Robust estimation in time series factor model

CAPM robust regression

Maastricht University
Bachelor thesis in Econometrics & Operations Research
Name: Alessandro Tenderini
Student ID: i6221893
Supervisor: Wilms I. and Raymaekers J.
Date submitted: July 1, 2022

Abstract

This thesis evaluates the performance of the robust mOpt estimator in the capital asset pricing model (CAPM) in comparison to the ordinary least square (OLS). The latter is indeed highly sensitive to the occurrence of outliers, and abnormal stock returns are a frequent occurrence in finance. A robust estimator, which downweights outlying observations, should thus provide a better measure of systematic risk that reflects the general structure of the stock returns disclosed by the bulk of the data. Therefore, I compare the CAPM's robust and non-robust versions across different stock market sectors. I analyze the OLS and mOpt beta estimates for 220 stocks from 11 stock market sectors. The results show that the robust estimator leads to significantly different estimates and that these robust estimates tend to be smaller than those derived from the OLS. This indicates a trend of abnormal returns that generally rotate the CAPM line counterclockwise, resulting in an OLS CAPM line steeper than the robust version. The results also suggest that returns of stocks belonging to the same sector exhibit similar features.

Contents

1 Introduction

The efficient estimation of some parameters is held in high regard in the field of finance. The market beta estimated in the capital asset pricing model (CAPM), for instance, plays a prominent role as it is used to estimate an asset's expected return. Hence, relying on financial models that generate reliable and efficient estimates is crucial because untrustworthy estimates pose a major threat to further financial and managerial analysis. In fact, there are a number of issues with estimating betas in the financial reality. The capital asset pricing model, despite some of its acknowledged deficiencies, is still a widely used model to estimate the systematic risk of an asset. The stock's excess return is regressed on a market index excess return as follows

$$E(R_i) = R_f + \beta_i(E(R_M) - R_f)$$

In classic statistics, such regression is usually evaluated through the Ordinary Least Square estimator. However, it is generally recognized that the ordinary least squares (OLS) method does indeed present a variety of statistical drawbacks. Particularly, it is highly sensitive to the occurrence of outliers and deviations from the normal distribution. Dealing with the former, it is straightforward that the presence of atypical observations can completely alter OLS estimates, as they are based on the minimization of a quadratic function of the residuals. Regarding departures from normality, the OLS is still the best linear unbiased estimator even without such assumption about the error terms' distribution. It is also still consistent. However, with normally distributed error, the OLS corresponds to the maximum-likelihood estimator. Hence, the derived estimates are the best among both linear and nonlinear estimates. Moreover, satisfying such assumption allows to perform reliable statistical inferences as hypothesis tests, confidence intervals, and prediction intervals.

In the context of CAPM and financial returns, problems with OLS beta estimates derive from the enormous complexity of the underlying sector. For instance, Blume [6] and Levy [23] showed that OLS beta estimates over time exhibit low correlation. Hence, least-square (LS) estimation of an asset's beta based on historical data has poor predictive power in the present. Additionally, both Hamada [12] and Galai and Masulisthe [10] showed that the theoretical determinants of betas, as the firm's leverage, are likely to change over a period of five years, which is a widely used interval in regressions. On the other hand, Levy[23] demonstrated that regressing a CAPM model over a more reasonable period of one-year results in unreliable estimates with low R squared and substantial errors.

Dealing with the statistical distribution of financial returns, Mandelbrot [25] and Fama [8] showed that stock returns exhibit a fat-tailed distribution compared to a normal distribution, leading to outliers. Such a feature is supported by reasonable real-world explanations that are addressed in the research by Chan and Lakonishok [4]. Lakonishok and Vermaelen [22] showed that small firms engaged in stock repurchase by tender offers usually exhibit large negative returns before the announcement as they are underpriced. Jensen [17] demonstrated that firms involved in takeovers have extreme returns before the announcement of a change in corporate management. Indeed, firm-specific news plays a prominent role in the financial context. Roll [3] showed that stock returns are generated by a mixture of distributions. He claimed that these returns are interspersed with news-driven extreme values.

Another source of outlying observations is the irrational overreaction phenomenon, which usually describes an extreme emotional response to news regarding "losers" and "winners" stocks. The former refers to stocks that have performed poorly in recent years, while the latter describes stocks that have experienced positive returns. Typically, firms ranked as extreme losers in the past years tend to substantially outperform prior-period winners over the next few years. Chopra, Lakonishok and Ritter [5] showed that such "Winner-Loser reversal" phenomenon, driven by an irrational behavior by investors, results in a large number of outliers that make OLS beta estimates on "loser" or "winner" stocks inefficient and unreliable.

Therefore, if we want to have a measure of systematic risk that reflects the general structure of the stock return disclosed by the bulk of the data, a robust approach is statistically preferable. Indeed, while the OLS is highly altered by the presence of outliers, robust estimators downweight outlying observations. Overall, a robust approach to statistical modeling seeks to develop methods that generate reliable parameters not only when the data follows a certain distribution exactly, but also when the data does not.

Genton and Ronchetti [11] proposed a criticism to robust approach in finance which claims that abnormal returns are important observations. They argue that such criticism does not hold if we aim to derive an estimate that provides a good fit to the majority of the historical data. However, this criticism is reasonable in a prediction perspective. Indeed, if extreme returns are genuine outlying observations rather than errors, then they'll probably show up again in the future. Hence, a robust approach that places less weight on such legitimate observations will result in biased estimates. However, it is a classic bias-variance trade-off scenario since an OLS regression in the previously-mentioned scenario will provide no bias at the expenses of huge variance in the estimates.

In this thesis, the CAPM's robust and non-robust versions are applied to different stock market sectors in order to evaluate the performance of the mOpt robust estimator in comparison to the OLS. The paper is organized as follows. Section 2 provides a selected overview of the literature on robust theory during the last century. Section 3 delves into the concepts and theories that underpin the methods I will be using throughout the study. The research's financial context is introduced in Section 4. Section 5 will cover the analysis as well as the discussion of the results. Finally, in section 6, I will draw some conclusions on the impact of using a robust estimator in finance.

2 Literature review

2.1 History of robust statistics

Robust statistics was developed to cope with the problems resulting from the approximate nature of typical parametric models of classic statistics. Classic procedures, such as OLS, are optimal, i.e., the most efficient, when the underlying assumptions of the model are perfectly met. However, it is well known that models only represent an approximation to reality and that real data often deviate from the model's stochastic assumptions. Robust statistics deals with departures from the model's stochastic assumptions through methods that generate reliable and efficient results in a neighborhood of the model.

The focus of classic statistics was to find the most efficient model, though, efficiency was defined under strong and rarely met data assumptions. Tukey [35] was the first one to question the optimality of classic procedures under non-satisfied assumptions. He demonstrated how even small deviations from the model's stochastic assumptions result in a substantial loss of efficiency for optimal procedures. His work played a prominent role in the development of robust theory as he was able to show why we need robust statistics.

A few years later, Huber [16] introduced the minimax approach, bringing about a breakthrough in the classic notion of optimality. The best model was no longer the most efficient under strict assumptions. Conversely, the goal was a model whose efficiency remains high in a neighborhood of the model. Indeed, he was looking for the method with the best worst performance under a wide range of conditions rather than restricting the analysis to classical assumptions. Moreover, he proposed a first solution to the robustness problem. He introduced the M-estimator, a class of estimators that played a key role in the following development of new robust methods.

In the same period, Hampel [14] [13] proposed the concept of breakdown point, which provides an intuitive understanding of the robustness of an estimator. It refers to the smallest fraction of data which can be given arbitrary values without making the estimator arbitrary large. For instance, the breakdown point of the OLS estimator is $\frac{1}{n}$ since one single abnormal observation can have an arbitrary large impact on the OLS estimate. Hampel [15] subsequently continued his research into functional analytic tools in order to derive new robust indicators, and he introduced the concept of influence function which describes the effect of an infinitesimal contamination, i.e., data from another distribution, on the estimate.

Overall, the work of Tukey [35], Huber [16] and Hampel [13] paved the way of robust theory. Following them, several studies and researches have focused on the improvement of existing robust techniques, the invention of new estimators, and the application of robust methods in different sectors, such as finance.

When it comes to financial models, Knez and Ready [18] showed that the inherent complexity of financial markets, as well as the vast network of factors involved, brings about departures from the model's assumptions. Thus, it is reasonable to consider financial models as approximations of the financial reality and develop methods that are robust to outlying values.

Cornell and Dietrich [6] investigated deviations from OLS stochastic assumptions in financial returns as well as the instability of short-term estimated CAPM beta. They evaluated a MAD estimator, which is less sensitive to abnormal observations, to deal with the presence of fat-tailed residuals, concluding that it has almost the same stability features. Hence, they were unable to find any arguments supporting MAD's superiority over LS.

