RISK RETURN EFFICIENCY:

A DATA ENVELOPMENT ANALYSIS ON THE CROSS SECTION OF STOCK RETURNS

Alexandru C. Badoi

Erasmus University Rotterdam

Erasmus School of Economics

This dissertation is submitted for the degree of Financial Economics

November 2016

DECLARATION

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification.

Alexandru C. Badoi

Signed: ______________________

Date: ________________________

ABSTRACT

PURPOSE:

This study uses the Data Envelopment Analysis (DEA) CCR model as a value asset selection criterion to construct ten portfolios and investigate their relative performance.

METHODOLOGY:

The Fama French (2015) factor model is used to decompose the systematic risk into five components. The CCR DEA is then applied to the risk factors and asset returns with the goal of selecting superior investment options from a large US sample. Some of the empirical complications with implementing the CAPM model are also addressed through the implementation of Quantile Regression.

FINDINGS AND VALUE ADDED:

This paper provides the groundwork for future DEA research in the context of risk and return of financial assets. Moreover, significant alphas are generated while optimizing the cross-section of factors and returns indicate towards the existence of an anomaly among the formed portfolios at daily frequency.

I thank my mum and Derek for supporting me during my studies and to Dr. Alexander Loudon for providing the computing power necessary to calculate the DEA analysis.

1 INTRODUCTION

Asset pricing in Financial Economics is mainly concerned with identifying market inefficiencies and the drivers of return. This paper aims to analyze specific cases; that is undervalued stocks with strong fundamentals and overvalued stocks with weak fundamentals. "Value Investing" was first suggested by Ben Graham and David Dodd in the 1930s and it generally involves buying securities that are underpriced as determined by some form of fundamental analysis. The term "value" was never mentioned by Graham himself and his idea is still subject to miss-interpretation, the most common being that he is simply suggesting cheap or low priced stocks.

One thing that should be widely understood is that security prices tend to diverge from their fundamental value quite often. Investors have traditionally gotten over-excited about stocks which provide high returns, often ignoring fundamentals and further bidding prices of already overvalued companies. Conversely, wide-spread pessimism about a stock could potentially drive the price so low that it overstates the company's risks and understates the prospects of positive future returns. This crowd behavior of investors can lead to exploitable conditions in these so called "value stocks" and this is a view defined as "Contrarian Investing". This involves purchasing distressed stocks with good fundamentals irrespective of the market's opinion and selling them once they converge with their intrinsic value.

The introduction of the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966; Black, 1972) sparked a stream of empirical research surrounding the idea that there is a positive linear relationship between stock returns and the CAPM beta. Soon after it came to light that there are other risks associated with holding securities with certain fundamental characteristics. It was found that the CAPM beta became insignificant in explaining returns of portfolios sorted on size and the ratio of book value to market value (B/M) of firms. Fama and French (1993) formalized this idea into a three factor model and concluded that adding the combination of size, and B/M performs better in explaining the variation in the cross section. Recently it was found that the aforementioned three factors cannot explain returns of portfolios ranked on profitability and investment and the model has been extended to a total number of five risk factors. Their empirical research shows how the five Fama-French model incorporates most of the information contained in accounting ratios traditionally used in fundamental analysis. (Fama and French, 2015).

Frazzini (2014) adds to this factor literature by suggesting that investors are willing to pay a premium for what he defines as being high quality stocks. According to his definition these stocks are safe, profitable, growing and well managed. The Fama French (2015) value factor (B/M) is a good proxy for identifying what we would traditionally call value investment opportunities however there is a very clear distinction between the two: Frazzini sorts stocks on certain quality characteristics, irrespective of its price, while Fama and French value factor sorts stocks on price irrespective of quality. It seems natural that both of the ideas above should be combined and this will be the focus of this paper. This concept is not new:

> "Investment must always consider the price as well as the quality of the security."
>
> *Benjamin Graham*, *Security Analysis: Principles and Technique*

Traditional forms of Fundamental analysis involve ratio analysis such as debt – to – equity, price – to – earnings, current ratio, dividends and discounted cash flow analysis. This is however one of the most time consuming forms of analysis and in the fast paced trading style of the 21st century obtaining accurate key information at a rapid pace is highly important. There is a growing need for the implementation of methods which can help a value investor select fundamentally superior stocks. At the time this paper was written there were 3181 companies listed on the National Association of Securities Dealers Automated Quotations (NASDAQ), 3216 on the New York City Stock Exchange (NYSE) and 373 on the American Stock Exchange (AMEX). Governments, Funds, Investment Managers and other parties involved in the financial markets could benefit from quantitative methods which could screen and analyze fundamental characteristics in this ever increasing number of stocks.

"An investment operation is one that can be justified on both qualitative and quantitative grounds."

Benjamin Graham, Security Analysis: Principles and Technique

This paper is related to a variety of methodologies. It applies a Data Envelopment Analysis as a cross-sectional benchmarking tool on the Fama French factor model to measure and rank the Fundamental "Quality" in stocks. This paper is not concerned with the technicalities involved in construction those factors, instead we measure the fundamental strength of a firm via proxies provided by the Fama-French factor portfolios. I complement the literature by showing that investors pay a premium for stocks or baskets of stocks which are characterized by large and positive Fama-French factor coefficients i.e. have strong fundamentals. Thus a new source of alpha is identified in the market. Following high negative return days long investors take advantage of the oversold conditions found in stocks with strong fundamentals and open positions. Similarly, investors which are holding short positions in the same securities will take advantage of the high negative return days and cover their short by buying the amount they shorted. Thus a combination of short covering and speculation from the long side after days of large negative returns contributes to a constant Alpha on the second day. The opposite effect would have been expected about fundamentally weak stocks after days of high positive returns. In this case however I find that the strategy continues to be biased to the long side and also generates a positive alpha. I suggest that a "short squeeze" effect might be taking place. It makes sense that large short positions could be open on what this paper identifies as being fundamentally weak stocks. Thus after a day of high positive returns, margin calls and short covering for safety might be taking place.

To summarize, the Fama - French five factor model will be used to measure the fundamental strength of stocks and in order to get a most accurate estimate of those risk coefficients this paper uses a quantile regression approach. The factors are then used as inputs into the Charnes (1978) DEA CCR Model. This is a data oriented approach used to evaluate the performance of DMUs (Decision Making Units) which convert multiple data inputs into multiple data outputs. This study assumes that stocks are DMUs which convert inputs (the extended CAPM factor coefficients) into outputs (returns). I will later show in the methodology the usefulness of DEA at benchmarking fundamentally strong stocks versus their weaker peers.

Specifically, the purpose of this study is to analyze returns on contrarian portfolios:

- Baskets of stocks characterized by large/positive Fama French risk coefficients (strong fundamentals) after days of large negative returns.
- Baskets of stocks characterized by small/negative Fama French risk coefficients (weak fundamentals) after days of large positive returns.

The remainder of this paper is structured as follows: Section 2 will provide an overview of the theoretical background which underpins the three quantitative models used in this paper, specifically the extended CAPM, the CCR Data Envelopment Analysis model and Quantile Regression. Section 3 summarizes the data and methodology. Section 4 presents the results, proofs and Section 5 concludes the paper and suggest some topics for further study

2 THEORETICAL BACKGROUND

This section will present the reader with the theoretical background on the quantitative methods used in this study. It begins by evaluating the developments leading up to the Fama – French (2015) five factor model. It will then briefly review some theoretical elements of Quantile Regression which be used to estimate the systematic risk components associated with each security in this study. We then look at the underpinnings of DEA modelling which will later be applied in the Methodology section as a period by period benchmarking tool.

