Gastwirth	0.0000
Tukey L-Estimator	0.0000
Ten Percent Trimmed	0.6264
Five Percent Trimmed	0.0033
Huber M-Estimator	0.0009
Tukey M-Estimator	0.0003

Table 4.10: Wilcoxon Signed Ranks Test

Siegel-Tukey	
Estimator	P-Value
Least Absolute errors	0.238
Five Quantile	0.075
Gastwirth	0.0016
Tukey L-Estimator	0.0004
Ten Percent Trimmed	0.58
Five Percent Trimmed	0.0033
Huber M-Estimator	0.3628
Tukey M-Estimator	0.327

Table 4.11: Siegel-Tukey Alternating Ranks Test

of robust estimation are different from those of least squares but we did point to two possible reasons for this: the persistent measurement error problem and the actual scatter of observations reduces a component of the estimated variance calculation. The χ^2 test shows the coefficients to be the same when the sample includes one coefficient from each regression. Again it is not clear whether we are "rejecting' the "null" that robust estimators are the same as least squares. The Wilcoxon test suggests the robust results are different but this result may depend on the lower estimated variances associated with the robust estimators.

For this Category the estimators are examined to see if they display any unusual features. The Tukey M-Estimator, Least Absolute Errors and the Ten Percent Trimmed estimator stand out here as they reject the null more than the others.

4.10 Criterion VII and VIII

Variance Differences-Category Three

Criterion–VII *Percentage of the estimated slope coefficients for which the robust estimated variance is lower than the estimated least squares variance.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is greater than or equal to fifty percent.
- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an four. This criterion receives a low "grade" as we suspect the estimated variances may be understated.

Criterion–VIII *Average percentage change in the estimated variance for those instances out of the many estimated slope coefficients for which the estimated robust variance is smaller than the least squares variance* **versus** *the average percentage change in the estimated variance for those instances out of the many cases in which the ordinary least squares estimated variance is smaller. The median percentage change is also calculated.*

Percentage lower estimated variance		
Criterion–VII	Percentage	Sample Size
Least Absolute Errors	80	229
Five Quantile	100	57
Gastwirth	98	222
Tukey L-Estimator	99.5	221
Ten Percent Trimmed	87	136
Five Percent Trimmed	86	49
Huber M-Estimator	95	214
Tukey M-Estimator	98	210

Table 4.12: Percentage of robust estimators with a lower estimated variance

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above fifteen and above ten percent for the two possible outcomes (each of robust or least squares smaller) for each of the average and the median.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an four. This criterion receives a low "grade" as we suspect the estimated variances may be understated.

Having dealt with the size of estimates and the size compared to the original standard deviation, Table 4.12 considers the estimated variance. Caution must be exercised in interpreting the tables dealing with variances as the sample size is small. For instance, least squares never has a lower estimated variance than that of the Five Quantile Estimator. The percentage of cases in which the robust estimator has a lower estimated variance is very high. Again the Least Absolute Errors estimator stands out not because it has the highest percentage but because it has the lowest percentage with a lower estimated variance. These lower estimated variances are almost too good to be true. One suspects the estimates of the robust variances are not very good. All the robust estimators exceed the subjective level. In the next chapter, we turn to examine a possible source of the lower variances which Chapter

Average Percentage Change when robust estimated variance lower		
Criterion–VIII	Robust lower estimated variance	Median
Least Absolute Errors	63(229)	64
Five Quantile	97(57)	99
Gastwirth	88(222)	85
Tukey L-Estimator	89(221)	90
Ten Percent Trimmed	28(136)	24
Five Percent Trimmed	28(49)	21
Huber M-Estimator	35(214)	34
Tukey M-Estimator	43(210)	42

Table 4.13: Average Percentage Change when robust estimated variance lower

Average Percentage Change when least squares' estimated variance lower		
Criterion–VIII	Least squares' lower estimated variance	Median
Least Absolute Errors	98(45)	45
Five Quantile	-(0)	-
Gastwirth	43404(4)	1577
Tukey L-Estimator	934(1)	934
Ten Percent Trimmed	208(18)	8
Five Percent Trimmed	0.49(7)	0.18
Huber M-Estimator	643(11)	13
Tukey M-Estimator	2882(4)	2240

Table 4.14: Average Percentage Change when least square's estimated variance lower

Two indicated could lie in the method used to estimate the standard deviations of the estimate. At least, this is so for the estimators from the econometrics package SHAZAM (**White** (1978)) .

As the ultimate concern is with hypothesis tests one would want to know whether the lower robust variance is a lot lower than those instances when the least squares' variance is lower. These results are reported in Tables 4.13 and 4.14. The problem here is so few of the cases have a lower least squares' variance. Thus the sample sizes are small (just one for the Tukey L-Estimator) and inferences about the size of the smaller least squares' variance is just not possible. Further percentage changes are calculated to agree with percentage changes in other criteria but as estimated variances can be very small, the percentage changes can become very large. This effect is certainly evident in Table 4.14. We would do well to ignore this table. Overall, the robust estimators have more cases with lower estimated variances and using the more reliable results of Table 4.13 all estimators exceed the subjective level of fifteen percent. A suspicion is the programme SHAZAM (**White** (1978)) is producing this effect. The fall in the variance is larger for the four quantile estimators (the first four of the table) compared to the BMDP (**Dixon** (1990)) estimators: the last two estimators in the table. Notice the percentage change for the Trimmed Estimators is the lowest. SHAZAM (**White** (1978)) calculates these correctly. We correct for the problem of underetimated variances in this chapter by considering randomization tests and in the next chapter explore the source of these lower estimated variances with Monte Carlo work.

Summary-Category Three

This has been judged to be the least important category relative to the others as we suspect the estimated variances are underestimated. Thus although most of the robust variances are lower, this we think is an irregularity. This possible underestimation is investigated in more detail in the next chapter. It is not possible based on variances to say whether we are in position to make another decision on our crudely set up "null" that least squares and robust estimators are the same. As there does seem to be a problem with standard deviations we do not employ the non-parametric statistics for this category. We are to check the accuracy of the estimated variances in the next chapter. The Tukey L-Estimator and the Huber stand out in this Category. The former as it has the largest percentage changes and the latter as it, like its performance in Category One, has the lowest percentage changes. A pattern that begins to manifest itself after three categories is the Huber M-Estimator is different from the rest and tend to be "conservative" in terms of percentage changes for the criteria.

4.11 Criterion IX

Forecast Differences-Category Four

Criterion–IX *Percentage of the forecasting equations for which forecasts using robust estimates beat forecasts using least squares' estimates on a mean absolute percentage error criterion. Also calculated for the median absolute percentage error.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above seventy and below thirty percent. If the percentage is in some sense an intermediate one one might be equivocal in recommending either technique. We have set the zone of indifference at forty thus robust is different for percentages above seventy and below thirty.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an eight. This "grade" is given as forecasting has an important role in econometrics and should be ranked higher than say the criterion concerned with the size of coefficients.

The estimators were also compared on their ability to forecast known values, out of the the sample used to estimate the equations, in Tables 4.15 and 4.16. Ten percent of the observations for each data set were forecast but the equation for forecasting is based on the remaining ninety percent of the observations. For only five of the robust estimators was the forecast exercise possible, because to forecast known values one has to reduce the sample and with fewer observations the methods used for robust estimates do not have enough observations to calculate all the quantiles. Using the mean absolute percentage error criterion for about fifty percent of the forecasts were the robust estimators able to outperform least squares. All five estimators fell in the "equivocal range" and thus on this basis we conclude the robust estimators are not different. This is also the case when we consider the median absolute percentage error instead of the mean absolute percentage error in Table 4.16. These lie within

Percentage when robust forecast beats OLS			
Criterion–IX	Percentage when robust forecast beats OLS mean	Sample Size	Percentage when robust forecast beats OLS median
Least Absolute Errors	41	63	46
Ten Percent Trimmed	55	31	45
Five Percent Trimmed	48	11	46
Huber M-Estimator	46	66	39
Tukey M-Estimator	43	65	39

Table 4.15: Percentage when robust forecast beats least squares (me(di)an percentage error criterion)

Average improvement in MAPE				
Criterion–X	Mean improvment in MAPE if robust forecast beats OLS	Sample Size	Mean improvment in MAPE if OLS forecast beats robust	Sample Size
Least Absolute Errors	12	26	24	37
Ten Percent Trimmed	11	15	24	16
Five Percent Trimmed	23	5	30	6
Huber M-Estimator	13	30	18	36
Tukey M-Estimator	9	28	54	37

Table 4.16: Average improvement in MAPE when robust (OLS) forecast beats OLS (robust)

Average improvement in MedianAPE				
Criterion–X	Improvement in MedAPE if robust forecast beats OLS	Sample Size	Improvement in MedAPE if OLS forecast beats robust	Sample Size
Least Absolute Errors	6	28	15	33
Ten Percent Trimmed	10	13	7	18
Five Percent Trimmed	8	5	4	6
Huber M-Estimator	25	24	62	42
Tukey M-Estimator	30	24	52	31

Table 4.17: Average improvement in MedianAPE when robust (OLS) forecast beats OLS (robust)

the seventy to thirty percent range which we decided meant the robust estimates are not different.

4.12 Criterion X

Criterion–X *Average improvement in the mean absolute percentage error for those instances out of the forecasting equations for which the robust forecast is superior* **versus** *the improvement in the average mean absolute percentage error for those instances out of the forecasting equations for which the ordinary least squares forecast is superior. This is also done for the improvement in the median absolute percentage error.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this improvement is higher for the robust estimators.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an eight and should be ranked higher than say the criterion concerned with the size of coefficients. This "grade" is given as forecasting has an important role in econometrics.

In Table 4.16 notice no robust estimators exceed the subjective requirement that they beat least squares. Looking at Tables 4.17, only the trimmed estimators beat least squares. On the basis of this criterion and the subjective requirement, the robust estimators are not different from least squares. Robust does beat least squares in a number of cases but the improvement is not enough to be judged different from least squares. **Fair** (1974) notes that for forecasting purposes the choice of an estimator, robust or otherwise is not that important. We reach a similar conclusion here, as least squares seems to do better in terms of a lower forecast errors (as defined) but not in enough cases to be judged different from least squares. For the Median Absolute Percentage Error section of Criterion X it is the case the Trimmed Least Squares Estimators do better than least squares. This result is in line with **Hallman and Kamstra** (1989) and **Stigler** (1977).

Summary-Category Four

The robust estimators did not manage to outperform least squares on the basis of forecasting. This is a somewhat disappointing given the use of forecasts by others to evaluate robust estimators. The pattern from comparing estimators, for this category, is one in which Least Absolute Errors joins the Huber (as usual) in the "conservative" effects group. The Ten Percent Trimmed Estimator exhibits the largest changes in this the Forecasting Category.

4.13 Criterion XI

Hypothesis Testing Differences-Category Five

Criterion-XI *Percentage of the tested coefficients for which robust estimation changes the hypothesis that the author deemed to be of interest.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is above five percent. It is this low as hypothesis testing is so fundamental to empirical work and anyone using the results of empirical work would be worried if five percent of the results did not stand up to scrutiny in much the same spirit in which a five percent level is chosen for a Type I error.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an ten, it is one of the most important.