Subsequently, Chan and Lakonishok [4] examined efficiency gains over OLS using a variety of robust methods. They focused their investigation on CAPM beta estimates and selected robust candidates such as Tukey's trimean, Minimum Absolute Deviations (MAD), Trimmed Regression Quantile (TRQ), and Gastwirth estimator. The results demonstrated that moving from OLS to robust techniques can lead to both significant efficiency gains under deviations from stochastic assumptions and acceptable normal efficiency. In fact, there is only a small efficiency loss of around 10 to 20 percent with normally distributed residuals, while an efficiency improvement of roughly 80 percent takes place when residuals follow a student-t distribution with three degrees of freedom. Overall, working with the actual distribution of residuals and excess market returns, based on both daily and monthly observations, shows significant benefits from employing robust approaches. These results were encouraging for further researches on robust applications in finance.

Recently, factor models and robust statistics were merged in Martin and Xia's research [29]. They applied both least square and robust estimators to CAPM, Fama-French 3-factor, and Fama-French-Carhart 4-factor models. Consequently, they were able to compare the performance of each factor model's robust and non-robust versions. They showed that, in contrast to the LS estimator, the robust estimator successfully provided a good fit to the bulk of the data for all factors models. Overall, they presented strong arguments in favor of the use of a specific robust estimator, the mOpt, in the context of time series factor models.

2.2 Outliers

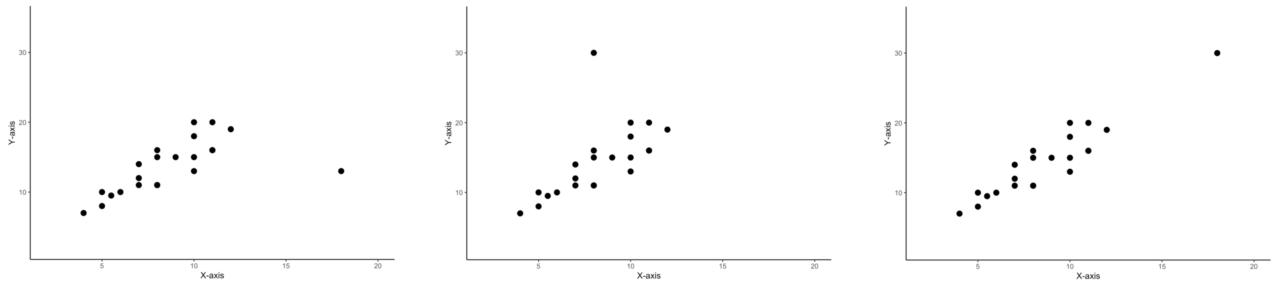
An outlier is an atypical observation which stands out from the bulk of the data, or in some way deviates from the data's overall pattern. However, due to the fact that such a statistical word lacks a rigorous definition, there is no unique and objective approach to determine whether or not a data point is an outlier. Usually, once a model is defined together with the assumption on the underlying distribution of the data, outliers are those points deviating from that distribution. Barnet and Lewis [1] gave outliers a broader interpretation. They defined outliers as observations that are inconsistent with the rest of the data and they introduced the notions of extreme values and contaminants. The former represent the sample's largest and smallest values, whereas the latter are data points from a different distribution.

If we consider a simple regression model and the estimated residuals

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad\qquad r_i = \hat{\beta}_0 + \hat{\beta}_1 x_i - y_i$$

we can classify as regression outliers those points with large standardized residuals.

If we define vertical outliers as regression outliers that lie close to predictors' values of the bulk of the data, we can introduce two additional types of data points: leverage points and influence points. Leverage points are the outliers in the x-direction. Hence, an observation (x_i, y_i) is a leverage point if the x_i is well separated from the bulk of the observed x. Such a definition does not take into account the y_i value, so a leverage point does not have to necessarily be a vertical outlier. Indeed, when a leverage point presents a small residual, meaning that it lies close to the regression line, it's called a good leverage point. On the other hand, those points whose deletion would noticeably change the regression coefficients are called influent points. As a result, influent points can be either bad leverage points, i.e., leverage points with normal y-value, or vertical outliers with normal x-values.



(a) bad leverage point (b) vertical outliers (c) good leverage point

Figure 1: Vertical outliers, leverage points and influential points

2.3 Outliers detection

In classic statistics, a measure for outliers' detection is the Mahalanobis distance

$$MD_i = \sqrt{(x_i - T(X))C(X)^{-1}(x_i - T(X))^T}.$$

Where T(X) is the sample mean and C(x) is the covariance. However, Rousseeuw and van Zomeren [33] showed that the non-robust nature of this measure might bring about the masking effect, by which multiple adjacent outliers do not necessarily have large MD. Therefore, replacing T(X) and C(X) with robust estimators provides a solution to this problem. A candidate for such a role is the minimum covariance determinant estimator (MCD) which is characterized by a high breakdown point [32]. It looks for a half-data subset, the covariance matrix of which presents the smallest determinant. Then, T(X) becomes the mean of this subset containing half of the data, C(X) the corresponding sample covariance, and we now have a robust measure for outliers 'detection. In my research, X is a univariate vector since the CAPM model only includes the excess market return as a single regressor.

Moreover, since I am computing robust distance only on the predictor, this method serves as leverage points discovery rather than outliers detection.

3 Methodology

3.1 Robust estimator

Throughout my study I will be using the mOpt robust estimator. However, in order to analyze such estimator, some additional robust estimators that are necessary for its computation need to be introduced.

3.1.1 M-estimator

M-estimators are a large class of robust estimators introduced by Huber (1964). They are obtained through the M-estimation method which consists in minimizing a function of the scaled residuals

$$\sum_{i=1}^{n} \rho\left(\frac{r_i(\hat{\beta})}{\hat{\sigma}}\right) = min \qquad r_i(\hat{\beta}) = \hat{\beta}_0 + \hat{\beta}_1 x_i^{'} - y_i = \hat{\beta} x_i^{*'} - y_i$$

where $\hat{\sigma}$ is a robust scale estimate of the residuals and $x_i^{*'} = (1, x_i^{'})$. Typically, such robust scale is computed as the median absolute deviation (MAD) of the residuals as follows

$$1.483 med(|r_i - med(r)|)$$

Initially, a preliminary scale estimate can be computed from the residuals of a least absolute deviation (LAD) regression, defined as

$$\sum_{i=1}^{n} |r_i| = min$$

In contrast to ordinary least squares (OLS), which places more emphasis on large residuals by squaring the residuals, such a method gives equal weights to all observations, making it robust to outliers.
If the function ρ is differentiable, we can differentiate the equation with respect to β and solve for the root of the derivative as follows

$$\sum_{i=1}^{n} \psi\Big(\frac{r_i(\hat{\beta})}{\hat{\sigma}}\Big) x_i^* = 0$$

When the ρ-function is differentiable, the M-estimator is said to be of ψ-type and it can be easily computed via iterated re-weighted least square (IRWLS) after defining a weighting function $W(x)$ as follows

$$W(x) = \begin{cases} \frac{\psi(x)}{x}, & \text{if } x \neq 0. \\ \psi(x)', & \text{if } x = 0. \end{cases} \qquad w_i = W\big(\tfrac{r_i}{\hat{\sigma}}\big)$$

$$\sum_{i=1}^{n} w_i r_i x_i^* = \sum_{i=1}^{n} w_i (y_i - x_i^{*'} \hat{\beta}) x_i^* = 0$$

As a result, it is intuitive that the nature of an M-estimator depends on the ρ-function and its derivative ψ, which we can choose based on the desired levels of bias and efficiency for our estimator. The LS estimate is a special case of M-estimator where ρ is quadratic and ψ is linear. Also, the LAD estimate is another special case where ρ is the absolute value function. The first ψ-function introduced by Huber [16], the "Huber function", is defined as

$$\rho_{huber}(x) = \begin{cases} \frac{1}{2}x^2, & \text{if } |x| \leq k \\ k\big(|x| - \frac{k}{2}\big), & \text{if } |x| > k \end{cases}$$

$$\psi_{huber}(x) = \begin{cases} x, & \text{if } |x| \leq k \\ sign(x)k, & \text{if } |x| > k \end{cases}$$

and the corresponding plot is shown in Figure 2.

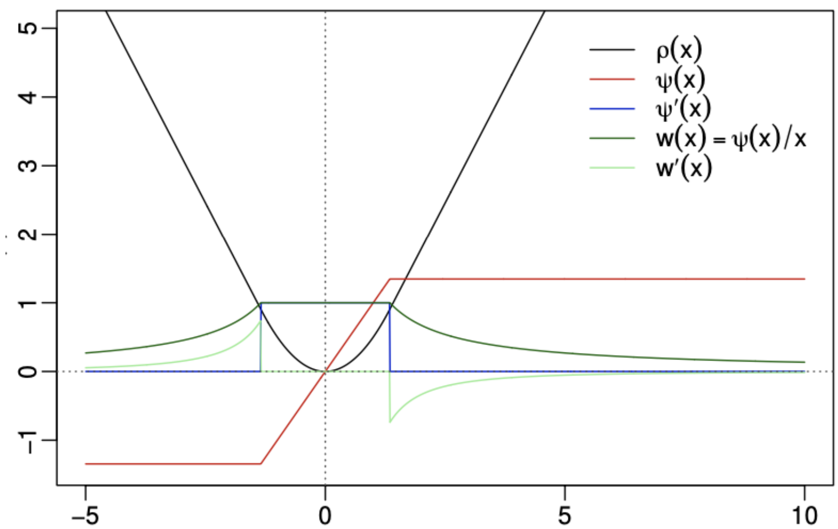


Figure 2: Huber functions. From [19](Figure 1, p.3).

