ROBUST REGRESSION, HCCM ESTIMATORS, AND AN EMPIRICAL BAYES APPLICATION

A THESIS PRESENTED BY MEHMET ORHAN
TO
THE INSTITUTE OF
ECONOMICS AND SOCIAL SCIENCES
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS
FOR THE DEGREE OF Ph. D. OF
ECONOMICS

BILKENT UNIVERSITY
May, 1999

ROBUST REGRESSION, HCCM ESTIMATORS, AND AN EMPIRICAL BAYES APPLICATION

A THESIS PRESENTED BY MEHMET ORHAN

TO

THE INSTITUTE OF

ECONOMICS AND SOCIAL SCIENCES

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF Ph. D. OF

ECONOMICS

BILKENT UNIVERSITY

May, 1999

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph. D. of Economics.

Prof. Dr. Asad Zaman

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph. D. of Economics.

Assist. Prof. Dr. Mehmet Caner

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph. D. of Economics.

Assist. Prof. Dr. Savaş Alpay

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph. D. of Economics.

Assist. Prof. Dr. Süheyla Özyıldırım

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Ph. D. of Economics.

Assoc. Prof. Dr. Turan Erol

Approved by the Institute of Economics and Social Sciences

Director:

ABSTRACT

ROBUST REGRESSION, HCCM ESTIMATORS, AND AN EMPIRICAL BAYES APPLICATION

MEHMET ORHAN
Ph. D. OF ECONOMICS
Supervisor: Prof. Dr. Asad Zaman
May 1999

This Ph.D. thesis includes three topics of econometrics where the chapters of the whole study are devoted to robust regression analysis, research on the estimators for the covariance matrix of a heteroskedastic regression and finally an application of the Empirical Bayes method to some real data from İstanbul Stock Exchange. Some robust regression techniques are applied to some data sets to show how outliers of a data set may lead to wrong inferences. The results reveal that the former studies have gone through some wrong results with the effect of the outliers that were not detected. Second chapter makes a thorough evaluation of the existing heteroskedasticity consistent covariance matrix estimators where the Maximum Likelyhood estimator recently promoted to the literature by Zaman is also taken into consideration. Finally, some empirical study is carried out in the last part of the thesis. The firms of ISE are categorized into sectors and some estimation is done over an equation which is very common and simple in the finance literature.

Key Words: Heteroskedasticity, Breakdown Point, Least Median of Squares, Outlier, Robust Distance, Empirical Bayes.

ÖZET

KATI REGRESYON, HUKM TAHMİN EDİCİLERİ, VE BİR AMPİRİK BAYES UYGULAMASI

MEHMET ORHAN

Doktora Tezi, İktisat Bölümü

Tez Yöneticisi: Prof. Dr. Asad Zaman

May 1999

Bu doktora tezi üç ekonometri konusunu içermektedir ki bunlardan ilki katı regresyon analizine, ikincisi heteroskedastik regresyonda kovaryans vektörü tahmin edicileriyle ilgili araştırmalara, ve sonuncusu da İstanbul Borsası'ndaki gerçek verilerin kullanıldığı Ampirik Bayes yöntemine ayrılmıştır. Değişik katı regresyon teknikleri avantajlı ve sakıncalı taraflarıyla incelenmiş ve katı regresyon analizinin katkılarıyla daha önceden yapılmış bazı çalışmalar gözden geçirilmiştir. Sonuçlar ortaya çıkarmıştır ki daha önceki çalışmalarda dikkate alınmayan bazı dışgözlemler yanlış neticelere yol açmışlardır. İkinci kısım, mevcut heteroskedastisitiye uygun kovaryans matrisi (HUKM)tahmin edicilerinin teferruatlı ve kapsamlı bir değerlendirmesini bazı karşılaştırma kriterlerine göre yapmıştır. Zaman tarafından literatüre kazandırılan bir tahmin edici de dikkate alınmıştır. Son olarak, ampirik bir çalışma yapılmıştır. İstanbul Borsası'ndaki firmalar sektörlere sınıflandırılmış ve bunlar üzerinde finans literatüründe çok yaygın ve basit bir denklem kurularak bazı katsayı vektörü tahminleri yapılmıştır.

Anahtar Kelimeler: Heteroskedastiklik, Kırılma Noktası, En Küçük Kareler Medyanı, Dışgözlem, Katı Mesafe, Ampirik Bayes.

Acknowledgements

I would like to express my gratitude to Prof. Dr. Asad Zaman for his valuable supervision and for providing me with the necessary background in econometrics. Special thanks go to Assist. Prof. Erdem Başçı for his helps and advices throughout my Ph.D. studies. I also would like to thank professors Turan Erol, Savaş Alpay, Mehmet Caner, and Süheyla Özyıldırım for their valuable comments, and contributions as the members of the jury and Kıvılcım Metin and Kıvanç Dinçer who kindly accepted to be the members of the evaluation committee. I also thank my wife, Zeynep, and my little son Mustafa Fethi, although he has done nothing else than smiling.

Contents

1 Robust Regression Analyses with Applications

1.1 Introduction

One might expect to see some reasonable and realistic results even when some of the data points deviate from the usual assumptions of classical regression analysis, but the classical regression method is very sensitive to the outliers. Indeed, the least squares method is currently the most popular approach for estimation. There are several reasons for this, two of which are the ease of calculation and the tradition that shaped the current literature.

Real data sets containing outliers are very common situations. So many data sets contain outliers as a result of mistakes in recording or observing the data or some exceptional observations that might take place. It is possible that the estimates become totally incorrect and the outliers themselves are hidden, which means that it becomes impossible to detect the existence of the outliers for ever. To solve this problem out, robust statistical techniques (RRT) [1] have been developed. These techniques give more trustworthy results when the data are contaminated and may let us identify the outliers to some extend.

The goal of positive breakdown regression is to be robust against the possibility of one or more unannounced outliers that may be seen anywhere in the data. The outliers may be in the response variable as well as the regressors themselves. The positive breakdown regression became more popular in the eighties although there was a huge amount of previous work about the detection and the neutralization of the outliers via different methods that have their own positives and negatives.

Let's suppose that we have a simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \epsilon_i \tag{1}$$

for i=1,2, ..., n where y stands for the response variable (dependent variable) and x stands for the independent regressors (explanatory variables). β_0 denotes the constant term, or the vertical intercept. The classical theory assumes that the error term, ϵ follows a Gaussian distribution with mean 0 and variance σ^2. The main objective is to make some inferences about the vector of coefficients, β. The Ordinary Least Squares (OLS) residual

[1] From now on Robust Regression Techniques will be abbreviated by RRT.

for the i^{th} row of observations, e_i, is given by

$$e_i(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k) = y_i - (\hat{\beta}_0, \hat{\beta}_1 x_{i1}, \ldots, \hat{\beta}_k x_{ik}) \tag{2}$$

More precisely speaking, the objective of the LS method is to minimize the sum of squares of the residuals $e_i(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$. More formally,

$$minimize_{(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)} \sum_{i=1}^{n} e_i^2 \tag{3}$$

The main idea is to make all of the residuals as small as possible so that the sum of their squares should be minimized. Indeed, the observations that deviate from the bulk of the data are penalized by taking the square of the distance from the line. LS simply wants to place a line among the regression points in such a way that the cumulative squares of the distances is minimized. The main motivation behind such a preference is that the method lets one to compute the vector of coefficients directly and explicitly from the data by a simple formula.

After such an initiation Gauss was able to introduce the distribution which is world famous by his name, the Gaussian distribution, as the one for which LS is optimal. More recently, people began to realize that actual data often do not satisfy his assumptions, sometimes with dramatic deviations from them leading to some serious mistakes of the estimation procedure.

In the terminology, regression outliers are observations that do not obey the linear pattern formed by the majority of the data. It is difficult to make a good analysis of how things are shaped for robust regression because the mentioned outliers do usually affect the trend of the data in such a way that one can never be sure about the whole picture without working on the outliers. In most cases outliers are not the mistakes but they are the cases which represent the data coming from extraordinary conditions. But some recording or reading errors of the data are also possible. Regardless of the source of the outliers the conclusion is that one has to detect and work on them very carefully to make some correct inferences.

We say that an observation (x_i, y_i) is a leverage point when its regressor lie outside of the majority of the regressors. Indeed, the term leverage comes from mechanics, because such a point pulls the LS solution towards it. The LS method estimates σ from the residuals, e_i using the formula:

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} e_i^2 \tag{4}$$

where k is apparently the number of regressors. Once the estimate for variance is calculated one can obtain the standardized residuals, $e_i/\hat{\sigma}$. It is also common to calculate these values and label the observations for which this figure exceeds 2.5, or less than -2.5 as the regression outliers. The logic behind is that values generated by Gaussian Distribution are rarely larger than 2.5 or less than -2.5, whereas the other observations are considered to obey the model. In simple regression models, where the number of regressors is small, the detection of the outliers may be possible even by observing the plot of the regressors and the regressand, but in multiple regression, where k is large, the detection by eye is no longer possible and the residual plot mentioned about above become an important tool. Since most of the regressions done by the economists and even the econometricians are done routinely, many results must have been affected or even determined by the outliers and this may have remained unnoticed.

1.1.1 Breakdown Value

In any data set, one can displace the LS fit as much as he wants by simply moving a single data point (x_i, y_i) enough far away. This statement can be experimented by any statistical package by changing one of the observations. The statement is true for both single and multiple regression. On the other hand, it is possible to find some robust regression methods that can resist some of the outliers.

The breakdown value can be considered as a superficial but useful measure. The concept was first introduced by Hampel [34] and is applied to the finite sample setting by Donoho and Huber [21]. It is a rough but useful measure of robustness. Let's use the latter version. Consider a data set $Z = (x_{i1}, \ldots, x_{ik}, y_i; i = 1, \ldots, n)$ and a regression estimator ET. Applying ET to Z yields a vector $(\hat{\beta}_0, \ldots, \hat{\beta}_k)$ of regression coefficients.

Now consider all possible contaminated data sets Z' obtained by replacing any m of the original observations by arbitrary points.

This yields the maximum bias

$$maxbias(m; ET, Z) := max_{Z'}|ET(Z') - ET(Z)| \quad (5)$$

where | . | is the Euclidean norm. If m outliers can have an arbitrarily large effect on ET, it follows that $maxbias(m; ET, Z) = \infty$, hence ET(Z') becomes useless. Therefore, the breakdown value of the estimator ET at the data set Z is defined as

$$\epsilon_n^*(ET, Z) := min\{\frac{m}{n}; maxbias(m; ET, Z) = \infty\} \quad (6)$$

In other words, it is the smallest fraction of contamination that can cause the regression method ET to run away arbitrarily far from ET(Z). For many estimators $\epsilon_n^*(ET, Z)$ varies only slightly with Z and n, so that we can denote its limiting value (for $n \to \infty$) by $\epsilon^*(ET)$.

How does the notion of breakdown value fit in with the use of statistical models such as (1)? We essentially assume that the data from a mixture of which a fraction $(1 - \epsilon)$ was generated according to (1), and a fraction ϵ is arbitrary (it could even be deterministic, or generated by any distribution). In order to be able to estimate the original parameters $(\beta_0, \ldots, \beta_k)$, we need that $\epsilon < \epsilon^*(ET)$. For this reason ϵ^* is sometimes called breakdown bound.

For least squares we know that one outlier may be sufficient to destroy the regression. Its breakdown value is thus $\epsilon_n^*(ET, Z) = 1/n$ hence $\epsilon^*(ET) = 0$. The estimators where $\epsilon^*(ET) > 0$, will be called positive-breakdown methods.

1.1.2 Positive-Breakdown Regression

Let us first consider the simplest case (k=0) in which the model (1) reduces to a univariate location problem $y_i = \beta_0 + e_i$. The LS method (3) yields the sample average $ET = \hat{\beta}_0 = E_i(y_i)$, E standing for the expected value or the average, with again $\epsilon^*(ET) = 0\%$. On the other hand, it is easily verified that the sample median $ET := med_i(y_i)$ has $\epsilon^*(ET) = 50\%$, which is the highest breakdown value attainable. Because for a larger fraction of contamination, no method can distinguish between the original data and the replaced data. The further the contamination is disseminated, the worse the situation is. Estimators ET with $\epsilon^*(ET) = 50\%$, like the univariate median, will be called high-breakdown estimators.

The first high-breakdown regression method was the repeated median estimator proposed by Siegel [92] in 1982. It computes univariate medians in a hierarchical way. For simple regression, it is described in the entry Repeated Median Line Method. Its asymptotic behaviour was obtained by Hössjer et al [43], and for algorithms and numerical results see Rousseeuw et al [83, 85]. But in multiple regression where ($k \geq 2$) the repeated median estimator is not equivariant, in the sense that it does not transform properly under linear transformations of $(x_{i1}, \ldots, x_{ik})$.

However, it is possible to construct a high-breakdown method which is still equivariant. It is instructive to look at (3). This criterion should logically be called least sum of squares, but for historical reasons (Legendre's terminology) the word *sum* is rarely mentioned. Now let us replace the sum by a median. This yields the least median of squares method (LMS), defined by

$$minimize_{\hat{\beta}_0,\ldots,\hat{\beta}_k} med_i r_i^2 \tag{7}$$

[78] which has a 50% breakdown value. The LMS is clearly equivariant because (7) is based on residuals only.

Another method is the least trimmed squares method (LTS) proposed in (Rousseeuw [76, 78]). It is given by

$$minimize_{\hat{\beta}_0,\ldots,\hat{\beta}_k} \sum_{i=1}^{h} (r^2)i : n \tag{8}$$

where $(r^2)1 : n \leq (r^2)2 : n \leq, \ldots, \leq (r^2)n : n$ are the ordered squared residuals (note that the residuals are first squared and then ordered). Criterion 8 resembles that of LS but does not count the largest squared residuals, thereby allowing the LTS fit to steer clear of outliers. For the default setting $h \approx n/2$ we find $\epsilon^* = 50\%$, whereas for larger h we obtain $\epsilon^* \approx (n-h)/n$. For instance, putting $h \approx 0.75n$ yields $\epsilon^* = 25\%$, which is often sufficient. The LTS is asymptotically normal unlike the LMS, but for $n \leq 1000$ the LMS still has the better finite-sample efficiency. Here we will focus on the LMS, the LTS results being similar.

When using the LMS regression, σ can be estimated by

$$\hat{\sigma} = 1.483\left(1 + \frac{5}{n-k-1}\right)\sqrt{med_i r_i^2} \tag{9}$$

where r_i are the residuals from the LMS fit, and $1.483 = \Phi^{-1}(3/4)$ makes $\hat{\sigma}$ consistent at Gaussian error distribution. The finite-sample correction factor $\left(1 + \frac{5}{n-k-1}\right)$ was obtained from simulations. Note that the LMS scale estimate $\hat{\sigma}$ is itself highly robust. Therefore, we can identify regression outliers by their standardized LMS residuals $r_i/\hat{\sigma}$.

In regression analysis inference is very important. The LMS by itself is not suited for inference because of its low finite-sample efficiency. This can be resolved by carrying out a reweighted least squares, RLS, step. To each observation i one assigns a weight w_i based on its standardized LMS residual $r_i/\hat{\sigma}$, e.g. by putting $w_i := w(|r_i/\hat{\sigma}|)$ where w is a decreasing continuous function. A simpler way that is followed in this study many times, but still effective, is to put w_i if $|r_i/\hat{\sigma}| \leq 2.5$ and $w_i = 0$, otherwise. But simplicity brings some trouble of not qualifying the point as a good leverage one. Either way, the RLS fit $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)$ is then defined by:

$$minimize_{(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2,\ldots,\hat{\beta}_k)} \sum_{i=1}^{n} w_i r_i^2 \tag{10}$$

which can be computed quickly. The result inherits the breakdown value, but is more efficient and yields all the usual inferential output such as t-statistics, F-statistics, and R^2 statistics, and the corresponding p-values. The p-values assume that the data with $w_i = 1$ come from the model (1) whereas the data with $w_i = 0$ do not. Another approach which avoids this assumption is to bootstrap the LMS, as done by Efron and Tibshirani [23]. The LMS and the RLS are computed with the program PROGRESS by Rousseeuw and Leroy [84]. Indeed, the RLS does nothing more than running OLS over the data set avoiding the observations with 0 weights assigned by LMS.

1.1.3 Detecting Leverage Points by Eye

In the typical regression model a data point $(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)$ with outlying $x_i(x_{i1}, x_{i2}, \ldots, x_{ik})$ plays a crucial role, because a slight change of the coefficients estimated may give case i a large residual. Therefore, the LS method gives priority to approaching such a point in minimizing the objective function.

Detecting outliers in the k-dimensional data set X is not trivial. Especially where k is greater than two when we can no longer have the opportunity of inspection by eye.

A classical approach to the solution of the problem is to compute the Mahalanobis Distance defined as:

$$MD(x_i) = \sqrt{(x_i - \bar{X})(Cov(X))^{-1}(x_i - \bar{X})^t} \tag{11}$$

for each x_i. Here $\bar{X}$ is the sample mean of the data set where the $Cov(X)$ is the sample covariance matrix. This distance tells us how far away x_i from the mass of the data relative to the size of the mass is. It is well known that this approach suffers from the masking effect, by which the multiple outliers do not necessarily have a large Mahalanobis Distance.

One of the most commonly used statistic to discover the leverage points has been the diagonal entries of the hat matrix. Indeed, these entries are equivalent to the Mahalanobis Distances since,

$$h_{ii} = \frac{MD^2_{x_i}}{n-1} + \frac{1}{n} \tag{12}$$

Therefore, the diagonal entries of the hat matrix are masked when the distances are masked.

One can play with the elements in the square root formula of the (11) equation to have some more reliable diagnostics.

The Minimum Volume Ellipsoid proposed by Rousseeuw [77, 79] proposes an ellipsoid with the minimum volume to include some certain percentage of the data. One can refer to [84] in order to have some more detailed information about the technique.

Since the MVE estimator, the robust distances $RD(x_i)$ and the one-step reweighted estimates (14) depend only on the x-data, they can also be computed in data sets without a response variable y_i. This makes them equally useful to detect one or several outliers in an arbitary multivaraite data set. For some examples see [86], page 634, and [10].

The MVE and the $RD(x_i)$ can be computed by the software available from the supervisor or the author of this thesis, as well as the LTS subroutine written in Gauss.

1.1.4 Diagnostic Display

Combining the notions of regression oultliers and leverage points, we see that four types of observations may occur in regression data:

regular observations with internal $\mathbf{x_i}$ and well-fitting y_i

vertical outliers with internal $\mathbf{x_i}$ and non-fitting y_i

good leverage points with outlying $\mathbf{x_i}$ and well-fitting y_i

bad leverage points with outlying $\mathbf{x_i}$ and non-fitting y_i

In general, good leverage points are beneficial, since they can improve the precision of regression coefficients. Bad leverage points are harmful because they can change the least squares fit drastically. In the coming applications one of the best techniquies is to detect the regeression outliers with standardized LMS residuals and leverage points which are diagnozed by robust distances. Indeed, Rousseeuw and van Zomeren proposed a display which plots robust residuals versus robust distances [87] where the cutoffs at the [-2.5,2.5] band and the $\sqrt{\chi^2_{p,0.975}}$ are bordered by horizontal and vertical lines. With the help of such a display, the four types of points categorized above are determined automatically. One can play with the band length and the critical values of the $\sqrt{\chi^2_{p,.}}$ to be more robust or loose to such points of outliers.

1.1.5 Applications

Although there are some applications of positive-breakdown methods, there have been quite a few substantive applications performed where the use of LMS and/or MVE has made a difference.

The main obstacle preventing the wide, common, and frequent applications of high breakdown methods was the difficulty and slowness of computation, but the invention of powerful computers enabled such computations available. For instance, there are several intensive users of LMS in financial markets, where profits can be made by finding majority patterns and detecting subgroups that believe in another way. In management science, the LMS has been applied to measures of production efficiency by Seaver and Trinatis

[91]. The LMS regression is being used in chemistry after the publication of Massart et al [66]. Also, the LMS is an essential component of a new system for connecting optical fiber cables implemented at NIST, see Wang et al [106]. In large electric power systems, Mili et al [70] modified positive-breakdown methods to estimate the system's state variables. Faster algorithms needed to be constructed to allow real-time estimation.

Positive-breakdown methods have opened new possibilities in the rapidly evolving field of computer vision. The LMS has been used for analyzing noisy images, Meer et al [68], for interpreting color omages Drew [22], for discontinuity-preserving surface reconstruction, Sinha and Schunk [95], for extracting geometric primitives, Roth and Levine [75], Stewart [99], for robot positioning Kumar and Hanson [53], and for detecting moving objects in video from a mobile camera, Thompson et al [104], Abdel-Mottaleb et al [1]. The MVE was applied to image segmentation Jolion et al [46]. Chork [10] used the MVE to analyze data on surface rocks in New South Wales, for which concentrations of several chemical elements were measured. Outliers in this multivariate data set revealed mineralizations, yielding targets for mining prospection. A larger study in Finland carried out MVE-based factor analysis, Chork and Salminen [11]. The same methods apply to environmetrics, since mineralizations in geochemistry are similar to contaminations of the environment.

1.1.6 Other Robust Methods

The earliest systematic theory of robust regression was based on M-estimators Huber [44], [45] given by

$$minimize_{\hat{\beta}_0,\ldots,\hat{\beta}_k} \sum_{i=1}^{n} \rho(r_i/\hat{\sigma}) \tag{13}$$

where $\rho(t) = |t|$ yields L^1 regression (see Method of Least Absolute Values) as a special case. For general ρ one needs a robust $\hat{\sigma}$ to make the M-estimator equivariant under scale factors. This $\hat{\sigma}$ either needs to be fixed in advance or estimated jointly with $(\hat{\beta}_0, \ldots, \hat{\beta}_k)$, see Huber [45], page 179. Scale equivariance holds automatically for R-estimators, Jureckova [49], and L-estimators Koenker and Portnoy [52]. The breakdown value of all M-, L-, and R- estimators is 0% because of their vulnerability to bad leverage points.

Zaman [112] makes a thorough appreciation of the robust methods where he mentions the need for the consequently invented estimators, with the drawback of them in chapter

5.

The next step was the development of generalized M-estimators (GM- estimators) with the purpose of bounding the influence of outlying $(x_{i1}, \ldots, x_{ik})$ by giving them a small weight. This is why GM-estimators are often called bounded influence methods. A survey is given in Hampel et al [35]. Both M- and GM-estimators can be computed by iteratively rewighted LS or by the Newton-Raphson algorithm. Unfortunetly the breakdown value of all GM-estimators goes down to zero for increasing k, when there are more opportunities for outliers to occur.

In the special case of simple regreesion (k=1) several earlier methods exist, such as the Brown-Mood line, the robust-resistant line of Tukey, and the Theil-Sen slope. Their breakdown values are derived in Rousseeuw and Leroy ([84] Section 2.7).

For multiple regression the LMS and the LTS described above were the first equivariant methods to attain a 50% breakdown value. By choosing h in (8), any positive breakdown value between 0% and 50% can be set as well. Their low finite-sample efficiency can be improved by carrying out one-step RLS fit (10) afterwards. Another approach is to compute one-step M-estimators starting from LMS as proposed by Rousseeuw [78], which also maintains the breakdown value and yields the same asymptotic efficiency as the corresponding M-estimator. In order to combine these advantages with those of the bounds influence approach, it was later proposed to follow the LMS or LTS by a one-step GM-estimator of the Mallows type, see Simpson et al [93], the Schweppe type, see Coakley and Hettmansperger [13], or the Hill-Ryan type, see Simpson and Yohai [94]. For tests and variable selection in this context see Markatou and He [59] and Ronchetti and Staudte [74].

A different approach to improving on the efficiency of the LMS and the LTS is to replace their objective functions by a more efficient scale estimator applied to the residuals r_i. This yielded the class of S-estimators, see Rousseeuw and Yohai [89]. An S-estimator is the $(\hat{\beta}_0, \ldots, \hat{\beta}_k)$ which minimizes an M- estimator $S(r_1, \ldots, r_n)$ given by

$$\frac{1}{n}\sum_{i=1}^{n}\rho\Big(\frac{r_i}{S(r_1,\ldots,r_n}\Big) \tag{14}$$

with bounded ρ. The breakdown value of the S-estimator $(\hat{\beta}_0, \ldots, \hat{\beta}_k)$ depends on k and n, and can be as high as 50%. Altough S-estimators are not M- estimators, they happen to have the same expression for their influence function, hence they have the same asymptotic efficiency. Anologuous situations already occur in univariate location where the trimmed mean L-estimator and the Huber-type M-estimator happen to possess the same infuence function while their breakdown values are different.