2.1 The Fama – French Extended Model

The Capital Asset Pricing Model (CAPM) was developed independently by Sharpe (1964), Lintner (1965), Mossin (1966) and Black (1972) and is based on the idea that markets should only compensate investors for the systematic component of risk which cannot be diversified away by just holding a broad portfolio of assets. Bounded by a set of assumptions[1] it implies that investors only hold two kinds of assets – identical risky assets (which coincide with the market portfolio) and risk free assets. If the market portfolio is mean-variance efficient we can derive the major conclusion of this model: that the expected return of any asset is solely determined by the sensitivity of its returns to the market returns, namely the market beta. Assuming a linear relationship between the market beta and an asset's returns, we can model CAPM as follows:

$$r_A - r_f = \alpha + \beta_A(r_M - r_f) + \varepsilon \qquad [1]$$

1. All investors: 1) Are maximizers of economic utility; 2) Are rational and risk averse; 3) Diversify across a broad range of investments; 4) Are price takers and can't influence prices; 5) Can lend and borrow any amounts at the risk free rate; 6) Incur no transaction costs; 7) Invest in securities which are infinitely divisible; 8) Have homogenous expectations; 9) Benefit from information symmetry. Glen, A. (2005)

Where: r_A and r_M are the retrurn of the asset and the market portoflio respectivelly, β_A is the regression slope coefficent i.e. the sensitivity of the asset to the market portfolio; α is the regression intercept or return with no associated risk; ε is the Standard Error of the regression.

In contrast to the original CAPM model, Fama and French (1992, 1993) have concluded that stocks with certain characteristics tend to do better than the market as a whole: Small cap (the size effect) and stocks with a low ratio of market value to book value (the value effect). Two decades later the same two scholars brought forward an additional two factors namely profitability and investment which they found significant in explain an asset returns (Fama and French, 2015). The extended CAPM is show in Eq. [2] below:

$$r_A - r_f = \alpha + \beta_A(r_M - r_f) + s_A SMB_A + h_A HML_A + \varphi_A RMW_A + c_A CMA_A + \varepsilon \qquad [2]$$

Where s_A , h_A, φ_A , c_A capture the asset's sensitivity to the four additional risk factors. The SMB (long small cap firms and short large cap firms) represents "size" risk. This comes from the idea that firms with a low market capitalization generate superior returns and are also more sensitive to extreme economic events due to their undiversified nature. The HML (long high B/M and short low B/M) represents the higher risk associated with "value" stocks. RMW (long most profitable and short least profitable) is the profitability factor which is a proxy for stocks with high future earnings. Finally also in agreement with the findings of Titman, Wie and Xie (2004), CMA (long conservative and short aggressive) is a proxy for stocks whose management invests the company's cash in a conservative matter.

2.2 Quantile Regression

This paper aims to get a most accurate estimation of those through the application of a Quantile Regression Methodology. Traditional methods of quantifying CAPM such as ordinary least squares (OLS) do so by assuming a linear relationship across the mean of the distribution. In other words, the upper and lower tails of the distribution are not assessed so the assets are assumed to behave as their mean would predict.

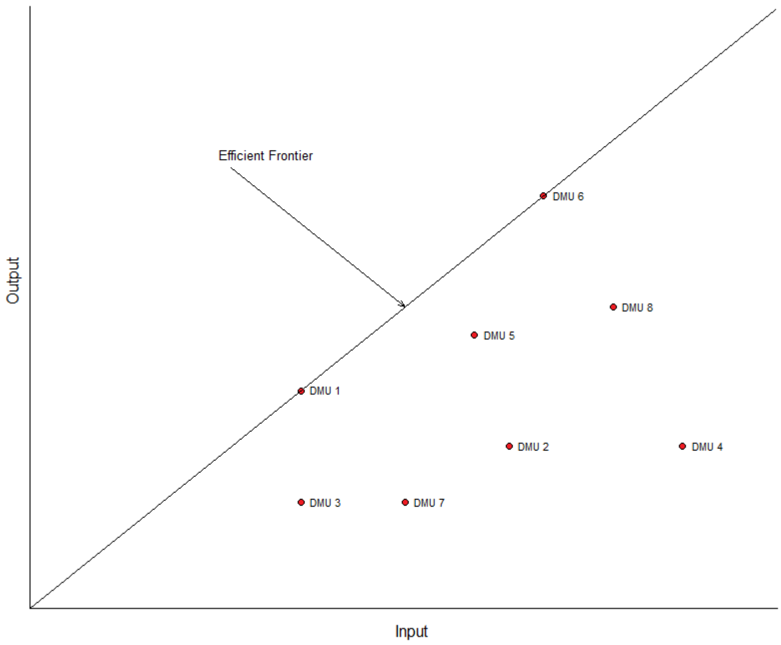
One method which helps alleviate some of these problems is the quantile regression approach which is an extension of the OLS estimation (Koenker and Bassett, 1978). A linear regression coefficient gives us an estimate of the change in the dependent variable produced by one-unit change in the independent variable associated with that coefficient. Quantile regression coefficients on the other hand give us an estimate of the change in the specified quantile of the dependent variable produced by one-unit change in the independent variable. Thus, this method allows us to model the return performance of firms which underperform or over-perform as predicted by the conditional mean of the firm's returns.

Studies suggest that using quantile regression to model the skewed and fat tailed distribution of stock returns may be much more appropriate: Buchinsky (1998) finds that quantile estimators are more efficient than OLS when the distribution is non-normal. This is confirmed by other academics who further test the CAPM in the context of quantile regression. They conclude that it is indeed superior to OLS and that this methodology alleviates other statistical problems encountered in many of the CAPM studies such as errors in variables, omitted variable bias and sensitivity to outliers. (Barnes and Hughes, 2002; Gerrans, Singh and Powell, 2009)

2.3 Data Envelopment Analysis and the CCR Model

Various definitions of Data Envelopment Analysis can be found in literature. Cooper et. al. (2007) provides a very intuitive explanation of a DEA frontier and compares it to a regression line. In the case of a regression, the line goes through the "middle" of the data points with the ones above being considered "superior" and the ones bellow being inferior". One can measure the degree of inferiority or superiority by calculating the deviations from this line and that is the distance from the data-point to the line. On the other hand, the DEA frontier line attaches to the performance of the most superior data point and then measures the deviations of others from this line as shown by Figure 1.

Figure 1



There is a fundamental difference between statistical approaches such as regression and frontier analysis such as DEA: The former reflects an "average" or "central tendency" behavior of observations while the latter deals with best performance and the benchmarks all performances by deviations from this line, making it especially suited for the ranking of financial assets. A special case of DEA is the one input – one output model which can be expressed as follows:

$$e = \frac{Data\ Output}{Data\ Input} \quad [3]$$

I now expand Eq. [3] into a multi input – multi output model. By attaching weights to the numerator and denominator this DEA model will allow us to optimize the above ratio such that e is maximized.