In our view the most important criterion is the extent to which a robust estimate changed a test of interest to the original researcher. As not all coefficients are of interest, the number of coefficients is smaller than the number used in other criterion of comparison. Given that just under one half of the **studies** report significant or mostly significant results and the lower estimated variances (with t tests done in the usual manner) will result in higher t values, some of these t values may cross critical values. If this is the case tested coefficients may become significant. We know from our discussion of Criterion VII that the robust estimators provided lower variances. An examination of the Least Absolute Errors Estimator's results showed that the estimated slope coefficients increased (the decreases are consistent with **Lichtenberg and Siegel** (1991)) in forty-nine percent of the cases. As a consequence of these two facts we expect the percentage of significant results becoming insignificant to be relatively small and the percentage of insignificant results becoming significant to be relatively large. The right hand columns of Table 4.18 show the the percentage of hypothesis tests that changed from being significant results to being insignificant. Care is taken to perform the original hypothesis test, be it a t or F test. Five **studies** used the F-Test. The percentage of hypothesis tests originally insignificant becoming significant is also calculated and reported in the left hand column of Table 4.18. As expected the percentage becoming insignificant is smaller than the percentage becoming significant for all estimators. Only three estimators did not exceed the five percent cutoff and on this basis we would say robust estimators are making some difference. The high percentages becoming significant are as a result of the *lower* estimated variances of the robust estimators. Also some of these are based on very small samples: insignificant results are not reported as often as significant results. In Table 4.19 we sum all changes and calculate these as a percentage of the sum

Percentage of Coefficients Becoming S(Ins)ignificant				
Criterion–XI	Percentage Changing From Insignificant To Significant	Sample Size	Percentage Changing From Significant To Insignificant	Sample Size
Least Absolute Errors	48	50	10	133
Five Quantile	60	5	15	48
Gastwirth	67	42	3	134
Tukey L-Estimator	72	39	2	135
Ten Percent Trimmed	42	33	8	78
Five Percent Trimmed	73	11	11	28
Huber M-Estimator	19	43	4	113
Tukey M-Estimator	33	40	11	118
Randomization	21	57	43	145
Randomization (Residual)	14	57	48	145

Table 4.18: Percentage Becoming S(Ins)ignificant

of reported insignificant and significant coefficients. These percentages indicate we have a different result on the criterion with the robust estimators. Some caution must be exercised as we know the percentage becoming significant is affected by the estimated variances of the robust estimators being too low which produces some spurious significant results. We reëxamine these percentages in the next chapter but the percentages of Table 4.19 are sufficiently high in relation to the five percent requirement for one to conclude the robust estimators are different.

Randomization methods are also used to see if results become significant or insignificant. The two methods are are generally consistent, reflecting **Kennedy's** (1993) results. Comparing this result with the same percentage for each of the robust methods it is interesting the randomization techniques are the ones that reverse significant results to a much greater extent than the robust methods. The opposite is true for the changing of insignificant results to significant results. If anything one can conclude from this economic data should be subjected to closer scrutiny with randomization tests in conjunction with robust methods, as more results became insignificant with the randomization methods than is the case with the robust methods. These

Percentage Sig/Insig	
Criterion–XI	Percentage
Least Absolute Errors	20
Five Quantile	33
Gastwirth	17
Tukey L-Estimator	18
Ten Percent Trimmed	18
Five Percent Trimmed	29
Huber M-Estimator	8
Tukey M-Estimator	16
Randomization	37
Randomization (Residual)	38

Table 4.19: Percentage Becoming Significant and Insignificant

reversals arise from the concern, already mentioned, that the estimated standard errors of the robust estimators are too low. This has the effect of making significant results more significant or not allowing insignificant results to remain so and this is one cause of the difference between the randomization and robust methods. In the next chapter, this issue is explored in more detail using Monte Carlo methods.

Randomization and Estimated Variances

As we are already suspicious of the lower estimated variances, this anomaly in the changes from significant to insignificant (especially in the light of the randomization results) indicates more work needs to be done on the estimated variances. This we do in the next chapter where the various estimators of the standard deviation are examined using Monte Carlo methods. Randomization tests have the advantage over the robust methods in that the calculation of the variance covariance matrix is avoided. Given the power of modern computers the randomization method may be a better alternative to the estimates of standard errors provided by the computer packages SHAZAM (**White** (1978)) and BMDP (**Dixon** (1990)).

Criterion–XII	Percentage With Bigger P-Value
Least Absolute Errors	18.7(229)
Five Quantile	1.75(57)
Gastwirth	4.5(222)
Tukey L-Estimator	4.5(221)
Ten Percent Trimmed	16.1(137)
Five Percent Trimmed	20.4(49)
Huber M-Estimator	17.8(214)
Tukey M-Estimator	17.6(210)

Table 4.20: Percentage With Bigger P-Value

4.14 Criterion XII

Criterion–XII *Percentage of times the P-value gets bigger using robust regression.*

- In my subjective judgement, on this criterion the robust estimators are judged "different" from least squares to a "significant" degree if this percentage is less than, or equal to, ten and equal to, or above, ninety percent. For each coefficient there is a P-Value produced by least squares and also by the robust estimators. If many of the P-values are smaller with least squares then many will be larger with the robust estimator. Based on this we have to decide if robust is "different" from least squares. So, if only thirty percent of the P-values are larger with a robust estimator this means seventy percent are smaller with least squares and it is the "seventy" which gains attention. The problem is how do you set these bounds? This is not an easy question to answer but on a subjective basis they have been set at ten and ninety.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve seven as the results have been covered in the tests of hypotheses.

Finally the changes in the P-Value are examined and shown in Table 4.20. On the basis of the subjective levels only three estimators fall below the ten percent (or ninety

Average Percentage Change When Bigger P-Value		
Criterion–XIII	Average: Change When Bigger P-Value	Median
Least Absolute Errors	559(43)	192
Five Quantile	118(1)	118
Gastwirth	165(10)	76
Tukey L-Estimator	116(10)	59
Ten Percent Trimmed	2064(22)	99
Five Percent Trimmed	557(10)	243
Huber M-Estimator	154(38)	29
Tukey M-Estimator	1367(37)	107

Table 4.21: Average and Median Percentage Change When Bigger P-Value

for least squares) subjective level. For these three the actual number of increases is small as Table 4.21 indicates. Thus we may want to pay less attention to these three estimators. If they are ignored, on the basis of this criterion, the robust estimators are not "different" from least squares. The reason for this is by the time we analyse this criterion we have already established the estimated variances of the robust estimators are problematic.

4.15 Criterion XIII

Criterion–XIII *Average percentage change in the P-value when it gets larger. Also the average percentage point change. Also calculated are the medians of these changes. Further, all of this is calculated for the P-value becoming smaller.*

- In my subjective judgement, on this criterion the robust estimators are thought "different" from least squares to a "significant" degree if this percentage is above eighty percent for the means and above sixty percent for the medians.

- In my subjective judgement, on a scale of one to ten, in my subjective judgement, this criterion is important enough to deserve a seven.

The average and median percentage change is calculated when the P-Value is bigger. The changes are based on small sample sizes and not too much attention should be

Average Percentage Change When Smaller P-Value		
Criterion–XIII	Average: Change When Smaller P-Value	Median
Least Absolute Errors	83(128)	100
Five Quantile	99(26)	100
Gastwirth	96(162)	100
Tukey L-Estimator	94(161)	100
Ten Percent Trimmed	69(67)	72
Five Percent Trimmed	69(19)	67
Huber M-Estimator	62(120)	68
Tukey M-Estimator	65(119)	71

Table 4.22: Average and Median Percentage Change When Smaller P-Value

Average Percentage Point Change When Bigger P-Value		
Criterion–XIII	Average: Point Change When Bigger P-Value	Median
Least Absolute Errors	0.08(43)	0.05
Five Quantile	0.06(1)	0.06
Gastwirth	0.11(10)	0.12
Tukey L-Estimator	0.09(10)	0.02
Ten Percent Trimmed	0.08(22)	0.03
Five Percent Trimmed	0.09(10)	0.03
Huber M-Estimator	0.06(38)	0.01
Tukey M-Estimator	0.12(37)	0.07

Table 4.23: Average and Median Percentage Point Change When Bigger P-Value

given to Table 4.21, although all exceed the subjective level with the exception of the Tukey L-Estimator and the Huber at the median. Also comparisons with Tables 4.20 do not mean much due to the sample sizes involved. The results are consistent with many of the coefficients having lower estimated standard errors as already seen in criteria above and thus we do not want to inadvertently reject the crude "null" for it is the lower estimated variances producing this result.

4.16 Criterion XIV

Criterion–XIV *Percentage of times the P-value changes by less than five percent or five percentage points. Also ten and twenty percent or ten and twenty*

Average Percentage Point Change When Smaller P-Value		
Criterion–XIII	Average: Point Change When Smaller P-Value	Median
Least Absolute Errors	0.08(128)	0.02
Five Quantile	0.06(26)	0.01
Gastwirth	0.08(162)	0.03
Tukey L-Estimator	0.08(161)	0.02
Ten Percent Trimmed	0.06(67)	0.02
Five Percent Trimmed	0.05(19)	0.01
Huber M-Estimator	0.04(120)	0.01
Tukey M-Estimator	0.05(119)	0.01

Table 4.24: Average and Median Percentage Point Change When Smaller P-Value

percentage points.

- In my subjective judgement, on this criterion the robust estimators are thought "different" from least squares to a "significant" degree if this percentage is above five percentage points for the means and above two percentage points for the medians.

- In my subjective judgement, on a scale of one to ten, this criterion is important enough to deserve an seven.

As a P-Value contains information at the percentage point level it is not enough to look at the percentage change. Thus the average and median percentage point changes are calculated when the P-Value is bigger in Table 4.23. For the Least Absolute Errors Estimator the average percentage point change is eight. Against a benchmark of five for a P-Value, this appears to be large. We know that many coefficients did not become insignificant thus it must be the case that some of the initial P-Values must have been well below the benchmark. A problem with Table 4.23 is the small sample sizes but on the basis of the subjective levels of five and two percentage points the crude "null" is rejected (excluding the Huber at the median) and robust estimators are different. The average percentage point change for a smaller P-Value in Table 4.24 shows in seven cases the change is smaller than when the P-Value increases. As many more coefficients became significant this suggests the P-Values of these coefficients

Percentage Points Bigger P-Value Changes Less Than Stated			
Criterion–XIV	Five Point	Ten Point	Twenty Point
Least Absolute Errors	58	63	67
Five Quantile	0	100	100
Gastwirth	20	40	100
Tukey L-Estimator	60	60	90
Ten Percent Trimmed	68	77	86
Five Percent Trimmed	50	80	80
Huber M-Estimator	71	89	97
Tukey M-Estimator	43	59	78

Table 4.25: Percentage Bigger P-Value

could not have been too far above the cutoffs used to determine the significance of the original coefficients. In terms of the subjective levels the Five-Percent Trimmed is not different at the mean and only four estimators are different at the median. So again we can conclude that most of robust estimators are different on this criterion.

For P-Values it makes sense to consider percentage point changes as these contain the information of interest to those interested in the marginal significance level. Thus our interest is focussed on the percentage point changes and specifically we work out, for both increases and decreases in the P-Value, changes less and more than five, ten and twenty percentage points. At the twenty percentage point level we note in Table 4.25 for the smaller P-Value four estimators have percentages larger than those for a bigger P-Value (Table 4.26) at twenty percent. This reflects the results of Table 4.21 where the percentage changes for bigger P-Values are *higher*. Although on a percentage point basis (Tables 4.23 and 4.24) the difference in percentage changes is muted as with percentage points while most estimators have larger percentages for bigger P-Values, the differences are not as pronounced as the comparison of Table 4.21 and 4.22. As we already know the percentage changes are large for changes in the P-Values, the division into five, ten and twenty percent is provided (See Tables 4.27 and 4.28) but the number of coefficients in these three cases is small.

Percentage Points Smaller P-Value Changes Less Than Stated			
Criterion–XIV	Five Point	Ten Point	Twenty Point
Least Absolute Errors	66	73	82
Five Quantile	61	73	81
Gastwirth	65	71	86
Tukey L-Estimator	66	75	86
Ten Percent Trimmed	70	82	88
Five Percent Trimmed	73	84	95
Huber M-Estimator	80	89	93
Tukey M-Estimator	78	87	93

Table 4.26: Percentage Smaller P-Value

Percentage Bigger P-Value Changes Less Than Stated			
Criterion–XIV	Five Percent	Ten Percent	Twenty Percent
Least Absolute Errors	0	2	5
Five Quantile	0	0	0
Gastwirth	0	0	10
Tukey L-Estimator	10	10	10
Ten Percent Trimmed	0	0	20
Five Percent Trimmed	0	0	10
Huber M-Estimator	13	21	32
Tukey M-Estimator	5	11	22

Table 4.27: Percentage Bigger P-Value

Percentage Smaller P-Value Changes Less Than Stated			
Criterion–XIV	Five Percent	Ten Percent	Twenty Percent
Least Absolute Errors	1	3	5
Five Quantile	1	36	58
Gastwirth	0	0.6	1
Tukey L-Estimator	0	0	1
Ten Percent Trimmed	2	2	7
Five Percent Trimmed	0	0	0
Huber M-Estimator	1	8	13
Tukey M-Estimator	4	8	16

Table 4.28: Percentage Smaller P-Value

Summary-Category Five

For the most important criterion of direct testing of hypotheses we find that robust estimators do make some difference and we can reject our "null" that they do not. Partly due to the reservation about the estimated standard errors and partly from a concern to verify our results we also relied on randomization tests to confirm two findings. Robust estimators make a difference with hypothesis tests, and that, indeed there is a problem with the standard error estimates of the robust estimators. An analysis of the P-Values confirms that robust estimators are different but there are problems with estimated standard errors. As to which estimator is distinctive no clear pattern emerges except the already well established "conservative" nature of the Huber Estimator is in evidence.