Going further in this direction has led to the introduction of even more efficient positive-breakdown regression methods, including MM-estimators, Yohai [110], τ-estimators , Yohai and Zamar [111], and generalized S-estimators Croux et al [14].

Multivariate M-estimators have a relatively low breakdown value. (see Hampel et al [35], page 298). Together with the MVE estimator, Rousseeuw [77, 79] also introduced the minimum covariance determinant estimator (MCD), which looks for the h observations of which the empirical covariance matrix has the smallest possible determinant. Then T(X) is defined as the average of these h points, and C(X) is a certain multiple of their covariance matrix. The motivation for the MCD are given by Davies [18] and Butter et al [8]. S-estimators were extended to the multivariate scatter framework in (Rousseeuw and Leroy [84], Davies [17]). The breakdown value of one-step reweighted estimators (14) was obtained by Lopuhaä and Rousseeuw [55], whereas Davies [19] studied one-step estimators.

All positive-breakdown estimators, for regression as well as multivariate location and scatter, have some unconventional features that distinguish them from zero-breakdown methods (Rousseeuw [81]).

1.1.7 Maxbias Curve

There is a growing interest in the maxbias curve, which plots the worst-case bias (5) ET of an estimator as a function of the fraction $\epsilon = m/n$ of contamination. It is increasing in ϵ, and is usually drawn for the population case. The maxbias curve of an estimator was considered in Hampel et al [35], where it was mentioned that its tangent at $\epsilon = 0$ is related to the influence function of T, and that it has a vertical asymptote at $\epsilon = \epsilon^*(ET)$. Therefore, the maxbias curve measures both local robustness (breakdown value) and global robustness (breakdown value). There has been much work on finding estimators with low maxbias curve: for univariate location Huber [45], for univariate scale Martin and Zamar [65] Rousseeuw and Croux [82], and for residual-based regression Martin et al [64], where the LMS turns out to be optimal.

The maxbias curves of the LMS, S-, τ-, and generalized S- estimators were compared by Croux et al [14]. Lower bounds for maxbias curves were obtained by He and Simpson [40]. The research for multivariate scatter methods with low maxbias curve led to new types of projection estimators (Maronna et al [62]). Related projection methods for regression were proposed by Maronna and Yohai [63].

1.1.8 Algorithms

The basic resampling algorithm for approximating the LMS (described fully in Rousseeuw and Leroy [84]) considers some k+1 observations, computes the coefficients $\tilde{\beta}_1, \ldots, \tilde{\beta}_k$ that fit these points precisely, and evaluates the objective function (7) for these coefficients. This is repeated often, and the solution with lowest objective function is kept. For small data sets we can consider all subsets of k+1 observations. For larger data sets we randomly draw many (k+1)-subsets so that with high probability at least some of them are outlier-free. This algorithm can be speeded up on a parallel computer as in Kaufman et al [50]. Also the MVE can be aproximated using (k+1)-subsets (Rousseeuw and Leroy [84]). It is even possible to combine the LMS and MVE algorithms (Dallal and Rousseeuw [15], Hawkins and Simonoff [39]).

Several modified algorithms were proposed for computing these and other positive-breakdown estimators for regression or multivariate location and scatter. These developments include exact algorithms for the LMS (Souvaine and Steele [97], Stromberg [100]) as well as rough approximations (Rousseeuw and van Zomeren [87]). One can also carry out local improvements by means of M-steps (Ruppert [90]), by interchanging points (Hawkins [38]), or by sequentially adding data points (Hadi [33], Atkinson [3]). Rousseeuw [80] constructs a relatively small collection of (k+1)-subsets which is still sufficient to maintain the exact breakdown value. Finally, Woodruff and Rocke [108] incorporate simulated annealing, genetic algorithms and tabu search.

1.1.9 Other Models

Positive-breakdown regression methods such as LMS can be extended to zero-intercept model (see [84]), as well as to models with several intercepts [88]. Rousseeuw and Leroy [84], Chapter 7, applied the LMS and other high breakdown RRTs to autoregressive time series, to orthogonal regression, and to directional data. Other extensions were to nonparametric regression, nonlinear regression (Stromberg and Ruppert [102], Stromberg [101]), and logistic regression (Christmann [12]).

In multivariate analysis one can replace the classical covariance matrix by a positive-breakdown scatter matrix, e.g., for discriminant analysis, correlation matrices, principal components, and factor analysis (see e.g., Chork and Salminen [11]). More research needs to be done for these and other situations.

1.2 Application to a Growth Model

This part of the thesis includes some applications of the robust regression techniques to some data sets from previous econometric studies. The first data set is from De Long and Summers [20] . They employ the data from the United Nations Comparison Project and the Penn World Table. Their main claim is that there is a strong and clear relationship between national rates of machinery and equipment investment and productivity growth. Equipment investment has far more explanatory power for national rates of productivity growth than other components of investment, and outperforms many other variables included in cross country equations accounting for growth. Some justification of intuition is also given to persuade people that the idea is very plausable. The claim is that this association between growth and the equipment investment is causal, that is, the higher equipment investment drives faster growth, and that the social return to equipment investment in well-functioning market economies is on the order of 30 percent per year.

1.2.1 Model and the Data

A simple regression is used to support the claim. The explanatory variable is the GDP per worker growth (GRW), the regressors are the constant term (co.), labor force growth (LFG), relative GDP gap (GAP), equipment investment (EQP) and the non-equipment investment (NEQ). Cross-section data are used for 61 countries for which data are available. See Table 1, for the data set in detail. Although different time periods and sets of countries are selected for regressions -and this leads to many different regressions to be carried out- we focused on the main regression where the time period is the longest, from

I	GRW	LFG	GAP	EQP	NEQ	31	0.0535	0.0117	0.7484	0.1223	0.2464
1	0.0089	0.0118	0.6079	0.0214	0.2286	32	0.0146	0.0346	0.9415	0.0462	0.1268
2	0.0332	0.0014	0.5809	0.0991	0.1349	33	0.0479	0.0282	0.8807	0.0557	0.1842
3	0.0256	0.0061	0.4109	0.0684	0.1653	34	0.0236	0.0064	0.2863	0.0711	0.1944
4	0.0124	0.0209	0.8634	0.0167	0.1133	35	-0.0102	0.0203	0.9217	0.0219	0.0481
5	0.0676	0.0239	0.9474	0.1310	0.1490	36	0.0153	0.0226	0.9628	0.0361	0.0935
6	0.0437	0.0306	0.8498	0.0646	0.1588	37	0.0332	0.0316	0.7853	0.0446	0.1878
7	0.0458	0.0169	0.9333	0.0415	0.0885	38	0.0044	0.0184	0.9478	0.0433	0.0267
8	0.0169	0.0261	0.1783	0.0771	0.1529	39	0.0198	0.0349	0.5921	0.0273	0.1687
9	0.0021	0.0216	0.5402	0.0154	0.2846	40	0.0243	0.0281	0.8405	0.0260	0.0540
10	0.0239	0.0266	0.7695	0.0229	0.1553	41	0.0231	0.0146	0.3605	0.0778	0.1781
11	0.0121	0.0354	0.7043	0.0433	0.1067	42	-0.0047	0.0283	0.8579	0.0358	0.0842
12	0.0187	0.0115	0.4079	0.0688	0.1834	43	0.0260	0.0150	0.3755	0.0701	0.2199
13	0.0199	0.0280	0.8293	0.0321	0.1379	44	0.0295	0.0258	0.9180	0.0263	0.0880
14	0.0283	0.0274	0.8205	0.0303	0.2097	45	0.0295	0.0279	0.8015	0.0388	0.2212
15	0.0046	0.0316	0.8414	0.0223	0.0577	46	0.0261	0.0299	0.8458	0.0189	0.1011
16	0.0094	0.0206	0.9805	0.0212	0.0288	47	0.0107	0.0271	0.7406	0.0267	0.0933
17	0.0301	0.0083	0.5589	0.1206	0.2494	48	0.0179	0.0253	0.8747	0.0445	0.0974
18	0.0292	0.0089	0.4708	0.0879	0.1767	49	0.0318	0.0118	0.8033	0.0729	0.1571
19	0.0259	0.0047	0.4585	0.0890	0.1885	50	-0.0011	0.0274	0.8884	0.0193	0.0807
20	0.0446	0.0044	0.7924	0.0655	0.2245	51	0.0373	0.0069	0.6613	0.0397	0.1305
21	0.0149	0.0242	0.7885	0.0384	0.0516	52	0.0137	0.0207	0.8555	0.0138	0.1352
22	0.0148	0.0303	0.8850	0.0446	0.0954	53	0.0184	0.0276	0.9762	0.0860	0.0940
23	0.0484	0.0359	0.7471	0.0767	0.1233	54	0.0341	0.0278	0.9174	0.0395	0.1412
24	0.0115	0.0170	0.9356	0.0278	0.1448	55	0.0279	0.0256	0.7838	0.0428	0.0972
25	0.0345	0.0213	0.9243	0.0221	0.1179	56	0.0189	0.0048	0.4307	0.0694	0.1132
26	0.0288	0.0081	0.6457	0.0814	0.1879	57	0.0133	0.0189	0.0000	0.0762	0.1356
27	0.0452	0.0305	0.6816	0.1112	0.1788	58	0.0041	0.0052	0.5782	0.0155	0.1154
28	0.0362	0.0038	0.5441	0.0683	0.1790	59	0.0120	0.0378	0.4974	0.0340	0.0760
29	0.0278	0.0274	0.9207	0.0243	0.0957	60	-0.011	0.0275	0.8695	0.0702	0.2012
30	0.0055	0.0201	0.8229	0.0609	0.1455	61	0.011	0.0309	0.8875	0.0843	0.1257

Table 1: De Long and Summers growth data on 61 countries

1960 to 1985 and all 61 countries are included.

The regression is carried out by using Ordinary Least Squares method. The results obtained from the OLS regression are given in Table 2. The claim that the equipment share of investment very crucial is supported by the results of the regression. Some other points deserve attention. First of all, the coefficient of the labor force growth is negative which means as GDP per worker growth is increasing the labor force growth is decreasing. Another point is the significance of the regressors. The t-statistics are not listed in the original article. These statistics reveal that labor force growth and the non-equipment share are not significant. This fact should have been considered by the authors.

But these results are more than enough to prove that the most important driver of

var.	coef.	s.e.	t-val.	p-val.
con	-0.0143	0.0103	-1.391	0.170
LFG	-0.0298	0.1984	-0.150	0.881
GAP	0.0203	0.0092	2.208	0.031
EQP	0.2654	0.0653	4.064	0.000
NEQ	0.0624	0.0348	1.791	0.079

Table 2: De Long and Summers data set, OLS, $R^2 = 0.788$, F-val=41.6

var.	coef.	s.e.	t-val.	p-val.
con	-0.02306	0.00899	-2.56440	0.01315
LFG	0.10040	0.17215	0.58290	0.56238
GAP	0.02230	0.00797	2.78277	0.00741
EQP	0.28279	0.05595	5.05444	0.00010
NEQ	0.09147	0.03038	3.01071	0.00396

Table 3: De Long and Summers data set, RLS, $R^2 = 0.843$, F-val=57.9

growth is equipment investment and the other causes of growth are far below the equipment investment.

1.2.2 RLS and LMS

Same data and the same regression equation are used in some robust regression techniques to understand how much results obtained correct and reliable are. Several robust regression techniques are run the first of which is the Reweighted Least Squares (RLS) based on LMS. The same table is arranged for this technique also.

The RLS simply assigns some weights to the cases of regression and then reruns OLS. The weights are based on the LMS. For this regression weights assigned to the cases turned out to be all one except for two cases belonging to Cameroon and Zambia leading to an average weight of 0.97. It is obvious that the data belonging to these countries were having high standardized LMS residuals. The standardized residuals of LMS for these two countries are 2.57 and -4.65.

The consequences of delivering 0 weights to only these two countries is apparent over the table for RLS. The impact of eliminating these small number of data points is high on the regression statistics. The R^2 statistics has risen from 0.78 to 0.84. The F-statistic is also improved from 41.6 in OLS to 57.9 in RLS. Now each of the regressors but the labor force growth becomes significant, that is, the constant term and the non-equipment investment alter their significance.

The Least Trimmed Squares will also be applied to the same data as well as the Minimum Volume Ellipsois method. One basic drawback of the minimum Ellipsoid Method is about its coverage. The method is applied to the regressors only where the regressand may well be contaminated by outliers. And another prominent drawback of the method is that it just detects the outlying observations and does nothing about qualifying them as good or bad leverage points and beyond that, the method assigns weights to the cases according to the robust distances calculated. The weights just consider whether the distances exceed the corresponding χ^2 critical values. Indeed, some cases exceeding these critical values may be very precious good outliers that should never be eliminated by means of assigning zero weights.

1.2.3 Least Trimmed Squares

Yet another prominent robust regression technique applied is the Least Trimmed Squares. A Gauss program is written to perform this technique. One important question about the application of the technique is to decide which percentage of the data to trim. Indeed, a parameter is assigned to this percentage in the program. Although so many percentages are tried several of them are chosen to be reported. Tables 4, and 5 are arranged to display the summary statistics for the LTS where the trimmed percentages begin from 5 percent and goes until 20 percent with equal increments in percentage.

Table 4 suggests that all the regressors but the labor force growth are significant, and the t-statistic for this variable is not as small as the one obtained by OLS. There are some

var.	coef.	s.e.	t-val.	p-val.
con	-0.0261	0.0087	-2.991	0.004
LFG	0.1552	0.1671	0.929	0.357
GAP	0.0238	0.0077	3.096	0.003
EQP	0.3111	0.0551	5.646	0.000
NEQ	0.0895	0.0292	3.063	0.003

Table 4: De Long and Summers data set, LTS, 5 % trim, $R^2 = 0.518$, F-val=14.2

improvements for the significance of all of the regressors. The coefficient for the labor force growth again turns out to be positive.

The trimmed percentage is increased to 10, 15, and 20 percents to note the additional effects of eliminating some more of the data provided that the objective of the LTS is satisfied.

When the trimmed percentage insreases, there arises a trade-off in between the data lost by being trimmed and the sum of squares of OLS residuals of the remaining data. Here by incresing the trimmed data percentage from 5 % to 10 % some more cases are lost and there is some more improvement for the R^2 and the F-statistics of regression. Some more significance for all of the regressors is achieved as well.

One important result of applying the the LTS algorithm is that we are now able to see the labor force growth among the significant regressors. We were not able to observe this as a consequence of the robust regression techniques we have been trying so far. One can comment on the sign of this regressor depending upon which portion of the business cycle and the marginal productivity of labor the economy is.

1.2.4 Minimum Volume Ellipsoid Method on Growth Data

The last technique to be discussed about is the Minimum Volume Ellipsoid Technique. The software using the technique is fed by the regressors of a data set and then determines the cases to be included and excluded. One important point that deserves attention is that

var.	coef.	s.e.	t-val.	p-val.
con	-0.0297	0.0079	-3.781	0.000
LFG	0.2455	0.1519	1.616	0.112
GAP	0.0255	0.0068	3.728	0.000
EQP	0.3303	0.0491	6.726	0.000
NEQ	0.0887	0.0260	3.414	0.001
LTS10%	R^2 :	0.604	F-val.:	19.05
con	-0.0326	0.0070	-4.645	0.000
LFG	0.2091	0.1326	1.576	0.122
GAP	0.0306	0.0060	5.081	0.000
EQP	0.3808	0.0453	8.407	0.000
NEQ	0.0787	0.0232	3.391	0.001
LTS15%	R^2 :	0.690	F-val.:	26.16
con	-0.0358	0.0063	-5.718	0.000
LFG	0.2973	0.1194	2.489	0.017
GAP	0.0293	0.0054	5.438	0.000
EQP	0.3636	0.0429	8.468	0.000
NEQ	0.1037	0.0229	4.530	0.000
LTS20%	R^2 :	0.757	F-val.:	34.3

Table 5: De Long and Summers data set, LTS, 10, 15, and 20 % trims

the points excluded outside of the ellipsoid may be good or bad leverage points and if the good leverage points are let out, this is a very big loss for the success of the regression.

The software used may try all possible combinations to fix the minimum volume ellipsoid or it may select a huge number of combinations. The first one is complete enumeration and results in the best possible performance of the minimum volume of the ellipsoid, but this requires so much computation time and a powerful computer. On the other hand, the compensation of the huge number of combinations is at its saving for computation time.

1.3 Detection of Good and Bad Outliers

Good leverage points are very precious since they manipulate the regression line towards where it has to, but bad ones are at least as bad to compensate the advantages of the good leverage points. This fact makes the detection of the characteristics of the data points extremely important. Two main statistics play crucial roles in analyzing these points. The main purpose of detecting the outliers is not to eliminate them. Some RRTs eliminate some of them and some of them delete all of the outliers regardless of whether they are useful or harmful for the regression.

The two criteria we will follow heavily depends on the robust distances and the standardized residuals. The standardized residuals are LMS residuals divided by their standard errors. Since these are supposed to follow the Gaussian Distribution they will hardly be out of the [-2.5,2.5] tolerance band. So we suspect the cases that are outside this band. The second criterion is the robust distances of the cases. Each robust distance should be less than the $\chi^2_{k,.}$ where k is the number of regressors the and percentages may replace (.) depending on the sensitivity of the researcher. If the robust distance exceeds this critical value and the standardized residual is out of the band then the case is marked as a bad leverage point. But if it stays in the tolerance band while the χ^2 critical value is exceeded then that particular point proves to be a good leverage point. Now we are going to analyze the data sets of the examples with the above method.

The MVE subroutine prepared by Rousseeuw and Leroy, simply checks for the robust distances calculated only and then assigns a weight of 0 or 1 depending on the size of the robust distance. But this approach suffers from not taking the regressand into account, since the program does not even require the input of the regressand.

var.	coef.	s.e.	t-val.	p-val.
con	-0.222	0.0934	-2.379	0.021
LFG	0.045	0.1771	0.253	0.802
GAP	0.025	0.0082	2.982	0.004
EQP	0.283	0.0581	4.859	0.000
NEQ	0.085	0.0314	2.703	0.009

Table 6: Regression statistics for De Long and Summers data which considers both robust distances and the standardized residuals, R^2 is 0.84 and F-val is 55.8

Finally, a new method that checks for both the standardized LMS residual and the robust distance at the same time is applied to the data. Only one of the 13 points removed by MVE subroutine is qualified as a bas leverage point and is removed. The procedure expalined above is applied and the only such country to be removed is found to be Zambia. The results of regression when only this country removed is displayed in Table 6. Notice that there arises some alterations in the significance of the constant term and the non-equipment investment term, the t-values improve from -1.39 to -2.38 for the constant and from 1.79 to 2.70 for the non-equipment term. GAP is also rescued from being borderline significant (t-value is 2.21) to significant (now t-value is 2.98) by using the new method instead of OLS.

Table 7 lists the standardized residuals by OLS and LMS, and also the robust distances of the MINVOL. Note that there are some differences between the OLS and the LMS standardized residuals.

1.4 Gray's Data Set on Aircrafts

Gray [31] had made a regression to figure out the cost of building the airctafts from 1947 to 1969. The response variable is the cost whereas the regressors are the weight of the plane, maximal thrust, lift-to-drug ratio and the aspect ratio. The data set contains 23 years' data, and no constant is included as a regressor. The same procedure is followed.

I	OLS	LMS	RD	W	31	0.40	-0.29	4.68	0
1	-0.67	-1.16	3.78	0	32	-0.71	-0.92	2.63	1
2	0.08	0.22	1.40	1	33	1.45	1.24	2.50	1
3	0.25	0.26	3.61	0	34	0.10	-0.09	4.66	1
4	-0.13	0.05	1.40	1	35	-1.74	-1.40	1.22	1
5	1.48	1.37	7.30	0	36	-0.36	-0.16	1.42	1
6	1.12	0.98	2.69	1	37	0.69	0.35	1.31	1
7	1.93	2.52	1.19	1	38	-1.00	-0.45	1.40	1
8	-0.12	-0.3	5.05	0	39	0.41	0.18	2.50	1
9	-1.21	-2.2	4.06	0	40	0.93	1.52	1.15	1
10	0.58	0.54	1.29	1	41	-0.09	-0.32	3.21	1
11	-0.38	-0.4	1.40	1	42	-1.66	-1.64	0.78	1
12	-0.35	-0.6	3.23	1	43	0.06	-0.39	3.22	1
13	0.09	0.03	0.58	1	44	1.03	1.45	0.81	1
14	0.43	0.00	1.23	1	45	0.33	-0.21	1.40	1
15	-0.51	-0.17	1.24	1	46	0.98	1.29	0.97	1
16	-0.23	0.44	1.33	1	47	-0.16	0.06	1.63	1
17	-1.09	-1.94	2.84	1	48	-0.20	-0.05	1.05	1
18	-0.01	-0.19	1.93	1	49	0.08	0.00	1.65	1
19	-0.33	-0.58	2.18	1	50	-1.08	-0.92	0.92	1
20	0.89	0.60	2.10	1	51	1.51	1.94	2.77	1
21	0.04	0.55	1.14	1	52	-0.06	0.00	1.61	1
22	-0.44	-0.36	1.53	1	53	-1.14	-1.22	4.74	0
23	1.58	1.65	2.87	1	54	0.87	0.88	1.67	1
24	-0.69	-0.78	1.37	1	55	0.74	1.04	0.67	1
25	1.34	1.67	1.01	1	56	-0.07	0.23	3.70	0
26	-0.23	-0.51	1.21	1	57	-0.04	0.00	7.38	0
27	0.45	0.02	4.20	0	58	-0.34	0.00	5.27	0
28	0.79	0.80	2.39	1	59	0.27	0.56	3.81	0
29	0.91	1.24	0.84	1	60	-3.42	-4.43	3.32	1
30	-1.65	-1.96	1.40	1	61	-1.68	-2.06	1.40	1

Table 7: Minimum Volume Ellipsoid, standardized residuals by OLS and LMS, and weights assigned by the MVE subroutine

We first run OLS and then we will keep on applying the RRTs starting by RLS, ans LTS. We will finalize this section also by the joint consideration of the robust distances and the standardized LMS residuals. Data are provided in Table 8.

The results of the OLS regression are given in the following Table 9. The coefficient of determination is very high and this seems to be a good way of explaining the cost in terms of the regressors. All the regressors prove to be significant, and the F-statistics for regression is so high to claim that all coefficients of the regression are far from being equal to 0 simultaneously.

There may be some outliers that cannot be detected by the OLS. The residuals over scale with respect to the OLS are searched and it is detected that there are no outliers. That is, all the standardized residuals reside in the band covered by 2.5 standard deviations around 0. Table 10 is prepared to display the standardized residuals by OLS and LMS for the current data. Note that all OLS standardized residuals stay in the [-2.5,2.5] tolerance band, leading to the idea that the data contains no outliers. According to the LMS standardized residuals there are three cases which are not covered by the band for 1960, 1962, and 1968. OLS is unaware of this and does not consider these cases as outliers. All cases are in the band as long as OLS regression is used.

1.4.1 LMS and RLS Based on LMS

Note that the t-values of the reweighted least squares show that some of the regressors proved to be significant according to the LS now loose their significance. There are some substantial changes of the regression statistics compared to the LS regression. The main difference is from the detection of the leverage points. Although some improvement is seen in the coefficient of determination, R^2, the F-statistics for the regression is subject to some smaller values. The coefficients are having substantial changes according to the comparison of the two Tables 9 and 11.