It is actually a mix of LS and LAD estimates since the ρ-function exhibit a quadratic behavior in the central interval [-k, k] and switches to an absolute value function outside that range. However, the Huber estimator, as well as LS and LAD, is based on an unbounded ρ-function. Martin, et al. [28] showed that M-estimators based on such unbounded functions are indeed resistant to vertical outliers, but they are not robust to leverage points. Therefore, we require a bounded ρ-function with a corresponding redescending ψ-function in order to construct M-estimator robust to leverage points. A ψ-function is redescending if it is non-decreasing near the origin but it goes to 0 far from the origin. The Tukey's bisquare function is a bounded ρ-function defined as

$$\rho_{bisquare}(x) = \begin{cases} \frac{k^2}{6}\{1 - \left[1 - \left(\frac{e}{k}\right)^2\right]^3\}, & \text{if } |x| \leq k \\ \frac{k^2}{6}, & \text{if } |x| > k \end{cases}$$

$$\psi_{bisquare}(x) = \begin{cases} \left[1 - \left(\frac{e}{k}\right)^2\right]^2, & \text{if } |x| \leq k \\ 0, & \text{if } |x| > k \end{cases}$$

and Figure 3 shows the related graph.

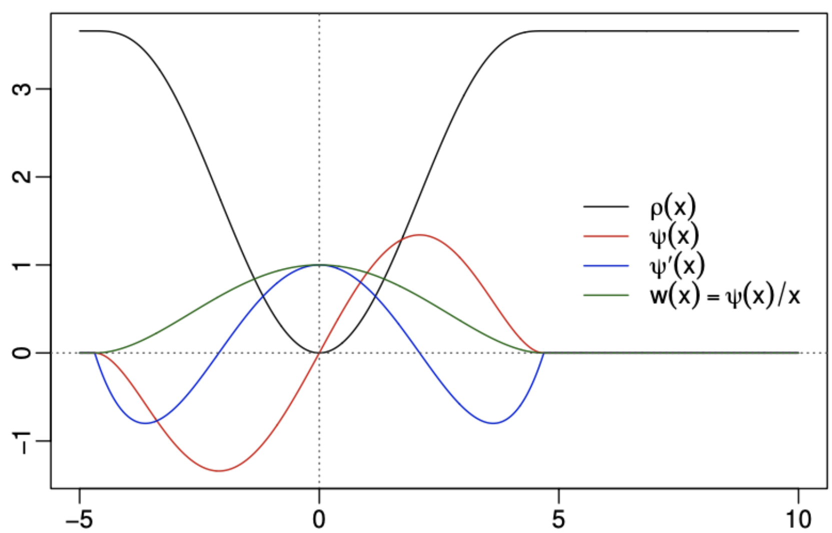


Figure 3: Bisquare functions. From [19](Figure 2, p.4).

3.1.2 MM-estimator

The MM-estimator introduced by Yohai [36] is a robust estimator which is characterized by both the highest possible breakdown point of 0.5 and high efficiency under normal distribution. It is based on the M-estimation method of IRWLS. However, we can't use the LAD residuals to compute an initial scale estimate because the LAD estimator is not robust to leverage points. Thus, we require an initial estimate that is unreliant on a previously computed scale and that is robust to all types of outliers: the S-estimator by Rousseeuw and Yohai [31].

The S-estimate of beta is the value of beta that minimized the scale of residuals $\hat{\sigma}(r(\hat{\beta}))$ meaning $\hat{\beta}_s = \arg\min_\beta \hat{\sigma}(r(\hat{\beta}))$ where $\hat{\sigma}(r)$ is defined by the collection of betas that satisfy

$$\sum_{i=1}^{n} \rho\left(\frac{r_i(\hat{\beta})}{\hat{\sigma}}\right) = k$$

where k is the expected value of ρ at a normal distribution.

Therefore, I am going to use an MM-estimator derived by taking an S-estimator, based on a Tukey's bisquare function, as a starting point for the M-estimation procedure based on a mOpt function with different constant. Specifically, I initially choose a Tukey bisquare function with c_1=1.548 in order to derive an initial M-estimate with the highest possible breakdown point of 0.5 and low normal efficiency of 0.287. Indeed I want consistency of the initial $\hat{\sigma}(\beta)$ estimate. Then, the second M-estimation procedure based on a mOpt function presents a tuning constant c_2 which determines the normal distribution efficiency of the final estimator. The final estimate presents the same breakdown point as the initial estimate because of the redescending mOpt ψ-function and fixed initial scale.

3.1.3 mOpt-estimator

The mOpt-estimator is a robust estimator introduced by Yohai and Zamar [37]. Such estimator presents the nice feature of minimizing the maximum estimator bias across a Tukey-Huber family of distributions, given a constraint of high normal distribution efficiency. Yohai and Zamar's original optimal bias estimator was actually the Opt estimator. However, as discussed in the work from Martin and Konis [21], this estimator presented some computational limitation in its ψ-function, which is why I use the mOpt, a slightly modified version. Let's first introduce the original family of optimal robust ψ-functions by Yohai and Zamar [37] given by

$$\psi_{opt}(x) = sign(x)\left(|x| - \tfrac{a}{\phi(x)}\right)^{+} = \begin{cases} x - sign(x)\frac{a}{\phi(x)}, & \text{if } x \in (lower, upper) \\ 0, & \text{otherwise.} \end{cases}$$

where a is the tuning constant controlling the bias-efficiency trade-off of the estimator. The optimal ψ-functions for a few normal distribution efficiencies are shown in Figure 4.

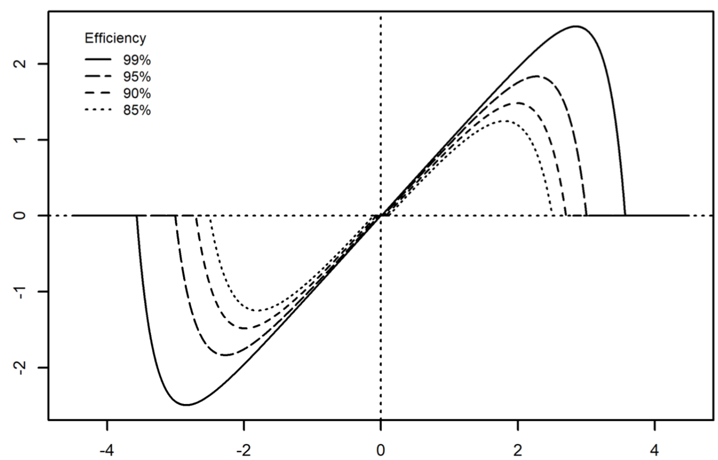


Figure 4: Optimal ψ-functions. From [20](Figure 1, p.6)

The weight function associated with the optimal psi function is

$$w_{opt}(x) = \frac{\psi_{opt}(x)}{x}$$

and the plot of optimal weight functions for various normal distribution efficiencies is shown in Figure 5.

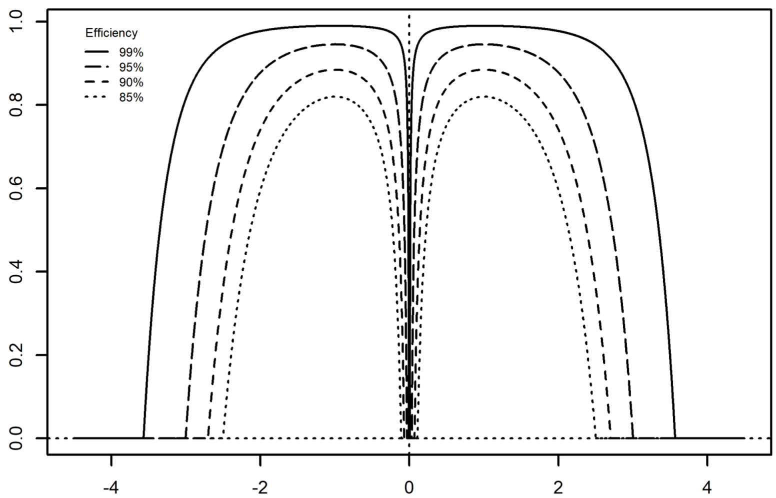


Figure 5: Optimal weight-functions. From [20](Figure 4, p.9).