4.17 Outliers in the Studies' Data Sets

For those studies and data sets that had a change in the hypothesis test results occurring when using robust estimators, it is of interest to ask if standard diagnostics testing should have led the original researcher to question using least squares. This is done by checking the data sets for outliers using four indicators of outliers used and recommended by others. The four measures are:

1. The measure h_t or the leverage, which is large for observations far from the mean of all the observations, and used by **Judge** *et al* (1988);

2. The studentized residual which is a residual that is calculated to avoid the propensity of least squares to underestimate the residual, and used by **Judge** *et al* (1988);

3. The contribution of any observation to the least squares estimates of the coefficient vector, captured by the measure DFBETAS, recommended by **Chaterjee and Hadi** (1986);

4. Finally we consider the contribution of any observation to the predictions of the model as given by DFFITS, also recommended by **Chaterjee and Hadi** (1986). This measure is affected by large residuals and leverage points.

For each of these measures there are recommended "critical values" and if any of the measures exceed the critical value, it indicates there may be problems in the regressors or the residuals. Exceeding a critical value does not mean the observation is useless; it becomes something to be explained. For instance, is there something unusual say about April, leap years or individuals with low incomes or high incomes?

If any data set gave a different result for Least Absolute Errors, the Tukey L-estimator or either M-Estimator the above four measures. These four estimators were chosen as they were likely to have been known to the original researchers had they wished to use them. The DFBETAS measure is calculated for each observation, for each coefficient and the intercept. Thus the two (split into two to fit into the text) tables (they are the two tables, Tables 4.29 and 4.30, with the captions indicating parts one and two) of the results reports the number of measures that exceeded the "critical" value and the number of observations except for DFBETAS where we have the number of observations times the number of coefficients plus the constant. The effect of this can be seen in the two tables where in the column DFBETAS, are some large numbers in the denominator. In the other columns the denominator is the number of observations. Each numerator is the number of measures that exceeded the critical value. It is remarkable that out of the one hundred and forty-four cells in the both tables, only five cases have a zero numerator. So if any of the original researchers cry foul and say "robust methods were not commonly used when the study was done", we have done it for them. The overall conclusion on the basis of these measures seems to be the original researchers would have had cause to take a closer look at their data as was hoped in Chapter One. Indeed, some of them probably did so but at least we have given those that chose not to use robust methods (and we cannot know how many) the benefit of checking for problems in the data.

4.17.1 The Extent of Outliers

We want to know if the studies, for which we have obtained changes in the hypothesis tests, have more outliers than the studies that did not have hypothesis tests change. For this we need an "index" of the extent of outliers calculated for a division of all data sets into two groups: first, hypothesis test changes and second, no change in hypothesis tests. This "index" is constructed from the DFBETAS measures in Tables 4.29 and 4.30 as this measure captures the effect of outliers on coefficients and further is closely related to the first two measures. Also **Swinton and King** (1991) found it a useful method for finding outliers.

The construction of the "index" proceeds as follows:

1. The data sets were divided into the two groups, mentioned above;
2. For each data set the critical value ($\frac{2}{\sqrt{T}}$) of DFBETAS is calculated;
3. The actual DFBETAS are calculated for each data set. There is a DFBETA for each estimated parameter;
4. The critical value is subtracted from the actual absolute value of DFBETA. The positive values from this procedure (the number of positive values is the number of outliers) are averaged over all parameters for a data set;
5. The averaging procedure is repeated for all the data sets;
6. At this stage we have the average of the extent to which the identified outliers are different from the critical value, divided into the two groups;

7. Finally the means (the "index") of these averages are calculated and the data sets which had hypothesis tests change have a mean of 0.4 and the data sets which did not have hypothesis tests change had a mean of 0.35.

8. This "index" calculates the extent of "outlierness" but does not take into account the size of the data set. If a large data set produces a small number of outliers with large "outlierness" we would want an index to increase with the "outlierness" but also decrease if the number of outliers is small relative to the total possible outliers, a size of the data set effect. In other words the "index" should fall with fewer possible outliers but rise if, of the fewer possible, there are, on average, really "large" outliers. The "index" we have constructed thusfar does not capture this effect.

9. To improve the "index" each measure of "outlierness" from a data set is weighted by the inverse of the sample size for a data set to capture the size of the data set effect. When this is done, the weighted means corresponding to those above are 0.012 and 0.018. The sample size weighted "index" is higher for the data sets which did not have hypothesis tests change. The opposite being true for the unweighted index above.

These results do provide some support for thinking the identification of outliers can help identify those observations that exert "influence" on the outcomes of regression results, at least on the basis of the first "index". The results do not support, nor were they designed to, the notion that such outliers be accommodated or deleted. Further the variances of the indices as we have constructed them indicate these particular indices are unable to discriminate between the two groups of data sets: hypothesis test changes and no hypothesis changes.

4.17.2 A LOGIT Analysis of Outliers

A logit analysis of the effect of outliers is possible. Divide the data sets into two groups, those that had a hypothesis test change and those that did not, creating a *qualitative response* variable which we call **HC**. This is done for the Least Absolute

Outliers: Original Data				
Study	e_t^*	h_t	DFBETAS	DFFITS
Barkley	2/45	3/45	21/135	2/45
Christenson	2/26	3/26	8/130	1/26
Christenson	2/26	3/26	18/130	4/26
Kirkpatrick/Yamin	1/22	2/22	10/66	3/22
Kirkpatrick/Yamin	1/22	4/22	15/66	5/22
Kirkpatrick/Yamin	2/22	3/22	6/66	3/22
Kirkpatrick/Yamin	1/22	3/22	4/66	2/22
Kirkpatrick/Yamin	1/22	3/22	12/66	4/22
Kirkpatrick/Yamin	2/22	3/22	15/66	5/22
Brada/Graves	2/25	3/25	9/100	2/22
Brada/Graves	1/25	0/25	8/100	2/25
Walsh	2/16	0/16	4/48	1/16
Lewis	2/18	2/18	1/64	1/18
Lewis	1/23	1/23	5/69	1/23
Coghlan	2/33	1/33	10/165	1/33
Owen	2/30	1/30	10/90	5/30
Owen	2/30	3/30	10/120	2/30
Ball	0/34	1/34	8/102	3/34
Bodvarsson	1/42	6/42	14/168	4/42

Table 4.29: Outliers in Data Sets: Part One

Outliers: Original Data				
Study	e_t^*	h_t	DFBETAS	DFFITS
Chiang	8/115	13/115	15/230	10/115
Chiang	6/115	7/115	17/130	6/115
Kim	1/40	5/40	8/160	1/40
Laderson/Bombara	4/36	3/36	22/252	0/36
Laderson/Bombara	2/36	2/36	13/216	4/36
Laderson/Bombara	2/36	2/36	19/216	3/36
Laderson/Bombara	3/36	2/36	11/180	4/36
Lahiri it et al	4/65	8/65	37/455	5/65
Scut/Van Bergeijk	2/32	1/32	18/192	3/32
Lucas/Rapping	3/36	3/36	21/288	5/36
Merrick	6/123	5/123	108/1722	10/123
Mroz	12/428	23/428	163/2996	17/428
Wheelock	5/109	10/109	38/654	12/109
Nerlove/Waugh	3/45	3/45	20/225	4/45
Rassekh/Wibratte	3/60	4/60	29/360	7/60
Saurman	2/52	1/52	7/104	3/52
Saurman	4/52	0/52	6/104	3/52
Saurman	4/52	1/52	7/104	4/52
Saurman	1/52	2/52	8/104	4/52

Table 4.30: Outliers in Data Sets: Part Two

Errors estimator using an estimate of the standard errors suggested in the literature and not those produced by the **SHAZAM** in-house method. Possible explanatory variables to include are:

- The number of estimated coefficients (least squares and Least Absolute Errors) differing by more more than two least squares' standard deviations, **M2S**;
- The number of estimated coefficients (least squares and Least Absolute Errors) differing by more more than one least squares' standard deviation;
- The sample size for each regression;
- The number of explanatory variables;
- The sample size less the number of estimated parameters, **TMK**;
- The ratio of the number of estimated parameters to the sample size;
- The number of outliers chosen on a DFBETAS basis, **NO**;
- The percentage of outliers relative to the sample size and number of estimated parameters;
- The average difference between an outlier and the "critical value", a measure of "outlierness", **EO**;
- An indicator capturing whether the residuals from a least squares' regression tested as normally distributed or not, **N**;

The variables included in the logit estimation have the **boldface** variable names. This subset is chosen as multicollinearity problems are likely to be avoided and also this specification covers:

- One of the fourteen criteria-**M2S**-we expect this to have a positive effect on **HC**;
- The sample size less the estimated parameters-**TMK**-we expect this to have a negative effect on **HC**;

Logit Results			
Variable Name	Estimated Coefficient	Standard Error	T-Ratio
M2S	3.4387	1.26	$2.74^{0.05}$
TMK	-0.04	0.02	$2.51^{0.05}$
NO	0.06	0.03	$2.09^{0.05}$
EO	-2.28	1.56	$1.45^{0.1}$
N	1.04	0.63	$1.65^{0.1}$
0.05-five percent significance level			
0.1-ten percent significance level			

Table 4.31: Results from the Logit Analysis

- The number and extent of outliers-**NO** and **EO**-we expect this to have a positive effect on **HC**;

- Whether the least squares' residuals come from a normal distribution-**N**-we expect this to have a negative effect on **HC** as the variable takes a value one if the residuals are normal and zero if they are not normal.

The results are presented in Table 4.31. Two variables **EO** and **N** have the incorrect sign but are only significant at the ten percent level. The other three variables are all significant at the five percent level. Of interest here, as we are concerned with outliers, is the variable associated with the number of outliers (**NO**) is positively and significantly related to **HC** the variable capturing whether a data set had a hypothesis test change. If the logit analysis is repeated but for hypothesis tests originally *significant* but becoming *insignificant* the *only* change to the results in Table 4.31 is the variable **EO**, or the extent of outliers, is significant at the five percent level, and **NO** is still positively and significantly related to **HC**.

4.18 Conclusions

1. The percentage of the regressions in which at least one estimate slope coefficient changes by more than one least squares' standard deviation is not above the

subjective level for all of the estimators. Refer to Table 4.5. One is getting a signal that is easy to interpret. The robust estimators do not make a difference. This result is confirmed at two standard deviations.

2. Coefficients are changing a great deal with the Robust Estimators. Looking at Table 4.4 all the the estimators exceeded the subjective cut-off of fifteen percent for percentage changes more than thirty. Also, all exceeded the cut-off of ten for percentage changes more than fifty. These are stunning results at least for these criteria. On the basis of percentage changes alone all estimators exceeded the expected cut-offs. There does seem to be a mixed message from the first category: the first two criteria show a difference but the more important second two criteria do not. Due to the importance attached to standard deviations, we are of the opinion robust estimators are not different from least squares.

3. The Vth and VIth criterion and to some extent the XIth are other attempts to determine if the robust analysis give different results from least squares. Considering Table 4.7 it appears the robust results differ from that of ordinary least squares. As the econometrics literature has suggested, at least since 1978, such differences be investigated we find it difficult to believe this investigation was undertaken and not reported on publication. Care is taken to ensure that nearly all the studies made no mention of robust results. Only one did correct for outliers. In every study that gave different results on the hypothesis test criterion we found evidence of problematic data points using widely accepted and available measures compared with the number of outliers in the studies that did not have results change. Another test (based on randomly selected coefficients) or the VIth criterion produced the result that least squares and robust estimates are not different for the robust estimators. There may be problems with the F-test and thus more importance is attached to the second test, suggesting the conclusion robust estimates are not different from least squares.