Year	Cost	Asp.	Lift.	Weight	Thrust
1947	2.76	6.3	1.7	8176	4500
1948	4.76	6.0	1.9	6699	3120
1949	8.75	5.9	1.5	9663	6300
1950	7.78	3.0	1.2	12837	9800
1951	6.18	5.0	1.8	10205	4900
1952	9.50	6.3	2.0	14890	6500
1953	5.14	5.6	1.6	13836	8920
1954	4.76	3.6	1.2	11628	14500
1955	16.70	2.0	1.4	15225	14800
1956	27.68	2.9	2.3	18691	10900
1957	26.64	2.2	1.9	19350	16000
1958	13.71	3.9	2.6	20638	16000
1959	12.31	4.5	2.0	12843	7800
1960	15.73	4.3	9.7	13384	17900
1961	13.59	4.0	2.9	13307	10500
1962	51.90	3.2	4.3	29855	24500
1963	20.78	4.3	4.3	29277	30000
1964	29.82	2.4	2.6	24651	24500
1965	32.78	2.8	3.7	28539	34000
1966	10.12	3.9	3.3	8085	8160
1967	27.84	2.8	3.9	30328	35800
1968	107.10	1.6	4.1	46172	37000
1969	11.19	3.4	2.5	17836	19600

Table 8: Gray's data

var.	coef.	s.e.	t-val.	p-val.
Asp.	-4.442	0.7780	-5.710	0.00002
Lift	2.482	1.1595	2.140	0.04552
Weight	0.003	0.0005	7.666	0.00000
Thrust	-0.002	0.0005	-4.119	0.00058

Table 9: Gray's Aircraft Data, OLS, $R^2 = 0.937$, F-val=70.97

Year	OLS	LMS	1958	-1.81	-0.44
1947	0.88	0.00	1959	-0.18	0.41
1948	1.19	0.25	1960	0.01	-2.94
1949	1.27	0.94	1961	-0.11	0.14
1950	-0.82	0.00	1962	0.13	3.85
1951	-0.19	-0.17	1963	-1.50	-0.30
1952	-0.72	-0.11	1964	-0.30	2.03
1953	-0.49	-0.29	1965	0.61	2.26
1954	0.78	0.41	1966	0.92	0.00
1955	-0.13	1.20	1967	-0.37	1.18
1956	-0.97	1.60	1968	2.21	11.04
1957	-0.40	1.94	1969	-0.31	0.00

Table 10: Gray's Aircraft Data, OLS and LMS standardized residuals

var.	coef.	s.e.	t-val.	p-val.
Aspect	-1.99517	0.67576	-2.952	0.0094
Lift.	2.40440	1.86995	1.286	0.2168
Weight	0.00160	0.00040	4.019	0.0010
Thrust	-0.00065	0.00036	-1.817	0.0880

Table 11: Gray's Aircraft Data, RLS, $R^2 = 0.942$, F-val=64.94

var.	coef.	s.e.	t-val.	p-val.
Aspect	-2.61821	0.64751	-4.04353	0.00076
Lift.	1.90141	0.79402	2.39467	0.02773
Weight	0.00223	0.00040	5.63467	0.00002
Thrust	-0.00106	0.00038	-2.82496	0.01122

Table 12: Gray's Aircraft Data, LTS 5 percent trim, $R^2 = 0.938$, F-val=67.98

1.4.2 The LTS on Gray's Aircraft Data

First around 5 percent of the data are eliminated and the OLS regression is run with the remaining data. The coefficient of determination and the F-statistics are close to the ones obtained by OLS, but there are some crucial changes in the coefficients. See Table 12.

When the trimmed percentage increases to 10, both statistics for regression say that this is a more successful regression than the OLS. Indeed, one needs a bencmark to show that the robust regression technique is doing better than OLS and the only two such criteria that we are using are the coefficient of determination, R^2 and the F-statistics both of which reveal that now, the LTS is doing better than OLS when some of the cases are eliminated.

The improvement still continues when 20 percent of the data are removed. Note that the statistics belonging to the robust regression techniques are close to each other and

var.	coef.	s.e.	t-val.	p-val.
Aspect	-1.88404	0.550351	-3.42337	0.00348
Lift.	1.35277	0.62742	2.15609	0.04664
Weight	0.00170	0.00035	4.91687	0.00015
Thrust	-0.00058	0.00032	-1.83241	0.08557

Table 13: Gray's Aircraft Data, LTS 10 percent trim, $R^2 = 0.951$, F-val=77.41

var.	coef.	s.e.	t-val.	p-val.
Aspect	-1.96283	0.45502	-4.31377	0.00084
Lift.	1.32201	0.50754	2.60476	0.02181
Weight	0.00199	0.00029	6.84605	0.00001
Thrust	-0.00081	0.00026	-3.05932	0.00913

Table 14: Gray's Aircraft Data, LTS 20 percent trim, $R^2 = 0.973$, F-val=118.22

substantially different than the ones by OLS.

1.4.3 Minimum Volume Ellipsoid Method Applied

The subroutine is run to obtain the Minimum Volume Ellipsoid and thereby the outlying cases. The subroutine itself assigns zero weights to two cases but these may be good or bad outliers. To detect whether they are harmful or useful to the appropriateness of the regression, the robust distances are considered over the standardized residuals. The band for the standardized residuals was already determined as [-2.5,2.5] and the 0.975 percent critical value for the χ^2 distribution is 11.14. So the robust distance calculated must be more than 3.34 and the standardized LMS residual should be less than -2.5 or higher than 2.5 for the point to be regarded as a bad leverage point. These, and may be the ones near the boundary, must be eliminated from the data set. Although 1960, 1966, and 1968 have their robust distances greater than 3.34, 1960, and 1968 have their standardized residuals

var.	coef.	s.e.	t-val.	p-val.
Aspect	-2.921	0.688	-4.248	0.001
Lift.	4.190	2.060	2.034	0.058
Weight	0.002	0.000	4.602	0.000
Thrust	-0.001	0.000	-2.917	0.010

Table 15: Gray's Aircraft Data, Outliers removed using standardized residuals and robust distances together, $R^2 = 0.941$, F-val=67.96

outside the tolerance band. So they are eliminated. Then OLS is run over the remaining data and the results in Table 15 are obtained.

Note that there is a slight improvement in the fit of the regression line according to the coefficient of determination, and the coefficients are subject to changes.

1.5 Augmented Solow Model

Nonneman and Vanhoudt [72] introduces human capital to the Mankiw, Romer, and Weil's 1992 study [60] on augmented Solow model. The augmented Solow Model suggests

$$ln(Y_t/Y_0) = \beta_0 + \beta_1 ln(Y_0) + \beta_2 ln(S_k) + \beta_3 ln(N) \qquad (15)$$

where Y is real GDP per capita of working age, S_k is average annual ratio of domestic investment to real GDP, and N is annual population growth,n, plus 5 percent.

Nonneman and Vanhoudt uses the data in Table 16 all throughout their paper. Sometimes they are changing the regression equation and sometimes they are playing with the regressors included but the data set does not change. Their main objective is to apply the augmented Solow model introduced by Mankiw, Romer and Weil to the OECD countries in a better way.

	I	Y_{85}	Y_{60}	S_k	S_h	S_τ	n
Canada	1	23060	12361	0.2542	0.106	0.0125	0.0197
USA	2	25014	16364	0.2397	0.119	0.0255	0.0154
Japan	3	17669	4648	0.3658	0.109	0.0240	0.0124
Austria	4	16646	7827	0.2828	0.080	0.0110	0.0036
Belgium	5	16876	8609	0.2645	0.093	0.0140	0.0045
Denmark	6	19406	10515	0.2915	0.107	0.0110	0.0058
Finland	7	17776	8630	0.3852	0.115	0.0120	0.0076
France	8	18546	9650	0.2972	0.089	0.0205	0.0099
Germany	9	17969	9819	0.3095	0.084	0.0245	0.0050
Greece	10	9492	3164	0.2885	0.079	0.0020	0.0070
Ireland	11	12054	5454	0.2877	0.114	0.0080	0.0105
Italy	12	16055	7086	0.3139	0.071	0.0095	0.0064
Netherlands	13	16937	10008	0.2789	0.107	0.0205	0.0138
Norway	14	22107	8977	0.3494	0.010	0.0145	0.0068
Portugal	15	7925	2965	0.2608	0.058	0.0035	0.0060
Spain	16	11876	4916	0.2817	0.080	0.0045	0.0090
Sweden	17	20826	11364	0.2636	0.079	0.0225	0.0031
Switzerland	18	22428	14532	0.3142	0.048	0.0230	0.0084
Turkey	19	51500	2884	0.2323	0.055	0.0020	0.0271
UK	20	17034	10004	0.2067	0.089	0.0225	0.0033
Australia	21	20617	12824	0.3128	0.098	0.0105	0.0200
New Zealand	22	17319	13569	0.2680	0.119	0.0095	0.0170

Table 16: Nonneman and Vanhoudt data on Augmented Solow model

var.	coef.	s.e.	t-val.	p-val.
Const.	2.9759	1.0205	2.913	0.009
$ln(Y_0)$	-0.3429	0.0565	-6.070	0.000
$ln(S_k)$	0.6501	0.2020	3.218	0.005
$ln(N)$	-0.5730	0.2904	-1.973	0.064

Table 17: Augmented Solow Model, OLS, $R^2 = 0.746$, F-val=17.7

1.5.1 OLS and LMS

If there is no robust regression technique applied, the OLS regression gives the coefficients and regression results tabled in 17

Table 18 orders the OLS and the LMS standardized residuals, as well as the weights assigned by the MVE method. These weights are taken into consideration and the cases penalized by zero weights are eliminated from the data set, and OLS is run over the remaining ones to see the results obtained by the MVE algorithm. These weights by the MVE algorithm are assigned according to the robust distances (RD) of the corresponding cases. Note that cases 1 and 19 are assigned 0 weights and these are the cases with the highest robust distances. Indeed, the MVE just checks whether the RDs are exceeding the critical χ^2 values or not, and the cases exceeding the critical values are addressed as the bad leverage points and assigned zero weights. It is no coincidence that these are the cases with maximum robust distances.

RLS based on LMS leads to the coefficient of determination equal to 0.971 which is higher than 0.746 of OLS, there is some improvement in terms of the F-statistic of regression also. There are some considerable changes in the coefficients of the regressors as well. One of the main differences between the RLS and the OLS is the significance of population growth in the regression equation. The neoclassical theory of growth claims that the growth rate of population is effective in determining the GDP growth and so does the enogeneous growth theory in the short run, so the result obtained by RLS is more plausible. The cases deleted by the LMS algorithm are 1, 2, 3, 14, and 19. See Table 19.

Country	OLS	LMS	RD	W	Country	OLS	LMS	RD	W
Canada	1.82	3.05	0.902	0	USA	1.06	3.22	3.832	1
Japan	2.40	4.85	1.920	1	Austria	-0.02	0.00	0.749	1
Belgium	0.01	0.00	0.615	1	Denmark	-0.30	0.38	0.496	1
Finland	-1.21	0.00	2.396	1	France	-0.01	1.15	0.902	1
Germany	-0.9	-0.48	0.969	1	Greece	0.40	-0.44	0.901	1
Ireland	-0.23	-0.23	0.999	1	Italy	-0.09	0.44	0.902	1
Netherlands	-0.29	0.81	2.151	1	Norway	0.65	2.51	1.666	1
Portugal	-0.22	-2.06	0.902	1	Spain	0.17	0.00	0.749	1
Sweden	0.12	0.49	0.902	1	Switzerland	-0.98	0.41	0.913	1
Turkey	-1.38	-2.82	5.734	0	UK	0.45	-0.25	2.494	1
Australia	-0.20	2.34	3.064	1	New Zealand	-1.22	-0.08	3.284	1

Table 18: Augmented Solow Model, OLS and LMS standardized residuals, and weights assigned by MVE

var.	coef.	s.e.	t-val.	p-val.
Const.	2.4949	0.5193	4.805	0.001
$ln(Y_0)$	-0.4437	0.0260	-17.054	0.000
$ln(S_k)$	0.3208	0.0807	3.975	0.002
$ln(N)$	-0.9037	0.1575	-5.737	0.000

Table 19: Augmented Solow Model, RLS, $R^2 = 0.971$, F-val=122.6

The LTS subroutine is run over the same data set. 5, 10, 15, and 20 percent trim's regeression results are tabled in the same fashion. The only country deleted for 5 percent trim is Canada, 10 percent trim only adds Japan, and 15 percent trim adds USA to the deleted countries. Finally 20 percent trim deleted Australia additionally. Table 20 lists all results for the LTS. Notice that the coefficient of variation is increasing as more data points are deleted. The more data points eliminated according the LTS objective, the more successful the fit is, finally 20 percent fit leads to much better results than OLS.

The Minimum Volume Ellipsoid tried all combinations possible and found out that Canada, and Turkey should be eliminated to have a better regression. The results obtained here are similar to the ones obtained by other robust regression techniques. The most stimulating drawback of MVE seems to be its rejecting the significance of population growth. Refer to Table 21 for regression results.

The robust distances of MVE and the standardized LMS residuals are simultaneously considered to identify the bad leverage point. The only such point is from Turkey. Removing Turkey's data gives the tabled results in Table 22. Notice that the population growth turned out to be insignificant again.

1.6 Benderly and Zwick's Return Data

J. Benderly, and B. Zwick's data set from AER [4] aims to explain the return on common stocks by output growth and inflation over 1954-1981 period. Indeed, they would like to make some contribution to the original article by Fama [26] on the significance of inflation to determine the real stock returns. The regression equation concentrated on is

$$R_t = \beta_0 + \beta_1 G_t + \beta_2 I_t \tag{16}$$

where R is the real stock return, G is the output growth in percentage, and I stands for inflation in percentage again. Here R is measured using Ibbotson-Sinquefeld data base, G is measured by real GDP, and P is measured by the deflator for personal consumption expenditures.

The data are given in Table 23

var.	coef.	s.e.	t-val.	p-val.
Const.	1.8730	0.8506	2.202	0.042
$ln(Y_0)$	-0.3010	0.0454	-6.632	0.000
$ln(S_k)$	0.3955	0.1721	2.298	0.035
$ln(N)$	-0.7108	0.2287	-3.108	0.006
LTS5	R^2 :	0.780	F-val.:	20.1
Const.	1.5979	0.6933	2.305	0.035
$ln(Y_0)$	-0.3255	0.0375	-8.679	0.000
$ln(S_k)$	0.4568	0.1105	3.250	0.005
$ln(N)$	-0.9093	0.1953	-4.656	0.000
LTS10	R^2 :	0.864	F-val.:	34.0
Const.	1.7573	0.5951	2.953	0.010
$ln(Y_0)$	-0.3624	0.0349	-10.370	0.000
$ln(S_k)$	0.5670	0.1271	4.462	0.000
$ln(N)$	-1.0154	0.1715	-5.919	0.000
LTS15	R^2 :	0.900	F-val.:	44.9
Const.	1.4047	0.5764	2.437	0.029
$ln(Y_0)$	-0.3822	0.0337	-11.338	0.000
$ln(S_k)$	0.5337	0.1180	4.520	0.000
$ln(N)$	-1.1850	0.1804	-6.571	0.000
LTS20	R^2 :	0.916	F-val.:	51.1

Table 20: Augmented Solow Model, LTS 5, 10, 15 and 20 percent

var.	coef.	s.e.	t-val.	p-val.
Const.	4.2295	1.1941	3.542	0.003
$ln(Y_0)$	-0.4234	0.0568	-7.457	0.000
$ln(S_k)$	0.5746	0.1855	3.099	0.007
$ln(N)$	-0.3530	0.3428	-1.030	0.318

Table 21: Augmented Solow Model, MVE, $R^2 = 0.838$, F-val=27.7

var.	coef.	s.e.	t-val.	p-val.
Const.	4.8712	1.2027	4.050	0.001
$ln(Y_0)$	-0.4222	0.0601	-7.026	0.000
$ln(S_k)$	0.4968	0.1906	2.607	0.018
$ln(N)$	-0.0921	0.3266	-0.282	0.781

Table 22: Augmented Solow Model, both robust distances and standardized LMS residuals are considered, $R^2 = 0.809$, F-val=23.9

Year	R	G	I	Year	R	G	I
1954	53.0	6.7	-0.4	1955	31.2	2.1	0.4
1956	3.7	1.8	2.9	1957	-13.8	-0.4	3.0
1958	41.7	6.0	1.7	1959	10.5	2.1	1.5
1960	-1.3	2.6	1.8	1961	26.1	5.8	0.8
1962	-10.5	4.0	1.8	1963	21.2	5.3	1.6
1964	15.5	6.0	1.0	1965	10.2	6.0	2.3
1966	-13.3	2.7	3.2	1967	21.3	4.6	2.7
1968	6.8	2.8	4.3	1969	-13.5	-0.2	5.0
1970	-0.4	3.4	4.4	1971	10.5	5.7	3.8
1972	15.4	5.8	3.6	1973	-22.6	-0.6	7.9
1974	-37.3	-1.2	10.8	1975	31.2	5.4	6.0
1976	19.1	5.5	4.7	1977	-13.1	5.0	5.9
1978	-1.3	2.8	7.9	1979	8.6	-0.3	9.8
1980	22.2	2.6	10.2	1981	-12.2	-1.9	7.3

Table 23: Data set of Benderly and Zwick

Applying OLS to the data gives the initial regression results in Table 24. These OLS regression results are parallel to the ones reported in the original paper. The small R^2 value compared to that of Gray's data suggests a less successful regression by OLS. Both t-values and the p-values prove that the only significant regressor according to OLS is the percentage growth of the coming year whereas both the constant term and the inflation term are very obviously insignificant.

A glance at Table 25 for the standardized residuals of OLS reveal that no such residual falls outside the tolerance band. So all of the cases seem to be obeying the general trend of regression, or otherwise the masking effect is influencing the general trend of the data and the outliers are so pulling the regression line towards themselves that they all look

var.	coef.	s.e.	t-val.	p-val.
Const.	-3.586	8.581	-0.418	0.679
Growth	4.778	1.368	3.492	0.002
Inflation	-1.046	1.145	-0.913	0.370

Table 24: Return on stocks, OLS and RLS, $R^2 = 0.558$, F-val=10.5

close to that line. Although the LMS standardized residuals also show no standardized residuals outside the same band there are two such residuals outside the tolerance band of two standard deviations from 0, belonging to years 1979 and 1980. Note that the OLS residuals for the above two years are among the maxima of them. One can mark these as the points outside the band provided that the there will be a narrower band of length 4 instead of five.

RLS uses the consequences of LMS and is designed to remove the data for years falling outside the wider band but since there is no such year detected, RLS gives the same results as OLS and therefore is not reported.

LTS subroutine is run to make the mandatory deletes from the original data set to minimize the sum of squared residuals of the remaining data set. Instead of dealing with the percentages, some certain number of observations are deleted this time, namely 1, 2, 3, and 4 of the data set are removed consequently and the results are displayed in Table 26. These are denoted by LTS1, LTS2, and etc.

Indeed, one can suspect about the existence of outliers in any data set, and the detection of such points from the whole set of points is another problem. Different RRT's may suggest different points as outliers and attempts to remove them. LMS uses the minimization of the median of squares as the criterion and has no obligation to blame some of the points as outliers. The same is true for the MVE also, but LTS can determine any number of points as the ones that adds more than the others to minimize the sum of squared residulas, therefore there is no limitation to the number of observations to be deleted from the initial set of observations. One can even run the program to mark three fourths of the initial data for this purpose. The superiority of LTS comes from the lower

Year	OLS	LMS	Year	OLS	LMS
1954	1.61	1.35	1968	0.10	0.47
1955	1.68	1.52	1969	-0.25	0.25
1956	0.11	0.37	1970	-0.56	-0.11
1957	-0.34	0.00	1971	-0.61	-0.23
1958	1.22	1.20	1972	-0.33	0.00
1959	0.37	0.48	1973	-0.53	0.26
1960	-0.55	-0.31	1974	-1.11	0.00
1961	0.19	0.22	1975	1.02	1.39
1962	-1.61	-1.26	1976	0.09	0.46
1963	0.08	0.19	1977	-1.81	-1.10
1964	-0.57	-0.43	1978	-0.19	0.52
1965	-0.83	-0.55	1979	1.59	2.27
1966	-1.28	-0.84	1980	1.60	2.29
1967	0.38	0.56	1981	0.54	1.16

Table 25: The standardized residuals of OLS and LMS for the return on stocks

var.	coef.	s.e.	t-val.	p-val.
Const.	-6.80	8.24	-0.82	0.418
Growth	5.44	1.33	4.09	0.000
Inflation	-0.52	1.11	-0.47	0.643
LTS1	R^2 :	0.553	F-val.:	14.8
Const.	-14.08	8.30	-1.70	0.103
Growth	6.25	1.28	4.87	0.000
Inflation	0.35	1.10	0.32	0.754
LTS2	R^2 :	0.613	F-val.:	18.2
Const.	5.84	7.33	0.80	0.435
Growth	4.22	1.13	3.75	0.001
Inflation	-3.30	1.08	-3.04	0.006
LTS3	R^2 :	0.702	F-val.:	25.9
Const.	10.11	7.03	1.44	0.166
Growth	3.46	1.09	3.16	0.005
Inflation	-4.04	1.05	-3.83	0.001
LTS4	R^2 :	0.744	F-val.:	30.5

Table 26: Return on stocks data, LTS, up to 4 years data are removed one by one

RRT	Year
LTS1	1977
LTS2	1955,1977
LTS3	1962,1979,1980
LTS4	1962,1975,1979,1980
MVE	1974,1979

Table 27: Suggested cases to be deleted by MVE and LTS

var.	coef.	s.e.	t-val.	p-val.
Const.	-3.35	8.26	-0.41	0.689
Growth	4.79	1.33	3.59	0.001
Inflation	-1.20	1.22	-0.98	0.336

Table 28: Return on stock's data, MVE, $R^2 = 0.576$, F-val=10.4

breakdown value it has.

To make comparison of the data points, the deleted observations by LTS and MVE are listed in Table 27. Note that the years with the highest LMS standardized residuals belong to 1979 and 1980. These two years are detected by LTS3.

Notice that year 1977, suggested to be deleted by LTS2, is not included in the delete list of LTS3, and LTS4. Such things may happen theoretically since the selection by both LTS2 and LTS3-LTS4 can be correct as long as the objective of the least trimmed squares technique is concerned. But we use LTS as a RRT technique and it does not sound very plausible to accept or reject case year 1977 as an outlier by the same technique.

Since MVE does not even require the response variable as an input, it lacks a complete analysis of robust regression. MVE only determines the cases far away from the bulk of the regressors. This is the main reason behind MVE's selecting 1974 as a special selection for itself and this year is not detected by the other LTSs. Table 28 is the regression results of OLS after removing the years with zero weight assigned by MVE. The other year assigned

Year	W	MD	MD2	RD	RD2	Year	W	MD	MD2	RD	RD2
1954	1	5.10	26.04	0.96	0.92	1968	1	0.24	0.06	0.84	0.70
1955	1	1.88	3.52	1.37	1.87	1969	1	1.45	2.11	1.79	3.20
1956	1	1.05	1.11	0.96	0.92	1970	1	0.26	0.07	0.77	0.59
1957	1	2.00	4.02	1.71	2.94	1971	1	1.17	1.36	0.73	0.53
1958	1	1.23	1.51	0.76	0.58	1972	1	1.18	1.40	0.71	0.50
1959	1	1.45	2.10	1.06	1.12	1973	1	1.58	2.51	2.58	6.65
1960	1	1.17	1.36	0.85	0.72	1974	0	2.31	5.35	3.61	13.05
1961	1	1.32	1.75	0.96	0.92	1975	1	1.59	2.52	1.42	2.02
1962	1	0.88	0.77	0.54	0.30	1976	1	1.26	1.58	0.96	0.92
1963	1	1.05	1.11	0.64	0.41	1977	1	1.38	1.90	1.34	1.79
1964	1	1.33	1.77	0.94	0.88	1978	1	1.39	1.93	2.10	4.42
1965	1	1.19	1.42	0.67	0.45	1979	0	1.93	3.72	3.13	9.82
1966	1	0.60	0.36	0.68	0.46	1980	1	2.26	5.12	3.01	9.04
1967	1	0.69	0.48	0.33	0.11	1981	1	2.02	4.10	2.73	7.45

Table 29: Return on stock's data, Robust and Mahalanobis distances, and the weights assigned by MVE

0 weight is 1979.

The essence of MVE is the original robust distance calculation it presents. These distances are able to detect the outliers much better than the Mahalanobis distances as Table 29 suggests. The table is directly copied from the MVE subroutine output.