As shown in the graph, the weight functions exhibit a strange behavior close to the origin, where they drop to a value of zero. Marina et al. [27] proposed a necessary condition to ensure convergence of the iterative re-weighted algorithm, namely that w(x) must be non-increasing for $x > 0$. Consequently, such Opt weight function is inappropriate for our estimation, which is why we adopt the modified Opt psi function and its related new weights. The mOpt ψ-function is a smooth approximation of the previous Opt function which is now linear at the origin. As a result, the corresponding weight function is now non-increasing and IRWLS convergence is guaranteed.
The mOpt and Opt ψ-functions are displayed in Figure 6.

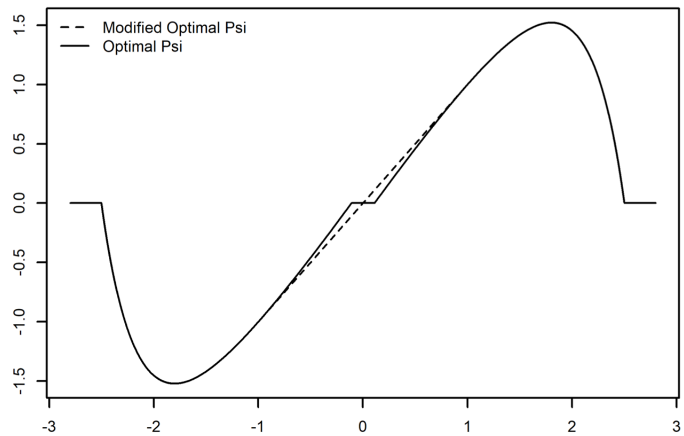


Figure 6: mOpt ψ-functions. From [20](Figure 5, p.12).

and the new weight functions for various normal distribution efficiencies are shown in Figure 7.

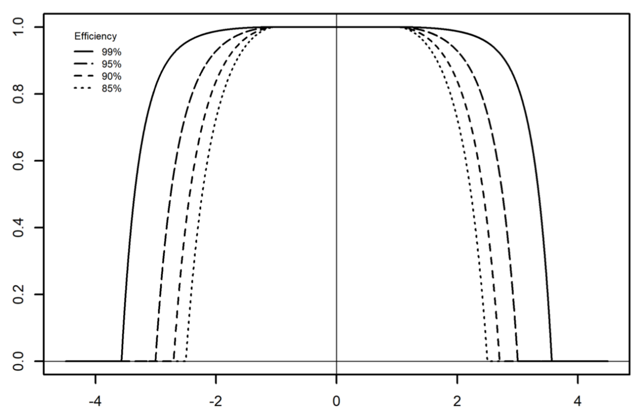


Figure 7: mOpt weight-functions. From [20](Figure 6, p.13).

In order to obtain the new mOpt functions, we modify the weight function by connecting the two maximum values, which lie at $x = 1$ and $x = -1$, with a horizontal line. As a result, we eliminate the issue of the “sink hole” at the origin and we use the new weights to derive the modified ψ-function

$$\psi_{mOpt}(x) = sign(x)\big(|x| - \tfrac{a}{\phi(x)}\big) = \begin{cases} x, & \text{if } x \in [-1,1] \\ \big(\tfrac{\phi(1)}{\phi(1)-a}\big)\big(x - sign(x)\tfrac{a}{\phi(x)}\big), & \text{if } |x| \in (1, upper) \\ 0, & \text{otherwise.} \end{cases}$$

The mOpt estimator is obtained by implementing an MM-estimator based on the mOpt ψ-function.

Let's now consider the bias-efficiency trade-off, which is sensitive to changes in the support interval $(lower, upper)$ and a. The support interval is actually determined by a as the set of positive values of x such that $g(x) = \phi(x)x - a > 0$. Therefore, as we increase a, the support interval shrinks, efficiency decreases, and we face an improvement in robustness toward outliers. In the bias-efficiency trade-off, we usually look for an estimator with a specific level of normal distribution efficiency, that is the inverse of the asymptotic variance of the mOpt estimator at a standard normal distribution. Efficiency is also defined by Martin and Xia [29] as the ratio of the least squares estimator's asymptotic variance to that of the mOpt estimator, generally represented as a percentage. In our analysis, I set and $a = 0.0132$ ($upper = 3$) in order to ensure 95% normal distribution efficiency.

3.2 Outliers detection

As part of my investigation, I will apply outlier detection techniques on a bivariate dataset. Particularly, I am going to follow a robust approach to detect outliers, leverage points and influential points for a simple linear regression $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$. Such a method relies on a plot with robust distance on the x-axis and standardized residuals on the y-axis as in figure 8.

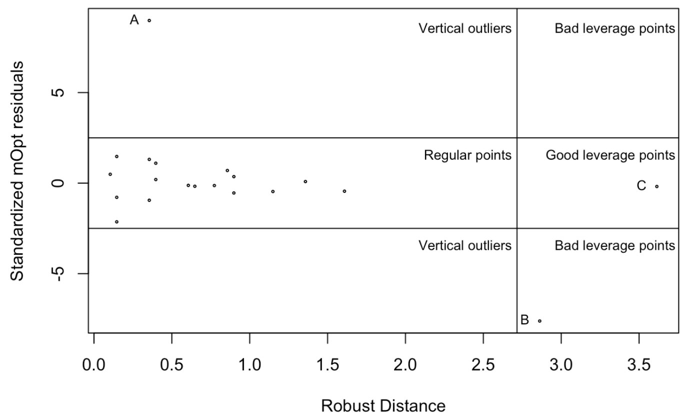


Figure 8: Robust residuals against robust distances

The robust distance is Rousseeuw and van Zomeren's [33] robust version of the Mahalanobis distance, and it is exclusively computed on the unidimensional regressor x. The standardized residuals, on the other hand, are the residuals of a robust regression, in this case the mOpt-regression, that are scaled by a robust estimate of scale. The cutoff point on the x-axis is $\sqrt[2]{\chi_{1,0.975}} = 2.24$ while on the y-axis I use a tolerance band $[-2.5, 2.5]$, which can be changed based on the robustness-normal efficiency trade-off chosen. As a result, the graph is divided into six sections and the nature of a point (y_i, x_i) can be assessed according to its location. Figure 9 helps to identify different kinds of points.

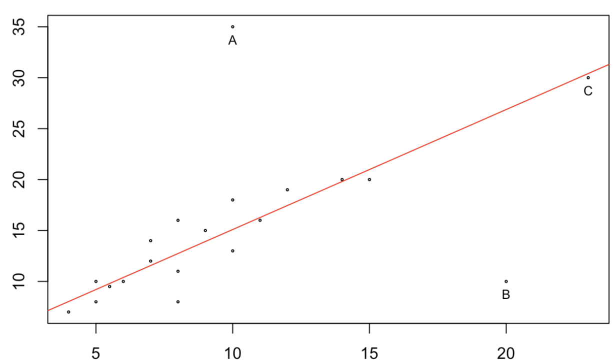


Figure 9: mOpt regression

Looking at figures 8 and 9 together, we can better understand the nature of points A, B and C. Point A has a standardized residual larger than 2.5 and normal $x - values$, thus, it is a vertical outlier above the fitted line. Point C is a leverage point since its $x - value$ stands out from the bulk of the observed x_s resulting in a large robust distance. However, it also has a relatively large $y - value$ that makes (x_c, y_c) fit to regression line. Hence, point C is a good leverage point. Point B is a leverage point as well, but it differs from point C in that it has a large standardized residual due to its distance from the regression line. Overall, only point A and B are influential observations or regression outliers.

An overview of the different types of points in each of the six locations is shown in Figure 8.

4 Financial application

4.1 CAPM and betas

The Capital Asset Pricing Model describes the relationship between the expected return of an asset and its systematic risk [2]. Such financial model was developed by William Sharpe [34], Jack Treynor [9], Jhon Linter [24] and Jan Mossin [30], and it is based on Harry Markowitz's portfolio theory [26]. It was the first asset-pricing model to provide accurate return and risk predictions and it was built on the notion that not all risks should have an impact on asset prices. The CAPM indicates that a security's expected return is equal to the risk-free return plus a risk premium based on the security's beta. Such a relation is

$$E(R_{it}) = R_{ft} + \beta_i(E(R_{Mt}) - R_{ft})$$

where $R_{it} = \frac{P_t - P_{t-1} + D_t}{P_{t-1}}$ is the stock return with the initial stock price P_{t-1}, the ending stock price after one period P_t and the received dividends in period t, D_t. Then, R_{ft} is the risk free rate and R_{Mt} is the market return at time t. Since I am interested in the estimation of beta, I run the CAPM regression as follows

$$(R_{it} - R_{ft}) = \beta_0 + \beta_1(R_{Mt} - R_{ft}) + \epsilon_{it}$$

I am going to apply this asset pricing model to the American stock market over a three-year period from 01/01/2018 to 31/12/2021 with weekly observations. I chose weekly stock returns over monthly returns as the former should in principle present more outliers, which makes a robust analysis more interesting. In the next sections, I am going to estimate the regression beta both with LS and mOpt estimator.