We assume there are n DMUs (Decision Making Units) to be evaluated. Each DMU consumes varying amounts of m different inputs to produce s different outputs. Specifically, DMU_j consumes amount X_{ij} of input i and produces amount Y_{rj} of output r. We further assume that X_{ij} , $Y_{rj} \geq 0$ and that each DMU has at least one positive input and one positive output value. In the CCR model the objective is to determine the weights u , v using linear programing so as to maximize the composite efficiency score e of each decision making unit. This is defined by the ratio:

$$Max:\ e_j\,(u, v) = \frac{Weighted\ Sum\ of\ Outputs}{Weighted\ Sum\ of\ Inputs} = \frac{\sum_r u_r\, Y_{rj}}{\sum_i v_i\ \ X_{ij}} \qquad [4]$$

$$st.\ \frac{\sum_r u_r\, Y_{rj}}{\sum_i v_i\ \ X_{ij}}\ \leq\ 1 \qquad \forall\, j \qquad [5]$$

$$u_r\ , v_i\ \ \geq 0 \qquad \forall\, i, r \qquad [6]$$

The DEA model presented above was redesigned by Charnes and Cooper and Rhodes (1978) such that it selects a representative solution (i.e., the solution (u, v) for which $\sum_{i=1}^{m} v_i\, X_{ij} = 1$). The fractional DEA model presented above can now be expressed in linear form under the name CCR DEA in the following way:

$$Max:\ e_j = \sum_{r=1}^{s} u_r\, Y_{rj} \qquad [7]$$

$$\sum_{r=1}^{s} u_r\, Y_{rj}\ - \sum_{i=1}^{m} v_i\, X_{ij} \leq 0 \qquad [8]$$

$$\sum_{i=1}^{m} v_i\, X_{ij} = 1 \qquad [9]$$

$$u_{r\,,}\, v_i\ \ \geq 0 \qquad [10]$$

According to Cooper et. al. (2007) DMU_j is considered to be CCR efficient if $e_j = 1$ and there exists at least one optimal u_r, v_i with $u_r, v_i \geq 0$. Similarly, DMU_j is considered to be CCR inefficient if $e_j < 1$.

The subgroup of relatively efficient DMU_j where $e_j = 1$ serve as the basis for the determination of the efficient frontier and establish a benchmark against whom all other DMUs are measured. For $e_j < 1$, this indicates the possibility of constructing a superior DMU_j utilizing an equal or lower quantity of inputs than DMU_j which is analyzed. (Lopes et al, 2008). For $e_j < 1$ one can measure the degree of inefficiency of DMU_j by computing $(1 - e_j)$.

The original CCR model which is presented above is said to be input oriented. In relation to the orientation of inputs or outputs, it is important to note that if DMU_j is efficient in the input oriented model it will also be efficient in the output oriented one, however for the inefficient DMUs the efficiency variable e_j can be different. (Bhat et al. 2001).

The application of DEA to financial markets only emerged in the last decade and literature directly relating to asset selection or financial security analysis is scarce. Thus, apart from the research topic chosen here, one of the objectives of this paper is to also demonstrate how DEA can be used to analyze and rank financial assets. I proceed to outline some of the developments which inspired this paper:

At first the model was proposed to study the effect of financial ratios on stock prices. This concept is based on the idea that numbers in financial statements reflect the performance and efficiency of the company. Powers and McMullen (2000) use the CCR DEA model to select a group of desirable securities from a group of 185. In their DEA analysis they use Earnings per Share (EPS), Price – Earnings per Share (P/E) Ratio, CAPM Beta and Standard Deviation as inputs and 1, 3, 5 and 10 year return as output. Fourteen securities within their asset universe were found to be efficient but the authors don't compare the performance of the efficient and inefficient portfolios thus we cannot know for sure if those assets were indeed superior in terms of risk and return performance.

A study also done on the Brazilian stock market replicates the study done by Powers and McMullen (2000) and presents a similar approach using quarterly data over a period of 22 quarters. DEA CCR analysis is performed using price to earnings ratio, Beta and return volatility as inputs and earnings per share, and the last 12, 36 and 60 month return as outputs. (Lopes et. Al. 2008). In the later paper the DEA portfolio yields a mean return of 12.33% versus 7.03% in the XBrX 100 Brazilian stock Index. This may appear highly profitable however upon close inspection of the paper it is evident that ex-ante information was used when rebalancing the portfolio. In other words, the DEA analysis is performed at time 1 and the resulting efficient stocks are invested in the portfolio also beginning with time 1. This methodology has no investable value since future information is used to achieve the outperformance however it shows the potential benefits of using the CCR DEA methodology.

Gardijan and Kojic (2012) use the CCR DEA model to benchmark between 78 stocks listed on the Zagreb Stock Exchange in Croatia over a period of 4 years at monthly frequency. Expected Monthly return calculated as the 52-month historical average return is used as output and return variance, 95% VaR and CAPM Beta as inputs. The CCR DEA portfolio was found to underperform the market.

The paper by Pätäri et al. (2010) is closely related to the topic of value investing. The authors are attempting to create value portfolios by using stock price as an input parameter and Earnings/Share, Dividend/Share and Book Value/Share as outputs. They do however remove all stocks with negative ratios from their sample. As the methodology section will later explain this is probably because the CCR DEA model only accepts positive inputs and output parameters. I get around this issue by standardizing and rescaling the variables such that they all lie between 0 and 1. The authors find that the DEA portfolios have a marginally higher return.

The inspiration for this paper follows from the paper of Sing and Allen (2010) and their research methodology is very similar to what will follow in this paper. They were the first ones to decompose the variance and propose using the three factor CAPM model; that is market, size and value all estimated using a Quantile Regression Methodology. Risk factors are used as inputs and asset return as output for the DEA process. Although their paper is more of a model application study, it provides us with valuable insight into the potential of using this type DEA analysis. It is unclear from their paper whether the focus is on creating superior portfolios or comparing regression methodologies. They present us with two portfolios and show that DEA efficient assets using Quantile Regression estimates outperform DEA efficient assets that use OLS Regression estimates. They only present the reader with their "best" portfolios and do not benchmark against the market or other firms within their sample. Their study also suffers from small sample bias. Their data extends over 4 years of monthly data covering just the 30 stocks contained in the DJIA index.

DEA analysis measures the relative rather than absolute performance. In other words, if such an analysis was to be done on the worst performing stocks in the market the model would pick the "best of the worse" and still yield relatively efficient stocks. Due to this, it is obvious how this type of analysis would not perform well within a limited sample size. This can create further complications when you entirely remove all stocks with negative performance ratios as in the paper by Pätär et al. (2010) where the DEA CCR will pick "the best of the best" instead of the best overall. Besides using slightly different inputs and output, the later argument may provide a potential explanation to the vastly different results used in the studies presented above.

This paper aims to settle the above issues through the use of higher frequency data, a longer study period and a much larger and diversified sample. Moreover, most of the above authors have used DEA as a black box and have not specifically explained the effects of DEA analysis on the underlying portfolios, something this paper will also show.

3 DATA AND METHODOLOGY

This section will first explain the raw data collection and preparation process. It then describes how the DEA input data is estimated using Quantile Regression. This is followed by the data standardization and re-scaling methodology and finally all of the above are applied in the context of our analysis.