If one considers a narrower strip for the limitation to the standardized LMS residual, that is, if a band of 2 standardardized deviations around 0 instead of 2.5 is preferred, and a χ^2 critical value for 95 % instead of 97.5 % is imposed then two years may be registered for leading to bad leverage points, 1979, and 1980. See Tables 29 and 25. So the simultaneous analysis of the robust distances and the standardized LMS residuals, marks 1979 and 1980 and eliminating these two points and running OLS over the remaining data points give the results in Table 30. Maybe these points reflect the economic crisis at the beginning of eighties.

var.	coef.	s.e.	t-val.	p-val.
Const.	3.26	7.82	0.42	0.681
Growth	4.31	1.22	3.54	0.002
Inflation	-2.97	1.16	-2.56	0.018

Table 30: Return on stock's data, both LMS standardized residuals and robust distances considered, $R^2 = 0.682$, F-val=16.4

Two points deserve attention: now the inflation is significant, which is reasonable as also Fama had found, and the fit is better than the one by OLS -if we consider the coefficient of determination and the F-value of regression. So just removing the suspicious points makes a more successful and reasonable regression.

1.7 Tansel's Study on Cigarette Demand in Turkey

Tansel [103] reports the results of a comprehensive study on cigarette demand in Turkey, where he uses four different models. The regressand is the cigarette consumption in Turkey per adult, C, and the regressors are the constant term, co., income, I, and price, P, for all models. The additional regressors are lagged consumption, $C - l$, two dummies for years, D1, and D2, for Model 1, $C - l$ and the first dummy for Model 2, the two dummies for Model 3, and $C - l$ the first dummy, secondary and tertiary, S and T, enrolment ratios for Model 4. All variables are in logarithms.

This time the joint analysis of considering both the robust distances and the standardized LMS residuals is called NEW. That is, as explained before, the bad leverage points are the ones with a standardized LMS residual outside the [-2.5,2.5] tolerance band, and at the same time with a robust distance greater than the tabled critical χ^2 value. NEW is supposed to yield better results since the standardized residual used by it has 50 % breakdown value, that is it is very robust to outliers, and the same is true for the robust distance since the formula for the Mahalanobis Distance is substituted by a new one where the covariance matrix is more robust than the ordinary covariance matrix again.

OLS and NEW are run for all these four models and the results are reported in Table 31, as well as the coefficient of determination and the F-values of regressions. The other RRTs are also run but not reported here.

Finally, Table 32 is about the indices of observations detected by NEW and LTS as the bad leverage points for the different models of Tansel. Note that the observations detected are similar.

var.	coef.	s.e.	t-val.	p-val.	var.	coef.	s.e.	t-val.	p-val.
co.	-3.019	0.8455	-3.570	0.002	co.	-1.599	0.7880	-2.030	0.056
I	0.447	0.1206	3.706	0.001	I	0.232	0.1140	2.032	0.056
P	-0.172	0.0881	-1.952	0.064	P	-0.085	0.0755	-1.126	0.274
C-1	0.311	0.1731	1.797	0.086	C-1	0.700	0.1732	3.993	0.001
D1	-0.087	0.0264	-3.287	0.003	D1	-0.086	0.0214	-4.042	0.001
D2	-0.050	0.0441	-1.139	0.267	D2	-0.027	0.0384	-0.707	0.488
O1	R^2	0.88	F-val	33.6	N1	R^2	0.93	F-val	53.2
co.	-2.676	0.7953	-3.365	0.003	co.	-2.441	0.8671	-2.816	0.010
I	0.395	0.1225	3.514	0.002	I	0.232	0.1233	2.927	0.008
P	-0.192	0.0870	-2.203	0.038	P	-0.085	0.0890	-2.036	0.054
C-1	0.429	0.1389	3.094	0.005	C-1	0.700	0.1592	3.039	0.006
D1	-0.088	0.0265	-3.299	0.003	D1	-0.086	0.0268	-3.225	0.004
O2	R^2	0.877	F-val	41.1	N2	R^2	0.88	F-val	40.3
co.	-4.240	0.5272	-8.043	0.000	co.	-4.040	0.5071	-7.965	0.000
I	0.628	0.0692	9.081	0.000	I	0.600	0.0667	8.990	0.000
P	-0.218	0.0884	-2.462	0.022	P	-0.166	0.0872	-1.909	0.069
D1	-0.101	0.0264	-3.821	0.001	D1	-0.082	0.0266	-3.071	0.006
D2	-0.098	0.0368	-2.661	0.014	D2	-0.132	0.0388	-3.408	0.003
O3	R^2	0.87	F-val	37.5	N3	R^2	0.93	F-val	53.2
co.	-7.544	2.9490	-2.558	0.018	co.	-2.265	2.6622	-0.851	0.406
I	0.905	0.3196	2.831	0.010	I	0.287	0.2993	0.958	0.351
P	-0.152	0.0864	-1.763	0.092	P	-0.087	0.0742	-1.171	0.257
C-1	0.389	0.1374	2.832	0.010	C-1	0.697	0.1718	4.058	0.001
D1	-0.111	0.0293	-3.798	0.001	D1	-0.119	2.0245	-4.866	0.000
S	-0.194	0.1739	-1.117	0.277	S	0.085	0.1509	0.564	0.580
T	-0.134	0.0701	-1.916	0.069	T	-0.116	0.0567	-2.049	0.055
O4	R^2	0.898	F-val	30.7	N4	R^2	0.95	F-val	52.6

Table 31: Regression statistics for Tansel's cigarette consumption data, OLS and NEW are abbreviated by O and N, respectively

RRT M1	Indices	RRT M2	Indices
LTS1	20	LTS1	20
LTS2	15,20	LTS2	15,20
LTS3	15,20,27	LTS3	15,20,27
LTS4	14,15,20,27	LTS4	14,15,20,27
LTS5	7,14,15,20,27	LTS5	7,14,15,20,27
NEW	20,27	NEW	27
RRT M3	Indices	RRT M4	Indices
LTS1	1	LTS1	20
LTS2	1,23	LTS2	20,27
LTS3	1,23,24	LTS3	15,20,27
LTS4	1,23,24,27	LTS4	6,15,20,27
LTS5	1,25,26,28,29	LTS5	6,7,15,20,27
NEW	25,29	NEW	20,22,27

Table 32: Suggested cases to be deleted by different RRTs, M1 stands for Model 1

2 HCCM Estimators

2.1 Introduction

An important assumption of the classical linear regression model is that the disturbances entering the regression are homoskedastic, that is, they all have the same variance, σ^2. If this is not the case, we have the situation of **heteroskedasticity**.

Since this is a classical assumption of the classical linear regression model, it need not be guaranteed in practice. So one has to be careful about the nature of heteroskedasticity, the importance of its detection, its consequences, and the remedies to recover the problems forwarded by it.

Although we assume that the variances of the disturbance terms may be different we assume that they are pairwise uncorrelated throughout this part of the thesis. In short, we assume

$$\Sigma = E(\epsilon\epsilon') = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

where the σ terms stand for the variances of the disturbances.

Heteroskedasticity arises in numerous applications, both from cross-section, and time series data, especially from finance literature. But cross-section data revealed more heteroskedasticity than time series ones. Heteroskedasticity may also be a consequence of data aggregation. Regardless of the source and type of heteroskedasticity, OLS stays no more preferable. The usual formulae to estimate the variances of the OLS estimators are generally unbiased. One cannot tell whether the bias is upward (positive), or downward (negative). The bias arises from the fact that $\hat{\sigma}^2_{OLS}$ is not an unbised estimator of $\hat{\sigma}_i^2$ any more. Furthermore, the usual confidence intervals and hypothesis tests based on t and F distributions are not reliable. So, every possibility arises in drawing wrong conclusions if conventional hypothesis-testing procedures are employed.

If the type of heteroskedasticity is known with certainty, the OLS estimator is undesirable, and one should use generalized least squares. However, the exact type of heteroskedasticity is most of the times unknown, so GLS cannot be used properly either.

The performance of the variance-covariance matrix esimators of the vector of coefficients being used is very much dependent on the variances of the noise terms. It is well known that Ordinary Least Squares (OLS) estimator is extremely good in estimating this variance-covariance matrix in a homoskedastic regression, but when one of the crucial and classical assumptions of the OLS is broken by setting the variance of the error terms to different numbers, OLS performance becomes very poor. And the homoskedasticity assumption of OLS is not very plausable in many cases. So the problem becomes very serious when one uses OLS in a heteroskedastic regression setting. The intention of this part of the thesis is to make a comprehensive study to evaluate, and discuss about prominent estimators of literature, and introduce two more of them. Simulation design to compare the estimators will be explained and the results of the simulation-based comparisons will be reported. Some more information will be given on research about the biases of some of the prominent estimators of the current literature.

2.1.1 Tests for Heteroskedasticity

Heteroskadasticity may cause some serious consequences if the regression is based on least squares, and one cannot even understand that the regression they are using heteroskadastic is without applying some reliable tests. There are several tests developed and suggested to be confident about heteroskedasticity of the data.

The test hypothesis can be expressed in terms of the following claims

$$H_0 : \sigma_i^2 = \sigma^2, \qquad H_1 : Not H_0 \tag{17}$$

The correct covariance matrix for the least squares estimator is

$$Cov(\hat{\beta}_{OLS}) = (X'X)^{-1}(X'\Sigma X)(X'X)^{-1} \tag{18}$$

for which White's estimator is

$$Cov_{Wh}(\hat{\beta}) \;=\; (X'X)^{-1}X'EX(X'X)^{-1} \tag{19}$$

where E is a diagonal matrix of squares of OLS residuals, and the conventional OLS estimator is

$$Cov\hat{\beta}_{OLS} = \hat{\sigma}^2(X'X)^{-1} \tag{20}$$

White [107] has found a statistical test based on this observation. A simple operational version of his test is carried out by obtaining nR^2 in the regression of e_i^2 on a constant and all unique variables in $X \otimes X$. The statistic is asymptotically distributed as χ^2 with $k-1$ degrees of freedom where k is the number of regressors.

The White test is extremely general. In order to carry it out one need not make any specific assumptions about the nature of heteroskedasticity. Although, this seems to be a very advantageous benefit of White's invention, there is a potential shortcoming. The test may reveal heteroskedasticity, but it may as well identify some specification error, such as the omission of the x^2 term from the simple regression see [105]. Little can be said about the power of this test except for some specific cases. One further drawback of the test is about the consequences of running the test, i.e. it does not suggest anything after the rejection of the homoskedasticity null.

We can obtain a more powerful test by narrowing our focus, the Goldfeld-Quandt [30] test is more general. For the Goldfeld-Quandt test, the assumption is that the observations can be divided into two groups in such a way that under the assumption of homoskedasticity, the disturbance variances would be the same in the two groups, while under the alternative, the disturbance variances would differ systematically. The most favourable case for this test is groupwise heteroskedasticity. By ranking the observations based on the level of assumed heteroskedasticity, one separate the observations into those with high and low variances. The test is applied by dividing the observations into two groups with sizes T_1, and T_2. In order to obtain statistically independent variance estimators, the regression is then estimated separately with the two sets of observations. The test statistic is

$$F[T_1 - k, T_2 - k] = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \tag{21}$$

where the disturbance variance of the first sample is assumed larger. Under the null hypothesis of homoskedasticity, this has an F-distribution with $T_1 - k$, and $T_2 - k$ degrees of freedom. The statistic obtained from the sample must be compared to the tabled F-statistic's critical values.

Goldfeld and Quandt suggest to select some of the observations in the middle of the sample to be omitted to increase the power of their test. However, the more observations dropped, the less the power of the test will be, since the degrees of freedom for the F-statistic will be smaller. And the number of observations to be dropped from the middle of the sample depends on the subjectivity of the applier. Harvey and Philips [37] suggest that no more than one third of the sample size should be dropped. The Goldfeld-Quandt statistic has F distribution under the null hypothesis, and the nominal size of the test is correct. And if the null is incorrect, it will follow the F-distribution for only large samples. The separation of the sample and the number of observations to delete from the middle make the Goldfeld-Quandt test less powerful.

Breusch-Pagan [7] have suggested a Lagrange Multiplier test of the hypothesis

$$\sigma_i^2 = \sigma^2 f(\alpha^0 z_i) \tag{22}$$

where z is the vector of exogeneous variables. The model is homoskedastic if $\alpha = 0$. Under the null hypothesis of homoskedasticity, LM is asymptotically distributed χ^2 with degrees of freedom equal to the number of variables in z. It is claimed that the Breusch-Pagan test is sensitive to the assumption of normality. Koenkar and Basset [51] suggest to replace the denominator of LM by a more robust estimator of the variance of the disturbance term

$$V = \frac{1}{n} \sum_{i-1} n(e_i^2 - \frac{e'e}{n}) \tag{23}$$

The modified statistic will have the same asymptotic distribution of Breusch-Pagan statistic, but absent normality, it provides a more powerful test [32].

One other quite old test for heteroskedasticity is by Glejser [29]. After obtaining the residuals from the original model, Glejser, suggests regressing the absolute values of residuals on X which is thought to be closely associated with the heteroskedastic variance σ_i^2. Some suggested forms of the regression includes the constant and X, or constant and $\sqrt{X}$, or the constant with the inverse of X.

2.2 Introduction of the Estimators

In the standard regression analyses, $y = X\beta + \epsilon$, where y is an $n \times 1$ vector of observations of the dependent variable, X is the $n \times k$ matrix of regressors, and ϵ is the $n \times 1$ vector of the errors terms, the OLS estimator (OLS) for the vector of coefficients is:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y,$$

and the distribution for this esimate is:

$$\hat{\beta}_{OLS} \sim N(\beta, \sigma^2(X'X)^{-1}))$$

OLS method estimates the covariance matrix by:

$$\begin{aligned} Cov_{OLS}(\hat{\beta}) &= \hat{\sigma}^2(X'X)^{-1}, \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^{n}(Y_i - x_i\hat{\beta})^2}{n-k} \end{aligned} \tag{24}$$

Since OLS fails when heteroskedasticity is introduced, White [107] developed a good method where there were earlier studies made by Eicker [25]. The heart of the problem is to find Cov $\hat{\beta}_{OLS}$, which is equal to

$$\begin{aligned} \text{Cov } \hat{\beta} &= \text{Cov } [(X'X)^{-1}X'Y] \\ &= \text{Cov } [(X'X)^{-1}X'(X\beta + \epsilon)] \\ &= \text{Cov } [(X'X)^{-1}X'X\beta] + \text{Cov } [(X'X)^{-1}X'\epsilon] \\ &= 0 + (X'X)^{-1}X'\Sigma X(X'X)^{-1} \end{aligned}$$

where Σ is the covariance matrix of the disturbance terms as stated in previous section: $\Sigma = E(\epsilon\epsilon') = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2,)$

All terms in Cov $\hat{\beta}_{OLS}$ are known except Σ. White estimates this Σ matrix by simply replacing the σ_i^2 terms by squares of the OLS residuals, that is, White's estimator (Wh) for the covariance term is:

$$Cov_{Wh}(\hat{\beta}) = (X'X)^{-1}X'EX(X'X)^{-1} \tag{25}$$

where $E = diag(e_1^2, e_2^2, \ldots, e_n^2)$, e_i being the i^{th} residual of regression. This was a very useful finding since one need not specify the correct type of heteroskedasticity to use the estimator developed by White.

Some time later White's estimator was proved to be biased by Chesher and Jewitt [9] and some others as follows:

$$\begin{aligned} E(e_i^2) &= m_i' \Sigma m_i \\ &= \sigma_i^2 - 2\sigma_i h_i' h_i + h_i' \Sigma h_i \end{aligned}$$

where $M = I - X(X'X)^{-1}X'$, and $H = I - M$. σ_i, m_i, and h_i are the $(i,i)^{th}$ entries of Σ, M, and H, respectively. In the above expression $-2\sigma_i h_i' h_i + h_i' \Sigma h_i$ is the bias term.

There appeared several attempts to recover for the bias term once it became available. Most of such attempts were towards approximating this bias term and removing it. Indeed, White's estimator did not have any problems regarding consistency but it suffered from bias.

Hinkley [41] attempted to correct White's esimator by simply premultiplying it by a factor to make some sort of a degrees of freedom correction and obtained the following estimator (Hi):

$$Cov_{Hi}(\hat{\beta}) = \frac{n}{n-k}(X'X)^{-1}X'EX(X'X)^{-1} \tag{26}$$

Instead of premultiplying White's estimator Horn, Horn & Duncan [42] divided the squares of the OLS residuals by the corresponding entries of the hat matrix, H, and obtain their estimator, (HD). That is:

$$Cov_{HD}(\hat{\beta}) = (X'X)^{-1}X'\tilde{E}X(X'X)^{-1}$$

$$\tilde{E} = diag(\tilde{e}_1^2, \tilde{e}_2^2, ..., \tilde{e}_n^2) \tag{27}$$

$$\tilde{e}_i^2 = \hat{e}_i^2/(1 - h_{tt}) \tag{28}$$

Two bootstrap methods which were first discovered by Efron (1982) and developed later by Freedman (1984) are also included in the study, denoted by (BO). The first one resamples the (y_i, x_i) pairs and gets the (y_i^*, x_i^*), for each resampling $\hat{\beta}_{OLS}$ is obtained for this pair and is called $\hat{\beta}^*$. The corresponding covariance matrix estimator is:

$$Cov_{BO}\hat{\beta} = \frac{1}{M}\sum_{j=1}^{M}(\hat{\beta}_j^* - \bar{\beta}^*)(\hat{\beta}_j^* - \bar{\beta}^*)' \tag{29}$$

M is the simulation sample size, and $\bar{\beta}^*$ is the average of all $\hat{\beta}_j^*$'s.

The second bootstrap method first obtains the OLS residuals, resamples on them with replacement, and obtains the randomly ordered residual vector, e^*. The new vector of coefficients are calculated from $Y^* = X\hat{\beta} + e^*$, and the vector of coefficients, $\hat{\beta}^{**}$, are calculated for each resampling. The covariance estimate is calculated similar to the above one. Since the estimator converges to the OLS estimator asymptotically, it is not included in the tables.

Wu (1986) introduces another bootstrap idea in which he figures out the coefficient vector, $\hat{\beta}_{Wu}$, of $Y^* = X\hat{\beta} + \frac{e_i}{1-h_i}t_i^*$. Here Wu states that any (t_i^*), i.i.d. samples from a finite population $(a_j)_{j=1}^S$ with $\sum_{j=1}^S a_j = 0$, and $\frac{1}{S}\sum_{j=1}^S a_j^2 = 1$ would work. For the simulation we carried on our selection is

$$a_j = \frac{e_j - \bar{e}}{\sqrt{\frac{1}{n}\sum_{j=1}^n (e_j - \bar{e})^2}}$$

which proves to satisfy the conditions stated above. The corresponding covariance matrix is calculated similar to the other bootstrap methods mentioned above. Wu states that his method is equivalent to the method by Horn, and Horn & Duncan [42] when the parameter of interest is linear. Since the covariance estimate for our study requires linearity, his claim for equivalence holds, and the simulations carried out justify his claim for sufficiently large simulation sample sizes, but not included here to save space.

One of the prominent and well-performing estimators is by jackniving (Ja). The logic behind jacknife method is to drop one of the observations each time and calculate the estimator n times and the variabiliy of the recomputed estimates will be used to get the variability of the original one. See Efron [24], Wu [109], and MacKinnon and White [58]. After tedious manipulations the jacknife estimator turns out to be:

$$
\begin{aligned}
Cov_{Ja}\hat{\beta} &= \frac{n-1}{n}(X'X)^{-1}[X'\Omega^{*}X - \frac{1}{n}X'u^{*}u^{*'}X)](X'X)^{-1} \\
\Omega^{*} &= diag(u_1^{*2}, u_2^{*2}, ..., u_n^{*2})
\end{aligned}
\tag{30}
$$

where u^* is a vector of u_i^*'s, $u_i^* = \frac{e_i}{(1-h_{tt})}$.

Another estimator tries to detect whether the regression is heteroskedastic first, and suggests OLS if heteroskadasticity is not detected and suggests White's estimator, otherwise, see [48]. It is called Pre-test OLS, (PO) but is old fashioned now.

Finally, another way of making a better estimator is through estimating the bias of OLS estimator and simply subtracting it (BC). The estimator attempts to use the formula stated for bias of the OLS estimator and tries to fix the bias by replacing the unknown variance terms of bias by their OLS estimates with the hope of estimating the bias so well that there will be some improvement. This is somewhat true since our simulations reveal that this bias correction idea works better than OLS.

Previos studies in the literature reveal that Horn and Horn & Duncan's estimator, Wu's bootstrap, and jacknife method dominates the others but the jackniving does slightly better.

Two other methods, invented by Zaman, are also included in the study. These are James Stein (JS) and Maximum Likelihood (ML) estomators.

2.3 Random Coefficients Model

The key difference between the classical regression models and the random coefficients model stems from the variance of the noise terms. That is,

$$\sigma_t^2 = x_t \Omega x_t'$$

where Ω is equal to LL'. This can be obtained by allowing some more flexibility to the regression setting. If one assumes

$$y_i = X_i \beta_i \tag{31}$$

$$\beta_i = \beta + \nu_i \tag{32}$$
$$E[\nu_i] = 0 \tag{33}$$
$$E[\nu_i \nu_i'] = LL' \tag{34}$$
$$\tag{35}$$

combining the terms gives

$$y_i = X_i\beta + X_i\nu_I \tag{36}$$
$$= X_i\beta + w_i \tag{37}$$
$$E[w] = 0 \tag{38}$$
$$E[w_i w_i'] = X_i \Omega X_i' \tag{39}$$
$$= X_i LL' X_i' \tag{40}$$
$$\tag{41}$$

L is selected to shape Ω throughout the studies. L is allowed to come from Cauchy distribution to make it as free as possible.

2.4 Simulation Design and Data Sets

A GAUSS program is coded to run the simulations towards evaluating the performance of the prominent estimators in a random coefficients model addressed above. Several evaluation criteria are used to assess the estimators.

The Monte Carlo sample size is 500 and the simulation is run for 100 different L matrices. The Monte Carlo sample size for the bootstrap is 250.

Different criteria are used to assess all aspects of the success of the estimators from the literature. First criteria may be called the χ^2_{loss}. $(\hat{\beta} - \beta)'(\hat{C})^{-1}(\hat{\beta} - \beta)$ is assumed to follow χ^2 distribution with $k-1$ degrees of freedom. The percentage of times this statistic exceeds the 99 % critical values of the χ^2 table are considered. The above statistics are calculated for all estimators and the percentage of times it exceeds 99 % χ^2 critical values are obtained. Then the absolute value of the difference of this number from 1 % are calculated and the median is reported in the tables. This statistic is expected to exceed the critical value only 1 % of the times, and when 1 % is subtracted one ends up with

0, for a perfect estimator and sufficiently large simulation sample size. Consequently, the greater this number, the less successful the estimator is.

Entropy-loss is the second criterion we used which is defined as:

$$E_{loss} = trace(\hat{C} - C^{-1}) - ln(abs(det(\hat{C}C^{-1}))) - k \tag{42}$$

where trace returns the sum of the diagonal entries of a matrix, abs stands for absolute value, det denotes the determinant of a matrix and ln is the natural logarithm. If the estimator is perfect in the sence that it can hit the true value then, the first term of the summation, the trace, returns k, the logarithm component returns 0, and when k is subtracted from the previous two components of the E-loss, one gets 0.

Third criterion is the quadratic-loss, where the deviations from the correct figures are penalized by the squares of the difference. Namely,

$$Q_{loss} = tr((\hat{C} - C)^2) \tag{43}$$

White and MacKinnon use a very convenient statistic to compare the existing estimators of 1980's which they call the quasi-t statistic. One can refer to [58] to follow their reference. We adopted a similar method and used a similar criteria. First we obtained the critical values where 99 % of the random numbers are divided to the left side of the t-distribution density. We referred to Monte Carlo simulation to obtain the critical values and compared those figures to the tabulated ones from t and Normal distributions. Then we realized that the t-critical values are the most fitting ones. The percentage by which the estimators exceed the critical values are calculated out of the simulations carried on over the data sets for all entries of the vector of coefficients. Then the maximum of the absolute difference for each run of Monte Carlo for different L are calculated and are recorded as the t-loss statistic. That is,

$$t_{loss} = max(abs(\frac{\hat{\beta}_i}{\hat{C}_{ii}})), i = 1, 2, \ldots, k \tag{44}$$

Finally, the percentage of the times the 99 % critical values are passed are averaged for each $\frac{\hat{\beta}_i}{\hat{C}_{ii}}$, and these are also listed at the last columns of the tables. For all estimators, these numbers would preferably be around 1 %, and the deviation of these from 0.01 can be accounted for the failure. 1 % critical values are used for the χ^2, and t losses as well as the t-statistic because 1 % critical values are more sensitive to discriminate among the performance of the estimators.