4.2 Data

The analysis relies entirely on data from the Yahoo Finance and Kenneth French databases. I restricted my analysis to 220 American stocks listed on the NYSE, AMEX, or NASDAQ. As in research from the Corporate Finance Institute [7], I used the S&P sectors method, which sorts publicly traded companies into 11 sectors: Information, Technology, Health Care, Financials, Consumer Discretionary, Communication Services, Industrials, Consumer Staple, Energy, Utilities, Real Estate and Material. As a result, by selecting the first 20 American stocks in each sector based on market capitalization, I added a new interesting layer of analysis to my study. As I am running all my analysis in R, I imported the daily stock prices and computed the weekly returns via the quantmod library, which retrieves data from the Yahoo Finance database. As for the market factors, I obtained the data from the open data library of Kenneth French. The market return is the value-weighted return of all CRSP American firms listed on NYSE, AMEX, or NASDAQ and the risk-free rate is the weekly adjusted one-month Treasury bill rate.

5 Analysis of the results

In the first section of the analysis, I am going to investigate whether the mOpt estimator used in the robust regression does actually generate different results from the classic OLS regression. Afterward, I will conduct a robust outliers analysis. Finally, assuming that the robust estimates are significantly different, I am going to compare the predictive performance to see whether the robust estimator truly outperforms the OLS.

5.1 In sample evaluation

I run the CAPM regression both with the LS and mOpt estimator for each stock, ending up with 440 beta estimates. Table 1 displays the mean and standard deviations of the collections of mOpt and LS estimates for each sector.

Sector	LS average	mOpt average	LS SD	mOpt SD
Telecomunication	1.14	1.11	0.381	0.346
Energy	1.33	1.26	0.388	0.297
Utilities	0.840	0.562	0.235	0.293
Basic_materials	1.08	1.07	0.325	0.401
Industrial	1.09	1.05	0.319	0.211
Consumer_staples	0.697	0.666	0.290	0.231
Consumer_discrectionary	0.934	0.926	0.300	0.266
Real_estate	1.03	0.644	0.349	0.211
Financial	1.10	0.993	0.264	0.244
Healthcare	0.849	0.841	0.269	0.202
Technology	1.07	1.09	0.218	0.202

Table 1: Mean and standard deviation of beta estimates

Stocks in the energy sector tend to be the most volatile compared to the market according to both the robust and non-robust estimator, with average estimates of 1.33 and 1.26. Conversely, the two estimators lead to different results when I consider the sector with the lowest systematic risk. The mOpt estimator identifies the consumer staples sector as the least volatile with a mean of 0.697, whereas the LS estimator leads to the lowest average value of 0.562 in the utilities sector. Overall, except from the financial and real estate sector, both estimators lead to the same conclusions as regard being more or less volatile than the market. Indeed, stocks in the basic materials, energy, industrial, technology and telecommunication sectors have an average beta value greater than 1 which indicates that the securities are overall more volatile than the market. On the other hand, stocks in consumer discretionary, consumer staples, healthcare and utilities sectors have an average estimate of less than 1. Dealing with the variation of estimates, the mOpt estimator results in less dispersed estimates compared to the LS estimator for all sectors except for basic materials and utilities. This is a reasonable outcome as the downweighting of uncorrelated extreme observation among stocks in the same sector should lead to lower dispersion between estimates. In the case of basic materials, the standard deviation is larger for the robust estimator and it is even the largest of all standard deviations, with a value of 0.4. The real estate represents an interesting sector as the robust estimates suggest that underlying stocks tend to be slightly more volatile than the market, with an average of 1.03, whereas LS estimates indicate that stocks tend to be significantly less volatile than the market with a value of 0.64.

In Figure 10 the mOpt-betas are shown against the LS-betas for all stocks grouped by sector

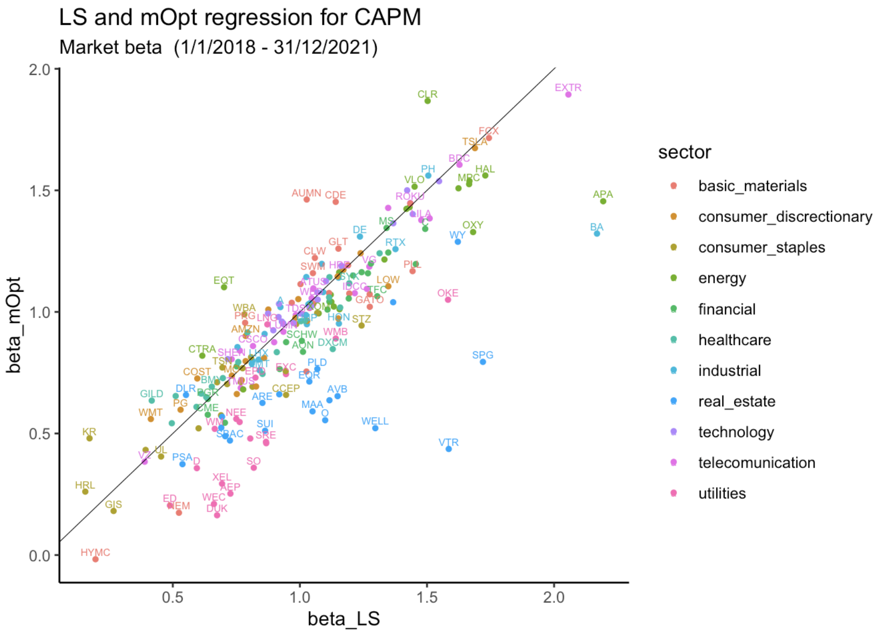


Figure 10: mOpt beta estimates against LS beta estimates for each stock

The diagonal line depicts the set of points where the robust regression yields the same estimate as the traditional LS regression. Therefore, the further a point is from the line, the greater the disparity between the two estimations. As we can see, the performances of stocks in the same sector tend to be similar, both in terms of market beta level and also in terms of differences between estimators. Indeed, Figure 10 shows various clusters of same-colored points, which indicate stocks from the same sectors with similar behavior. Such figure is actually a visualization of Table 1 since the standard deviation of LS estimates is the dispersion on the x-axis of same-colored points whereas the standard deviation of mOpt estimates corresponds to the variation on the y-axis. I also provide the boxplots of the difference between robust and non-robust estimates for each sector in Figure 11, and for the overall set of stocks in Figure 12. The horizontal lines in both figures represent the median.

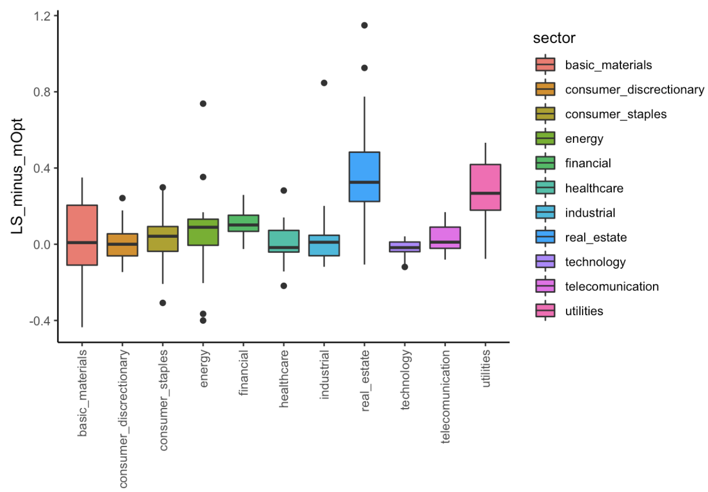


Figure 11: Boxplot of LS beta minus mOpt beta by sector

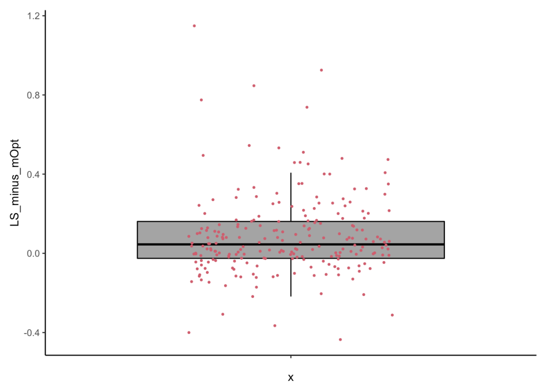


Figure 12: Boxplot of LS beta minus mOpt beta

In Figure 12, red dots represent the difference in estimates for each stock and they are spaced out along the x-axis for better visualization. We can see several points behind the whiskers that represent extreme differences. Most of them lie above the third quartile and represent stocks with LS betas that are significantly larger than mOpt estimates. In order to get such greater non-robust estimates, abnormal weekly returns must have happened for these stocks. Moreover, both the mean of 0.086 and the median of 0.045 suggests the tendency of the robust estimates to be smaller than non-robust ones. From Figure 11 there seem to be some interesting differences between the two estimators, especially for stocks in the utilities and real estate sectors. Additionally, we can notice that the 4 greatest differences, among those identified in Figure 12, belong to stocks in the industrial, energy and real estate sector.