4. Estimated variances are much lower with the Robust estimators. These results were completely unexpected and not believable. This is especially so as the randomization methods indicated results becoming insignificant, see Table 4.18, rather than significant which the lower estimated variances produced for the Xth criterion. Also looking at Table 4.13 one notices the estimated variances change erratically from one estimator to another. Something is not right here and although one might be led to conclude robust estimators are different on the basis of estimated variances, such a conclusion would be premature. In fact coupled with Chapter Two where it is suggested the method of calculating the standard errors may produce aberrant results, a Monte Carlo study is needed to compare methods of estimating the standard errors. This is done in the next chapter.

5. Robust Estimators show some promise when pitted against least squares in a forecasting race, but they hardly overwhelm least squares. The thesis offers a seventy and thirty percent split as the percentage of regressions that the robust methods must beat least squares on a mean absolute percentage error criterion and thus be regarded as different. None of them do. Most are above forty percent, though. The forecasting criterion seems to show some promise, although one cannot say the robust estimators are very different on this basis. Least squares does have the largest improvement in the forecast error but not in a enough cases that it would be judged different from the robust estimators on the basis of our cutoff level.

6. Robust estimators do result in significant results becoming insignificant. If one looks at Table 4.18, more than five percent of the coefficients, that were originally significant, became insignificant. Also the highest percentage becoming significant is eleven percent, excluding the Five Quantile and Five Percent Trimmed estimator due to small sample sizes. There may be some doubt about the hypothesis testing results, especially given the concern expressed about the lower estimated variances. Also the randomization tests produced the opposite result to the robust estimators for this criterion as Table 4.18 shows forty-three percent

of studies had coefficients become insignificant and twenty-one percent of the studies had coefficients become significant for one method of randomization.

7. Overall, what emerges from this chapter is that the criteria for determining if robust estimators make a difference, present a set of results from which it is possible to distill the conclusion that robust results do give different coefficient estimates in terms of size and there are enough changes in hypothesis tests to be concerned, despite our initial supposition that they would not make a difference. But the criteria related to estimated variances in conjunction with other criteria, especially the randomization tests pointed towards the estimated variances used here as being suspect and possibly underestimated. It is this problem the next chapter addresses and redoes the hypothesis testing criterion-XI-with an additional two methods of calculating the standard errors.

8. Also going through all of the tables for all of the methods and determining which of the robust estimators are different from the others points to two, based on larger sample sizes, namely the Huber M-Estimator and Least Absolute errors, that we felt stood out from the others. While it is difficult to establish the importance of this finding, there is some duty on our part to recommend one or more for general use. We do this solely on the basis of which stood out or could be grouped with others as being different from the rest. For this reason the recommendation must be treated with some caution: this is not a Monte Carlo Study. Based on the performance in the categories here, we would suggest using the Least Absolute Errors and Huber Estimator. If there is still an objection to the estimated standard errors associated with these estimators, use one of the Randomization methods instead.

4.19 Studies

- W. Andrews and C. Christenson. Some economic factors affecting safety in underground bituminous coal mines. *Southern Economic Journal*, 42(3):364–376, 1974.

- Stuart Allen. Klein's price variability terms in the US demand for money. *Journal of Money, Credit and Banking*, 14(2):524–530, 1982.
- Russel Boyer and Charles Adams. Forward premia and risk premia in a simple model of exchange rate determination. *Journal of Money, Credit and Banking*, 20(4):632–644, 1988.
- R. Ball. Some econometric analysis of the long term rate of interest in the United Kingdom, 1921-1961. *Manchester School Of Economics and Social Studies*, 33(1):45–96, 1965.
- Andrew Barkley. The determinants of the migration of labor out of agriculture in the United States, 1940-1985. *American Journal of Agricultural Economics*, 72(3):567–573, 1990.
- J. Brada and R. Graves. The slowdown in Soviet defense expenditures. *Southern Economic Journal*, 54(4):969–984, 1988.
- Orn Bodvarsson. Monitoring with no moral hazard: The case of small vessel commercial fishing. *Eastern Economic Journal*, 13(4):421–434, 1987.
- Jason Benderly and Burton Zwick. Inflation, real balances, output, and real stock returns. *American Economic Review*, 75(5):1115–1123, 1985.
- Rosanne Cole *et al.* Quality-adjusted price indexes for computer processors and selected peripheral equipment. *Survey of Current Business*, 66(1):41–50, 1986.
- Laurits Christensen and William Greene. Economies of scale in US electric power generation. *Journal of Political Economy*, 84(4):655–676, 1976.
- Timothy Cook and Thomas Hahn. The information content of discount rate announcements and their effect on market interest rates. *Journal of Money, Credit and Banking*, 20(2):167–180, 1988.

- Thomas Chiang. The forward rate as a predictor of the future spot rate–a stochastic coefficient approach. *Journal of Money, Credit and Banking*, 20(2):213–231, 1988.

- Gregory Chow. Technological change and the demand for computers. *American Economic Review*, 57(5):1117–1130, 1967.

- R. Coghlan. Bank competition and bank size. *Manchester School Of Economics and Social Studies*, 43(2):173–179, 1975.

- L. A. Dick-Mireaux. The interrelationship between cost and price changes, 1946-1959. *Oxford Economic Papers (NS)*, pages 267–292, 1961.

- Sebastian Edwards. Money, the rate of devaluation, and interest rates in a semi-semiopenopen economy: Columbia, 1968-82. *Journal of Money, Credit and Banking*, 17(1):59–68, 1985.

- Alvin Egbert. An aggregate model of agriculture-empirical estimate and some policy implications. *American Journal of Agricultural Economics*, 51(1):71–86, 1969.

- Paul Farris. Export supply and demand for US cattle hides. *American Journal of Agricultural Economics*, 53(4):643–646, 1971.

- Stefan Gerlach. World business cycles under fixed and flexible exchange rates. *Journal of Money, Credit and Banking*, 20(4):621–632, 1988.

- Edward Gramlich. Models of inflation expectations formation. *Journal of Money, Credit and Banking*, 15(2):155–173, 1983.

- Richard Guldin. Predicting costs of eastern national forest wilderness. *Journal of Leisure Research*, 13(2):112–128, 1981.

- David Hakes. The objectives and priorities of monetary policy under different Federal Reserve chairmen. *Journal of Money, Credit and Banking*, 22(3):327–337, 1990.

- Hendrik Houthakker. Some calculations of electricity consumption in Great Britain. *Journal of the Royal Statistical Society*, 114:351–371, 1951.

- Lawrence Klein and Arthur Goldberger. *An Econometric Model of the United States, 1929-1952.* North-Holland, 1955.

- Inchul Kim. Exchange market pressure in korea: An application of the Girton-Roper monetary model: A note. *Journal of Money, Credit and Banking*, 17(2):258–263, 1985.

- Colin Kirkpatrick and M Yamin. The determinants of export subsidiary formation by US transnationals in developing countries: An inter-industry analysis. *World Development*, 9(4):373–382, 1981.

- Mark Ladenson and Kenneth Bombara. Entry in commercial banking. *Journal of Money, Credit and Banking*, 16(2):165–174, 1984.

- Stephen Lewis. Government revenue from foreign trade. *Manchester School of Economics and Social Studies*, 31:39–47, 1963.

- Robert Lucas and Leonard Rapping. Real wages, employment, and inflation. *Journal of Political Economy*, 77(5):721–754, 1969.

- Kajal Lahiri, Christie Teigland, and Mark Zaporowski. Interest rates and the subjective probability distribution of inflation forecasts. *Journal of Money, Credit and Banking*, 20(2):233–248, 1988.

- John Merrick. Financial market efficiency, the decomposition of "anticipated" versus "unanticipated money growth, and further tests of the relation between money and real output: A note. *Journal of Money, Credit and Banking*, 15(2):222–232, 1983.

- Thomas Mroz. The sensitivity of an empirical model of married woman's hours's of work to economic and statistical assumptions. *Econometrica*, 55(4):765–799, 1987.

- Marc Nerlove. Returns to scale in electricity supply. In Carl Christ, editor, *Measurement in Economics*, pages 167–198. Stanford University Press, California, first edition, 1963.

- Marc Nerlove and Frederick Waugh. Advertising without supply control: Some implications of a study of the advertising of oranges. *Journal of Farm Economics*, 43(4):813–837, 1961.

- John Owen. The demand for leisure. *Journal of Political Economy*, 79(1):56–76, 1971.

- Kristian Palda. *The Measurement of Cumulative Advertising Effects.* Prentice Hall, 1964.

- Allen Paul. The pricing of binspace-a contribution to the theory of storage. *American Journal of Agricultural Economics*, 52(7):1–12, 1970.

- Rati Ram. Economic development and income inequality: Further evidence on the U-curve hypothesis. *World Development*, 16(11):1371–1376, 1988.

- Farhad Rassekh and Barry Wilbratte. The effect of import price changes on domestic inflation: An empirical test of the ratchet effect-a note. *Journal of Money, Credit and Banking*, 22(2):263–267, 1990.

- Lynne Schneider *et al.* Government regulation of cigarette health information. *Journal of Law and Economics*, 24(3):575–612, 1981.

- David Saurman. Transactions costs, foreign exchange demands, and the expected rates of change of exchange rates. *Journal of Money, Credit and Banking*, 14(1):20–32, 1982.

- Frederick Schut and Peter Van Bergeijk. International price discrimination: The pharmaceutical industry. *World Development*, 14(9):1141–1150, 1986.

- Richard Schmalensee. *The Economics of Advertising.* North-Holland, 1972.

- Edmund Sheehey. The neutrality of money in the short run: Some tests: A note. *Journal of Money, Credit and Banking*, 16(2):237–241, 1984.
- D. Smyth. Stop-go and United Kingdom export of manufactures. *Bulletin Oxford University Institute of Economics and Statistics*, 30(7):25–36, 1968.
- Brandon Walsh. Health education and the demand for tobacco in Ireland, 1953-76: A note. *Economic and Social Review*, (2):147–152, 1980.
- David Wheelock. Member bank borrowing and the fed's contractionary monetary policy during the great depression. *Journal of Money, Credit and Banking*, 22(4):409–426, 1990.
- Arlington Williams. The formation of price forecasts in experimental markets. *Journal of Money, Credit and Banking*, 19(1):1–18, 1987.

Chapter 5

Afraid of Heights: A Monte Carlo Study

5.1 Introduction

In Chapter Two is noted the problem of estimating the variance covariance matrix with the quantile estimators. Specifically, one needs to select ordered residuals to estimate the **height** of a density. The selection relies on a value "d". There are complicated ways to determine "d". In fact, there are three ways available in SHAZAM (**White** (1978)) but only for the Least Absolute Errors estimator, which has θ equal to $\frac{1}{2}$. One is SHAZAM's in-house method. The in-house method uses $\frac{T-k}{6}$ when T is large enough. The second is the method of **Bofinger** (1975). Finally there is the method of **Siddiqui** (1960).

The formulæ for the latter two methods are

$$d_{bof} = T^{-\frac{1}{5}} \left[\frac{4.5\varphi^4(\Phi^{-1}(\theta))}{\left(2\Phi^{-1}(\theta)^2 + 1\right)^2} \right]^{\frac{1}{5}}$$

$$d_{sid} = T^{-\frac{1}{3}} z_{\alpha}^{\frac{2}{3}} \left[\frac{1.5\varphi^2(\Phi^{-1}(\theta))}{\left(2\Phi^{-1}(\theta)^2 + 1\right)^2} \right]^{\frac{1}{3}}$$

where φ, Φ and Φ^{-1} are the normal density, the cumulative normal and the inverse of the cumulative normal. For least absolute errors θ is equal to $\frac{1}{2}$. SHAZAM (**White** (1978)) uses 1.96 for z in the **Siddiqui** (1960) formula. As in other chapters θ is the quantile.

In Chapter Four the estimated variances of the robust estimators were a lot lower than the estimated variances for least squares. For instance the Least Absolute Errors estimator gave seventy-seven percent of the coefficients a lower estimated variance three times lower than least squares when Least absolute Errors produced a lower estimated variance. These lower estimated variance fall into the category ""if it is too good to be true" it probably is" and thus it is necessary to ascertain the source of these lower variances.