Three different data sets are used from the literature, the first of which is the famous data set used by White and MacKinnon [58] which uses quarterly data on the rate of growth of the real U.S. disposable income and the U.S. treasury bill rate. The simulation results from his data set are listed in Table 1. The second data set is from Cohen et al. (1993) that comprises 79 observations about the number of hours needed to splice x pairs of wires. The analyses made by them reveal that the explanatory variables present heteroskedasticity. Finally, the third data set is taken from the graduate textbook of Judge et al.

2.5 Simulation Results

All the simulation results are listed in Tables 1, 2, and 3. In terms of the χ^2_{loss} it is very apparent that ML outperforms all the others by leading to a loss of 0.008, where the second best is HD by 0.016 as long as the first data set is concerned. The nearest loss is two times the loss encountered by ML, the others are even worse. The same comment may be true for the Q_{loss} as well because ML Q_{loss} is a bit more than half of the others. Coming to the t_{loss} one can say that ML is among the best ones but has performed not more successfully than Hi, HD, and BC. Finally, if we check how much the last three columns are to 1 % we realize that ML has done differently for different entries of the $\hat{\beta}$ vector but the overall performance of ML seems to be the best among all. OLS is the worst without any question. This is natural because its basic assumption of homoskedasticity is broken. The James Stein estimator does second best for the E and Q losses, but does worse according to the other criteria.

The second data set also reveals the best result of the ML as long as the χ^2 loss is concerned. The χ^2 loss by ML is about 25 percent less than the others. The same is true for the E_{loss} also, where the E_{loss} by ML is 0.086 and the smallest other is 0.108. The Q_{loss} say the same thing but the difference is not that big now. The best ones of the remaining set are jacknife, HD, and Hinkley's estimators. Bias-corrected estimator scores the best in terms of the quasi-t losses.

	χ^2	E_{loss}	Q_{loss}	t_{loss}	β_0	β_1	β_2
OLS	0.041	0.459	0.00872	0.030	0.01050	0.0312	0.0200
Wh	0.025	0.374	0.00885	0.010	0.01236	0.0196	0.0154
Hi	0.018	0.364	0.00921	0.008	0.01050	0.0172	0.0128
HD	0.016	0.370	0.00971	0.008	0.01016	0.0159	0.0121
Ja	0.020	0.373	0.00940	0.010	0.01136	0.0174	0.0134
ML	0.008	0.199	0.00707	0.008	0.01122	0.0138	0.0107
BC	0.018	0.398	0.01007	0.008	0.01032	0.0162	0.0124
BO	0.018	0.354	0.00905	0.010	0.01140	0.0164	0.0131
PO	0.034	0.418	0.00912	0.018	0.01104	0.0220	0.0168
JS	0.024	0.327	0.00858	0.016	0.01036	0.0208	0.0148

Table 33: First data set, 1 % critical values are used

Similar comments are valid for the third data set.

One interesting result of our simulations is that the jacknife is not the best of the remaining estimators. The James Stein estimator introduced in this paper does the second best as long as the first three columns of the table are concerned, but does substantially worse when the remaining columns are taken into account.

2.6 Bias of the Eicker-White Estimator under Simplifying Assumptions

Heteroskedasticity consistent covariance matrices are being used very widely. In so many applications people are not reporting the ordinary standard errors they used to report. Instead, they are first checking for heteroskedasticity and then possibly reporting the heteroskedasticity corrected standard errors. One might use one of the prominent estimators developed so far and has to prefer one of them. In this part of the study, the aim is to find the bias terms of the estimators to make comparison over the ranges they perform better under some simplifying assumptions. The assumptions can be summarized as follows:

	χ^2	E_{loss}	Q_{loss}	t_{loss}	β_0	β_1
OLS	0.041	0.323	0.03253	0.023	0.01016	0.03606
Wh	0.010	0.110	0.02794	0.006	0.01090	0.01532
Hi	0.008	0.108	0.02848	0.006	0.01016	0.01442
HD	0.008	0.111	0.02952	0.006	0.01002	0.01362
Ja	0.008	0.111	0.02881	0.006	0.01076	0.01456
ML	0.006	0.086	0.02460	0.006	0.01086	0.01246
BC	0.008	0.114	0.02999	0.006	0.01004	0.01344
BO	0.008	0.117	0.02908	0.006	0.01142	0.01476
PO	0.016	0.181	0.02910	0.008	0.01046	0.01800
JS	0.018	0.167	0.02791	0.010	0.01008	0.02014

Table 34: Second data set, 1 % critical values are used

	χ^2	E_{loss}	Q_{loss}	t_{loss}	β_0	β_1	β_2
OLS	0.041	0.421	0.00104	0.021	0.01112	0.01580	0.03250
Wh	0.038	0.476	0.00137	0.014	0.01376	0.01760	0.02238
Hi	0.030	0.461	0.00143	0.010	0.01112	0.01418	0.01862
HD	0.026	0.473	0.00155	0.008	0.01076	0.01266	0.01700
Ja	0.031	0.476	0.00148	0.010	0.01248	0.01488	0.01904
ML	0.014	0.225	0.00092	0.008	0.01122	0.01382	0.01074
bc	0.028	0.513	0.00160	0.008	0.01098	0.01332	0.01706
bo	0.026	0.438	0.00135	0.010	0.01200	0.01386	0.01842
PO	0.036	0.485	0.00129	0.017	0.01150	0.01510	0.02652
JS	0.027	0.391	0.00119	0.014	0.01094	0.01394	0.02446

Table 35: Third data set, 1 % critical values are used

1. X is T $\times$ 2.

2. First column of X is 1.

3. $\sum_{t=1}^{T} \sigma_t^2 = T$ Noise terms

4. $\sum_{t=1}^{T} x_t = 0$

5. $\sum_{t=1}^{T} x_t^2 = T$

The second column of X is indexed by $x_1, x_2, \ldots, x_T$. Here, most of the assumptions can be satisfied by making simple manipulations over the regressors, or selecting the regressors in the given conditions. The notation for some of the expressions are determined as follows:

1. $\sum_{t=1}^{T} x_t^3 = ST$

2. $\sum_{t=1}^{T} x_t^4 = KT$

3. $\sum_{t=1}^{T} x_t^5 = GT$

4. $\sum_{t=1}^{T} x_t^6 = LT$

5. $M(X) = \frac{1}{T} \sum_{t=1}^{T} x_t = 0$

6. $M(X^2) = \frac{1}{T} \sum_{t=1}^{T} x_t^2$

7. $M(X, \sigma^2) = \frac{1}{T} \sum_{t=1}^{T} x_t \sigma_t^2$

And the rest of the moments are used in the text is similarly.

For our simple case the variance-covariance matrix is:

$$(X'X)^{-1}X'\Sigma X(X'X)^{-1} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} \sigma_t^2 & \frac{1}{T}\sum_{t=1}^{T} x_t\sigma_t^2 \\ \\ \frac{1}{T}\sum_{t=1}^{T} x_t\sigma_t^2 & \frac{1}{T}\sum_{t=1}^{T} x_t^2\sigma_t^2 \end{bmatrix}$$

For the above matrix the Eicker White estimator is:

$$(X'X)^{-1}X'\Sigma X(X'X)^{-1} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} e_t^2 & \frac{1}{T}\sum_{t=1}^{T} x_t e_t^2 \\ \\ \frac{1}{T}\sum_{t=1}^{T} x_t e_t^2 & \frac{1}{T}\sum_{t=1}^{T} x_t^2 e_t^2 \end{bmatrix}$$

where the e term stands for the OLS residuals.

Before coming to the calculation of the bias for the first entry of the covariance matrix, the preliminaries are:

$$\begin{aligned} e &= y - x\hat{\beta} && (45) \\ &= (I - X(X'X)^{-1}X')Y && (46) \\ &= (I - H)\epsilon && (47) \end{aligned}$$

$$\begin{aligned} e_t &= \epsilon_t - \sum_{t=1}^{T} h_{tj}\epsilon_j && (48) \\ &= (1 - h_{tt})\epsilon_t - \sum_{j=1, j\neq t}^{T} h_{tj}\epsilon_j^2 && (49) \\ Ee_t^2 &= Var(e_t) && (50) \\ &= (1 - 2h_{tt} + h_{tt}^2)\sigma_t^2 + \sum_{j=1, j\neq t}^{T} h_{tj}^2\sigma_j^2 && (51) \\ &= \sigma_t^2 - 2h_{tt}\sigma_t^2 + h_{tt}^2\sigma_t^2 + \sum_{j=1, j\neq t}^{T} h_{tj}^2\sigma_j^2 && (52) \\ &= \sigma_t^2 - 2h_{tt}\sigma_t^2 + \sum_{j=1}^{T} h_{tj}^2\sigma_j^2 && (53) \\ &= \sigma_t^2 - \frac{2}{T}(1 + x_t^2)\sigma_t^2 + \sum_{j=1}^{T} \frac{1}{T^2}(1 + x_t x_j)^2\sigma_j^2 && (54) \\ &= \sigma_t^2 - \frac{2}{T}\sigma_t^2 - \frac{2}{T}x_t^2\sigma_t^2 + \frac{1}{T^2}\sum_{j=1}^{T} \sigma_j^2 + 2x_t x_j\sigma_j^2 + x_t^2 x_j^2\sigma_j^2 && (55) \\ &= \sigma_t^2 - \frac{2}{T}\sigma_t^2 - \frac{2}{T}x_t^2\sigma_t^2 + \frac{1}{T} + \frac{2}{T}x_t M(x,\sigma^2) + \frac{1}{T}x_t^2 M(x^2,\sigma^2) && (56) \\ &= \sigma_t^2 - \frac{2}{T}(1 + x_t^2)\sigma_t^2 + \frac{1}{T}(1 + 2x_t M(x,\sigma^2) + x_t^2 M(x^2,\sigma^2)) && (57) \end{aligned}$$

Now the bias terms for the four entries of the variance-covariance matrix will be provided. Since the matrix is symmetric around the first diagonal axis, three bias terms will be calculated. The proofs for the manipulations are given in the appendix to the thesis.

Bias of Eicker-White estimator entry by entry of the covariance matrix is entry is

$$B_{11} = E\hat{C}_{11} - C_{11} \quad (58)$$

$$= -\frac{1}{T^2}(1 + M(x^2, \sigma^2)) \quad (59)$$

Now, coming to the calculation of the $B_{12} = B_{21}$

$$B_{12} = E\hat{C}_{12} - C_{12} \quad (60)$$

$$= (\frac{S}{T^2})M(x^2, \sigma^2) - \frac{2}{T^2}M(x^3, \sigma^2) \quad (61)$$

Finally, the bias of the last entry is

$$B_{22} = E\hat{C} - C \quad (62)$$

$$= \frac{1}{T^2} + \frac{2S}{T^2}M(x, \sigma^2) + \frac{K-2}{T^2}M(x^2, \sigma^2) - \frac{2}{T^2}M(x^4, \sigma^2) \quad (63)$$

$$(64)$$

2.7 Bias of H-HD estimator

The same assumptions are still holding. Recall that the Σ matrix in the middle of the covariance matrix is estimated by H-HD by dividing each entry of E-W by the corresponding entries of the hat matrix. So first we concentrate on an arbitrary entry of that matrix. Here we assume that

$$\frac{e_t^2}{1 - htt} = \alpha_t e_t^2 \;\; where \quad (65)$$

$$\frac{1}{1 - htt} = 1 + h_{tt} + h_{tt}^2 + h_{tt}^3 + \ldots + R \;\; (Taylor's \;\; expansion) \quad (66)$$

$$(67)$$

where R is the remainder of the approximation.

For the time being take $\alpha_t = 1 + h_{tt}$. So:

$$E\frac{1}{1-h_{tt}}e_t^2 = E\alpha_t e_t^2 \tag{68}$$

$$= E((1+h_{tt})e_t^2) \tag{69}$$

$$= Ee_t^2 + E(h_{tt}e_t^2) \tag{70}$$

Now let

$$\Lambda = Ee_t^2 = \sigma_t^2 - \frac{2}{T}(1+x_t^2)\sigma_t^2 + \frac{1}{T}(1 + 2x_t M(x,\sigma^2) + x_t^2 M(x^2,\sigma^2)) \tag{71}$$

$$\tag{72}$$

The bias of the first entry is:

$$B_{11}^{HD} = \frac{K-3}{T^3}M(x^2,\sigma^2) + \frac{2S}{T^3}M(x,\sigma^2) - \frac{2}{T^3}M(x^4,\sigma^2) \tag{73}$$

$$\tag{74}$$

For the bias of the off-diagonal entries

$$B_{12}^{HD} = B_{21}^{HD} \tag{75}$$

$$= \frac{S}{T^3} + (\frac{2K+T}{T^3})M(x,\sigma^2) + (\frac{ST+S+G}{T^3})M(x^2,\sigma^2) - (\frac{T+4}{T^3})M(x^3,\sigma^2) - \frac{2}{T^3}M(x^5,\sigma^2) \tag{76}$$

$$\tag{77}$$

Finally, for the last entry:

$$B_{22}^{HD} = (\frac{K+T+1}{T^3}) + M(x,\sigma^2)(\frac{2ST+2S+2G}{T^3}) + M(x^2,\sigma^2)(\frac{KT-T+K+L-2}{T^3}) - M(x^4,\sigma^2)(\frac{T+4}{T^3}) - M(x^6,\sigma^2)\frac{2}{T^3} \tag{78}$$

$$\tag{79}$$

2.8 Conclusion

This chapter of the thesis comprises two main issues: comparison of the main estimators for the covariance matrix of the heteroskedastic regression, and the determination of the bias terms for some of the estimators. Since this is some real research towards finding out some facts the study is not completed and still remains open for some further research. The research topics may be the comparison of the estimators under some more general assumptions, including some of such estimators, and more important is to find a better estimator, which outperforms all of them.

As long as the comparisons are concerned, ML has done the best. We examined all prominent variance-covariance matrix estimators for $\hat{\beta}$ that are relevant and simulated to observe their performance by using some data sets from econometrics literature. The ML and James Stein estimators promoted by Zaman are also included. Several different criteria are used to reveal different aspects of the estimators. ML estimator is found to compete well with the existing estimators of literature for estimating the variance-covariance matrix of $\hat{\beta}$ if not outperforms them, although it takes considerable more time to calculate it.

The issue about the bias terms of the estimators is also important since the econometrics literature is now more aware of the estimators in heteroskedastic regressions and people should use these estimators more consciously, if one figures out the ranges over which estimator is doing better than the others, at least econometricians should know which estimator to prefer over different ranges. But this seems like a tedious and time consuming job as one can easily understand by looking at the derivations at the appendix of the thesis.

3 Empirical Bayes Application to İstanbul Stock Exchange Data

The final part of the thesis is dedicated to some empirical application of the Bayesian approach. The initials of the theory is explained in its philosophy. The application requires some complicated manipulation of the theory where some inferiors of the ordinary Bayesian approach are somehow avoided.

The application uses some huge amount of daily data of the prices of all stocks processed in the İstanbul Stock Exchange. A simple relationship between the prices of stocks and the market index is used to run OLS and the Empirical Bayes procedure over different sectors and different time periods with different number of firms to evaluate which of the two approaches performs better. Two bencmark criteria are used. The estimated figures are compared to the already known numbers and the mean absolute deviation and the mean squared deviation are calculated for both methods.

The missing values of the data are substituted by the one-day-before values if the missing values are not coming consequently and frequently, otherwise, that portion of the data set is removed from the analyses.

This part of the thesis includes both the theory of the Empirical Bayes approach as well as the application supported by it, and some issues of the finance literature to better understand what the application refers to. So both of these are explained in the introductory part.

3.1 Introduction

3.1.1 Portfolio Risk and the Capital Asset Pricing Model

The stocks that are held in isolation are risky. The riskiness of the stocks held in portfolios[2] will be analyzed in this section via their beta coefficients. Such a relationship between a stock's price and its beta is drawn in finance literature. A stock held as part of a portfolio is less risky than the same stock held in isolation. This fact has been incorporated into

[2] A *portfolio* is used as a collection of investment securities throughout the chapter. If you owned some stocks of three different firms with the hope of observing some higher increases of these stocks then you are having a portfolio of three stocks.

a procedure for analyzing the relationship between risk and rates of return, the **Capital Asset Pricing Model,** or **CAPM**. The CAPM is an extremely important analytical tool in both financial management and investment analysis. Indeed, the 1990 Nobel Prize was awarded to Professors Harry Markowitz and William F. Sharpe, the developers of the CAPM. So many implications of the issue took place in the literature after the invention of CAPM. Lin, Chen, and Boot [54] investigate several important issues that feature the dynamic and stochastic behavior of beta coefficients for individual stocks and affect the forecasting of stock returns.

Jorion, Giovannini [47] provides two alternative estimation and testing procedures of a representative agent model of asset pricing which relies on a particular parametrization of non-expected-utility preferences. Smith [96] uses Intertemporal Capital Asset Pricing Model (ICAPM) as a vehicle to explain the predictability of excess returns in forward foreign exchange markets.

Handa, Kothari, and Wasley [36] perform multivariate tests of the CAPM using monthly and annual returns on market-value-ranked portfolios, and fail to reject the CAPM when annual holding period returns are used.

The plan for the research of this study included the possible selection of portfolios according to the estimates by Empirical Bayes and OLS. Some further comparisons of the two techniques can be realized when this part of the research will hopefully be finished later. That is, the two techniques will be considered according to the profitability of the portfolios they suggest.

3.1.2 Portfolio Risk and Return

Most financial assets are not held in isolation; rather, they are held as parts of portfolios. Most of the financial institutions such as banks, pension funds, insurance companies, mutual funds are required by law to hold diversified portfolios. Most of the individual investors hold stock portfolios, not the stock of only one firm just to diversify the risk coming from just one item of portfolio. Since the risk is diversified, from an investor's standpoint the fact that a particular stock goes up or down is not very important, indeed what is important is the return on his or her portfolio, and the portfolio's risk. The risk and return of an individual security should be analyzed in terms of how that security affects the risk and return of the portfolio in which it is held.

Companies are subject to risk because of strikes, failing marketing programs, the winning and losing of major contracts, and other events that are unique to a particular firm. Since these events are essentially random, their effects on a portfolio can be eliminated by diversification. Here the idea is that the bad events in one firm will be offset by good events in another. Market risk, on the other hand, stems from factors which systematically affect most firms, such as recession, inflation, and high interest rates, and even war. Since most stocks will tend to be negatively affected by these factors, systematic risk cannot be eliminated by diversification within a country, but the effect of it can be diminished by selecting a good balance of risk and return in the portfolio. This also depends on the risk aversion of the portfolio possessor.

We know that investors demand a premium for bearing risk; that is, the higher the riskiness of a security, the higher the expected return required to induce the investors to buy or hold it. However, if investors are primarily concerned with *portfolio risk* rather than the risk of the individual securities in the portfolio, the answer to the question of how should the riskiness of an individual stock be measured, is provided by the Capital Asset Pricing model (CAPM): the relevant riskiness of an individual stock is its contribution to the riskness of a well-diversified portfolio.

One other main concern is whether all stocks equally risky in the sense that adding them to well-diversified portfolio would have the same effect on the portfolio's riskiness are or not. The answer is negative. Different stocks will affect the portfolio differently, so different securities have different degrees of relevant risk. One should also want to know how the relevant risk of an individual stock can be measured? As we have seen, all risk except that related to broad market movements can, and presumably will, be diversified away. After all, the risk can never be eliminated completely. The risk that remains after diversifying is market risk, or risk that is inherent in the market, and can be measured by the degree to which a given stock tends to move up and down with the market. Beta is used to determine the required rate of return on a stock, given its market risk.

3.1.3 The Concept of Beta, β

The tendency of a stock to move with the market is reflected in its beta coefficient, β, which is a measure of the stock's volatility with respect to the volatility of an average stock. β is the key concept of the CAPM. There have been many studies under different conditions to make better estimates of beta, Mcdonald, and Nelson [67] have done it for thick tailed distributions for returns. Luoma, Martikainen, Perttunen, and Pynnonen [57] investigate beta estimation of different techniques for infrequently traded and inefficient stocks. Ferson, and Foerster [27] develop evidence on the finite sample properties of the Generalized Method of Moments (GMM) in the asset pricing context.

An average risk stock is defined as one that tends to move up and down in step with the general market as measured by some index. Such a stock will, by definition, have a β, of 1.0, which indicates that, in general, if the market moves up by 10 percent, the stock will also move up by 10 percent, while if the market falls by 10 percent, the stock will likewise fall by 10 percent. If $\beta = 0.5$, the stock is only half as volatile as market -it will rise and fall only half as much- and a portfolio of such stocks will be half as risky as a portfolio of $\beta = 1.0$ stocks. On the other hand, if $\beta = 2.0$, the stock is twice as volatile as an average stock, so a portfolio of such stocks will be twice as risky as an average portfolio. The value of such a portfolio could double, in a short time. The literature on risk is very wide.

If a stock with $\beta > 1$ is added to an average $\beta = 1$ portfolio, then β, and consequently the riskiness, of the portfolio will increase. Conversely, if a stock with $\beta < 1$ is added to an average-risk portfolio, then the portfolio's β and risk will decline. Because of this, since a stock's β measures its contribution to the riskiness of a portfolio, β is the theoretically correct measure of the stock's riskiness.

3.2 Bayes Method

The general method for the Bayesian calculations go as follows: let y be the vector of observations that presumably depends upon the unknown parameter of interest, θ. It is assumed that there is some prior information about θ in the form of a density $\pi(\theta)$. Then the joint density of y and θ can be written as: $f(y,\theta) = f(y|\theta)\pi(\theta) = \pi(\theta|y)m(y)$ What we need here for the Bayesian calculations is $\pi(\theta|y)$ which is the updated θ after receiving the data, y. $\pi(\theta|y)$ is usually called the posterior density. The way to calculate this parameter under the assumption of normality is given as follows: let y and θ be k-variate normal vectors where the density of y given θ, $f(y|\theta)$ is $N(\theta, \Sigma_{y|\theta})$, and the marginal density of θ is $N(\mu, \Sigma_\theta)$. Then the marginal density of y is $N(\mu, \Sigma_\theta + \Sigma_{y|\theta})$. The conditional density of θ given y is also normal with mean vector E$[\theta|y]$=$P^{-1}(\Sigma_{y|\theta}^{-1}y + \Sigma_\theta^{-1}\mu)$ and covariance matrix $Cov(\theta|y) = P^{-1}$ where $P = \Sigma_{y|\theta}^{-1} + \Sigma_\theta^{-1}$.

Similarly, if the data density is $\hat{\beta}|\beta \sim N(\beta, \sigma^2(X'X)^{-1})$, and the prior is $\beta \sim N(\mu, \Sigma_\beta)$, then the posterior density of β is multivariate normal with mean and covarance matrix

$E[\beta|\hat{\beta}] = (\frac{1}{\sigma^2}(X'X) + \Sigma_\beta^{-1})^{-1}(\frac{1}{\sigma^2}(X'X)\hat{\beta} + \Sigma_\beta^{-1}\mu)$

$Cov(\beta|\hat{\beta}) = (\frac{1}{\sigma^2}(X'X) + \Sigma_\beta^{-1})$

The Bayes method suffers from three difficulties that come from the very nature of it. First of all, the risk for Bayes estimator can be unbounded, and no one can dare such a risk. Second, the choice of the hyperparameters is important and this choice of the prior parameters may lead to the failure of Bayes estimation. Finally, the values selected for the prior may conflict the data [112], chapter 5. The Empirical Bayes method is specially designed to avoid these three difficulties by simply making the selection of the priors after looking at the data, and fix the values of the prior according to those of the data. The details are elaborated in the coming sections about the theory of Empirical Bayes procedure. Although there are different ways of implementing the Empirical Bayes method, all of them use the marginal density of the observations to bring estimates for the hyperparameters. The data density depends on the parameters, and the prior density of parameters depends on hyperparameters so that the marginal density of the observation depends on hyperparameters [112].