The previous visualizations are combined with the Wilcoxon Signed-Rank Test in order to see whether the differences in estimates between the two estimators are statistically significant. The null hypothesis of such a test is that the LS and mOpt estimates represent identical populations. Hence, if I get a p-value < 0.05 then the robust estimator does actually lead to a statistically significant different estimate. Table 2 contains the average differences by sector, along with the test's p-values and the standard deviations of differences.

	LS minus mOpt	p-value	SD
Total	0.0858	0.00000000464	0.206
telecomunication	0.0286	0.189	0.0797
energy	0.0639	0.123	0.238
utilities	0.278	0.00000954	0.177
basic_materials	0.0137	0.701	0.208
industrial	0.0428	0.674	0.206
consumer_staples	0.0312	0.231	0.145
consumer_discrectionary	0.008	0.841	0.103
real_estate	0.388	0.00000381	0.289
financial	0.105	0.0000267	0.0742
healthcare	0.00779	0.622	0.109
technology	-0.0241	0.0532	0.0467

Table 2: Wilcoxon Signed-Rank Test, mean and standard deviations of differences

I reject the null at a 5% significance level for stocks in utilities, real estate and financial sector, as well as for the entire sample of 220 stocks. These results back up the conclusions I previously drew from the boxplots. Indeed, both the utilities and real estate sectors exhibit pretty large average distances of 0.28

and 0.39 and really small p-values. The mOpt estimator leads to a decrease of 40% and 43.6% of the average estimates for stocks in the utility and real estate sectors respectively. Since the presence of outliers is the source of difference between the mOpt and LS estimators, these results might indicate a greater frequency of abnormal returns in these sectors. In addition, I can also identify the sectors where a robust approach does not provide substantial variations. Stocks in consumer discretionary, telecommunication and healthcare sectors are characterized by a combination of high p-values, low average differences and low standard deviations of the differences. These results indicate an insignificant effect of a robust approach in these sectors. Moreover, they suggest a similar robust effect across all the stocks in each of these sectors, which doesn't necessarily coincide with similar beta estimates. Indeed, Table 2 shows a low variation of differences in the telecommunications sector of 0.0797 whereas the LS estimates in the same sector have a standard deviation of 0.38, which is relatively high. Overall, based on the result for the whole sample of stocks with an average difference of 0.0858, I can conclude that the robust estimator does indeed lead to significantly different outcomes.

5.2 Robust outliers detection

In this section, I will examine the LS estimates propensity to typically exceed mOpt betas. I'm going to run an analysis on influence points in order to support the prior claim and I will also examine the outliers of an interesting stock. Figure 13 shows the LS and mOpt estimates for each stock and Figure 14 displays the differences between LS and mOpt estimates for each stock, grouped by sectors.

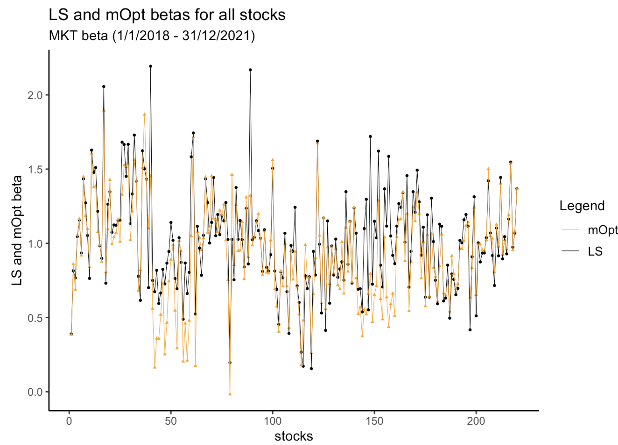


Figure 13: LS and mOpt betas

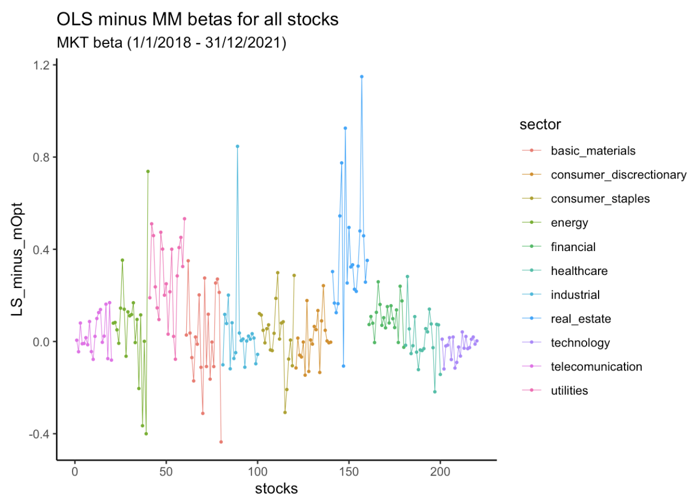


Figure 14: LS minus mOpt

As we can see from both Figure 13 and 14, the majority of the stocks present a LS-beta that is larger than the mOpt-beta. Indeed, 11.36% of the stocks present a difference between LS and mOpt estimates larger than 0.3 while only 2.27% of the stocks have a gap less than -0.3. Moreover, the average difference is 0.0858. This might point to a certain pattern of influential points, which can be both vertical outliers and bad leverage points.

Before looking at the frequency of different types of outliers over the whole sample, let's consider an interesting example of outliers detection in a single stock, "WELL". Welltower Inc. is a real estate investment trust investing in healthcare, thus it is strongly related to the covid-19 pandemic. Moreover, we've already seen how stocks in the real estate sector tend to have significantly higher LS estimates than mOpt estimates. From a CAPM regression on the weekly market excess return, I get an LS beta of 1.298 and a mOpt estimate of 0.522, yielding a significant difference of 0.775. Let's look at the standardized mOpt residuals against the robust distances and the CAPM regression lines in Figure 15 and 16.