3.1 Data

Because the DEA process benchmarks stocks according to the relative size of the extended CAPM regression slope coefficients, it was necessary to have a large variety of stocks in our asset universe. As previously mentioned in a small sample environment this analysis would still yield relatively superior investment options but since the universe is small, we can’t know for sure if they have any investable value. The universe here combines the three leading Standard and Poor US indices: S&P 500 (Large Cap), S&P (MidCap) 400 and S&P (SmallCap) 600, for a total of 1500 stocks which cover approximately 90% of the US market capitalization. Fifteen years of daily closing price data was downloaded from the DataStream database, covering the period November 2000 to October 2015 consisting of approximately 3700 observations. Finally, certain stocks were eliminated from the dataset due to missing price data which caused computational issues; with the final count being rounded to the nearest divisible by 10, resulting in 1150 stocks. In order to estimate the extended CAPM risk coefficients, daily factor portfolio data for the period including the 10 year US T-Bill to be used as the risk free asset was downloaded from the Fama and French online database. Daily returns were calculated and both datasets were matched by date with those from the Fama French database in order to eliminate the possibility of mismatch errors.

3.2 Regression Methodology

This step of the methodology has the sole purpose of obtaining input variables for the CCR DEA process, namely the extended CAPM coefficients presented in Eq. [2]. Quantile regression provides us with the advantage of being able to individually model a more accurate estimate of the distribution for each of our assets. This is contrasted by a one-fits-all Gaussian approach which predicts returns around the mean and is incapable of describing the extremes of the distribution. This can include anything from earning days, news announcements, acquisitions or periods of financial distress.

Rolling window quantile regressions were estimated with a window length of 200 such that the risk coefficients are estimated at each point in time. The regression coefficients are estimated at the following quantile levels: 5%, 25%, 50%, 75%, and 95%. Similarly, to the DEA paper by Sing and Allen (2010) which follow the approach proposed by Chan and Lakonishok (1992) a symmetric weighting scheme is used such that the resulting estimators have weights which are in the linear combination of quantile regression coefficients. Following the notation from Eq. [2] the coefficients at each point in time *t* are estimated like so:

$$\alpha_A = 0.05\alpha_{(0.05,t)} + 0.2\alpha_{(0.25,t)} + 0.5\alpha_{(0.5,t)} + 0.2\alpha_{(0.2,t)} + 0.05\alpha_{(0.95,t)} \quad [11]$$

$$\beta_t = 0.05\beta_{(0.05,t)} + 0.2\beta_{(0.25,t)} + 0.5\beta_{(0.5,t)} + 0.2\beta_{(0.2,t)} + 0.05\beta_{(0.95,t)} \quad [12]$$

$$s_t = 0.05s_{(0.05,t)} + 0.2s_{(0.25,t)} + 0.5s_{(0.5,t)} + 0.2s_{(0.2,t)} + 0.05s_{(0.95,t)} \quad [13]$$

$$h_t = 0.05h_{(0.05,t)} + 0.2h_{(0.25,t)} + 0.5h_{(0.5,t)} + 0.2h_{(0.2,t)} + 0.05h_{(0.95,t)} \quad [14]$$

$$\varphi_t = 0.05\varphi_{(0.05,t)} + 0.2\varphi_{(0.25,t)} + 0.5\varphi_{(0.5,t)} + 0.2\varphi_{(0.2,t)} + 0.05\varphi_{(0.95,t)} \quad [15]$$

$$c_t = 0.05c_{(0.05,t)} + 0.2c_{(0.25,t)} + 0.5c_{(0.5,t)} + 0.2c_{(0.2,t)} + 0.05c_{(0.95,t)} \quad [16]$$

For the purpose of robustness testing the coefficients are also estimated at quantiles 10%, 30%, 50%, 70% and 90% and then calculated using the same methodology as in Eq. [11-16].

3.3 Standardization Methodology

Although the DEA method is somewhat robust to processing information on different numerical scales via weighting we will proceed to standardize each variable. This procedure makes the data more balanced and reduces the risk of imprecision in the calculation. I first take the cross-section of regression slope coefficients estimated in section 3.2 (β_t , s_t , h_t , φ_t , c_t , α_t , R_t) and the Z scores are obtained for each variable like so:

$$\boldsymbol{Z_{ij}} = \frac{(\boldsymbol{X_{ij}} - \overline{\boldsymbol{X_J}})}{\boldsymbol{\sigma_j}} \qquad [17]$$

Where:

X_{ij} = The value of variable j for stock i

$\overline{X_J}$ = The mean of variable j for all stocks

σ_j = The standard deviation of the value of variable j across all stocks

The DEA objective function does not accept negative values and requires that the minimum inputs and outputs to be zero. This ties back to the paper by Pätär et al. (2010) where the authors remove a large portion of the universe based on negative input variables. Therefore, Z- scores are rescaled such that the minimum is equal to zero and the maximum to 1 while still preserving their relative scale. This is shown below in the two equations below:

$$RZ_{ij} = \text{Abs (Min } Z_j \text{)} + Z_{ij} \qquad [18]$$

$$MRZ_{ij} = \frac{RZ_{ij}}{Column\ Maximum} \qquad [19]$$

At the final stage of its preparation, the input data for the DEA process should look as shown in Exhibit [1].

3.4 DEA Methodology

The CCR model accepts unlimited amounts of outputs and inputs, the only obstacle being the computing power necessary to perform the calculations. The optimization problem must be solved using linear programing software. For this purpose, the package lpSolve for R has been used. The DEA R code used for the DEA calculations alongside with its implementation procedure required to compute Eq. [7 – 10] was taken from Pessanha et al. (2013).

If one was to use excess return as output and the one factor β_{CAPM} as input into a one input - one output DEA model, then there will be no weights or weights restrictions and the model will be expressed by Eq. [18] as shown below:

$$e = \frac{Output}{Input} = \frac{R_i - R_f}{\beta_{CAPM}} \tag{20}$$

In DEA terminology, the asset with the highest ratio would be considered the most efficient. The ratio shows that the most relatively efficient securities will have the lowest betas and the highest return. This is also a well-known ratio in finance; that popularized by Jack. L Treynor which extends the work of Sharpe. The Treynor ratio measures the returns earned in excess of that which could have been earned on an investment with no diversifiable risk. However according to this definition the returns are measured against what is assumed to be the only source of systematic risk and that is the Market portfolio. Today this definition can be considered to be flawed since four other sources of systematic risk have been identified.

This paper suggests the application of the CCR DEA methodology presented in Eq [7 – 10], in a multi input – singe output model with the goal to simultaneously sort portfolios based on their exposure to the market portfolio, and also their fundamental characteristics:

I define “Fundamentally Superior” stocks as ones which have a combination of high magnitude positive Fama - French factor coefficients. Conversely I define “Fundamentally Inferior” stocks as ones which have a combination of high magnitude negative Fama - French factor coefficients.