3.3 Empirical Bayes Method and the Model

Our main concern in this part is to find out the best regression that is capable of forecasting the prices of the stocks for the following days. We initially estimate these regressions with OLS, but so many of the regressions are imprecisely estimated, especially for the stocks with small number of firms and the cases where the number of observations is small.

The imprecision of OLS lead us to implement an Empirical Bayes procedure that generates substantially more precise estimates.

The model used is

$$Y_i = X_i\beta_i + \epsilon_i, \quad i = 1, ..., T \tag{80}$$

$$\beta_i = (\beta_{i1}, \beta_{i2})' \tag{81}$$

$$\epsilon_i = (\epsilon_{i1}, ..., \epsilon_{in_i})' \tag{82}$$

$$[\epsilon_i | \sigma_i^2] \sim N_{n_i}(0, \sigma_i^2) \tag{83}$$

For the given equations above i denotes the index for a stock, T is the number of stocks in that sector, n_i is the number of observations per stock, which ranges from 5 to 120 by construction.

More specifically, we used the following equation for estimation of all firms in the different sectors:

$$P_i = \beta_{i1} + \beta_{i2} R_m \tag{84}$$

$$P_i = lnP_k - lnP_{k-1}, k = 2, 3, \ldots, n_i \tag{85}$$

$$R_m = lnI_k - lnI_{k-1}, k = 2, 3, \ldots, n_i \tag{86}$$

In the above equations, n_i is the number of observations, that is the number of days for the estimation period. i is the İstanbul Stock Exchange Index, P_i is the price of the stock, and R_m, stands for the market return. The ratios of the logarithms are used to play with smaller numbers that are standardized.

We consider two approaches to estimating these sets of regressions. The first is OLS based on 1-4, so that,

$$[\hat{\beta}_i^{OLS} | \beta_i, \sigma_i^2] \sim N(\beta_i, \sigma_i^2 (X_i' X_i)^{-1}) \tag{87}$$

where

$$\hat{\beta}_i^{OLS} = (X_i'X_i)^{-1}X_i'Y_i \tag{88}$$

Note that the OLS approach of Equations 1-4 assumes that parameter estimates for one stock tell us nothing about the likely true parameter values for any other stock. While this is a standard conservative assumption, it should not be restrictive for us. Indeed, there is some more information embedded in the data which is ignored by OLS. The information is the likely coordinated action of the stocks within the same sector. The idea leading to the extra information employed by Empirical Bayes is that the stocks within the same sector are affected from the exogeneous shocks to that sector together, therefore they move similar to each other. This piece of information is used in our second approach, Empirical Bayes which assumes that the true parameter values for the individual stocks are related. In particular, the Empirical Bayes model is obtained by assuming that β_i has a normal prior distribution of the form

$$[\beta_i|(\theta, \Lambda)] \sim N(\theta, \Lambda) \tag{89}$$

The standard Bayesian approach now tries to specify the hyperparameters θ and Λ and use Bayes' rule for estimating the β_i's. This leads to the Bayesian estimator

$$\hat{\beta}_i^{Bayes} = D_i^{-1}(\sigma_i^{-2}X_i'X_i\hat{\beta}_i^{OLS} + \Lambda^{-1}\theta) \tag{90}$$

where

$$D_i = \sigma_i^{-2}X_i'X_i + \Lambda^{-1} \tag{91}$$

This estimator is a weighted average of the OLS estimate and the assumed prior mean where the weights are the estimated variances of the OLS estimate and the assumed prior variance. Note that the expression in the second squared bracket above is the OLS estimate of $\hat{\beta}$ pre-multiplied by the OLS estimate of covariance, $\hat{\sigma}^2(X'X)^{-1}$. Now we follow an Empirical Bayes method that allows θ and Λ to be estimated directly from the interstock distribution of the OLS parameters. In particular an initial estimate for θ is

$$\theta = [\sum_{i=1}^{T} \hat{\sigma}_i^{-2}X_i'X_i]^{-1}[\sum_{i=1}^{T} \hat{\sigma}_i^{-2}X_i'Y_i] \tag{92}$$

which is essentially a weighted average of the stock-specific OLS estimates where the weights are inversely related to the parameter's estimated variance. With this initial estimate of θ in hand, we then proceed to estimate θ and Λ with an iterative procedure.

While one can allow for all forms of Λ this makes computation intractable for our data, and that is why we restrict off-diagonal entries of Λ to be zero. This means we do not let any prior covariance across the coefficients, and this is called the D-Prior method of Empirical Bayes, D standing for the diagonal. We also tried another technique called, the g-prior but did not concentrate on it, since the results by d-prior dominated that. With this assumption and our initial estimate of θ we form an initial estimate of Λ via

$$\hat{\Lambda} = diag(\hat{\lambda}_1^+, \hat{\lambda}_2^+),\ where\ \ \hat{\lambda}_j^+ = max(0, \hat{\lambda}_j) \tag{93}$$

$$\hat{\lambda}_j = \frac{1}{T-1}\sum_{i=1}^{T}[(\hat{\beta}_{ij} - \hat{\theta}_j)^2 - \hat{\sigma}_i^2(X_i'X_i)_j]^+ \tag{94}$$

where i indexes stocks and j indexes regressors so that, for example, $(X_i'X_i)_j$ refers to the j^{th} diagonal element of $(X_i'X_i)$. In essence, each $\hat{\lambda}_j$ is an estimate of the interstock variance of parameter j, corrected for sampling error. We then reestimate β_i's with (90)-(91) and reestimate each element of θ with

$$\hat{\theta}_j = [\sum_{i=1}^{T}\hat{\sigma}_i^{-2}(X_i'X_i)_j + \hat{\lambda}_j^{-1}]^{-1}[\sum_{i=1}^{T}(\hat{\sigma}_i^{-2}(X_i'Y_i)_j + \hat{\lambda}_j^{-1}\hat{\beta}_{ij})] \tag{95}$$

Note that the calculation of (90-91) requires estimates of (93)-(94) and (95), and that (93)-(94) requires the estimates of (90)-(91) and (95) etc., so that solutions for θ, Λ and the β_i's must be solved iteratively. Fortunately, the number of the iterations we faced while running the coded program for this method did not exceed 10 most of the times, provided that we had started from good initial points.

With solutions to (90)-(91), (93)-(94) and (95) in hand, the estimated variance of the posterior distribution of the β_i's is computed as

$$Var(\beta_i^{EB}) = (\hat{\Lambda}^{-1} + [\hat{\sigma}_i^2(X_i'X_i)]^{-1})^{-1} \tag{96}$$

Note that the estimated variance of the Empirical Bayes estimator is smaller than the variance of the OLS estimator by construction. The increased precision is a result of the increased information introduced into the model.

3.4 Data and their Manipulation to Compare OLS and Empirical Bayes

Daily data belonging to İstanbul Stock Exchange from January the first of 1988 till the end of October 1995 are used to obtain the desired statistics. Firms are grouped into 21 sectors. The list of the sectors are given in Table 36 below as well as the number of firms in each sector.

To evaluate the performance of the methods, the data are split into two of different lengths. These are called the estimation and the forecast periods. We pretend as follows: the data in the forecast period are assumed unknown to the estimator, whereas the only data known are the ones in the estimation period. The true figures in the forecast period are then used to compare the estimated values with the true ones for different techniques. Table 37 below is designed to display the two periods expalined above, that is, the table expresses the different estimation and forecast periods in days. The stock exchange worked for 5 days of the week, so 20 days more or less corresponds to a week.

The comparisons are made for two kinds of stocks: initial public offerings, IPO's, and the firms that already existed in that particular sector, non-IPO's. We had set the initial dates of the regressions to the days where a new stock, IPO, is introduced to the sector. The data for the 21 sectors are arranged in a way to let the date at each joining firm initiates regressions. Many regressions are set via this way, where the length of the data were long enough to cover both the estimation and the forecast periods. If the data were not sufficient to handle this length then they were not included in the analyses.

In summary, we changed several things to make more detailed comparisons and we had to consider different aspects of the data. First, the estimation and the forecast periods were changed by which 10 different cases were obtained. Two different criteria are used: squared, and absolute discrepancies. The starting days of the regressions are set to dates where a new firm joined the stock exchange. All of these are carried out for the IPO's and the non-IPO's.

One may expect to observe $\beta_{i1} = 0$, and $\beta_{i2} = 1$. That is, the indiviual stocks are supposed to follow the market index with the intercept being equal to 0.

i	T_i	Stock
1	6	Leasing-Factoring
2	11	Holding
3	5	Investment
4	5	Insurance
5	5	Petroleum and its by Products
6	4	Plastic Products
7	4	Ceramics and Porcelain
8	15	Cement
9	7	Iron and Steel
10	3	Other Metals
11	11	Food and Alcoholic Drinks
12	17	Textiles
13	5	Beverage
14	6	Paper and Paper Products
15	5	Press and Publising
16	4	Fertilizer and Agricultural Products
17	6	Durable Goods
18	6	Electric Machinery
19	3	Metal Products and Machinery
20	3	Energy
21	3	Construction Material

Table 36: Sectors and the number of firms, or stocks, in each of them

Case	a	b	c	d	e	f	g	h	i	j
Estimation	5	10	15	20	20	40	60	80	100	120
Forecast	5	5	5	5	20	20	20	20	20	20

Table 37: Estimation and forecast periods in days

3.5 Comparison of Techniques According to Sectors

All the comparisons are made according to squared or absolute discrepancies of the predicted values around the true ones for IPO's, and the non-IPO's. The 10 different cases shaped due to estimation and forecast periods reveal the ability of the estimators at different number of data points included for estimation.

To evaluate the IPO's we took different estimation and forecast periods where the number of firms started from 3 and were increased gradually as new firms entered the sector. This gave an opportunity to observe the successes of the estimators when the number of stocks were changing. The capacity of the stock exchange enabled us to have 17 such newly joining firms only.

Four tables are designed to have a closer look at how successful Empirical Bayes and OLS estimators are doing. Tables 38, and 39 display the figures for IPO's. The tabled values are obtained as follows:

$$F = \frac{\sum_{i=1}^{n_i} \mid r_i^D \mid}{\sum_{i=1}^{n_i} \mid r_i^{OLS} \mid} \tag{97}$$

where $\mid . \mid$ stands for the absolute value, and r stand for the residuals of D-Prior and OLS. Similarly, Tables 40, and 41 display the same values for the non-IPO's. Letters at the top of the tables belong to cases explained in Table 37. People may have some idea about the future prices of the non-IPO's of a sector, but since there are no past prices of stocks for IPO's their behaviors can be estimated with more difficulty, and correct estimations about them may bring much more benefit. That is why the IPO's constitute an important part of the estimation.

It is reasonable to think that the two errors explained above should go parallel to each other, but they do not have the same formula of calculation and squared errors penalizes the discrepancy by the square of the difference, so the big differences are subject to huge numbers by squared residuals. Similarly, the small differences -less than 1- are made even smaller by taking the square.

By looking at the tables for IPO, Tables 39, and 38, one can say the Empirical Bayes technique does better when the estimation period is smaller. Note that the numbers in columns a-d tend to become smaller, as the estimation period moves from 1 week to four

	a	b	c	d	e	f	g	h	i	j
1	1.444	1.129	0.721	0.611	0.771	0.911	0.926	1.008	1.015	1.051
2	0.832	1.265	1.103	1.075	1.011	0.984	0.984	1.008	0.997	1.023
3	0.594	0.611	0.643	1.150	0.794	0.934	0.926	0.978	0.978	1.001
4	1.524	1.518	1.635	1.715	0.960	0.966	0.960	1.003	0.973	0.998
5	0.894	1.039	0.929	1.056	0.890	0.950	0.959	0.967	0.948	0.988
6	1.046	0.818	1.092	0.761	0.728	1.148	1.028	0.964	0.981	0.955
7	0.879	0.881	1.015	1.015	0.969	1.034	1.000	1.021	1.019	1.013
8	0.704	0.730	0.978	0.896	0.819	0.894	0.950	0.892	0.885	0.947
9	0.777	0.729	1.036	0.992	0.964	1.039	0.954	0.944	0.988	0.943
10	0.922	0.751	0.777	1.015	1.078	0.990	0.951	1.024	1.002	0.987
11	0.685	0.770	0.916	0.886	0.894	0.936	0.991	1.009	0.992	0.995
12	0.875	0.877	1.013	1.059	0.988	0.978	0.967	0.958	0.949	1.111
13	0.428	0.572	1.472	2.876	1.019	0.937	0.998	0.998	0.988	0.924
14	0.900	0.768	0.829	0.746	0.838	1.054	0.955	1.022	0.972	1.045
15	0.932	0.877	1.087	0.763	0.720	0.863	0.974	0.981	1.047	0.963
16	0.869	0.978	1.221	1.180	1.028	1.049	1.054	0.990	0.930	0.944
17	0.775	1.129	0.985	0.685	0.844	0.913	0.953	0.945	0.950	0.961
18	1.974	1.396	1.096	1.107	0.862	0.987	0.997	0.964	0.984	0.983
19	0.591	1.147	0.875	0.779	0.812	0.990	0.960	1.035	2.060	1.071
20	0.893	0.690	0.767	0.737	0.799	0.690	0.640	0.827	0.874	0.785
21	1.095	1.021	0.916	0.948	0.936	1.073	1.011	0.979	0.947	0.995
μ	0.935	0.938	1.005	1.050	0.892	0.968	0.959	0.977	1.023	0.985

Table 38: Mean for the ratio of Sum of Squared Residuals of D-Prior to OLS estimators for the 21 sectors for 10 different cases, IPO's

	a	b	c	d	e	f	g	h	i	j
1	1.107	1.111	0.819	0.757	0.850	0.968	0.951	1.002	1.010	1.032
2	0.917	1.116	1.020	1.022	1.010	0.983	0.984	0.990	0.984	1.006
3	0.795	0.749	0.805	1.017	0.880	0.969	0.951	0.981	0.991	0.998
4	1.170	1.322	1.295	1.315	1.053	0.977	0.991	0.990	0.990	1.004
5	0.929	0.972	0.941	1.029	0.942	0.965	0.962	0.969	0.969	0.982
6	1.011	0.890	1.014	0.858	0.831	1.069	0.994	0.971	0.998	0.979
7	0.954	0.885	0.992	1.020	0.976	1.026	0.989	1.001	1.011	0.999
8	0.796	0.826	0.949	0.893	0.864	0.935	0.991	0.948	0.926	0.983
9	0.890	0.841	0.989	0.962	0.969	1.006	0.965	0.970	1.002	0.978
10	0.979	0.861	0.893	0.926	0.989	0.984	0.962	1.018	1.011	1.012
11	0.780	0.877	0.958	0.890	0.923	0.957	1.007	1.015	1.022	0.994
12	0.906	0.953	0.955	0.987	0.951	0.977	0.962	0.968	0.960	1.036
13	0.635	0.672	1.031	1.616	1.247	0.917	0.989	1.006	0.992	0.908
14	0.945	0.865	0.864	0.912	0.883	1.033	0.959	0.996	0.981	1.010
15	0.975	1.022	1.115	0.943	0.834	0.939	0.975	0.976	1.039	0.982
16	0.936	1.029	1.147	1.102	1.021	1.041	1.034	0.978	0.914	0.962
17	0.867	1.037	1.014	0.821	0.895	0.960	0.966	0.963	0.985	0.985
18	1.348	1.111	0.935	0.828	0.783	0.944	0.965	0.950	0.984	0.977
19	0.765	1.021	0.913	0.858	0.878	0.978	0.968	1.001	1.435	1.025
20	0.946	0.860	0.885	0.862	0.886	0.799	0.753	0.893	0.905	0.829
21	1.054	1.002	1.002	0.999	0.912	1.056	1.005	0.928	0.971	0.998
μ	0.938	0.953	0.978	0.982	0.932	0.975	0.968	0.977	1.004	0.985

Table 39: Mean for the ratio of Sum of Absolute Residuals of D-Prior to OLS estimators for the 21 sectors for 10 different cases, IPO's

weeks gradually, and the forecast period is fixed. The same is true when the estimation period moves from 1 to 6 months for columns e to j. Similar comments are true for non-IPO's also. Regardless of whether we are looking for IPO's or non-IPO's and the estimation and forecast periods D-Prior dominates OLS, but the degree of domination changes due to some changes of the parameters.

One other point of interest may be the success of estimators, as the number of IPO's joining the sector are increasing, Table 42 is prepared to answer this question. Initially, Empirical Bayes dominates OLS, and the difference of the domination increases as the IPO starts to join more number of firms in the sector. And then the domination oscillates, but it is clear.

A table is prepared to lead to the histogram to display the percentage of the difference to the sum for the sum of absolute residuals for the ten cases, Table 43. The table also reveals that as the estimation period becomes longer the percentage of domination alleviates. See the zeros in the final columns of the tabel. The same is repeated for the IPO to give a rough idea. But only the fifth IPO's are included. There is nothing special to the fifth. The others are excluded just to save space. One can make the same comments for the IPO's also.

Several tables are prepared to have an insight about the number of cases belonging to the superiority of the estimators. Since the criteria are parallel to each other, it may be satisfactory to be content with just one of them, otherwise the number of tables would double.

Table 45 counts the number of times D-Prior has a smaller sum of absolute errors less than OLS ($<$), as well as the number of times OLS is better than D-Prior ($>$) for the initial public offerings. The same is repeated in Table 46 where the sum of absolute errors by D-Prior is at most 95 percent of that of OLS ($<$), and vice versa for the domination of OLS ($>$). Again these are all for IPO's.

The following two tables are doing the same thing for the non-IPO's.

The final four tables give better understanding of the domination of the techniques. One may see the averages in the previous tables but may still suspect about the side of the better estimator, since a particular substantial domination may affect the average very much, but this table omits that possibility.

	a	b	c	d	e	f	g	h	i	j
1	0.951	0.985	1.042	1.026	1.017	0.989	1.008	1.006	1.001	0.999
2	0.881	0.955	1.022	0.990	0.980	0.984	0.991	0.987	0.971	0.988
3	0.928	1.062	1.012	0.996	0.985	0.975	0.987	1.012	1.010	1.010
4	0.911	1.010	1.028	0.938	0.957	1.034	0.995	1.001	1.012	1.018
5	0.946	1.048	1.015	0.960	0.985	0.957	0.966	1.007	1.004	1.004
6	0.708	0.899	0.943	0.962	1.021	0.987	0.991	1.001	1.028	1.004
7	0.971	0.863	1.089	1.018	1.001	0.999	1.008	0.992	1.017	0.999
8	0.946	0.948	0.972	0.981	0.964	0.984	0.992	0.996	0.997	1.001
9	1.116	1.024	0.897	0.967	0.970	1.017	1.000	1.011	1.010	0.994
10	0.901	0.951	0.973	0.911	0.859	0.925	0.976	0.988	1.005	0.996
11	0.907	0.889	1.005	0.987	0.959	0.970	0.964	0.982	0.979	0.986
12	0.874	0.926	0.936	0.945	0.948	0.975	0.991	0.986	0.991	0.993
13	0.978	1.057	0.985	0.964	0.996	1.009	1.000	1.002	0.998	1.002
14	0.955	0.895	0.967	0.967	0.961	0.979	0.988	0.989	0.994	0.988
15	1.048	0.892	1.016	0.967	0.958	0.976	0.999	1.005	0.968	0.993
16	0.689	0.972	1.110	1.137	1.028	0.926	0.987	1.002	0.988	1.021
17	0.973	0.904	0.975	0.939	0.943	0.970	0.985	1.009	0.984	1.012
18	0.856	0.924	0.941	0.773	0.884	0.988	0.956	0.965	0.963	0.977
19	0.795	1.217	0.988	1.019	1.003	1.010	0.982	1.017	1.044	0.960
20	1.024	0.710	0.890	1.025	1.024	1.044	1.049	1.016	0.995	0.984
21	0.957	0.793	1.121	1.005	1.003	0.972	0.973	0.991	0.986	0.992
μ	0.920	0.949	0.997	0.975	0.974	0.984	0.990	0.998	0.997	0.996

Table 40: Mean for the ratio of Sum of Absolute Residuals of D-Prior to OLS estimators for the 21 sectors for 10 different cases, non-IPO's

	a	b	c	d	e	f	g	h	i	j
1	0.889	0.998	1.117	1.11	1.006	0.96	1.006	1	1.006	0.988
2	0.833	0.985	1.157	1.002	0.986	0.982	0.993	0.992	0.982	0.995
3	0.827	1.136	0.969	0.966	0.975	0.988	0.973	1.008	1.012	0.995
4	0.828	1.075	1.008	0.969	0.942	1.083	1.004	1.000	1.013	1.009
5	0.931	0.956	0.875	0.914	0.998	0.943	0.923	0.998	1.012	1.009
6	0.714	1.017	0.921	0.91	1.072	0.99	0.987	0.999	1.036	1.016
7	0.877	0.824	1.495	1.001	1.006	0.996	0.989	0.981	1.021	0.991
8	0.934	0.918	0.99	0.973	0.949	0.967	0.986	0.99	0.989	0.994
9	1.543	1.092	0.855	0.936	0.95	0.988	0.989	1.019	0.992	0.981
10	0.774	0.901	0.91	0.937	0.892	1.023	0.998	0.987	1.021	0.999
11	0.826	0.836	1.15	1.007	0.948	0.977	0.963	0.969	0.964	0.981
12	0.82	0.902	0.932	0.968	0.934	0.981	0.991	0.978	0.984	0.985
13	0.981	1.156	0.977	0.929	0.971	0.982	0.988	1.007	0.997	0.999
14	0.892	0.802	0.948	0.961	0.957	0.983	0.991	0.971	0.979	0.974
15	1.185	0.849	1.04	0.935	0.946	0.995	0.985	1.017	0.941	0.997
16	0.632	0.938	1.216	1.097	0.975	0.928	0.962	1.000	0.974	1.102
17	0.934	0.857	0.962	0.857	0.885	0.945	0.989	1.015	0.975	1.018
18	0.819	1.002	0.998	0.679	0.925	1.024	1.002	0.999	0.986	1.011
19	0.884	1.329	0.867	1.058	0.998	0.998	0.992	1.03	1.069	0.964
20	1.117	0.59	0.971	0.964	1.008	1.094	1.049	0.998	0.989	0.98
21	0.803	0.738	1.079	1.089	1.03	0.947	0.912	0.975	1.001	0.986
μ	0.907	0.948	1.021	0.965	0.969	0.989	0.984	0.997	0.997	0.999

Table 41: Mean for the ratio of Sum of Squared Residuals of D-Prior to OLS estimators for the 21 sectors for 10 different cases, non-IPO's

	a	b	c	d	e	f	g	h	i	j	μ
3	3.29	0.71	1.56	-0.39	3.09	-2.29	1.97	0.61	-0.52	-0.81	0.72
4	2.56	9.35	6.77	-0.52	4.56	0.06	1.13	0.77	1.26	1.54	2.75
5	23.68	9.10	-1.79	7.35	4.44	1.06	0.88	3.03	2.03	1.32	5.11
6	16.81	8.01	13.12	22.50	12.16	0.61	1.73	5.56	1.29	-0.50	8.13
7	8.48	6.90	-1.25	2.26	2.43	3.58	1.18	-0.02	0.18	0.31	2.41
8	17.58	20.95	0.53	3.94	6.99	3.61	13.55	1.35	0.83	-1.94	6.74
9	41.47	24.84	11.58	4.51	13.24	19.21	5.44	4.13	8.15	0.35	13.29
10	23.49	25.67	-7.83	5.42	14.00	14.26	6.46	6.07	12.32	6.66	1.065
11	12.24	4.95	-9.29	8.41	0.99	-2.03	2.70	-0.81	1.59	-0.87	1.79
12	36.42	8.50	3.46	-1.01	0.70	0.25	-2.10	-0.61	-0.87	-2.30	4.24
13	-1.65	27.71	40.50	13.14	28.59	12.30	-2.97	2.45	3.29	2.37	12.57
14	29.87	10.97	0.32	5.83	12.57	-0.74	-2.25	0.78	0.28	0.43	5.81
15	36.12	20.73	3.53	27.77	33.46	8.62	2.01	1.83	-1.61	1.06	13.35
16	27.63	9.47	-0.11	-1.58	0.73	0.41	-0.43	-1.83	1.40	1.26	3.70
17	3.31	-21.95	-1.45	-4.48	6.06	3.54	0.82	1.91	0.80	-0.18	-1.16
μ	18.75	11.06	3.98	6.21	9.60	4.16	2.01	1.68	2.03	0.58	6.01