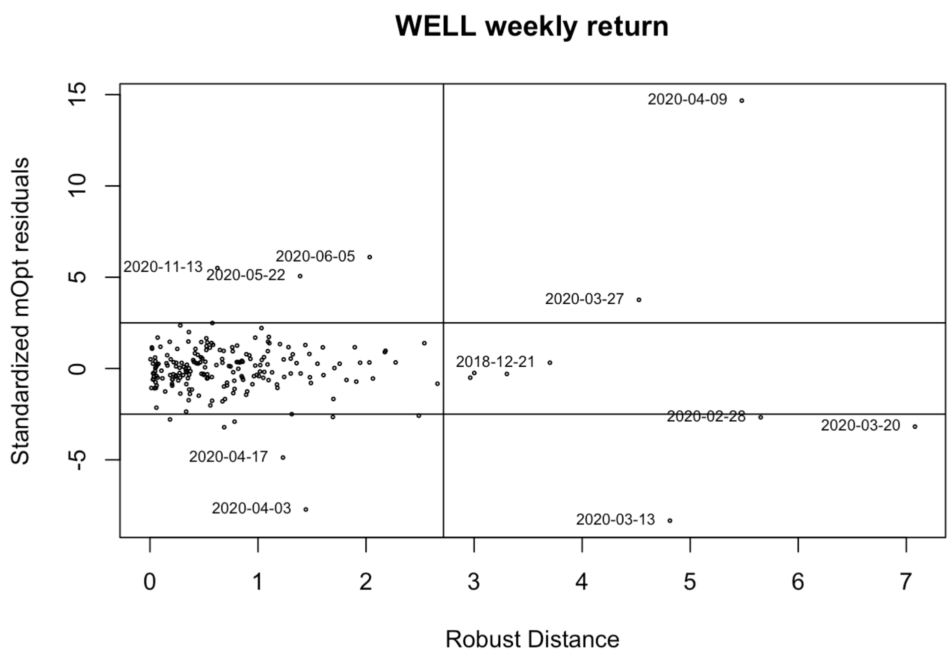


Figure 15: Outliers detection for WELL

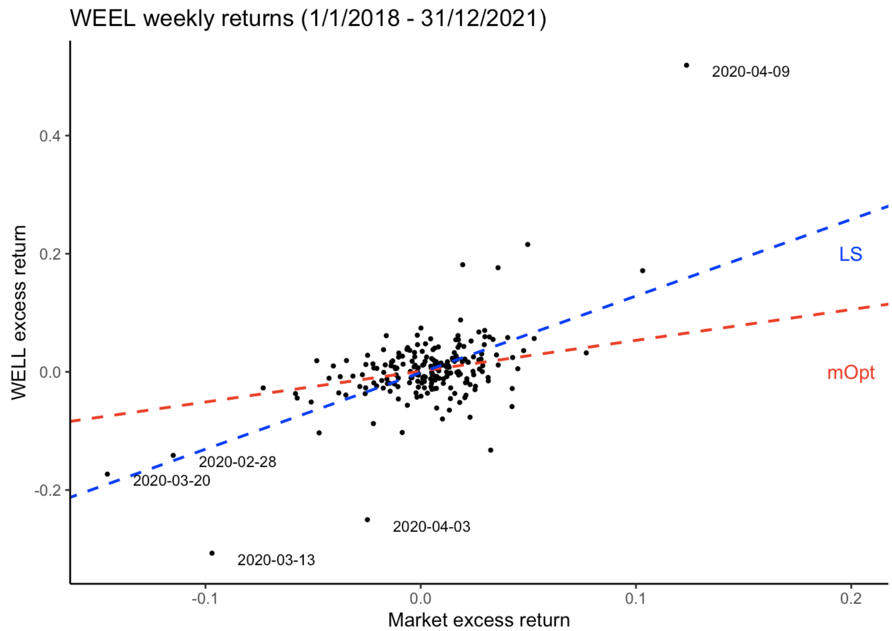


Figure 16: LS and mOpt CAPM regression for WELL

First, we can notice 9 leverage points, that represent dates when the weekly excess return of the market did not perform as usual. Particularly, the -14.6% market's weekly return on the 20th of march 2020 really stands out from the bulk of the data as it coincides with the stock market crash due to the covid-19 pandemic. Such a leverage point is also an influential observation as it crosses the cutoff value of regression outliers. However, it is just a weak bad leverage point since the standardized residual is not far from -2.5. Indeed, WELL's weekly return on that date was -17.3% and, because both returns are significantly lower than the majority of returns, that observation is not too far from the fitting line. On the other hand, the 9th of April 2020 is a really strong influential point. On that date, both the market and WELL stock perform better than usual. However, WELL's excess return was 52%, while the market excess return was just 12.4%, which is low relative to the former. As a result, such a bad leverage point stands far above the CAPM line and it breaks the least square estimator contributing to the significant difference of 0.775 between LS and mOpt estimates.

Dealing with the overall pattern outliers in the whole sample, I investigate why the LS CAPM regression line tends to be steeper than the mOpt regression line. Indeed, Table 2 shows the average differences by sector, and they all point to a higher LS estimate with the exception of the technology sector. This exception is interesting since it provides us a comparison element for the analysis that follows.

There have to be influential points that influence the LS CAPM line by rotating it counterclockwise with respect to the robust line. The detection of such observations is tricky. In fact, merely identifying high and low influential points is not sufficient since a high influential point with a large x value tends to make the line steeper, but a high influential point with a low x value has the opposite effect by making the line rotate clockwise. Therefore, I split influential points into those that have a positive market excess return and those that have a negative one. This analysis is rather speculative, but I assumed that such division could be reasonable since the average market excess return is 0.0032.

I divide weekly returns for each stock into two sets based on the sign of the market excess return, and then I look for observations with scaled residuals with an absolute value greater than 2.5. Therefore, I record the frequency and average intensity of four different types of outliers: high and low influential points for positive and negative x-values. Table 3 shows these results for each sector as well as for the entire sample of 220 stocks.

Influence points					
Sector	MKT	High mean	High count	Low mean	Low count
Total	+	3.61	932	-3.43	673
	-	3.42	455	-3.78	506
Telecomunication	+	3.61	88	-4.05	55
	-	3.51	35	-3.54	33
Energy	+	3.74	83	-3.2	33
	-	3.49	52	-3.88	32
Utilities	+	4.14	85	-3.17	39
	-	3.43	8	-4.8	56
Basic_materials	+	3.62	182	-3.36	171
	-	3.44	114	-3.78	101
Industrial	+	3.84	61	-3.59	47
	-	3.17	31	-3.72	39
Consume_staples	+	3.5	80	4.05	54
	-	3.39	45	-3.59	46
Consumer_discrectionary	+	3.93	61	-3.34	62
	-	3.65	56	-3.5	43
real_estate	+	4.93	79	-2.94	39
	-	3.33	5	-4.25	65
financial	+	3.6	75	-3.47	52
	-	3.08	22	-3.44	40
healthcare	+	3.34	70	-3.22	60
	-	3.47	35	-3.27	26
technology	+	3.25	68	-3.4	66
	-	3.5	52	-3.59	25

Table 3: Frequency and intensity of influential points

This analysis is based on the assumption that right-side high influential points and left-side low influential points make the LS line steeper than the mOpt line while the other two types of outliers have the opposite effect. Table 3 shows consistent results for the utilities and real estate sectors, which exhibit the largest average distances of 0.28 and 0.39. Indeed, stocks in the utility sector present 85 right-side high influential points with an average standardized residual of 4.14, and 39 right-side low influential points with an average intensity of -3.17. The left-side, on the other hand, is characterized by 8 high influential points with an average magnitude of 3.43, and 56 low influential points with an average intensity of -4.8.

If I consider the entire sample of 220 stocks, where the average difference of 0.0032 is pretty small, I still find interesting results. Such difference is explained by the presence of 932 right-side high influential points with an average value of 3.61 and 506 left-side low influential points with an average of -3.78, which outweigh the presence of 673 right-side low influential points with an average of -3.43 and 455 left-side high influential points with an average magnitude of 3.42. Indeed, when the market excess return is positive, high influential points are greater both in terms of frequency and intensity than low influential points, on the other hand, left-side high influential points are lower both in terms of frequency and magnitude than left-side low influential points.

Finally, let's consider the exceptional case of the technology sector with a negative average difference of -0.024. Table 3 shows consistent results with the slightly flatter LS line compared to the robust one. Indeed, for positive market excess return, 66 low influential points with an average standardized residual of -3.4 have a greater effect than 68 high influential points with an average of 3.25. Furthermore, for negative market excess return, even if the average magnitude of low influential points is greater in absolute value than the average intensity of high influential points, -3.59 and 3.5 respectively, the latter are more than twice the formers in terms of frequency.

5.3 Out of sample evaluation

As I previously mentioned, the dataset I'm working on includes observations up until 31/12/2021. Therefore, I create a new dataset with the same variables as before, ranging from 01/01/2022 until 22/04/2022 for a total of 16 new observations, and I can test the previous mOpt and LS models on the new values of the excess market return. Table 4 shows the trimmed root mean square error (RMSE) of the two models.

LS RMSE	mOpt RMSE
0.0322	0.0313

Table 4: Trimmed RMSE

I assumed that the classic RMSE was not the appropriate measure to compare the out-of-sample performance of the two models. Indeed, the robust estimator will poorly predict abnormal returns since it is actually constructed in a way to downweight such observations. Therefore, for each stock, I consider the 10% trimmed RMSE where I trim 10% of the largest squared prediction errors. This analysis relates to Genton and Ronchetti [11] criticism that abnormal returns are important observations in a prediction perspective. Hence, I am trying to evaluate the models' predictive power with respect to the bulk of the data by ignoring the inaccurate predictions of abnormal returns.

For the LS and mOpt estimators, Table 4 displays trimmed RMSE values of 0.0322 and 0.0313, respectively. As a result, the robust estimator's slightly lower RMSE denotes a better predictive power for the mOpt when discarding 10% of the worst predictions. However, such a difference is really small and could also be insignificant.

6 Conclusion

In a complex reality such as the financial market, it is reasonable to question the efficiency and reliability of classic statistical procedures. Indeed, the highly approximative nature of classic parametric models makes them possibly unreliable in contexts where the reality is significantly far from the approximation assumed by the model. In a financial context, the OLS version of the CAPM, which is highly sensitive to outliers, is affected by the frequent occurrence of abnormal returns. Such effect is non-negligible if we want to derive an estimate that provides a good fit to the bulk of the historical data. Therefore, in this thesis I evaluated a robust estimator, mOpt, in comparison to the OLS, as the former, which downweights extreme observations, should provide a better fit to the majority of the data. I applied our analysis to different stock sectors in order to assess differences in the effect of a robust approach.

I ran the mOpt and OLS versions of the CAPM on 220 stocks from 11 stock market sectors. I examined the average and the dispersion of both types of estimates for each sector in order to evaluate the stability of betas within stock market sectors, and how such stability changes with a robust approach. The results indicate an increase in stability and homogeneity of estimates across stocks in the same sector when switching from OLS to mOpt, except for the utilities and basic materials sectors. Further, I applied the same analysis on the differences between LS and mOpt estimates. Looking at the standard deviations, I found that the effect of a robust estimator is pretty homogeneous in the telecommunication, consumer discretionary, technology and financial sectors. This might indicate a correlation between abnormal returns in these sectors since the difference between OLS and mOpt estimates depends on such outlying observations. Then, the outcome of a Wilcoxon Signed-Rank Test defined the significance, at a 5% level, of the differences between the two types of estimate for stocks in utilities, real estate and financial sector, as well as for the entire sample of 220 stocks. Moreover, the robust approach results in higher LS estimates for all sectors, except for the technology stock sector, with an overall average difference of 0.086. Hence, I investigated what pattern of abnormal returns could account for the OLS CAPM line's generally greater slope when compared to the robust version. I divided observations according to the sign of the market excess return and I identified the frequency and average intensity of four types of influential points: right-side high and low regression outliers and left-side high and low regression outliers. The results are consistent with our assumptions that lower robust estimates are a consequence of a higher frequency and magnitude of right-side high and left-side low influential points.