Continuing with the notation of the standardized values from Eq. [19], the extended DEA model is shown in Eq. [21]:

$$Max\ e_o\,(u\,,v) \;=\; \frac{[MR(R_i - R_f)]u_1}{(MR\beta)v_1 + (MRs)v_2 + (MRh)v_3 + (MR\varphi)v_4 + (MRc)v_5} \qquad [21]$$

The objective function in Eq.21 aims to find the "best" set of weights u_1 , v_1 … v_5 for each asset that is analysed such that the efficiency variable e is maximized. The term "best", is used here to mean that e_i is maximized relative to all others when these weights are assigned to the risk factors of all assets in our universe. Up until this point it should be understood that if $e_i =$ 1, then security i sits on the efficient frontier estimated by the model, i.e. it provides the highest possible return for the lowest combination of regression slope coefficient values. We now provide a more robust interpretation of the variable e in the context of the above analysis:

Analysis Statements:

1. All else equal, if $e_i = 1$, then security i has a higher return relative to other assets at the time of the analysis.
2. All else equal, if $e_i = 1$, then security i is characterized by a lower combination of risk factor coefficients relative to all assets in our universe.

In order for one to understand how the model in Eq. [21] will perform in the context of our analysis, the characteristics of the extended CAPM regression coefficients, specifically magnitude and direction need to be evaluated. The market beta works under the assumption that there is a positive linear relationship between any asset and the market portfolio. More specifically, if an asset has a $\beta > 1$ it implies that it is said to be riskier than the market and for a $\beta < 1$ it is said to have a lower market risk. This also implies that for all assets in our investment universe the market beta will always be positive and will take values which will fluctuate above and below 1. Beta is unique in that regard as this characteristic does not hold for the other risk coefficients which are not bounded by a positive relationship. Negative factor loadings on SMB, HML, RMW and CMA are possible and according with their construction they would be characteristic of large firms with a high market valuation relative to book, who are unprofitable and invest aggressively. In order to show this variation of risk coefficients within our sample, OLS regressions are performed on each stock in the universe and their descriptive statistics are shown in Exhibit [2].

Because the DEA process is looking to achieve the maximum output (return) for the minimum input (factor coefficients), the objective function will assign a high e_i score to stocks with large negative factor loading and a high return. Similarly, a low e_i score will be assigned to stocks with large positive factor loadings and a low return. It is worth mentioning here how this paper differentiates itself from others presented at the end of Chapter 2. Most of the DEA literature applied to financial assets is focused mainly on stock selection rather than analysis of returns. Powers and McMullen (2000) use EPS and P/E ratio as two as their inputs alongside CAPM Beta and Standard Deviation. Using the reasoning presented above, it is clear how this may create complications in the analysis. The EPS and P/E variables can take negative values while CAPM Beta and Standard deviation can't. Thus their top portfolio where e_i is close or equal to 1, would contain securities with negative EPS, P/E and a low positive SD and Beta. Similarly, one of the variables used by Lopes e. Al. (2008) is EPS which also creates complication for reasons already mentioned.

To summarize, instead of using DEA analysis as an efficiency model, this paper uses DEA to rank and study the return of portfolios based on two attributes: a combination of high or low regression coefficients and daily return. High e_i stocks will be characterized by relatively large positive returns and relatively lower or even negative risk factor loadings. Conversely, low e_i stocks will be characterized by relatively large negative returns and relatively higher or positive risk factor loadings. This directly links to the research objectives of this paper. By aggregating portfolios based on the e_i score we aim to create:

- Baskets of stocks characterized by large/positive Fama French risk coefficients (fundamentally superior) after days of large negative returns

- Baskets of stocks characterized by small/negative Fama French risk coefficients (fundamentally inferior) after days of large positive returns.

3.5 Portfolio Formation

Based on everything so far, the benefits of using DEA analysis to analyse the cross section of stock returns become evident. A single efficiency score will be given to each security in our study which will allow us to create portfolio buckets based on the two attributes mentioned above, namely return and a combination of the five risk coefficients. Relatively speaking, stocks with higher returns and lower risk coefficients will be assigned a high score while stocks with low returns and high risk coefficients will be assigned a low score.

The DEA efficiency score e_i is estimated for each day and used as sorting criteria to create portfolio buckets. Each portfolio is composed by an investment in equal proportions in each selected stock, that is, all portfolios presented in this paper are equally weighted. Decile bucket portfolios are formed by ranking stocks on their previous period efficiency scores. Since there are 1150 assets in this study, the top 115 which have been found to have the highest efficiency scores are included in the top decile and similarly the smallest 115 efficiency scores for each period are included in the bottom decile. After the DEA efficiency scores have been calculated for one period, the investment in that group of assets is simulated to start at the beginning of the next period and end at the end of the same period. This procedure is repeated daily for the entire duration of the study.

4 RESULTS

The results section will proceed as follows: First I show that the two analysis statements presented in Section 3.4 are true; i.e. the model works as expected.

This means that all else equal:

1. The most efficient portfolio buckets will have the highest returns in the sample on the day of the DEA measurement is taken. Similarly, the least efficient bucket will have the lowest returns.

2. The most DEA efficient buckets will have the lowest combination of risk regression slope coefficients. Similarly, the least efficient bucket will have the highest combination of risk regression slope coefficients.

Secondly, all portfolio performance over the study period analyzed:

- Risk – return performance
- Fama – French factor decomposition
- Correlation Analysis
- Drawdown Performance

Finally I present robustness results. It was stated previously in the Methodology section that the quantile regression risk coefficients were estimated at two sets of confidence intervals:

- 5%, 25%, 50%, 75%, and 95%
- 10%, 30%, 50%, 70% and 90%

T- Test for differences between means are performed between each pair of decile yielded from the two sets of risk coefficients.

4.1 DEA Analysis Statements Proof

This section is dedicated to showing that the DEA methodology produces the desired results in terms of how it ranks stocks on both returns and risk factor coefficients. In order to prove Statement 1, additional data collection must be performed within each portfolio bucket. I am interested in the mean returns of each bucket on the day of the DEA analysis. If statement 1 holds true, these returns should be decreasing as we move from the most efficient bucket to the least efficient. The DEA calculation is done at time T – 1 and stocks are to be invested in at time T. Thus, for the purpose of this analysis all bucket returns at T – 1 are collected. 115 securities invested in over 3300 days translate into approximately 379,500 return observations per bucket portfolio; more than sufficient to yield significant results.

Figure 2

Return Distributions on DEA Day [T-1]

	100% - 90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%
Mean	1.80%	0.94%	0.64%	0.43%	0.25%	0.08%	-0.11%	-0.37%	-0.76%	-2.13%
Skewness	338.86	1.50	1.18	0.85	0.60	0.38	0.09	-0.20	-0.64	-1.87
Kurtosis	170807.77	11.67	10.86	10.16	8.88	8.56	8.09	7.19	7.04	9.91

Return Distributions on Investment Day [T]

	100% - 90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%
Mean	0.06%	0.03%	0.03%	0.04%	0.04%	0.05%	0.06%	0.07%	0.09%	0.16%
Skewness	402.57	1.91	0.93	0.25	2.67	0.10	0.39	0.71	0.91	9.51
Kurtosis	216309.28	94.16	44.65	26.68	285.64	26.39	19.87	31.62	37.21	1234.89

Figure 2 shows the distribution of mean returns across the ten portfolio buckets at both time T and T – 1. Looking at the first table (T – 1) the Mean calculations clearly show how on the day of the DEA analysis, a higher efficiency score e_i is associated with a higher mean return. The Skewedness and Kurtosis calculations also imply a much higher probability of returns above the mean for bucket 100 % - 90 % than it does for the lowest 10% - 0% but this is normal since this arrangement is done on purpose. When looking to maximize Eq. [21] the DEA model assigns a higher e_i to stocks with higher returns. Conversely, a lower e_i is assigned to stocks with a lower return. This confirms Statement 1.