Table 42: Mean for the ratio of the difference of D-Prior and OLS to OLS errors, for the 21 sectors for 10 different cases, IPO's

	a	b	c	d	e	f	g	h	i	j
-100-90	6	3	4	2	2	1	0	0	0	0
-90-80	1	0	1	0	0	0	0	0	0	0
-80-70	0	0	2	0	0	0	0	0	0	0
-70-60	1	0	1	0	0	0	0	0	0	0
-60-50	1	2	2	0	0	0	0	0	0	0
-50-40	6	1	5	3	0	0	1	0	0	0
-40-30	7	6	3	4	0	1	0	0	0	1
-30-20	11	15	14	6	1	1	0	0	0	0
-20-10	36	33	35	29	12	3	4	0	2	4
-10-0	98	136	179	181	185	238	231	232	262	255
0-10	213	211	217	261	318	316	338	344	309	320
10-20	60	79	63	64	48	19	7	6	8	1
20-30	52	43	24	13	15	3	1	2	2	0
30-40	29	21	13	7	3	2	3	0	3	1
40-50	18	13	4	8	1	1	0	2	0	2
50-60	20	12	9	4	1	0	0	0	0	1
60-70	9	8	3	2	0	1	1	1	0	0
70-80	5	0	2	1	0	1	0	0	1	1
80-90	10	2	2	2	1	0	1	0	0	1
90-100	4	2	4	0	0	0	0	0	0	0

Table 43: Histogram for the difference of OLS minus D-Prior divided by their sum in percentage for the non-IPO's

	a	b	c	d	e	f	g	h	i	j
-100-90	0	0	0	0	0	0	0	0	0	0
-90-80	0	0	0	0	0	0	0	0	0	0
-80-70	0	0	0	0	0	0	0	0	0	0
-70-60	0	0	0	0	0	0	0	0	0	0
-60-50	0	0	0	0	0	0	0	0	0	0
-50-40	0	0	1	0	0	0	0	0	0	0
-40-30	0	0	1	0	0	0	0	0	0	0
-30-20	0	0	0	1	0	0	0	0	0	0
-20-10	1	3	1	0	0	0	0	0	0	0
-10-0	2	2	4	6	4	4	6	5	4	7
0-10	6	7	8	5	10	13	11	11	12	9
10-20	0	2	0	2	1	0	0	0	0	1
20-30	2	1	1	2	1	0	0	0	1	0
30-40	1	0	1	0	1	0	0	0	0	0
40-50	2	1	0	0	0	0	0	1	0	0
50-60	1	0	0	0	0	0	0	0	0	0
60-70	1	0	0	1	0	0	0	0	0	0
70-80	0	0	0	0	0	0	0	0	0	0
80-90	0	0	0	0	0	0	0	0	0	0
90-100	1	1	0	0	0	0	0	0	0	0

Table 44: Histogram for the difference of OLS minus D-Prior divided by their sum in percentage for the IPO only for the fifth joining firms

	a	a	b	b	c	c	d	d	e	e	f	f	g	g	h	h	i	i	j	j
	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>
1	1	2	2	2	4	0	4	0	4	0	3	1	3	1	1	3	1	3	1	3
2	7	2	1	8	4	4	4	5	2	7	7	2	8	1	6	3	4	5	5	4
3	2	1	3	0	3	0	2	1	3	0	1	2	2	1	2	1	3	0	1	2
4	1	2	0	3	1	2	0	3	2	1	3	0	1	2	2	1	2	1	1	2
5	2	1	2	1	3	0	1	2	3	0	2	1	2	1	2	1	2	1	2	1
6	1	1	2	0	0	2	2	0	2	0	0	2	1	1	1	1	1	1	2	0
7	2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	7	3	8	3	7	5	10	2	9	3	6	7	9	4	8	4	9	4	6	7
9	3	2	4	1	4	1	1	4	4	1	3	2	3	2	3	2	4	1	3	2
10	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1
11	6	2	7	1	5	3	6	2	7	1	6	3	5	4	3	6	6	3	6	3
12	10	3	11	3	9	6	10	5	10	5	9	6	12	3	12	3	13	2	6	9
13	3	0	3	0	1	2	1	2	1	2	3	0	3	0	1	2	2	1	2	1
14	4	0	3	1	2	2	4	0	4	0	2	2	3	1	1	3	3	1	2	2
15	2	1	1	2	1	2	1	2	3	0	3	0	3	0	3	0	2	1	3	0
16	1	1	0	2	0	2	0	2	1	1	1	1	1	1	1	1	1	1	1	1
17	3	1	3	1	2	2	3	1	3	1	2	1	1	3	1	3	4	0	2	2
18	1	3	3	1	2	2	3	1	4	0	4	0	3	1	3	1	3	1	3	1
19	1	0	0	1	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1
20	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
21	0	1	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	1	0
sum	59	26	56	32	52	37	57	33	67	23	59	32	64	28	53	38	63	29	49	43

Table 45: Number of cases where D-Prior is better than OLS, IPO

	a	a	b	b	c	c	d	d	e	e	f	f	g	g	h	h	i	i	j	j
	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>
1	1	2	1	1	4	0	3	0	3	0	2	0	1	0	1	0	0	1	0	1
2	4	2	1	4	1	1	2	3	1	1	2	0	1	0	1	1	1	0	2	1
3	2	1	3	0	3	0	2	1	2	0	1	0	1	0	1	0	0	0	0	0
4	1	1	0	3	0	2	0	2	1	1	0	0	1	1	0	0	0	0	0	0
5	1	1	2	1	1	0	1	1	2	0	1	0	1	0	1	0	1	0	0	0
6	0	0	1	0	0	0	1	0	1	0	0	2	1	0	1	0	0	0	0	0
7	1	0	1	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
8	6	2	7	1	5	2	9	2	6	1	5	3	4	3	5	1	4	1	2	1
9	2	1	2	0	4	1	1	0	1	0	1	2	1	0	1	0	0	0	2	0
10	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
11	6	0	5	0	4	1	4	2	3	1	4	2	1	2	0	1	0	1	1	1
12	5	2	8	2	2	4	5	3	7	1	2	1	4	0	3	1	2	0	1	1
13	3	0	2	0	1	1	1	2	0	2	3	0	0	0	0	0	0	0	2	0
14	2	0	3	1	2	1	2	0	3	0	1	1	2	0	1	2	0	0	0	1
15	1	0	1	1	1	1	1	1	3	0	2	0	1	0	0	0	0	1	0	0
16	1	1	0	0	0	2	0	1	0	0	0	1	0	1	0	0	1	0	1	0
17	2	0	1	1	2	2	3	1	3	1	1	1	1	0	1	0	0	0	1	0
18	1	3	1	1	2	1	2	1	4	0	2	0	3	1	1	0	1	0	2	1
19	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
20	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
21	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0
sum	41	17	41	16	36	20	40	21	43	8	28	15	24	8	19	6	11	5	15	7

Table 46: Number of cases where D-Prior is at least 5 percent better than OLS, IPO

	a	a	b	b	c	c	d	d	e	e	f	f	g	g	h	h	i	i	j	j
	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>
1	11	3	9	5	8	6	8	6	7	7	9	5	7	7	7	7	7	7	8	6
2	36	18	30	24	27	27	29	25	33	20	30	21	29	25	33	18	26	20	23	24
3	7	2	4	5	6	3	5	4	6	3	6	3	6	3	2	7	2	6	4	5
4	6	2	4	4	4	4	7	2	8	1	6	3	6	3	5	4	4	5	3	6
5	8	1	5	4	6	3	5	4	7	2	7	2	8	1	5	4	5	4	4	5
6	4	1	4	1	4	1	4	1	2	3	3	2	4	1	1	4	1	4	1	4
7	3	2	5	0	3	2	3	2	2	3	3	2	1	4	3	2	2	3	3	2
8	64	36	64	38	58	45	56	48	62	42	62	42	52	52	53	51	55	48	58	44
9	12	7	9	11	16	4	16	4	16	4	11	9	11	9	8	12	10	10	11	9
10	2	0	1	1	1	1	2	0	2	0	1	1	2	0	2	0	1	1	1	1
11	36	15	40	13	34	19	33	20	38	15	29	25	31	23	26	28	26	27	27	27
12	104	24	83	48	88	46	84	49	93	40	83	51	95	40	88	46	73	60	78	56
13	5	4	5	4	5	4	5	4	4	5	5	4	4	5	4	5	5	4	4	5
14	8	6	10	4	8	6	8	5	10	4	9	5	8	6	9	5	6	8	8	6
15	4	5	7	2	5	4	6	3	7	2	7	2	5	4	4	5	6	3	6	3
16	4	1	4	1	0	5	2	3	2	3	5	0	3	2	2	3	3	2	2	3
17	6	8	10	4	10	4	8	6	9	5	9	5	7	6	7	7	10	4	5	9
18	10	4	10	4	9	5	13	1	14	0	9	5	10	4	10	4	12	2	11	3
19	2	0	1	1	1	1	1	1	1	1	0	2	2	0	1	1	1	1	2	0
20	1	1	2	0	2	0	1	1	1	1	0	2	0	2	0	2	1	1	2	0
21	2	0	2	0	1	1	1	1	1	1	2	0	2	0	1	1	2	0	1	1
sum	335	140	309	174	296	191	297	190	325	162	296	191	293	197	271	216	258	220	262	219

Table 47: Number of cases where D-Prior is better than OLS, non-IPO

	a	a	b	b	c	c	d	d	e	e	f	f	g	g	h	h	i	i	j	j
	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>	<	>
1	4	2	3	2	3	3	4	3	3	3	2	0	0	1	0	1	0	1	0	0
2	26	8	25	10	15	14	11	5	8	2	9	2	8	5	4	0	6	0	3	1
3	3	0	2	2	2	3	4	3	1	1	3	1	1	0	0	0	0	1	0	2
4	6	1	2	3	2	3	5	0	3	0	1	2	0	0	0	0	0	2	0	1
5	3	1	4	3	2	2	3	1	1	0	4	0	2	0	0	0	0	0	0	0
6	4	1	4	1	3	1	2	0	0	2	0	0	0	0	0	0	0	1	0	0
7	2	1	4	0	1	2	0	1	0	0	0	0	0	0	1	0	0	1	0	0
8	44	15	46	21	32	29	29	28	24	10	19	6	14	6	9	1	9	7	1	7
9	9	4	4	3	13	2	7	4	5	0	2	2	1	1	0	2	1	4	3	1
10	1	0	1	0	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0
11	31	9	33	10	20	10	18	10	16	1	11	6	7	3	5	3	9	7	5	6
12	72	11	58	18	59	24	55	18	51	9	27	12	14	8	16	2	12	4	6	5
13	3	2	2	3	3	3	3	4	3	4	0	1	1	1	0	0	0	0	0	0
14	5	4	8	3	5	3	7	4	4	0	3	1	2	0	1	0	3	0	1	0
15	4	3	4	1	0	1	3	0	5	0	3	1	0	1	2	1	2	0	0	0
16	4	1	2	0	0	2	0	3	2	2	2	0	0	0	0	0	1	1	0	1
17	4	5	8	1	7	3	5	1	6	0	3	0	2	1	0	2	2	0	0	1
18	9	4	7	3	8	4	12	1	10	0	6	4	6	0	6	1	5	0	1	1
19	2	0	0	1	1	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0
20	0	1	2	0	1	0	1	1	0	1	0	0	0	1	0	0	0	0	0	0
21	1	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sum	237	73	221	85	178	111	171	88	143	35	96	38	58	28	44	14	50	30	20	26

Table 48: Number of cases where D-Prior is at least 5 percent better than OLS, non-IPO

3.6 Concluding Remarks

A few points requires further mentioning:

1. In general, Empirical Bayes does better than OLS.
2. Empirical Bayes method is more successful when the estimation period is small.
3. The technique can successfully be used when there are groups of data moving together.
4. The IPO's joining a higher number of firms' sectors lead to more successful regression results of Empirical Bayes.
5. Hierchical Bayes may also be used to alleviate the problems faced by Bayes method in general, but its application requires more complicated calculations and computer program coding.

References

[1] Abdel-Mottaleb, M.,Chellapa, R., and Rosenfeld, A. *Binocular motion stereo with independently moving objects* Report CAR-TR-679, Center for Automation Research, University of Maryland, 1993.

[2] Amegbeto, K.N., and A.M. Featherstone *Risk, costs and the choice of market return index* Journal of Agricultural and Resource Economics, Vol 17, Iss 1, pp 80-87, 1992

[3] Atkinson, A.C. *Fast very robust methods for the detection of multiple outliers* J. Amer. Statist. Assoc., 89, 1329-1339,1994

[4] Benderly, J., B. Zwick *Inflation, real balances, output and real stock returns* American Economic Review, 75, 1115-1125, 1985

[5] Bessembinder, H. *Systematic-risk, hedging pressure, and risk premiums in futures markets* Review of Financial Studies, Vol 5, Iss 4, pp 637-667, 1992

[6] Bjornson, B., and R. Innes *Another look at returns to agricultural and nonagricultural assets* American Journal of Agricultural Economics, Vol 74, Iss 1, pp 109-119, 1992

[7] Breusch, T. and Pagan, A. *A simple test for heteroskedasticity and random coefficient variation* Econometrica, 47, 1287-1294, 1979

[8] Butler, R.W., Davies, P.L., and Jhun, M. *Asymptotics for the minimum covariance determinant estimator* Ann. Statist., 21, 1385-1400, 1993

[9] Chesher, A. and Jewitt, I. *The bias of a heteroskedasticity-consistent covariance matrix estimator* Econometrica, 55, 1217-1222, 1987

[10] Chork, C.Y. *Unmasking multivariate anomalus observations in exploration geochemical data from sheeted-vein tin mineralization near Emmaville, N.S.W.,Australia* J. Geochemical Exploration, 37, 205-223, 1990

[11] Chork, C.Y., and Salminen, R. *Interpreting geochemical data from Outokumpu, Finland: MVE-robust factor analysis* J. Geochemical Exploration, 48, 1-20, 1993

[12] Christmann, A. *Least median of weighted squares in logistic regression with large strata* Biometrika. 81, 413-417, 1994

[13] Coakley, C.W., and Hettmansperger, T.P. *A bounded influence, high breakdown, efficient regression estimator* J. Amer. Statist. Assoc., 88,872-880,1993

[14] Croux, C., Rousseeuw, P.J., and Hössjer, O. *Generalized S-estimators* J. Amer. Statist. Assoc., 89,1271-1281,1994

[15] Dallal, G.E., and Rousseeuw, P.J. *LMSMVE: a program for least median of squares regression and robust distances* Computers and Biomed Research, 25, 384-391, 1992

[16] Daveni, R.A., and A.Y. Ilinitch *Complex patterns of vertical integration in the forest products industry - systematic and bankruptcy risks* Academy of Management Journal, Vol 35, Iss 3, pp 596-625, 1992

[17] Davies, L. *Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices* Ann. Statist., 15, 1269-1292, 1987

[18] Davies, L. *The asymptotics of Roesseeuw's minimum volume elipsoid estimator* Ann., Statist., 20,1828-1843, 1992

[19] Davies, L. *An efficient Fréchet differentiable high breakdown multivariate location and dispersion estimator* J. Multivar. Anal., 40,311-327, 1992

[20] De Long, J.,B., and Summers, L.,H., *Equipment investment and economic growth,* Quarterly Journal of Economics, 1991, 106, 445-501

[21] Donoho, D.L., Huber, P.J. *The notion of breakdown point* In a Festschrift for Erich Lehmann, P. Bickel, K. Doksum, and J.L. Hodges, eds. Wadsworth, Belmont, CA, 1983

[22] Drew, M.S. *Robust specularity detection from a single multi-illuminant color image* Computer Vision, Graphics and Image Processing: Image Understanding, 59, 320-327, 1994

[23] Efron, B., and Tibshirani, R.J. *An introduction to the bootstrap* Chapman and Hall, New York, 1993

[24] Efron, B. *The jacknife, the bootstrap and other resampling plans* Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982

[25] Eicker, F. *Asymptotic normality and consistency of the least squares estimators* Annals of Mathematical Statistics, Vol 34, 447-456, 1963

[26] Fama, E.F. *Stock returns, real activity, inflation and money* American Economic Review, 71, pp 545-564, 1981

[27] Ferson, W.E., and S.R. Foerster *Finite-sample properties of the generalized-method of moments in tests of conditional asset pricing-models* Journal of Financial Economics, Vol 36, Iss 1, pp 29-55, 1994

[28] Freedman, D. A. and Peters, S. C. *Bootstrapping a regression equation: some empirical results* Journal of American Statistical Association, Vol 79, 97-106, 1984

[29] Glejser, H. *A new test for heteroskedasticity* Journal of American Statistical Association, Vol 64, 316-323, 1969

[30] Goldfeld, S. M., and Quandt R.E., *nonlinear methods of econometrics* North-Holland Publishing Company, Amsterdam, 1972

[31] Gray J.,B., *Graphics for regression diagnostics*, American Statistical Association Proceedings of the Statistical Computing Section, 1985, 102-107

[32] Greene, W. H. *Econometric analysis* MacMillan Publishing Company, NY, 1990

[33] Hadi, A.S. *Identifying multiple outliers in multivariate data* J. Royal Statist. Soc., Series B, 54, 761-771, 1992

[34] Hampel, F.R. *A general qualitative definition of robustness* Ann. Math. Statist., 42, 1887-1896,1971

[35] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. *Robust statistics: The approach based on influence functions* Wiley, New York, 1986

[36] Handa, P., S.P. Kothari, and C. Wasley *Sensitivity of multivariate tests of the capital asset-pricing model to the return measurement interval* Journal of Finance, Vol 48, Iss 4, pp 1543-1551, 1993

[37] Harvey, A., and Philips, G., *A comparison of the power of some tests for heteroskedasticity in the general linear model* Journal of Econometrics, 2, 307-316, 1974

[38] Hawkins, D.M. *The feasible set algorithm for least median of squares regression* Comp. Statist. Data Anal., 16, 81-101, 1993

[39] Hawkins, D.M., and Simonoff, J.S. *AS282: High breakdown regression and multivariate estimation* Applied Statist., 42, 423-432, 1993

[40] He, X., and Simpson, D.G. *Lower bounds for contamination bias: globally versus locally linear estimation* Ann. Statist., 21, 314-337, 1993

[41] Hinkley, D. V. *Jacknifing in unbalanced situations* Technometrics, Vol 19, 285-292, 1977

[42] Horn, S. D, Horn, R. A. and Duncan, D. B. *Estimating heteroskedastic variances in linear models* Journal of the American Statistical Association, Vol 60, 539-647, 1975

[43] Hössjer, O., Rousseeuw, P.J., and Croux, C. *Asymptotics of the repeated median slope estimator* Ann. Statist., 22, 1478-1501, 1994

[44] Huber, P.J. *Robust regression: asymptotics, conjectures and Monte Carlo* Ann. Statist., 1. 799-821, 1973

[45] Huber, P.J. *Robust statistics* Wiley, New York, 1981

[46] Jolion, J.M., Meer, P., and Bataouche, S. *Robust clustering with apllications in computer vision* IEEE Trans. Pattern Anal. and Machine Intell., 13, 791-802, 1991

[47] Jorion, P., and A. Giovannini *Time-series tests of a non-expected-utility model of asset pricing* European Economic Review, Vol 37, Iss 5, pp 1083-1100, 1993

[48] Judge, G.R, Hill, R., Griffiths, W., Lee, T., and Lutkepol, H. *An introduction to the theory and practice of econometrics* New York, Wiley, 1982

[49] Jurecková, J. *Nonparametric estimate of regression coefficients* Ann. Math. Stat., 42, 1328-1338, 1971

[50] Kaufman, L., Hopke, P.K., and Rousseeuw, P.J. *Using a parallel computer system for statistical resampling methods* Comput. Statist. Quarterly, 2, 129-141, 1988

[51] Koenker, R. and basset G. *Robust tests for heteroskedasticity based on regression quantiles* Journal of Econometrics, 50, 43-61, 1982

[52] Koenker, R., and Portnoy, S. *L-estimation fot linear models* J. Amer. Statist. Assoc., 82, 851-857, 1987

[53] Kumar, R., and Hanson, A.R. *Robust methods for pose determination* Proceedings of the NSF/ARPA Workshop on Performance versus Methodology in Computer Vision, 41-57, 1994

[54] Lin, W.T., Y.H. Chen, and J.C.G. Boot *The dynamic and stochastic-instability of betas-implications for forecasting stock returns* Journal of Forecasting, Vol 11, Iss 6, pp 517-541, 1992

[55] Lopuhaä, H.P., and Rousseeuw, P.J. *Breakdown points of affine equivariant estimators of multivariate location and covariance matrices* Ann. Statist., 19, 229-248, 1991

[56] Lubatkin, M., H. Merchant, and N. Srinivasan *Construct-validity of some unweighted product-count diversification measures* Strategic Management Journal, Vol 14, Iss 6, pp 433-449, 1993

[57] Luoma,M., T. Martikainen, J. Perttunen, and S. Pynnonen *Different beta-estimation techniques in infrequently traded and inefficient stock markets* Omega-International Journal of Management Science, Vol 22, Iss 5, pp 471-476, 1994

[58] MacKinnon, J. G. and White H. *Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties* Journal of Econometrics, Vol 29, 305-325, 1985

[59] Markatou, M., and He, X. em Bounded influence and high breakdown point testing procedures in linear models J. Amer. Statist. Assoc., 89, 543-549, 1994

[60] Mankiw, N., D. Romer, D. Weil, *A Contribution to the empirics of economic growth,* Quarterly Journal of Economics, CVII (1992), 407-437

[61] Maronna, R.A., W.A. Stahel and V.J. Yohai, *Bias-robust estimators of multivariate scatterBased on projections*, Journal of Multivariate Analysis, 1992, 42, 141-161

[62] Maronna, R.A., Stahel, W.A., and Yohai, V.J. *Bias-robust estimators of multivariate scatter based on projections* J. Multivar. Anal.. 42, 141-161, 1992

[63] Maronna, R.A., and Yohai, V.J. *Bias-robust estimates of regression based on projections* Ann. Statist., 21, 965-990, 1993

[64] Martin, R.D., Yohai, V.J., and Zamar, R.H. *Min-max robust regression* Ann. Statist., 17, 1608-1630, 1989

[65] Martin, R.D., and Zamar, R.H. *Bias robust estimation of scale* Ann. Statist., 21, 991-1017, 1993

[66] Massart, D.L., Kaufman, L., Rousseeuw, P.J., and Leroy, A. *Least median of squares:a robust method for outlier and model error detection in regression and calibration* Analytica Chimica Acta, 187, 171-179, 1986

[67] Mcdonald,J.B., and Nelson , R.D. *Beta-estimation in the market model - skewness and leptokurtosis* Communications in Statistics-Theory and Methods, Vol 22, Iss 10, pp 2843-2862, 1993

[68] Meer, P., Mintz, D., Rosenfeld, A., and Kim, D.Y. *Robust regression methods in computer vision: areview* Int. J. Computer Vision, 6, 59-70, 1991

[69] Messer, K. and White H. *A Note on computing the heteroskadasticity-consistent covariance matrix using instrumental variable techniques* Oxford Bulletin of Economics and Statistics, Vol 46, 181-184, 1984

[70] Mili, L., Phaniraj, V., and Rousseeuw, P.J. *Least median of squares estimation in power systems (with discussion)* IEEE Trans. on Power systems, 6, 511-523, 1991

[71] Mili, L., Cheniae, M.G., Vichare, N.S., and Rousseeuw, P.J. *Robust state estimation based on projection statistics* IEEE Trans. on Power systems, 10, 1995