Overall, I demonstrated that the mOpt estimator does actually lead to different results compared to the classic ordinary least square estimates of the CAPM model. I showed that comparing the classic and robust CAPM estimates uncovers interesting characteristics of stocks belonging to the same stock market sector. Moreover, I found out that such comparison also reveals an interesting trend of influential points which usually brings about a smaller robust beta.

More research could be conducted to overcome some of the limitations of the analysis conducted in this work. I considered only 20 stocks per stock market sector which is plausibly not enough to assess the nature of such sectors. Therefore, the same analysis on a larger number of stocks might reveal other interesting features shared by stocks in the same sector. Moreover, if the type of stock market sector actually defines a different trend of abnormal returns, it might be possible to evaluate the optimal robust factor model for each sector. In addition to a more in-depth analysis of the common nature of the returns of stocks belonging to the same sector, I could improve the outliers detection procedure. Indeed, I divided influential points according to the sign of the market excess return as its average value is 0.0032 and it seemed a reasonable choice. However, there might be a better procedure to identify those points that influence the OLS CAPM line by rotating it counterclockwise with respect to the robust line. For instance, I could identify regression outliers whose removal results in a lower LS estimate.

References

[1] Vic Barnett and Toby Lewis. *Outliers in Statistical Data.* Wiley, 3 edition, 1994.

[2] Peter J. Barry. Capital Asset Pricing and Farm Real Estate. *American Journal of Agricultural Economics*, 62(3):549–553, 1980.

[3] Raymond J. Carroll and A. H. Welsh. A Note on Asymmetry and Robustness in Linear Regression. *The American Statistician*, 42(4):285, 1988.

[4] Louis K. C. Chan and Josef Lakonishok. Robust Measurement of Beta Risk. *The Journal of Financial and Quantitative Analysis*, 27(2):265, 1992.

[5] Navin Chopra, Josef Lakonishok, and Jay R. Ritter. Measuring abnormal performance: Do stocks overreact? *Journal of Financial Economics*, 31:235–268, 1992.

[6] Bradford Cornell and J. Kimball Dietrich. Mean-Absolute-Deviation Versus Least-Squares Regression Estimation of Beta Coefficients. *The Journal of Financial and Quantitative Analysis*, 13(1):123, 1978.

[7] Corporate Finance Institute. The SP Sectors, 01 2022.

[8] Eugene F. Fama. The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1):34, 1965.

[9] Craig French. Jack treynor's 'toward a theory of market value of risky assets'. *SSRN Electronic Journal*, 12 2002.

[10] Dan Galai and Ronald W. Masulis. The option pricing model and the risk factor of stock. *Journal of Financial Economics*, 3(1-2):53–81, 1976.

[11] Marc G. Genton and Elvezio Ronchetti. *Robust Prediction of Beta*, pages 147–161. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[12] Robert S. Hamada. THE EFFECT OF THE FIRM'S CAPITAL STRUCTURE ON THE SYSTEMATIC RISK OF COMMON STOCKS. *The Journal of Finance*, 27(2):435–452, 1972.

[13] Frank R. Hampel. *Contributions to the theory of robust estimation.* PhD thesis, [Berkeley, 1968.

[14] Frank R. Hampel. A General Qualitative Definition of Robustness. *The Annals of Mathematical Statistics*, 42(6):1887–1896, 1971.

[15] Frank R. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974.

[16] Peter J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.

[17] Michael C. Jensen. Agency Cost Of Free Cash Flow, Corporate Finance, and Takeovers. *SSRN Electronic Journal*, 1999.

[18] Peter J. Knez and Mark J. Ready. On the robustness of size and book-to-market in cross-sectional regressions. *The Journal of Finance*, 52(4):1355–1382, 1997.

[19] M. K. Koller and M. M. Mächler. Definitions of -Functions Available in Robustbase. Technical report, 4 2022.

[20] Kjell Konis and R. Douglas Martin. Optimal Bias Robust Regression Psi and Rho Revisited. *SSRN Electronic Journal*, 2021.

[21] Kjell P. Konis and R. Douglas Martin. Optimal bias robust regression psi and rho revisited. *Political Methods: Quantitative Methods eJournal*, 2021.

[22] JOSEF LAKONISHOK and THEO VERMAELEN. Anomalous Price Behavior Around Repurchase Tender Offers. *The Journal of Finance*, 45(2):455–477, 1990.

[23] Robert A. Levy. On the Short-Term Stationarity of Beta Coefficients. *Financial Analysts Journal*, 27(6):55–62, 1971.

[24] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. In W.T. ZIEMBA and R.G. VICKSON, editors, *Stochastic Optimization Models in Finance*, pages 131–155. Academic Press, 1975.

[25] Benoit Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36(4):394, 1963.

[26] Harry M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press, 1959.

[27] Ricardo Maronna, Douglas Martin, and Victor Yohai. *Robust Statistics: Theory and Methods*. Wiley, 1 edition, 2006.

[28] R. D. Martin, V. J. Yohai, and R. H. Zamar. Min-max bias robust regression. *The Annals of Statistics*, 17(4):1608–1630, 1989.

[29] R. Douglas Martin and Daniel Z. Xia. Efficient bias robust regression for time series factor models. *Journal of Asset Management*, 23(3):215–234, 2022.

[30] Jan Mossin. Equilibrium in a capital asset market. *Econometrica*, 34(4):768–783, 1966.

[31] P. Rousseeuw and V. Yohai. Robust regression by means of s-estimators. In Jürgen Franke, Wolfgang Härdle, and Douglas Martin, editors, *Robust and Nonlinear Time Series Analysis*, pages 256–272, New York, NY, 1984. Springer US.

[32] Peter J. Rousseeuw. Multivariate estimation with high breakdown point. 1985.

[33] Peter J. Rousseeuw and Bert C. van Zomeren. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411):633–639, 1990.

[34] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, 19(3):425–442, 1964.

[35] John W. Tukey. A survey of sampling from contaminated distributions. 1960.

[36] Victor J. Yohai. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2):642–656, 1987.

[37] Víctor J. Yohai and Ruben H. Zamar. Optimal locally robust m-estimates of regression. *Journal of Statistical Planning and Inference*, 64(2):309–323, 1997.

List of Figures

List of Tables

A List of stocks

telecommunication	energy	utilities	basic_materials	industrial	consumer_staples
VZ	XOM	NEE	FCX	TMO	KO1
CSCO	CVX	DUK	NEM	RTX	PEP
TMUS	COP	SO	SCCO	UNP	UL
WBD	EOG	D	FAST	HON	SONY
HPE	PXD	WM	PKG	CAT	DEO
LUMN	OXY	EPD	GPK	LMT	BUD
ROKU	MPC	AEP	UFPI	DE	MDLZ
VG	VLO	SRE	HL	MMM	KDP
ATUS	DVN	EXC	CMP	BA	MNST
USM	PSX	WMB	CDE	CSX	ABEV
BDC	HES	KMI	PLL	ITW	STZ
LILA	HAL	RSG	SWM	GD	KHC
LILAK	BKR	XEL	NP	DCI	HSY
IDCC	MPLX	ET	CLW	NSC	GIS
TDS	CTRA	LNG	MSB	APD	KR
ESE	FANG	ED	GLT	EMR	WBA
EXTR	CLR	PEG	GATO	LHX	TSN
SHEN	MRO	WEC	GORO	ROP	FMX
ADTN	EQT	ES	HYMC	A	HRL
CNSL	APA	OKE	AUMN	PH	CCEP
consumer_discrectionary	real_estate	financial	healthcare	technology	
AMZN	PLD	JPM	JNJ	AAPL	
TSLA	CCI	BAC	DXCM	MSFT	
V	EQIX	WFC	UNH	GOOG	
PG	PSA	MS	PFE	GOOGL	
MA	O	SCHW	LLY	FB	
WMT	WELL	AXP	ABBV	NVDA	
HD	DLR	AMT	MRK	AVGO	
COST	SPG	SPGI	ABT	ADBE	
DIS	SBAC	GS	DHR	ORCL	
CMCSA	AVB	BLK	BMY	CRM	
NKE	EQR	C	AMGN	INTC	
MCD	WY	BX	MDT	AMD	
PM	ARE	MMC	CVS	TXN	
UPS	EXR	USB	ANTM	QCOM	
T	LEN	CME	SYK	IBM	
LOW	INVH	PNC	CI	INTU	
CHTR	VTR	PGR	GILD	AMAT	
PYPL	MAA	TFC	ZTS	NOW	
MO	DRE	AON	ISRG	ADI	
BKNG	SUI	ICE	BDX	MU	

Table 5: List of stocks for stock market sectors