Moving on to time T or the investment day, it is very interesting to show how the mean distribution across the buckets completely reverses, exception being bucket 100% - 90%. Here, we can observe decreasing means as we move from the least efficient buckets to the most with the highest being in 10% - 0% (0.0016). Surprisingly, high positive skewedness and excess kurtosis in bucket 100% - 90% imply a much higher frequency of returns above the mean than the rest of the portfolios. The data has however not been adjusted or cleaned of extreme outliers so the skew and kurtosis results display extreme values at both T and T -1.

<u>Figure 3</u>

Probaility of Up - Day [T]

	100% - 90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%
No. (+) Days	1822	1803	1810	1791	1808	1802	1828	1806	1827	1881
No. (-) Days	1517	1536	1529	1548	1531	1537	1511	1533	1512	1458
Prob (+)	54.57%	54.00%	54.21%	53.64%	54.15%	53.97%	54.75%	54.09%	54.72%	56.33%

Figure 3 also presents the probability of positive return across the buckets at time T. Within each bucket, the number of positive and negative returns is calculated. I obtain the probability simply by dividing the number of positive returns into the number of negatives. The bucket where stocks with the lowest efficiency score e_i are assigned has the highest probability of an "up" day with around 2% higher than the rest.

To prove *Statement 2*, I analyze the portfolio returns of each bucket through the extended CAPM factor decomposition presented in Eq. [2]. If *Statement 2* holds true we should observe lower or negative risk coefficients in the most efficient bucket which should become increasingly larger and positive as we approach the least efficient bucket.

To observe this effect, we turn our attention to the OLS regression heat maps presented in Exhibit [5]. I choose OLS instead of Quantile Regression of the results analysis because *P – Values* of the risk factor coefficients could not be shown using Eq. [11 – 16]. The tables show that for both quantile sets, the risk coefficients associated with each of RMRF, SMB, HML, RMW are increasing as we move from the most efficient to the least efficient portfolio. In the 5% -

95% portfolios, The Market beta (RMRF) is increasing as we move from the third highest and most efficient bucket (80% - 70%) where we have a value of 0.942 to the least efficient where we have a value of 1.255. Similarly, SMB increases without exception from 0.200 in the most efficient bucket to 1.011 in the least indicating that the last bucket contains mainly small cap stocks and conversely the first large cap. HML increases without exception from -0.161 to 0.578 indicating that value stocks are included in the last bucket while growth are included in the first. RMW performs almost the same as the former, ranging from -0.629 to 0.180 indicating that the most profitable stocks are included in the bottom decile. All slope coefficients mentioned up until this point are highly significant to the 1% level. Not the same can be said about the CMA factor where the factor arrangement looks completely random and non-significant. By decomposing portfolios in this way I show how the DEA model in Eq. [21] assigns a higher efficiency score e_i to securities with a higher combination of risk regression coefficients and conversely it gives a lower score e_i to securities with lower risk combinations, the exception being of course the investment factor CMA.

Although this is not the purpose of this paper, this methodology does also test the robustness of the factor data provided by the Fama French portfolio and their usefulness. The DEA methodology picks a different set of stocks each day and there should be no reason why after aggregation into portfolios those should not correlate with the CMA factor data. In the context of this analysis, it does not look robust or consistent throughout the results section of this paper.

Putting everything of the above together, I show that *Statement 1* and *Statement 2* are true and conclude this section with an important observation. The top or most efficient portfolio buckets include on average securities with relatively low or negative risk coefficients and relatively high and positive returns. Similarly, the least efficient portfolio buckets include on average securities with high and positive risk coefficients and relatively low and negative returns.

Results are consistent also at the second set of regression coefficients estimated at the 10%, 30%, 50%, 70% and 90% quantiles.

4.2 Decile Portfolio Performance

I begin here by looking at the Portfolio Summary Statistics Heat Maps in Exhibit [3] and the focus is on the 5% - 95% portfolios. Excluding what the model considers to be the two most efficient deciles, we can observe an increasing returns pattern. Beginning with 90% - 80%, the second most efficient decile, where we have an annualized value of 8% , returns are increasing to the least efficient where we have an impressive 41%. A similar effect is observed when looking at the standard deviation. A min of 0.20 is present in quantile 80% - 70% which is perfectly increasing to quantile 10% - 0% where we have a value of 0.31. This effect is also consistent with CAMP theory where a higher risk (as measured by standard deviation) is associated with higher returns translating into and increasing Annualized Sharpe Ratio from decile 80% through to 0%. This can also be observed in the Annualized Risk – Return Scatter Plot by looking at the extreme position of the 10% - 0% portfolio relative to the rest. The odd ones out are the two most efficient portfolios which do not fit the increasing pattern we observe in the others. A relatively high Sharpe Ratio of 0.63 is shown in decile 100% - 90% which is indeed an unexpected result.

Exhibit [4] charts the return performance over the investment period for all decile portfolios. The outperformance of the least efficient portfolios stands out. If $1 was invested in 2002 in the least efficient portfolio, it would now be worth close to $120. This is closely followed by the second least efficient portfolio where one would have seen an end investment value of just over $10. The most efficient portfolio comes in 4th place however it held second place up until mid-2006 and then fell into third and fourth in 2012 and 2014 respectively.

The returns can be explained using the extended CAPM decomposition in Eq. [2]. Since the arrangement of the risk regression coefficients across the buckets was presented in Section 4.1 I focus my attention on the last of the variables presented in the OLS regression tables in Exhibit [5] , namely the Alpha intercept. This is interpreted as being return which has no associated risk with either of our control variables. Since different buckets represent different levels of the efficiency score e_i, Alpha observations across the buckets are also a performance measure of this DEA methodology.

Also excluding the top two most efficient portfolios Exhibit [5] shows a perfectly increasing daily Alpha starting with -0.005% in decile 80% – 70% and ending with 0.109% in the last. Compounded, an Alpha of 0.102% a day is equivalent to around 31% a year.

Considering observations made in the previous section, it is shown that securities with larger negative returns and larger/ positive risk coefficients at time T-1 exhibit positive Alpha on the second day at time T. The magnitude of this effect decreases as we move up the buckets towards securities which at T – 1 have larger positive returns and lower/negative risk slope coefficients where we have an Alpha observation as low as -0.005% or approximately -1.2% a year.

The summary Table in Exhibit [7] shows averages of the worst five drawdowns within each decile portfolio over the entire investment period. While the Average Depth doesn't show any consistent patterns, it is very evident that a lower efficiency score e_i improves drawdown performance and conversely, a higher score worsens it. Average length of the drawdown is perfectly decreasing from an average of 407 days in the top decile to 132 in the last. Similarly the top decile drops for an average of 132 days versus 67 before bottoming and starting to recover. The recovery also takes an average of 337 days for the top decile portfolio which perfectly decreases to 65 in the bottom decile.