[72] Nonneman, W, and P. Vanhoudt, *A Further augmentation of the solow model and the empirics of economic growth for OECD countries*, Quarterly Journal of Economics, August 1996, vol 79, 943-953

[73] Montgomery, D.C., and Peck, A.E., *Introduction to linear regression analysis*, John Wiley and Sons, 1992, New York

[74] Ronchetti, E., and Staudte, R.G. *A robust version of Mallows's* C_p J. Amer. Statist. Assoc., 89, 550-559, 1994

[75] Roth, G., and Levine, M.D. *Extracting geometric primitives* Computer Vision, Graphics and Image Processing:Image Understanding 58, 1-22, 1993

[76] Rousseeuw, P.J. *Regression techniques with high breakdown point* Inst. Math. Statist. Bull., 12, 155, 1983

[77] Rousseeuw, P.J. *Multivariate estimation with high breakdown point* Inst. Math. Statist. Bull., 12, 234, 1983

[78] Rousseeuw, P.J. *Least median of squares regression* J. Amer. Statist. Assoc., 79, 871-880, 1984

[79] Rousseeuw, P.J. *Multivariate estimation with high breakdown point* In Mathematical Statistics and Applications, Vol. B, W. Grossmann, G. Pflug, I. Vincze and W.Wertz, eds. Reidel, Dordrecht, 283-297, 1985

[80] Rousseeuw, P.J. *A resampling design for computing high-breakdown regression* Stat. Prob. Letters, 18, 125-128, 1993

[81] Rousseeuw, P.J. *Unconventional features of positive-breakdown estimators* Stat. Prob. Letters, 19, 417-431, 1994

[82] Rousseeuw, P.J., and Croux, C. *Alternatives to the median absolute deviation* J. Amer. Statist. Assoc, 88, 1273-1283, 1993

[83] Rousseeuw, P.J., Croux, C., and Hössjer, O. *Sensitivity functions and numerical analysis of the repeated median slope* Computational Statistics, 10, 71-90, 1995

[84] Rousseeuw, P.J., and Leroy, A.M. *Robust regression and outlier detection* Wiley-Interscience, New York, 1987

[85] Rousseeuw, P.J., Netanyahu, N.S., and Mount, D.M. *New statistical and computaional results on the repeates median line* In New Directions in Statistical Data Analysis and Robustness, S. Morgenthaler, E. Ronchetti, and W.A. Stahel. eds/ Birkhäuser, Basel, 177-194, 1993

[86] Rousseeuw, P.J., van Zomeren, B.C. *Unmasking multivariate outliers and leverage points* J. Amer. Statist. Assoc., 85, 633-651, 1990

[87] Rousseeuw, P.J., van Zomeren, B.C. *A comparison of somequick algorithms for robust regrssion* Comput. Statist. and Data Anal., 14, 107-116, 1992

[88] Rousseeuw, P.J., and Wagner, J. *Robust regression with a distributed intercept using least median of squares* Comput. Statist. and Data Anal., 17, 65-76 ,1994

[89] Rousseeuw, P.J., and Yohai, V.J. *Robust regression by means of S-estimators* In Robust and Ninlinear Time Series Analysis Lecture Notes in Statistics No:26, J. Franke, W. Härdle and R.D. Martin, eds., Springer, New York, 256-272, 1984

[90] Ruppert, D. *Computing S-estimators for regression and multivariate location/dispersion* J. Comput. Graph. Stat., 1, 253-270, 1992

[91] Seaver, B.L., and Triantis, K.P. *The impact of outliers and leverage points for technical efficiency measurement using high breakdown procedures* Management Science, 1995

[92] Siegel, A.F. *Robust regression using repeted medians* Biometrika, 69, 242-244, 1982

[93] Simpson, D.G., Ruppert, D., and Carroll, R.J. *On one-step GM-estimates and stability of infrences in linear regression* J. Amer. Statist. Assoc., 87, 439-450, 1992

[94] Simpson, D.G., and Yohai, V.J. *Functional stability of one-step GM-estimators in linear regression* Technical Report 71, Department of Statistics, University of Illinois, Champaign, IL, 1993

[95] Sinha, S.S, and Schunck, B.G. *A two-stage algorithm for discontinuity-preserving surface reconstruction* IEEE Trans. on Pattern Anal and Machine Intell., 14, 36-55, 1992

[96] Smith, P.N. *Modeling risk premia in international asset markets* European Economic Review, Vol 37, Iss 1, pp 159-176, 1993

[97] Souvaine, D.L., and Steele, J.M. *Time- and space-efficient algorithms for least median of squares regression* J. Amer. Statist. Assoc., 82, 794-801, 1987

[98] *S-PLUS User's manual version 3.2* Statistical Sciences, Seattle: Statsci, a division of Mathsoft, Inc., 1993

[99] Stewart, C.V. *A new robust operator for computer vision: theoretical analysis*, Proceedings IEEE Conference on Computer Vision and Pattern Recognition, 1994, 1-8

[100] Stromberg, A.J. *Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression* SIAM J. Sci. Comput., 14. 1289-1299, 1993

[101] Stromberg, A.J. *Computation of high breakdown nonlinear regression* J. Amer. Statist. Assoc, 88, 237-244, 1993

[102] Stromberg, A.J., and Ruppert, D. *Breakdown in nonlinear regression* J. Amer. Statist. Assoc, 87. 991-997, 1992

[103] Tansel, A. *Cigarette demand, health scares and education in Turkey* Applied Ecocnics, 1993, 25, 521-529

[104] Thompson, W.B., Lechleider, P., and Stuck, E.R. *Detecting moving objects using the rigidity constraint* IEEE Trans. On Pattern Anal. And Machine Intell., 15, 162-166,1993

[105] Thursby, J. *Misspecification, heteroskedasticity, and the Chow and Goldfeld-Quandt tests* Review of Economics and Statistics, 64, 314-321, 1982

[106] Wang, C.M., D.F. Vecchia, M. Young, N.A. Brillant *Software for performing gray-scale measurements of optical fiber end faces*, Technical Report, 1370, National Institute of Standards and Technology, 1994, Gaithersburg, MD

[107] White, H. *A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity* Econometrica, Vol 48, 817-838, 1980

[108] Woodruff, D.L., and Rocke, D.M. *Heuristic search algorithms for the minimum volume elipsoid* J. Comput. Graph. Statist., 2, 69-95, 1993

[109] Wu, C. F. J. *Jacknife, bootstrap and other resampling methods in regression analysis* Annals of Statistics, Vol 14, No 4, 1261-1295, 1986

[110] Yohai, V.J. *High breakdown point and high efficiency robust estimates for regression* Ann. Statist., 15, 642-656, 1987

[111] Yohai, V.J., and Zamar, R.H. *High breakdown point estimates of regression by means of the minimization of an efficient scale* J. Amer. Statist. Assoc., 83, 406-413d, 1987

[112] *Statistical foundations of econometric techniques* Academic Press, 1996

4 Appendix to the Chapters

4.1 Proof for the Bias of the Estimators

4.1.1 Proof for the Bias of the Eicker-White Estimator

Bias of Eicker-White estimator for the first entry is

$$
\begin{aligned}
B_{11}^{EW} &= E\hat{C}_{11} - C_{11} && (98)\\
&= E\frac{1}{T^2}\sum_{t=1}^{T} e_t^2 - \frac{1}{T^2}\sum_{t=1}^{T}\sigma_t^2 && (99)\\
&= \frac{1}{T^2}(E\sum_{t=1}^{T} e_t^2 - \sum_{t=1}^{T}\sigma_t^2) && (100)\\
&= \frac{1}{T^2}(\sum_{t=1}^{T} Ee_t^2 - \sum_{t=1}^{T}\sigma_t^2) && (101)\\
&= \frac{1}{T^2}[\sum_{t=1}^{T}(\sigma_t^2 - \frac{2}{T}(1+x_t^2)\sigma_t^2 + \frac{1}{T}(1+2x_tM(x,\sigma^2) \\
&\quad +x_t^2M(x^2,\sigma^2)) - \sigma_t^2)] && (102)\\
&= \frac{1}{T^2}[\sum_{t=1}^{T}(-\frac{2}{T}(1+x_t^2)\sigma_t^2 + \\
&\quad \frac{1}{T}(1+2x_tM(x,\sigma^2) + x_t^2M(x^2,\sigma^2)))] && (103)\\
&= \frac{1}{T^2}[-2-2M(x^2,\sigma^2)+1+M(x^2,\sigma^2)] && (104)\\
&= -\frac{1}{T^2}(1+M(x^2,\sigma^2)) && (105)
\end{aligned}
$$

Now, coming to the calculation of the $B_{12}^{EW} = B_{21}^{EW}$

$$
\begin{aligned}
Ex_te_t^2 &= Var(\sqrt{x_t}e_t^2) && (106)\\
&= x_t(1-h_{tt})^2\sigma_t^2 + x_t\sum_{j=1,j\neq t}^{T} h_{tj}^2\sigma_j^2 && (107)\\
&= x_t(1-2h_{tt}+h_{tt}^2)\sigma_t^2 + \frac{x_t}{T^2}\sum_{j=1,j\neq t}^{T}(1+x_tx_j)^2\sigma_j^2 && (108)\\
&= x_t\sigma_t^2 - 2x_th_{tt}\sigma_t^2 + x_th_{tt}^2\sigma_t^2 + \frac{x_t}{T^2}\sum_{j=1,j\neq t}^{T}(1+x_tx_j)^2\sigma_j^2 && (109)
\end{aligned}
$$

$$= \quad x_t\sigma_t^2 - \frac{2}{T}x_t(1+x_t^2)\sigma_t^2 + \frac{x_t}{T^2}\sum_{j=1}^{T}(1+x_tx_j)^2\sigma_j^2 \tag{110}$$

$$= \quad x_t\sigma_t^2 - \frac{2}{T}x_t\sigma_t^2 - \frac{2}{T}x_t^3\sigma_t^2 + \frac{1}{T^2}x_t\sum_{j=1}^{T}(1+x_tx_j)^2\sigma_j^2 \tag{111}$$

$$= \quad x_t\sigma_t^2 - \frac{2}{T}x_t\sigma_t^2 - \frac{2}{T}x_t^3\sigma_t^2 + \frac{x_t}{T^2}[T + 2x_tTM(x,\sigma^2) + x_t^2TM(x^2,\sigma^2)] \tag{112}$$

$$= \quad x_t\sigma_t^2 - \frac{2}{T}x_t\sigma_t^2 - \frac{2}{T}x_t^3\sigma_t^2 + \frac{x_t}{T} + \frac{2}{T}x_t^2M(x,\sigma^2) + \frac{1}{T}x_t^3M(x^2,\sigma^2) \tag{113}$$

$$\tag{114}$$

Using the above derivation bias of the off-diagonal entries is:

$$B_{12}^{EW} = B_{21}^{EW} \tag{115}$$

$$= E(\frac{1}{T^2}\sum_{t=1}^{T}x_te_t^2) - \frac{1}{T^2}\sum_{t=1}^{T}x_t\sigma_t^2 \tag{116}$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}E(x_te_t^2) - \frac{1}{T^2}\sum_{t=1}^{T}x_t\sigma_t^2 \tag{117}$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}(1-\frac{2}{T})x_t\sigma_t^2 + (-\frac{2}{T})x_t^3\sigma_t^2 + \frac{x_t}{T} + (\frac{2}{T})x_t^2M(x,\sigma^2) + (\frac{1}{T})x_t^3M(x^2,\sigma^2) - \sum_{t=1}^{T}x_t\sigma_t^2 \tag{118}$$

$$= \frac{1}{T^2}[-2M(x,\sigma^2) - 2M(x^3,\sigma^2) + 2M(x,\sigma^2) + SM(x^2,\sigma^2)$$

$$= (\frac{S}{T^2})M(x^2,\sigma^2) - \frac{2}{T^2}M(x^3,\sigma^2) \tag{119}$$

The preliminary calculation for the bias of the final term goes as follows

$$Ex_t^2e_t^2 = Var(x_te_t) \tag{120}$$

$$= x_t^2\sigma_t^2(1-\frac{2}{T}) - x_t^4\sigma_t^2(\frac{2}{T}) + \frac{1}{T}x_t^2 + +\frac{2}{T}x_t^3M(x,\sigma^2) + +\frac{1}{T^2}x_t^4M(x,\sigma^2) \tag{121}$$

$$\tag{122}$$

Now coming to the bias calculation

$$
\begin{aligned}
B_{22}^{EW} &= E\widehat{C} - C && (123)\\
&= E[\frac{1}{T^2}\sum_{t=1}^{T} x_t^2 e_t^2] - \frac{1}{T^2}\sum_{t=1}^{T} x_t^2 \sigma_t^2 && (124)\\
&= \frac{1}{T^2}[\sum_{t=1}^{T} E x_t^2 e_t^2 - \sum_{t=1}^{T} x_t^2 \sigma_t^2] && (125)\\
&= \frac{1}{T^2}[\sum_{t=1}^{T} x_t^2 \sigma_t^2 [\frac{T-2}{T}] - x_t^4 \sigma_t^2 (\frac{2}{T}) + \frac{1}{T} x_t^2 + \\
&\quad \frac{2}{T} x_t^3 M(x,\sigma^2) + \frac{1}{T} x_t^4 M(x^2,\sigma^2) - x_t^2 \sigma_t^2 && (126)\\
&= \frac{1}{T^2}[-2M(x^2,\sigma^2) - 2M(x^4,\sigma^2) + 1 + \\
&\quad 2SM(x,\sigma^2) + KM(x^2,\sigma^2)] && (127)\\
&= \frac{1}{T^2} + \frac{2S}{T^2} M(x,\sigma^2) + \\
&\quad \frac{K-2}{T^2} M(x^2,\sigma^2) - \frac{2}{T^2} M(x^4,\sigma^2) && (128)\\
& && (129)
\end{aligned}
$$

4.1.2 Proof for the Bias of H-HD estimator

$$E\frac{1}{1-h_{tt}}e_t^2 = E\alpha_t e_t^2 \quad (130)$$
$$= E((1+h_{tt})e_t^2) \quad (131)$$
$$= Ee_t^2 + E(h_{tt}e_t^2) \quad (132)$$

$$E(1+h_{tt})e_t^2 = \Lambda + h_{tt}\sigma_t^2 - \frac{2h_{tt}}{T}(1+x_t^2)\sigma_t^2$$
$$+\frac{h_{tt}}{T}[1+2x_tM(x,\sigma^2)+x_t^2M(x^2,\sigma^2)] \quad (133)$$
$$= \Lambda + \frac{1}{T}(1+x_t^2)\sigma_t^2 - \frac{2}{T^2}(1+x_t^2)(1+x_t^2)\sigma_t^2$$
$$+\frac{1}{T^2}(1+x_t^2)[1+2x_tM(x,\sigma^2)+x_t^2M(x^2,\sigma^2)] \quad (134)$$
$$= \Lambda + \frac{1}{T}\sigma_t^2 + \frac{1}{T}x_t^2\sigma_t^2 - \frac{2}{T^2}\sigma_t^2 - \frac{4}{T^2}x_t^2\sigma_t^2$$
$$-\frac{2}{T^2}x_t^4\sigma_t^2 + \frac{1}{T^2} + \frac{2}{T^2}x_tM(x,\sigma^2) + \frac{x_t^2}{T^2}M(x^2\sigma^2)$$
$$+\frac{x_t^2}{T^2} + \frac{2}{T^2}x_t^3M(x,\sigma^2) + \frac{1}{T^2}x_t^4M(x^2,\sigma^2) \quad (135)$$
$$= \Lambda + \sigma^2(\frac{1}{T}-\frac{2}{T^2}) + \frac{1}{T^2} + x_t^2\sigma_t^2(\frac{1}{T}-\frac{4}{T^2}) + x_t^2\frac{1}{T^2} + \frac{2}{T^2}x_tM(x,\sigma^2)$$
$$+\frac{x_t^2}{T^2}M(x^2,\sigma^2) - \frac{2}{T^2}x_t^4\sigma_t^2 + \frac{2}{T^2}x_t^3M(x,\sigma^2) + \frac{1}{T^2}x_t^4M(x^2,\sigma^2) \quad (136)$$

Using the above derivation bias of the first entry is:

$$B_{11}^{HD} = (\frac{1}{T^2}\sum_{t=1}^{T}Ee_t^2(1+h_{tt})) - \frac{1}{T^2}\sum_{t=1}^{T}\sigma_t^2 \quad (137)$$
$$= \frac{1}{T^2}[\sum_{t=1}^{T}Ee_t^2(1+h_{tt}) - \sigma_t^2] \quad (138)$$
$$= \frac{1}{T^2}\sum_{t=1}^{T}\sigma_t^2 - \frac{2}{T}\sigma_t^2 - \frac{2}{T}x_t^2\sigma_t^2 + \frac{1}{T} + \frac{2}{T}x_tM(x,\sigma^2)$$
$$+\frac{1}{T}x_t^2M(x^2,\sigma^2) + \sigma_t^2(\frac{T-2}{T^2}) + \frac{1}{T^2} + (\frac{T-4}{T^2})x_t^2\sigma_t^2$$
$$+\frac{x_t^2}{T^2} + \frac{2}{T^2}x_tM(x,\sigma^2) + \frac{x_t^2}{T^2}M(x^2,\sigma^2) - \frac{2}{T^2}x_t^4\sigma_t^2$$
$$+\frac{2}{T^2}x_t^3M(x,\sigma^2) + \frac{1}{T^2}x_t^4M(x^2,\sigma^2) - \sigma_t^2 \quad (139)$$
$$= \frac{1}{T^2}[T-2-2M(x^2,\sigma^2)+1+0+M(x^2,\sigma^2)+\frac{T-2}{T}+\frac{1}{T}$$

$$+(\frac{T-4}{T})M(x^2,\sigma^2)+\frac{1}{T}+0+\frac{1}{T}M(x^2,\sigma^2)-\frac{2}{T}M(x^4,\sigma^2)$$
$$+\frac{2}{T}SM(x,\sigma^2)+\frac{K}{T}M(x^2,\sigma^2)-T] \quad (140)$$
$$= \frac{1}{T^2}[M(x^2,\sigma^2)(\frac{K-3}{T})+M(x,\sigma^2)(\frac{2S}{T})$$
$$+M(x^4,\sigma^2)(\frac{-2}{T})] \quad (141)$$
$$= \frac{K-3}{T^3}M(x^2,\sigma^2)+\frac{2S}{T^3}M(x,\sigma^2)-\frac{2}{T^3}M(x^4,\sigma^2) \quad (142)$$
$$(143)$$

For the bias of the off-diagonal entries

$$B_{12}^{HD} = B_{21}^{HD} \quad (144)$$
$$= E(\frac{1}{T^2}\sum_{t=1}^{T}x_t e_t^2(1+h_{tt}))-\frac{1}{T^2}\sum_{t=1}^{T}x_t\sigma_t^2 \quad (145)$$
$$= \frac{1}{T^2}\sum_{t=1}^{T}Ex_t(1+h_{tt})e_t^2-\frac{1}{T^2}\sum_{t=1}^{T}x_t\sigma_t^2 \quad (146)$$
$$= \frac{1}{T^2}[\sum_{t=1}^{T}x_t\sigma_t^2-\frac{2}{T}x_t\sigma_t^2-\frac{2}{T}x_t^3\sigma_t^2+\frac{x_t}{T}$$
$$+\frac{2}{T}x_t^2M(x,\sigma^2)+\frac{1}{T}x_t^3M(x^2,\sigma^2)+x_t\sigma_t^2(\frac{T-2}{T^2})+\frac{x_t}{T^2}$$
$$+(\frac{T-4}{T^2})x_t^3\sigma_t^2+\frac{x_t^3}{T^2}+\frac{2}{T^2}x_t^2M(x,\sigma^2)+\frac{x_t^3}{T^2}M(x^2,\sigma^2)$$
$$-\frac{2}{T^2}x_t^5\sigma_t^2+\frac{2}{T^2}x_t^4M(x,\sigma^2)+\frac{1}{T^2}x_t^5M(x^2,\sigma^2)-x_t\sigma_t^2)] \quad (147)$$
$$= \frac{1}{T^2}[TM(x,\sigma^2)-2M(x,\sigma^2)-2M(x^3,\sigma^2)+0$$
$$+2M(x,\sigma^2)+SM(x^2,\sigma^2)+M(x,\sigma^2)\frac{T-2}{T}+0+\frac{T-4}{T}M(x^3,\sigma^2)$$
$$+\frac{S}{T}+\frac{2}{T}M(x,\sigma^2)+\frac{S}{T}M(x^2,\sigma^2)-\frac{2}{T}M(x^5,\sigma^2)$$
$$+\frac{2}{T}KM(x,\sigma^2)+\frac{1}{T}GM(x^2,\sigma^2)-TM(x,\sigma^2)] \quad (148)$$
$$= \frac{1}{T^2}[\frac{S}{T}+(\frac{2K+T}{T})M(x,\sigma^2)+(\frac{ST+S+G}{T})M(x^2,\sigma^2)$$
$$-(1+\frac{4}{T})M(x^3,\sigma^2)-\frac{2}{T}M(x^5,\sigma^2)] \quad (149)$$
$$= \frac{S}{T^3}+(\frac{2K+T}{T^3})M(x,\sigma^2)+(\frac{ST+S+G}{T^3})M(x^2,\sigma^2)$$
$$-(\frac{T+4}{T^3})M(x^3,\sigma^2)-\frac{2}{T^3}M(x^5,\sigma^2) \quad (150)$$

$$ \tag{151} $$

Finally, the preliminary for the last entry:

$$
\begin{aligned}
& E\frac{1}{T^2}\sum_{t=1} Tx_t^2\alpha_t e_t^2 - \sum_{t=1} Tx_t^2\sigma_t^2 \\
= \; & \frac{1}{T^2}[\sum_{t=1} TEx_t^2\alpha_t(1+h_{tt}) - x_t^2\sigma_t^2]
\end{aligned} \tag{152}
$$

Again by following using the above calculation, a similar manipulation leads to

$$
\begin{aligned}
B_{22}^{HD} = \; & \frac{1}{T^2}[\sum_{t=1}^{T} x_t^2\sigma_t^2 - \frac{2}{T}x_t^2\sigma_t^2 - \frac{2}{T}x_t^4\sigma_t^2 + \frac{x_t^2}{T} \\
& + \frac{2}{T}x_t^3 M(x,\sigma^2) + \frac{1}{T}x_t^4 M(x^2,\sigma^2) + \frac{T-2}{T^2}x_t^2\sigma_t^2 \\
& \frac{x_t^2}{T^2} + \frac{T-4}{T^2}x_t^4\sigma_t^2 + \frac{x_t^4}{T^2} + \frac{2}{T^2}x_t^3 M(x,\sigma^2) \\
& + \frac{x_t^4}{T^2}M(x^2,\sigma^2) - \frac{2}{T^2}x_t^6\sigma_t^2 + \frac{2}{T^2}x_t^5 M(x,\sigma^2) \\
& + \frac{1}{T^2}x_t^6 M(x^2,\sigma^2) - x_t^2\sigma_t^2)] \qquad (153) \\
= \; & \frac{1}{T^2}[TM(x^2,\sigma^2) - 2M(x^2,\sigma^2) - 2M(x^4,\sigma^2) + 1 + 2SM(x,\sigma^2) \\
& + KM(x^2,\sigma^2) + M(x^2,\sigma^2)\frac{T-2}{T} + \frac{1}{T} + M(x^4,\sigma^2)\frac{T-4}{T} \\
& + \frac{K}{T} + M(x,\sigma^2)\frac{2}{T} + M(x^2,\sigma^2)\frac{K}{T} - M(x^6,\sigma^2)\frac{2}{T} + GM(x,\sigma^2)\frac{2}{T} \\
& + LM(x^2,\sigma^2)\frac{1}{T} - TM(x^2,\sigma^2)] \qquad (154) \\
= \; & (\frac{K+T+1}{T^3}) + M(x,\sigma^2)(\frac{2ST+2S+2G}{T^3}) + M(x^2,\sigma^2)(\frac{KT-T+K+L-2}{T^3}) \\
& - M(x^4,\sigma^2)(\frac{T+4}{T^3}) - M(x^6,\sigma^2)\frac{2}{T^3} \qquad (155)
\end{aligned}
$$

$$ \tag{156} $$