5.2 A Proposed Monte Carlo Study

Until the publication of Version 7 of SHAZAM (**White** (1978)) the only option available at low cost with robust estimation was to use the SHAZAM in-house method. The assumption made in this thesis is that others doing empirical work in economics used this option. To preserve consistency when trying to see if robust estimators make a difference, we have elected to do the same. In essence a level playing field is preserved. This does not relieve one of the duty to determine whether the different methods of estimating "d" are very different; this chapter attempts to fulfil this obligation.

Some interesting questions arise concerning "d". It is important to know if the SHAZAM (**White** (1978)) in-house method deviates from the two methods suggested in the literature. All the least absolute errors' estimated deviations in Chapter Four use the SHAZAM in-house method for the reason given in the above paragraph. Also as "d" is used as part of the procedure to estimate the variance of an estimator, we would like to know if there is a difference in the estimated variances produced by the three methods. This is a problem to solve using Monte Carlo methods as "d" is derived as part of a procedure assuming a large sample size. In reality, the data we have is based on small samples. So with a Monte Carlo study not only can we answer some important questions about "d" but discern those properties in a small sample environment.

Briefly, such a Monte Carlo study would have the following structure. For a given specification and particular error distribution, make 5000 Least Absolute Errors' estimates of the coefficient and 5000 estimates of the variance produced by the three ways to calculate "d". Work out the estimated variance of the 5000 estimates of the coefficient. Calculate the mean of each of the three sets of the 5000 estimated variances. Of interest is the comparison of the estimated variance from the former with the individual means of the latter.

It has been suggested by **Judge** *et al* (1980) that the estimator for the height of the density, which requires "d", may not be adequate when k/T is large where k is the number of coefficients plus the constant and T the sample size. Also one would want to use least absolute errors when there is a suspicion the error distribution is one with possible outliers as it is the maximum likelihood estimator for a fat-tailed distribution, namely, the double exponential or Laplace distribution. Consequently, for this Monte Carlo study a set of fixed independent variables is selected from a data set (**Benderley and Zwick**, see the references at the end of Chapter Four) that has a k/T equal to 3/23 and the possibility the errors have outliers.

5.3 Steps of the Monte Carlo Study

The following is not a flow chart but attempts to convey the steps of the Monte Carlo experiment. The exact SHAZAM (**White** (1978)) commands are given later.

1. Specification: $Y_t = \beta_0 + \beta_1 X_t + \beta_2 Z_t + \epsilon_t$

Notice we have three coefficients, counting β_0 and two independent variables fixed in repeated samples. Two independent variables is an advantage as anecdotal evidence indicates much Monte Carlo work includes only one independent variable. While we have two independent variables, we focus on β_2 for the purpose of investigating standard error estimators.

2. Initial Values

$$\boxed{\beta_0 = -3.1673 \ \beta_1 = 6.0633 \ \beta_2 = -2.4818}$$

Values chosen from an initial investigation of the data. Although these particular known values are immaterial, the estimates from the Five-Quantile estimator are used.

3. $\boxed{\text{Sample Size: } T = 23}$

The sample size is deliberately small as it is in such an environment, coupled with three coefficients, that the estimator of the height of the density might not be adequate. In examining robust estimators others (**Johnstone and Velleman** (1985)) have used 5000 thousand drawings of an error vector.

4. $\boxed{\text{Begin DO-LOOP}}$

Generate 23 ϵ_t, randomly from a $N(0, Var(1))$ with probability 0.9 and from $N(0, Var(9))$ with probability 0.1.

This creates a distribution of errors with "outliers" according to what is known as a mixed Gaussian sampling density. A fraction of the errors comes from a normal with a zero mean and unit variance. Mixed with these (ten percent) are normal errors having zero mean and variance *nine*. These variances (**Johnstone and Velleman** (1985)) have been used to investigate robust estimators.

$$\boxed{\text{Generate 23 } Y_t \text{ using } X_t,\ Z_t \text{ and } \epsilon_t.}$$

Estimate $\boxed{\beta^{lae}}$ and three versions of its estimated standard deviation namely $\boxed{se^{lae}_{d=sha}}$, $\boxed{se^{lae}_{d=bof}}$ and $\boxed{se^{lae}_{d=sid}}$.

With the data generated and the fixed design estimate the β's using least absolute errors. Take care to keep all the 5000 estimates of β_2 and the estimated standard errors.

5. $\boxed{\text{End DO-LOOP after 5000 iterations}}$

Now the process of collecting a large number of estimates of coefficients and standard deviations is complete.

6. Calculate the estimated standard deviation of the estimated βs.

With the 5000 drawings of the errors we have a small sample distribution of the β's and we estimate the standard deviation of the this distribution by finding the estimated standard deviation of the 5000 β^{lae}'s.

7. Calculate the mean of the βs' standard deviations.

We also have a distribution, for the three methods, of the estimated standard errors. Calculate the mean of each of these distributions by averaging the respective 5000 estimated standard errors.

5.4 The SHAZAM Programme

The following is the programme used to perform the Monte Carlo study. The $*$'s are not an integral part of the programme. An attempt is made to link the lines of the programme to the outline of the previous section.

```
* Set up data and output files.
file 11 tbl05.dat
file 12 results1.out
* Set the sample size.   See box 3. above.
sample 1 23
* Set up some counters.
genl av=0
genl sq=0
genl sesha=0
genl sebof=0
genl sesid=0
genl sum=0
* Read in the data.  We are to use qf and p only.
read(11) intercpt qf p rs
**************************************************************
```

```
* Get the Same Random Numbers Each Time The Programme Is Used *
***************************************************************
set ranfix
* Make sure you have enough memory.
par 10000
* Suppress useless output.
set nodoecho nowarn
* Begin the DO-LOOP.  See box 4. above.
do #=1,5000
* Select the errors as described in the first box below box 4. above.
* Bear in mind SHAZAM wants the standard deviation instead of the variance.
genr a=uni(1)
genr d=dum(a-0.9)
genr dd=(2*d)+1
genr ee=nor(1)
genr e=ee*dd
* The errors have been selected.
* Generate the dependent variable.
?genr rs1=-3.1673+6.0633*qf-2.4818*p+e
*************************************************
* Perform Three Regressions.  One for Each "d". *
* Save Estimates of coefficients.               *
* Save Estimates of standard errors             *
*************************************************
?robust rs1 qf p / lae coef=beta stderr=std1b
?robust rs1 qf p / lae diff=-1 stderr=std2b
?robust rs1 qf p / lae diff=-2 stderr=std3b
* Focus on the estimated coefficient of p.
* Accumulate the counters for means and standard errors.
gen1 sq=sq+((beta:2)**2)
gen1 sum=sum+(beta:2)
```

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.138832
Standard Deviation (SHAZAM)	0.099105
Standard Deviation (Bofinger)	0.145077
Standard Deviation (Siddiqui)	0.145077

Table 5.1: Monte Carlo Results-**Benderley and Zwick**

```
genl sesha=sesha+std1b:2
genl sebof=sebof+std2b:2
genl sesid=sesid+std3b:2
endo
* End of the DO-LOOP.   See box 5. above.
* Calculate an item for the standard error estimate.
genl av=sum/5000
* Calculate the means of the standard deviations.
genl msesha=sesha/5000
genl msebof=sebof/5000
genl msesid=sesid/5000
* Calculate the standard deviation of the coefficient for p.
genl sebeta=sqrt((sq-5000*av**2)/4999)
* Save the results.
write(12) sebeta msesha msebof msesid
stop
```

5.5 Monte Carlo Results

The tables present the results of the Monte Carlo study. To summarize, as shown in Table 5.1, the SHAZAM (**White** (1978)) in-house method substantially underestimates the standard error and the **Bofinger** (1975) and **Siddiqui** (1960) methods substantially overestimates the standard error by the same amount.

One can find the values of Table 5.1 in Figure 5.1 by reading the horizontal axis.

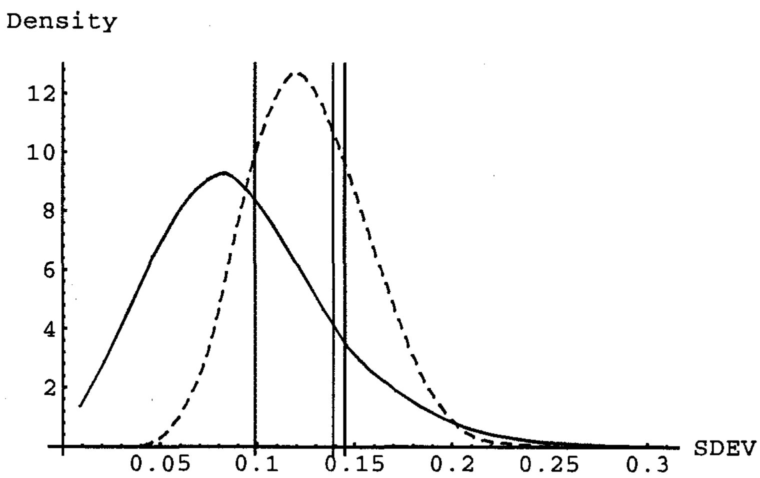

Figure 5.1: Empirical Densities-STDERR-Solid,SHAZAM-Dash,BOF and SID

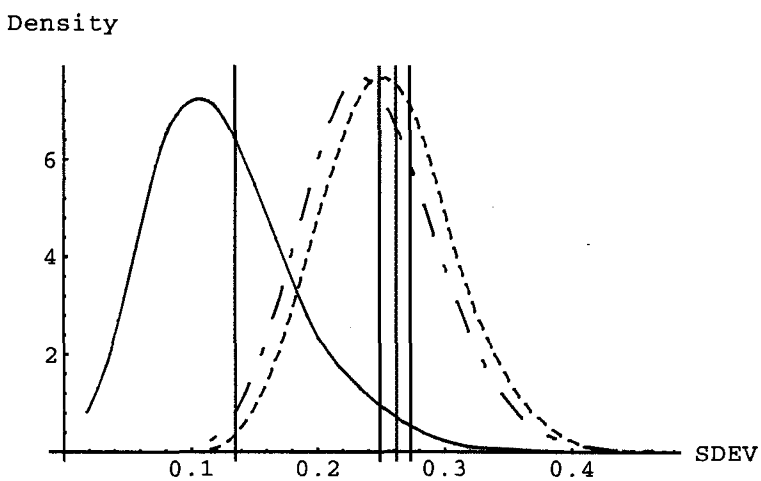

Figure 5.2: Empirical Densities-STDERR-Solid,SHAZAM-Dash,BOF and Dot Dash, SID

The numbers shown have darker ticks and as these are difficult to read a vertical line is drawn for each value in Table 5.1. For instance, the SHAZAM (**White** (1978)) method gave a value of 0.099105 and this is the first (left) vertical line in Figure 5.1. The three methods of estimating this standard deviation produce the distributions shown in Figure 5.1. The bold distribution is the SHAZAM (**White** (1978)) method. As the Bofinger and Siddiqui method gave the same distribution, only the Bofinger method is plotted with a bold line. The mean of the SHAZAM (solid) distribution lies at 0.099105, or the first vertical line. This is an underestimate of the value of 0.138832. The mean of the Bofinger and Siddiqui methods is 0.145077 (the last vertical line) and this is an overestimate. Notice all three estimates of 0.138832 (the middle vertical line) are biased but the latter two methods are closest. Almost identical results (0.098, 0.136, 0.144) were obtained when the number of replications is expanded to fifty-thousand.

Further, for each of the means of the estimated standard errors it is possible to estimate a standard error of the mean using a simple formula. This standard deviation of the mean is not to be confused with the standard deviations of the coefficient estimates. Here we are interested in the standard deviation of the *means* of the standard deviations shown on the horizontal axis of Figure 5.1. The "true" value of the mean value is 0.138832. If one adds (it lay below) for the SHAZAM (**White** (1978)) method and subtracts (they lay above) for the other methods, twice the estimated standard error *of the mean* to the mean values from the Monte Carlo study, it is possible to see "how close" are the three mean values using this measure of their standard deviation, to the "true" value. This calculation using the mean values and twice the estimated standard error of the mean values produces the upper value of the range as 0.1004 for the SHAZAM (**White** (1978)) method and the lower value of the range as 0.1441 for the Bofinger and Siddiqui methods. Thus we can have more confidence in the original mean values as the same conclusion is drawn when we look at the estimates of the mean values plus or minus two estimated standard deviations (of the mean values) away from the mean values.