Correlation analysis results are shown in Exhibit [6] in order to present the potential diversification benefits that this DEA methodology has to offer. I chose two methods: First Pearson Correlation Coefficient and then a non-parametric, Kendall Rank Correlation Tau Coefficient. They both display the same results in terms of relative correlation between the deciles however in the non-parametric method there is a larger dispersion between the top and bottom. Also very evident is the decreasing correlations going from top to bottom deciles suggesting increasing differences between portfolios ranked on the DEA score e_i. The best way to explain the diversification benefits is to simply look at the correlation between each of the portfolios with the Asset Universe. Interestingly, although not the top performer, the first decile, namely 100% - 90% displays the lowest correlation thus is the largest diversifier. The diversification benefits disappear as we approach the last decile, however, although still high the last (10% - 0%) does not fit in with the pattern.

The correlation result presented above directly links to the factor decomposition displayed in Exhibit [5] where we have the lowest R-Squared value. On average the Fama – French models seems to perform quite well in explaining the 10 portfolio returns, that is the value fluctuates between 90% and $\sim$ 97% which indicates a high explanatory power; that is returns of the decile portfolios are almost fully explained by the information in the factors. This is not the case however for the first decile where the R-squared is significantly lower with a value of 76%. Performance-wise, the portfolio performs almost equally to the market portfolio both having a Sharpe Ratio of ~0.6 as displayed in Exhibit [3] however, factor analysis indicates that these returns don't come from the market or any of the additional four factors; or better said 3, given the poor performance of CMA.

4.3 Robustness Testing

The results which have been discussed up until this point refer mainly to ones derived from using risk coefficients calculated at quantiles 5% and 95%. This is because Quantile Regression is not the focus on this paper; it was instead used to estimate a marginally more accurate input data for the DEA process. In order to show that the methodology is robust to changes in quantile levels, coefficients were also estimated at quantiles 10%, 30%, 50%, 70% and 90%. T – tests are performed between each decile pair in order to test for significant differences between their Means. Results presented in Exhibit [8] are highly insignificant suggesting that they are almost the same portfolios.

5 Conclusions

This paper started with the assumption that investors desire securities which are fundamentally superior. Not only should they be superior, but they should also be cheap. The Fama – French risk factors were chosen as proxies for the fundamental characteristics of firms and what literature so far suggests are the drivers of stock returns.

It makes sense that an investor should desire to take long positions in securities which have larger and positive factor loadings. While an investor may not directly use the Fama-French model but instead use metrics such as Book Value, P/E or ROI to take this decision, the five risk factors used are academically proved proxies of these fundamental characteristics of firms. Positive factor loadings would be representative of small firms which have a high B/M ratio, with positive expected future earnings and who invest conservatively. When added up as per Eq. [2], these factors should arithmetically contribute positively towards the final period return. Therefore, regardless of the complexity of the method used, it's intuitive that any one investor should desire long positions in such securities. Conversely securities which have negative factor loadings should not be desirable to investors. Negative risk factors are representative of large firms, with a low B/M, with negative expected earnings who invest aggressively. For the above reasons portfolios were sorted on the two attribute groups like so:

- Baskets of stocks characterized by large/positive Fama French risk coefficients (strong fundamentals) after days of large negative returns
- Baskets of stocks characterized by small/negative Fama French risk coefficients (weak fundamentals) after days of large positive returns.

The above is a very specific case and chooses to investigate how stock returns of such securities behave after extreme days. Portfolios composed of what I assume to be the most fundamentally superior stocks after days of large negative returns were found to exhibit positivize alpha on the second day. This alpha is found to be decreasing and then become negative as the fundamentals approach the average of all other firms and eventually become weaker (moving through the deciles). An unexpected result is however the positive Alpha found in the top decile which is composed of the weakest stocks after days of high positive returns.

One possible explanation for the alpha presented in the bottom decile would be that following highly negative days in stocks with strong fundamentals investors are simply "buying the dip". It would seem counter intuitive to sound investment practices to take long positions in securities with highly negative returns. Moreover implied volatility increases with lower returns, thus this would imply risk seeking behavior amongst investors. If we however consider the fact that these securities have superior relative fundamental characteristics i.e. large and positive risk factor loadings, they would be representative of a "contrarian" investment opportunity. It is also possible that investors which are holding short positions would take advantage of the high negative return days cover their short buy buying the amount they have shorted. Thus a combination of short covering and speculation from the long side after a highly negative day may generate alpha on the second day. Moving upwards in the deciles, long holders could take advantage of high gain days and exit positions by selling on the second day. Similarly, speculators from the short side may take advantage of high prices and initiate a short position. Thus the combination of selling and short initiation starts contributing to a decreasing alpha through the deciles.

Finally, we turn our attention to the positive Alpha present in the last and most efficient bucket 100-90. Looking at the risk factor loadings, it is clear that the stocks included in this portfolios were some of the weakest from a fundamental point of view at the time of the analysis. According to those, the most efficient decile is composed of large cap stocks with a low B/M ratio, negative future expected earnings, which spend the company cash on investments in an aggressive fashion. If I was a short trader, such companies would be at the top of my list. Although this would be much more difficult to prove and would require additional data not available at the time this paper was written it's very possible that such securities have a very high Short Interest Ratio, which is the ratio of tradable shares being shorted to shares in the market. The second characteristic of this decile is that it has some of the highest returns in the sample on the day of the analysis. Thus it's possible that a large number of margin calls takes place on the second day forcing short sellers out of their position, and pushing up the price enough to generate a constant Alpha throughout our investment period.

I conclude with a few words about DAE analysis. This is the first paper to demonstrate the power of DEA as a cross-sectional tool at the scale and complexity that is presented here. It shows how such optimization techniques can be used to sort portfolios on multiple attributes. Just to demonstrate, the size factor is one that is based on just one attribute, that is market cap of the firm. By simply ranking assets from largest to smallest in the 90s when quantitative finance was just an emerging field, Fama and French have developed an incredible easy to use tool (or time series) which works very well almost 25 years later. The same can be said about the other factors used in this paper. The free lunch, return with no associated risk or Alpha has however become more and more elusive as investors started to take advantage of these newly discovered market anomalies and increasingly complex analysis techniques are now needed to outperform the market. DEA offers the user the ability to simultaneously sort through as many variables as is desired. It is also worth mentioning that the DEA model used here is one of the simplest and that much more robust and complex models have become available since its inception in the 70s. The future is indeed exciting and I hope that the research presented here provides a stepping stone for the use of DEA in quantitative fundamental financial analysis.

LIMITATIONS AND TOPICS FOR FURTHER STUDY

I identify three limitations. First survivorship bias is present within the sample for two reasons. Non-surviving stocks are missing from the original dataset thus some of the worse performers have not been accounted for in DEA. I also exclude stocks which have had their IPO after the beginning of this study for due to computational issues. Secondly, the CMA factor is clearly a bad performer in the context of this analysis so it should probably be excluded or replaced with a more suitable one. Lastly, the study assumes that the strategy can capture the closing price which is not realistic. Market on Close orders could be used however this also means that the DEA calculation must be performed in a very short amount of time. Regarding further studies directly related to this strategy, I encourage future researchers to evaluate stocks with weak fundamentals after days of high positive returns by looking at the Short Interest ratio. This directly relates to the first decile presented here - 100% - 90% and the low R-Square associated with that portfolio.