As the Bofinger and Siddiqui methods gave the same answer for this data set,

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.2727677
Standard Deviation (SHAZAM)	0.1349326
Standard Deviation (Bofinger)	0.2624396
Standard Deviation (Siddiqui)	0.2493495

Table 5.2: Monte Carlo Results-**Barkley**

another data set (**Barkley**, see the references at the end of Chapter Three) with forty-five observations and k equal to eight is used with the same Monte Carlo procedure but selecting forty-five new errors. The empirical densities are plotted in Figure 5.2 but for this data set the Bofinger and Siddiqui methods gave different results and thus we are able to plot the density of the Siddiqui method with an alternate dot and dash. For this data set the mean of the 5000 β_2's is 0.2727677. This value can be found on the horizontal axis of Figure 5.2 just below 0.3 or the last (right) vertical line and what is noteworthy here is the "true" value lies above *all* the estimates unlike the "true" value in Figure 5.1. The mean of the Siddiqui density is 0.2493495 (the second vertical line) and is the density with the alternate dot and dash. The Bofinger method, the dash density, has a mean of 0.2624396, or the third vertical line. All estimates are still biased and all underestimate, unlike the previous data set, the standard deviation of the β_2's. The Bofinger method comes the closest to the standard deviation of theβ_2's.

Overall on the basis of the empirical distribution of β_2, the SHAZAM (**White** (1978)) in-house method fares rather badly as its estimated standard error is the furtherest of the three methods from the estimate standard deviation of the β_2's. On the basis of this Monte Carlo study it can be said the possibility of "d" influencing results, that we had anticipated in Chapter Two is realized. Also the mystery of the lower estimated variances of the robust methods is solved. The SHAZAM method is the culprit.

This result identifies one source of the lower robust estimated variance especially for the quantile estimators from SHAZAM (**White** (1978)). It raises the possibility that the percentage of coefficients becoming significant or insignificant is incorrect.

"Robust" HypothesisTest Results	
Criterion–XI	Percentage Sig/Insig
Least Absolute Errors	20
t tests	22
BOFINGER	26
SIDDIQUI	27

Table 5.3: "Robust" HypothesisTest Results

For this reason it was decided to repeat Criterion XI for the Least Absolute Errors Estimator but use both the Bofinger and the Siddiqui methods to compute the estimated standard errors for thirty-seven studies that use t-tests. We focus on t-tests as the results for these mirror the original results as Table 5.3 shows. The original results had twenty percent of coefficients changing from significant to insignificant and changing from insignificant to significant. For t tests this percentage is twenty-two percent. The Bofinger and Siddiqui methods reverse the relative magnitudes of the original changes and are almost identical. This unity serves to confirm these new results obtained with the Bofinger and Siddiqui methods. The SHAZAM (**White** (1978)) method should be avoided as it has been shown to underestimate the estimated variances and this results in incorrect conclusions being made when looking at many hypothesis tests. Note, however, that it remains the case that a substantial number of hypothesis test results are changed: we still must conclude that using robust methods does make a difference.

5.6 More Monte Carlo Results

To determine the nature of the relationship between k/T and the extent to which the estimated standard error departs from the "true" value, nine data sets were selected at random from those where a robust estimator gave a different result on the basis of comparing hypothesis tests. The k/T's of these nine data sets ranged from a low of 0.0174 to a high of 0.16 and we felt this more than enough spread to perform the experiment. Using a mixed distribution where the ϵ_t are randomly drawn from a $N(0, Var(1))$ with probability 0.75 and from $N(0, Var(9))$ with probability 0.25, we performed the same Monte Carlo experiment as outlined in the previous sections. For each data set the SHAZAM (**White** (1978)) estimated standard error is subtracted from the "true" value for the data set. This is called the SHAZAM ERROR and all were underestimates. Next the SHAZAM ERROR is regressed on k/T and the result showing nine data sets is plotted in Figure 5.3. Notice in Figure 5.3 all the data sets have positive and large SHAZAM ERROR above k/T equal to 0.065. In fact the two data sets with k/T at or below 0.065 have very small SHAZAM ERRORS: 0.0009 and 0.013 respectively. Although nine data sets is a small sample, this experiment does show how sensitive the estimated standard errors of the SHAZAM method are to the number of coefficients plus a constant relative to the sample size. In Figure 5.4 we plot the SHAZAM ERROR against $T - k$ and if the SHAZAM ERROR increases as k/T increases (Figure 5.3), the SHAZAM ERROR should decrease as $T - k$ rises. This is indeed the case. In Figure 3 for the SHAZAM (**White** (1978)) method to be satisfactory one needs a k/T less than 0.05 to obtain a low SHAZAM ERROR. To be fair to the SHAZAM (**White** (1978)) method, with the Huber M-Estimator, **Huber** (1973) finds one needs a k/T equal to 0.125 for the estimated variance covariance matrix to be satisfactory. Also one can infer from **Sposito** (In **Lawrence and Arthur** (1990)) that $1/T$-for the location case-must be as low as 0.01.

The Monte Carlo study is expanded to include three more distributions: the normal, the slash-a unit normal divided by an independent unit uniform which has an infinite variance and less peaked than the Cauchy distribution, and a mixed distribution where the ϵ_t are randomly drawn from a $N(0, Var(1))$ with probability 0.75 and from $N(0, Var(9))$ with probability 0.25. This is done for the **Benderley and Zwick** and **Barkley** data sets. The results are shown in Tables 5.4 to Tables 5.9 and the original results from Tables 5.1 and 5.2 are included for comparison. In every Monte Carlo experiment with the new distributions it is the case that the SHAZAM (**White** (1978)) method produces estimated standard errors below the "true" value. For the **Benderley and Zwick** data set it is still the case that the "true" value is between the the SHAZAM (**White** (1978)) and other estimates' levels. Again, for this

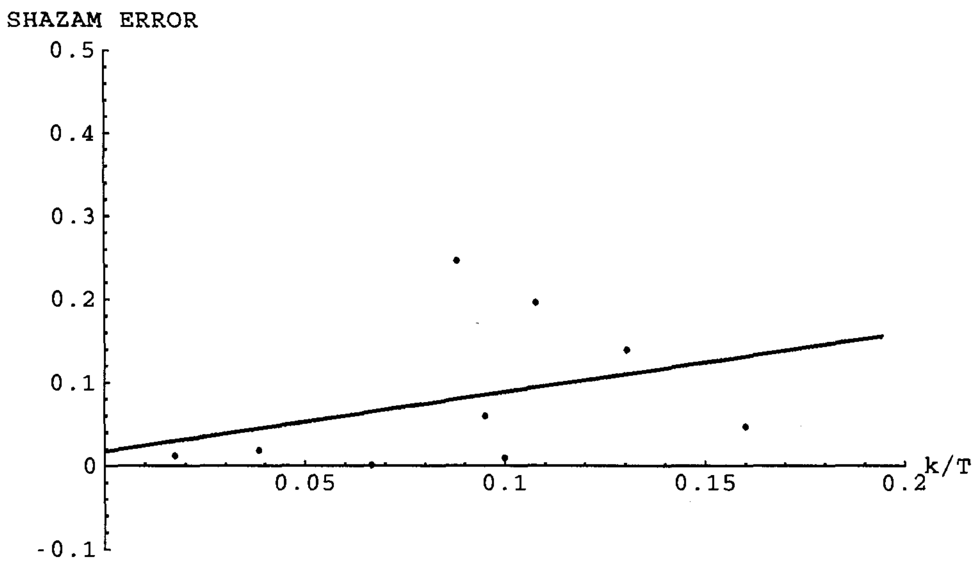

Figure 5.3: Relationship of k/T to the Difference between the SHAZAM estimated standard error and the "true" standard error: SHAZAM ERROR

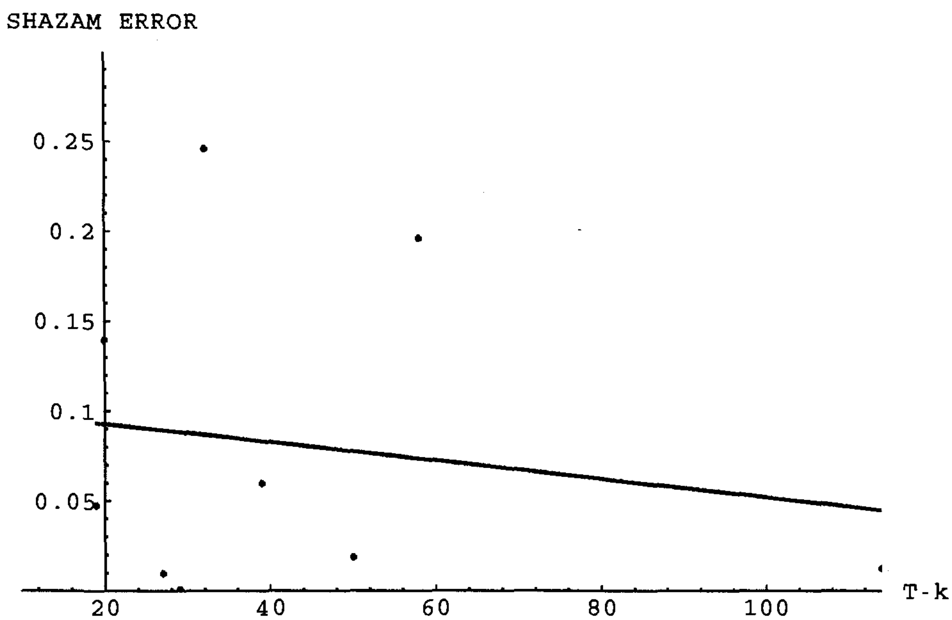

Figure 5.4: Relationship of $T-k$ to the Difference between the SHAZAM estimated standard error and the "true" standard error: SHAZAM ERROR

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.1245312(0.138832)
Standard Deviation (SHAZAM)	0.09092599(0.099105)
Standard Deviation (Bofinger)	0.1304525(0.145077)
Standard Deviation (Siddiqui)	0.1304525(0.145077)

Table 5.4: Monte Carlo Results-**Benderley and Zwick**-Normal

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.2486139(0.2727677)
Standard Deviation (SHAZAM)	0.1254389(0.1349326)
Standard Deviation (Bofinger)	0.2390932(0.2624396)
Standard Deviation (Siddiqui)	0.2273266(0.2493495)

Table 5.5: Monte Carlo Results-**Barkley**-Normal

data set and the new mixed distribution, the Monte Carlo experiment is performed with seventy-two thousand drawings of the error term with the following results (0.16, 0.11, 0.17) which agree with those obtained with five thousand. Also for **Barkley** the "true" value is still underestimated by all three methods. For the nine data sets analysed in the previous paragraph, all of them displayed the feature that the SHAZAM (**White** (1978)) method underestimated the "true" value. The initial results using a mixed distribution (the ϵ_t are randomly drawn from a $N(0, Var(1))$ with probability 0.90 and from $N(0, Var(9))$ with probability 0.10) are supported with many data sets and three new distributions. One should be careful if k/T is above 0.05 and one opts to use the SHAZAM (**White** (1978)) method to estimate standard errors.

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.1643315(0.138832)
Standard Deviation (SHAZAM)	0.1143268(0.099105)
Standard Deviation (Bofinger)	0.1742031(0.145077)
Standard Deviation (Siddiqui)	0.1742031(0.145077)

Table 5.6: Monte Carlo Results-**Benderley and Zwick**-Mixed (.75)

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.3127507(0.2727677)
Standard Deviation (SHAZAM)	0.155697(0.1349326)
Standard Deviation (Bofinger)	0.3078626(0.2624396)
Standard Deviation (Siddiqui)	0.2898554(0.2493495)

Table 5.7: Monte Carlo Results-**Barkley**-Mixed (0.75)

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.3284895(0.138832)
Standard Deviation (SHAZAM)	0.2120562(0.099105)
Standard Deviation (Bofinger)	0.3713633(0.145077)
Standard Deviation (Siddiqui)	0.3713633(0.145077)

Table 5.8: Monte Carlo Results-**Benderley and Zwick**-Slash

Monte Carlo	Standard Deviation
Standard Deviation β_2	0.9306974(0.2727677)
Standard Deviation (SHAZAM)	0.2849713(0.1349326)
Standard Deviation (Bofinger)	0.6114121(0.2624396)
Standard Deviation (Siddiqui)	0.5650521(0.2493495)

Table 5.9: Monte Carlo Results-**Barkley**-Slash

Chapter 6

Summary and Conclusions

On first discovering robust methods one is caught up in their revolutionary possibilities. For instance, the literature on robust methods does indicate that least squares is flawed. This opens up the possibility of trying to show robust methods **work better** with real data than does ordinary least squares. A literature review revealed that robust estimators had not made that much of an impact in economics despite theoretical developments in statistics and econometrics. This thesis attempts to determine if robust estimators do make a difference with economic data sets. If there was any bias initially, it fell on the side of the least squares. But before we could pit the robust methods against least squares many issues had to be covered.

First, we felt it imperative to work with a large number of data sets as this is a shortcoming of a literature which typically sees one data set as sufficient. Ninety data sets are used from forty-four empirical studies in economics, both cross-section and time-series. One advantage from using real data, rather than artificial data from, say, a Monte Carlo study, is the latter is indeed artificial and need not be anything like real data.

The next problem is how to pit robust and least squares against each other. This is easy, given the economist's penchant for hypothesis testing. We ask if the robust methods changed the conclusion of the hypothesis test of interest to the original researcher. Bear in mind we had mixed hopes for the robust estimators' ability to do this. As it turns out the robust estimators did reverse some results. This initial finding, on further investigation, is sensitive to the method used for calculating the standard errors of the estimate. After some Monte Carlo work, the the robust estimates of Least Absolute Errors, the one estimator for which we could obtain the "correct" calculation, still gave a different result, more in line with what had been obtained with randomization tests.

Given the problem with the SHAZAM standard errors for the robust estimators and that we redid the hypothesis tests with standard errors calculated with both of the new SHAZAM options for the least Absolute Errors estimator, we were able to pose two interesting questions. First, "Typically, what types of data sets had hypothesis tests change?". It appears that what might be termed "monetary data" exhibited more changes in hypothesis tests than did other types of data sets. Second, "Was there one data set, possibly familiar to many economists, that had results originally *significant* become *insignificant*?". Indeed, the **Lucas and Rapping** (1969) study of labour supply (specifically the determinants of the wage rate) had two of five coefficients singled out as important, on the basis of economic theory and originally reported as significant, become insignificant. While the reason for this change may warrant some further investigation, it is a data set from a widely cited and accepted study and the change in test results using robust analysis maybe of interest to many economists. Also, there seemed to be more problems with time series data, but this may reflect the preponderance of time series data.

To ensure the robust methods had been given a fair chance and to provide coverage of all aspects of what it means to **work better**, we also calculated the criteria for comparing estimates that others had provided in the literature. These results provided a mixed picture of the tendency for robust methods to be different from least squares. The sizes of coefficients are different but not relative to the least squares' standard deviation. Tests of differences produced mixed results but opting for the more important test we conclude robust estimators are not different. They do not make a difference when forecasting known values. As mentioned, some hypothesis tests are changed.

One problem that bedevils one is to try to see through the façade of the robust œuvre in order to recommend one for general use. What emerges is, using SHAZAM

(**White** (1978)), only Least Absolute Errors, with alternate estimates of the standard errors, is a viable practical option. The two M-Estimators from BMDP (**Dixon** (1990)) also suffer from the defect that the incorrect formulæ are used for the standard errors. However, if one looks at all the tables, in Chapter Four, both the Huber M-Estimator and Least Absolute Errors stand out from the others and we suggest both be used.

Another technique, closely related to robust methods in spirit, is a method of checking hypothesis tests called Randomization. We felt a two horse race (least squares and robust methods) to be insufficient. So we put the **data set** of data sets through the randomization process relying on two methods of randomization. The randomization methods were unanimous in reversing over thirty percent of the originally significant results. The direction of this result is partially supported by the robust estimators, after Monte Carlo results provided us with what are "better" methods of calculating robust standard errors. The conclusion from the exercise is while we should still use robust methods to vet our estimated equations, we should pay close attention to the methods of randomization which can also help in the same regard. Randomization tests have the advantage that the problems with estimating the variance covariance matrix are avoided.

Topics for Future Research

1. One criterion concerned itself with the difference of the robust estimate from that of least squares' relative to the the least squares' standard deviation. No adjustment is made here for the possible covariance between the coefficient estimates. Whether such an adjustment is necessary is a moot point and thus a promising avenue to pursue.

2. Related to the above is a possible need to correct the χ^2 test of Category Two, in Chapter Three and Four, for the covariance between the coefficient estimates. This has not been done here but is also another item for future research.

3. We discovered (Chapter Four) the estimates of standard errors from popular

econometrics packages to be flawed and Monte Carlo (Chapter Five) work confirmed this finding. Although we included four distributions and eleven data sets these could be increased and an even more comprehensive study performed.

4. Also, the Monte Carlo work could be expanded upon to allow an investigation of estimation of the standard errors using bootstrap methods.

5. Further, instead of using the econometric packages to provide standard errors, the correct formula could be employed for the Huber M-Estimator.

6. A final item for future work would be to link the the "index" of outliers to some of the criterion we have used to determine if robust estimators are different from least squares.

Bibliography

[And74] Andrews, D. F. (1974) 'A robust method for multiple linear regression.' *Technometrics* 16, 523–531

[Bax90] Baxter, Laurence A. (1990) 'Futures of statistics.' *The American Statistician* 44(2), 128–129

[Ben83] Benjamini, Yoav (1983) 'Is the t test really conservative when the parent distribution is long-tailed.' *Journal of the American Statistical Association* 78(383), 645–654

[Ber90] Berk, Richard A. (1990) 'A primer on robust regression.' In *Modern Methods of Data Analysis,* ed. John Fox and J. Scott Long, first ed. (Newbury Park, California: Sage Publications) chapter 7, pp. 292–324

[Ber91] Berndt, Ernst R. (1991) *The Practice of Econometrics: Classic and Contemporary* (Addison-Wesley)

[Bic76] Bickel, Peter J. (1976) 'Another look at robustness: A review of reviews and some new developments.' *Scandinavian Journal of Statistics* 3, 145–168

[BJ90] Bollen, Kenneth A., and Robert W. Jackman (1990) 'Regression diagnostics: An expository treatment of outliers and influential cases.' In *Modern Methods of Data Analysis,* ed. John Fox and J. Scott Long, first ed. (Newbury Park, California: Sage Publications) chapter 6, pp. 257–291

[Bof75] Bofinger, E (1975) 'Estimation of a density function using order statistics.' *Australian Journal of Statistics* 17, 1–7

[Boo80] Boos, Dennis D. (1980) 'A new method for constructing approximate confidence intervals from M-Estimates.' *Journal of the American Statistical Association* 75(369), 142–145

[Box53] Box, G (1953) 'Non-normality and tests on variances.' *Biometrica* 40, 318–335

[BMNW90] Butler, Richard J., James B. McDonald, Ray D. Nelson, and Steven B. White (1990) 'Robust and partially adaptive estimation of regression models.' *Review of Economics and Statistics* 72(2), 321–27

[Car79] Carroll, Raymond J. (1979) 'On estimating variances of robust estimators when the errors are asymmetric.' *Journal of the American Statistical Association* 74(367), 674–679

[CH86] Chaterjee, Samprit, and Ali Hadj (1986) 'Influential observations, high leverage points, and outliers in linear regression.' *Statistical Science* 1(3), 379–416

[CU88] Coate, M, and D Uri (1988) 'A simultaneous equations approach to modeling industry structure and economic performance.' *Metroeconomica* 39(2), 181–204

[Con91] Connolly, Robert A (1989) 'An examination of the robustness of the weekend effect.' *Journal of Financial and Quantitative Analysis* 24, 133–171

[CN88] Coursey, D, and H Nyquist (1988) 'Applications of robust estimation techniques in demand analysis.' *Applied Economics* 20, 595–610

[CH74] Cox, D, and D Hinkley (1974) *Theoretical Statistics* (Chapman and Hall)

[DE90] Darnell, Adrian C., and J. Lynne Evans (1990) *The Limits of Econometrics,* first ed. (Aldershot, England: Edward Elgar)

[dJdW85] de Jongh, P. J., and T. de Wet (1985) 'A Monte Carlo comparison of regression trimmed means.' *Communications in Statistics-Theory and Methods* 14(10), 2457–2472

[DWA86] Dewald, William, Jerry Thursby, and Richard Anderson (1986) 'Replication in empirical economics: The *The Journal of Money, Credit and Banking project.*' *American Economic Review* 76(4), 587–603

[Dix90] Dixon, W. J. (1990) *BMDP Statistical Software,* vol. 2 (Los Angeles: University of California Press)

[DT68] Dixon, W, and W Tukey (1968) 'Approximate behaviour of the distribution of winsorized t (trimming/winsorization 2).' *Technometrics* 10(1), 83–98

[E87] Edgington, Eugene (1987) *Randomization Tests* STATISTICS: Textbooks and Monographs, second ed. (New York: Marcel Dekker)

[EK82] Eddy, William F., and Joseph B. Kadane (1982) 'The cost of drilling for oil and gas: An application of constrained robust regression.' *Journal of the American Statistical Association* 77(378), 262–269

[Fai74] Fair, Ray C. (1974) 'On the robust estimation of econometric models.' *Annals of Economic and Social Measurement* 3(4), 667–77

[Fam63] Fama, Eugene (1963) 'Mandelbrot and the stable Paretian hypothesis.' *Journal of Business* pp. 420–429

[Fam65] Fama, Eugene (1965) 'The behavior of stock market prices.' *Journal of Business* pp. 34–105

[GT90] Geske, Robert, and Walter Torous (1990) 'Black-Scholes option pricing and robust variance estimation.' In *Options: Recent advances in theory* *and practice*, ed. Stewart Hodges, first ed. (New York: Manchester University Press) chapter 5, pp. 49–69

[GO72] Granger, C, and D Orr (1972) '"Infinite variance" and research strategy in time series analysis.' *Journal of the American Statistical Association* 67(328), 275–285

[GWK89] Granger, C, H White, and M Kamstra (1989) 'Interval forecasting: An analysis based upon ARCH-quantile estimators.' *Journal of Econometrics* 40, 87–96

[Gro76] Gross, Alan M. (1976) 'Confidence interval robustness with long-tailed symmetric distributions.' *Journal of the American Statistical Association* 71(354), 409–416

[Gro77] Gross, Alan M. (1977) 'Confidence intervals for Bisquare regression estimates.' *Journal of the American Statistical Association* 72(358), 341–354

[HL90] Hall, R, and D Lilien (1990) *MicroTSP User's Manual* (Irvine, California: Quantitative Micro Software)

[HS88] Hallman, J. and Mark Kamstra (1989) 'Combining algorithms based on robust estimation techniques and co-integrating restrictions.' *Journal of Forecasting*, 8, 189–198.

[Ham71] Hampel, Frank R. (1971) 'A general qualitative definition of robustness.' *The Annals of Mathematical Statistics* 42(6), 1887–1896

[Har77] Harvey, A. C. (1977) 'A comparison of preliminary estimators for robust regression.' *Journal of the American Statistical Association* 72(360), 910–913

[HS92] Hettmansperger, T, and S Sheather (1992) 'A cautionary note on the method of least median squares.' *The American Statistician* 46(2), 79–83

[HD82] Hill, Mary Ann, and W. J. Dixon (1982) 'Robustness in real life: A study of clinical laboratory data.' *Biometrics* 38, 377–396

[HH77] Hill, Richard W., and Paul W. Holland (1977) 'Two robust alternatives to least squares regression.' *Journal of the American Statistical Association* 72(360), 828–833

[HT75] Hinich, Melvin J., and Prem P. Talwar (1975) 'A simple method for robust regression.' *Journal of the American Statistical Association* 70(349), 113–119

[Hog74] Hogg, Robert V. (1974) 'Adaptive robust procedures: A partial review and some suggestions for future applications and theory.' *Journal of the American Statistical Association* 69, 909–927

[Hog77] Hogg, Robert V. (1977) 'An introduction to robust procedures.' *Communications in Statistics* (9), 789–794

[Hog79] Hogg, Robert V. (1979) 'Statistical robustness: One view of its use in applications today.' *The American Statistician* 33(37), 108–115

[HC78] Hogg, Robert V, and Allen Craig (1978) *Introduction to Mathematical Statistics*, fourth ed. (New York: Macmillian)

[HVRY90] Hogg V. Robert, Stephen J. Ruberg and Lianng Yuh (1990) *Robust data analysis* In Donald A. Berry, editor, *Statistical Methodology in the Pharmaceutical Sciences*, number 104 in Statistics: Textbooks and Monographs, chapter 12, pages 357–388. Marcel Dekker, New York, first edition.

[HW77] Holland, Paul W., and Roy E. Welsch (1977) 'Robust regression using iteratively reweighted least-squares.' *Communications in Statistics* (9), 813–827

[Hub72] Huber, Peter J. (1972) 'Robust statistics: A review.' *The Annals of Mathematical Statistics* 43(4), 1041–1067

[Hub73] Huber, Peter J. (1973) 'Robust regression: Asymptotics, conjectures and Monte Carlo.' *Annals of Statistics* 1(5), 799–821

[Hub77] Huber, Peter J. (1977) *Robust Statistical Procedures* CBMS-NSF Regional Conference Series In Applied Mathematics, first ed. (Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics)

[Hub81] Huber, Peter J. (1981) *Robust Statistics* Wiley Series In Probability And Mathematical Statistics, first ed. (New York: John Wiley)

[Igl83] Iglewicz, Boris (1983) 'Robust scale estimators and confidence intervals for location.' In *Understanding Robust and Exploratory Data Analysis*, ed. David C. Hoaglin, Frederick Mosteller, and John W. Tukey, first ed. Wiley Series In Probability and Mathematical Statistics (New York: John Wiley) chapter 12, pp. 404–431

[J+85] Judge, George G. et al. (1985) *The Theory and Practice of Econometrics* Wiley Series in Probability and Mathematical Statistics, second ed. (New York: John Wiley)

[JV85] Johnstone, I, and P Velleman (1985) 'The resistant line and related regression methods.' *Journal of the American Statistical Association* 80(392), 1041–1054

[JH83] Joiner, Brian L., and David L. Hall (1983) 'The ubiquitous role of f'/f in efficient estimation of location.' *The American Statistician* 37(2), 128–133

[JS68] Jennrich, R. I., and P. F. Sampson (1968) 'Application of stepwise regression to non-linear estimation.' *Technometrics* 10(1), 63–72

[J+82] Judge, George G. et al. (1988) *Introduction to the Theory and Practice of Econometrics* Wiley Series in Probability and Mathematical Statistics. (New York: John Wiley)

[Kaf82] Kafadar, Karen (1982) 'A biweight approach to the one-sample problem.' *Journal of the American Statistical Association* 77(378), 416–424

[Ken93] Kennedy, Peter (1993) 'Randomization tests in econometrics.' *Simon Fraser University, Department of Economics, Discussion Paper*

[Kle74] Klein, L (1974) *A Textbook of Econometrics,* second ed. (New Jersey: Prentice-Hall)

[Koe82] Koenker, Roger W. (1982) 'Robust methods in econometrics.' *Econometric Reviews* 1(2), 213–255

[Koe88] Koenker, R (1988) 'Asymptotic theory and econometric practice.' *Journal of Applied Econometrics* 3, 139–143

[KB72] Koenker, R, and G Basset (1978) 'Regression quantiles.' *Econometrica* 46, 33–50

[KB82] Koenker, R, and G Basset (1982) 'Tests of linear hypotheses and l_1 estimation.' *Econometrica* 50, 1577–1583

[LA90] Lawrence, K. D., and J. L. Arthur (1990) *Robust Regression: Analysis and Applications,* first ed. (New York: Marcel Dekker)

[Lax85] Lax, David A. (1985) 'Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions.' *Journal of the American Statistical Association* 80(391), 736–741

[LD75] Lehmann, E, and H D'Abrera (1975) *NONPARMETRICS, Statistical Methods Based on Ranks* (Oakland: Holden)

[LS91] Lichtenberg, Frank R., and Donald Siegel (1991) 'The impact of R and D investment on productivity-new evidence using linked R & D-LRD data.' *Economic Inquiry* 29(2), 203–229

[Lio82] Lioukas, Spyros K. (1982) 'Choice of travel mode and the value of time in Greece.' *Journal of Transport Economics and Policy* 16(2), 161–180

[LR92] Lott, W, and S Ray (1992) *Applied Econometrics: Problems with Data Sets* (Dryden)

[Mag91] Magee, Lonnie (1991) 'Edgeworth-adjusting test statistics for AR(1) errors.' *Communications in Statistics-Simulation* 20(4), 901–907

[MBR91] Magee, L, J Burbidge, and L Robb (1991) 'Computing kernel-smoothed conditional quantiles from many observations.' *Journal of the American Statistical Association* 86(415), 673–677

[MV91] Magee, Lonnie, and M Veall (1991) 'Selecting regressors for prediction using PRESS and White t statistics.' *Journal of Business & Economics and Statistics* 9(1), 91–96

[MVW84] Mahajan Vijay, Subash Sharma, and Yoram Wind (1984) 'Parameter estimation in marketing models in the presence of influential response data: Robust regression and applications.' *Journal of Marketing Research* 21, 268–277

[Mal79] Mallows, C (1979) 'Robust methods-some examples of their use.' *The American Statistician* 33(4), 179–184

[Man63a] Mandelbrot, Benoit (1963a) 'New methods in statistical economics.' *The Journal of Political Economy* 71(5), 421–440

[Man63b] Mandelbrot, Benoit (1963b) 'The variation of certain speculative prices.' *Journal of Business* 36, 394–419

[Man69] Mandelbrot, Benoit (1969) 'Long-run linearity, locally gaussian process, H-spectra and infinite variances.' *International Economic Review* 10(1), 82–111

[Mar79] Maritz, J. S. (1979) 'A note on exact robust confidence intervals for location.' *Biometrika* 66(1), 163–166

[Mir90] Mirowski, Philip (1990) 'From Mandelbrot to chaos in economic theory.' *Statistical Science* 1(3), 379–416

[MB76] Murdoch, J, and J Barnes (1976) *Statistical Tables for Science, Engineering and Management,* second ed. (Toronto: Macmillian)

[MS91] Mirowski, Philip, and S Sklivas (1991) 'Why econometricians don't replicate (although they do reproduce).' *Review of Political Economy* 3(2), 146–163

[NW86] Narula, Subhash C., and John F. Wellington (1986) 'An algorithm to find all regression quantiles.' *Journal of Statistical Computing and Simulation* 26, 269–281

[PKRF88] Patel, K R., G. S. Mudholkar, and J. L. Indrasiri Fernando (1988) 'Student's t approximations for three simple robust estimators.' *Journal of the American Statistical Association* 83(404), 1203–1210

[Pre75] Prescott, P. (1975) 'A simple alternative to Student's t.' *Applied Statistics* 24(2), 210–217

[Ram77] Ramsay, J. O. (1977) 'A comparative study of several robust estimates of slope, intercept, and scale in linear regression.' *Journal of the American Statistical Association* 72(359), 608–615

[RPE89] Rech Paul, Paul Schmidbauer, and Jamie Eng (1989) 'Least absolute regression revisited-a simple labeling method for finding a LAR line.' *Communications in Statistics-Simulation* 18(3), 943–955

[RR77] Relles, Daniel A., and William H. Rogers (1977) 'Statisticians are fairly robust estimators of location.' *Journal of the American Statistical Association* 72(357), 107–111

[RMB92] Robb, A, L Magee, and J Burbidge (1992) 'Kernel smoothed consumption-age quantiles.' *Canadian Journal of Economics* 25(3), 669–680

[RDMR82] Rocke, David M., George W. Downs, and Alan J. Rocke (1982) 'Are robust estimators really necessary?' *Technometrics* 24(2), 95–101

[Rou84] Rousseeuw, Peter (1984) 'Least median of squares regression.' *Journal of the American Statistical Association* 79(10), 871–880

[RC80] Ruppert, David, and Raymond J. Carroll (1980) 'Trimmed least squares estimation in the linear model.' *Journal of the American Statistical Association* 75(372), 828–838

[SH80] Schrader, Ronald M., and Thomas P. Hettmansperger (1980) 'Robust analysis of variance based upon a likelihood ratio criterion.' *Biometrika* 67(1), 93–101

[SH82] Shoemaker, Lewis H., and Thomas H. Hettmansperger (1982) 'Robust estimates and tests for the one- and two-sample scale models.' *Biometrika* 69(1), 47–53

[Sid60] Siddiqui, M (1960) 'Distribution of quantiles in samples from a bivariate population.' *Journal of Research National Bureau of Standards* (64), 145–150

[ST60] Siegel, S, and J Tukey (1960) 'A nonparametric sum of ranks procedure for relative spread in unpaired samples.' *Journal of the American Statistical Association* 55, 429–444

[Sim84] Simonoff, J (1984) 'A comparison of robust methods and detection of outliers techniques when estimating a location parameter.' *Communications in Statistics, A, Theory and Methods* 13(7), 813–842

[Sma86] Small, Kenneth A. (1986) 'Effects of the 1979 gasoline shortages on Philadelphia housing prices.' *Journal of Urban Economics* 19, 371–381

[Smi84] Smith, A (1984) 'Present position and potential developments: Some personal views Bayesian statistics.' *Journal of the Royal Statistical Society Series A* 147(7), 245–259

[Sti73] Stigler, S (1973) 'Simon Newcomb, Percey Daniell, and the history of robust estimation 1885-1920.' *Journal of the American Statistical Association* 27, 212–217

[Sti77] Stigler, Stephen M. (1977) 'Do robust estimators work with *real* data?' *The Annals of Statistics* 5(6), 1055–1098

[SCR88] Street, James O., Raymond J. Carroll, and David Ruppert (1988) 'A note on computing robust regression estimates via iteratively reweighted least squares.' *The American Statistician* 42(2), 152–154

[SW86] Swartz, S, and R Welsch (1986) 'Applications of bounded influence and diagnostic methods in energy modelling.' In *Model Reliability,* ed. D Belsey and R Welsch (Cambridge: MIT Press)

[SK91] Swinton, S, and R King (1991) 'Evaluating robust regression techniques for detrending crop yield data with nonnormal errors.' *American Journal of Agricultural Economics* 73(2), 446–451

[Tay74] Taylor, Lester D. (1974) 'Estimation by minimizing the sum of absolute errors.' In *Frontiers in Econometrics,* ed. Paul Zarembka (New York: Academic Press) chapter 6, pp. 169–190

[TTB86] Tiku, M, W Tan, and N Balakrishnan (1986) *Robust Inference* STATISTICS: textbooks and monographs, first ed. (New York: Marcel Dekker)

[WC77] Wegman, Edward J., and Raymond J. Carroll (1977) 'A Monte Carlo study of robust estimators of location.' *Communications in Statistics* (9), 795–812

[Wel87] Welsh, A. H. (1987) 'The trimmed mean in the linear model.' *The Annals of Statistics* 15(1), 20–36

[Whi78] White, Kenneth J. (1978) 'A general computer program for econometric methods-SHAZAM.' *Econometrica* pp. 239–240

[Wu85] Wu, Lawrence L. (1985) 'Robust M-estimation of location and regression.' In *Sociological Methodology 1985,* ed. Nancy Brandon Tuma, first ed. The Jossey-Bass Social And Behavioral Science Series (San Francisco: Jossey-Bass) chapter 10, pp. 316–388

[YM79] Yohai, Victor J., and Ricardo A. Maronna (1979) 'Asymptotic behavior of M-estimators for the linear model.' *The Annals of Statistics* 7(2), 258–268

[ZT70] Zeckhauser, Richard, and Mark Thompson (1970) 'Linear regression with non-normal error terms.' *The Review Of Economics And Statistics* 52, 280–286