6 TABLES AND FIGURES

6.1 Exhibit 1

In the table below the first column represents the DMU, or stock in our case. The next one represents the standardized return which will be used as an output variable and the following 6 columns represent the standardized input variables. It should be noted that an individual dataset is estimated for each period t, each having n assets (rows) in length.

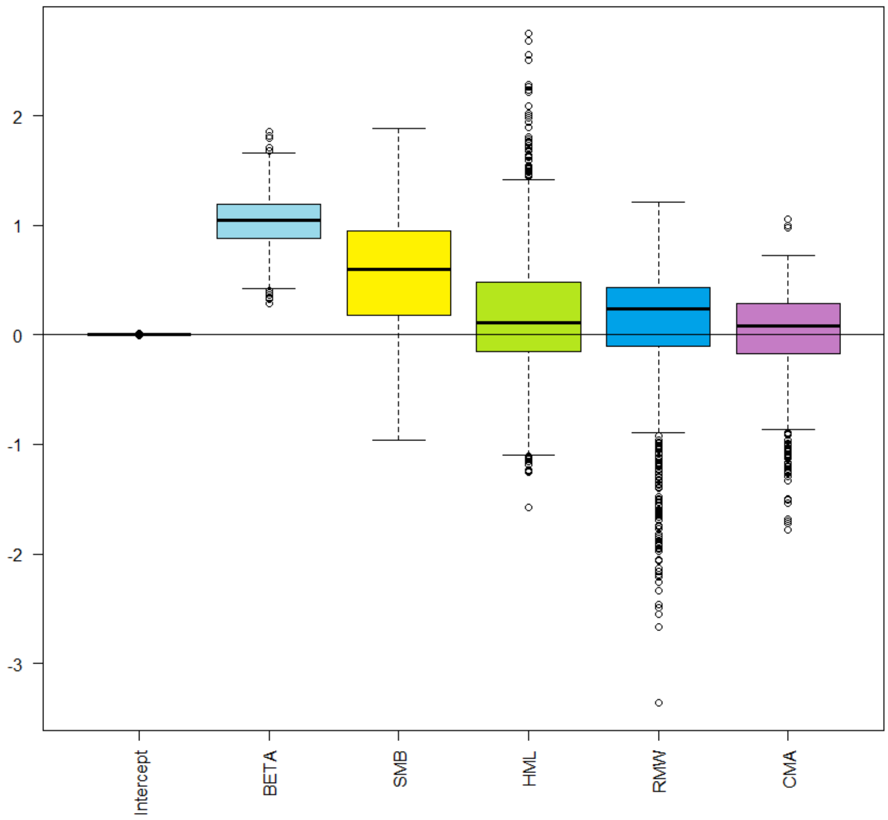
$\boldsymbol{DMU_{jt}}$	$\boldsymbol{R-rf}$	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{s}$	$\boldsymbol{h}$	$\boldsymbol{\varphi}$	$\boldsymbol{c}$
1_t	$MRReturn_{1,t}$	$MR\alpha_{1,t}$	$MR\beta_{1,t}$	$MRs_{1,t}$	$MRh_{1,t}$	$MR\varphi_{1,t}$	$MRc_{1,t}$
2_t	$MRReturn_{2,t}$	$MR\alpha_{2,t}$	$MR\beta_{2,t}$	$MRs_{2,t}$	$MRh_{2,t}$	$MR\varphi_{2,t}$	$MRc_{2,t}$
3_t	$MRReturn_{3,t}$	$MR\alpha_{3,t}$	$MR\beta_{3,t}$	$MRs_{3,t}$	$MRh_{3,t}$	$MR\varphi_{3,t}$	$MRc_{3,t}$
4_t	$MRReturn_{4,t}$	$MR\alpha_{4,t}$	$MR\beta_{4,t}$	$MRs_{4,t}$	$MRh_{4,t}$	$MR\varphi_{4,t}$	$MRc_{4,t}$
…	…	…	…	…	…	…	…
n_t	$MRReturn_{n,t}$	$MR\alpha_{n,t}$	$MR\beta_{n,t}$	$MRs_{n,t}$	$MRh_{n,t}$	$MR\varphi_{n,t}$	$MRc_{n,t}$

6.2 Exhibit 2

The summary statistics as well as a Box Plot of the DEA input variables, namely the risk regression slope coefficients are presented below. The Market Beta's unique characteristics become evident here, being the only risk factor bounded by positive extremes. This characteristic does not hold for all other risk coefficients which are not bounded by a positive relationship with the asset's returns. Negative factor loadings on SMB, HML, RMW and CMA are possible and are shown below.

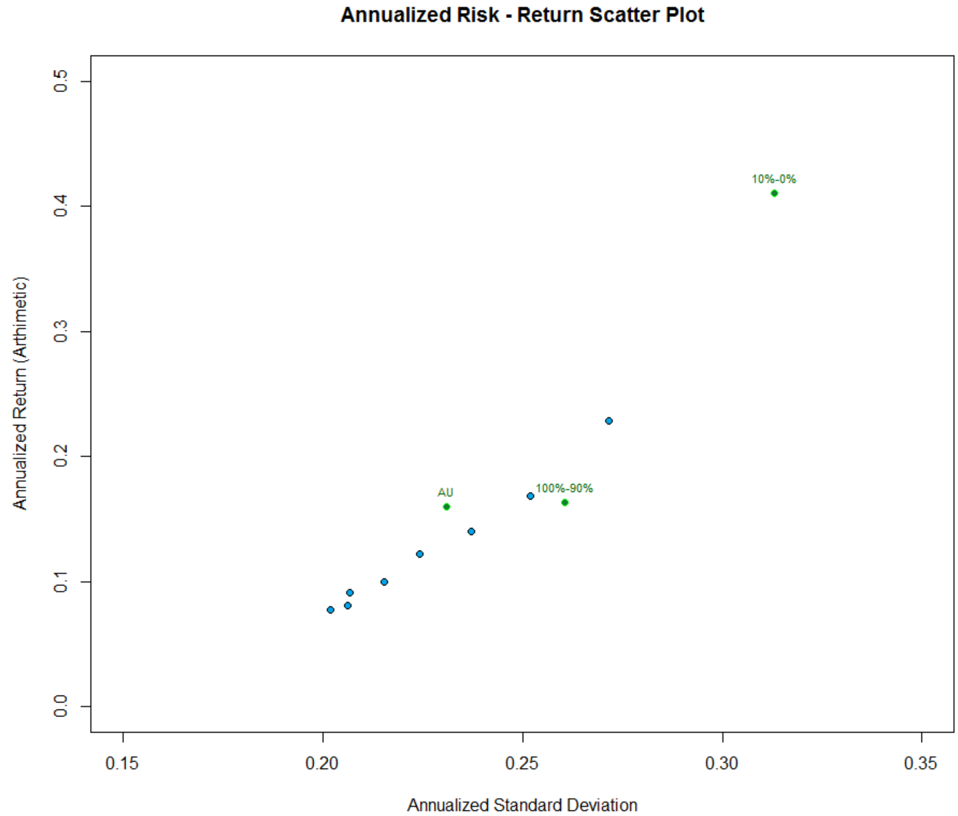
Risk Factor Loadings - Descriptive Statistics

	RMRF	SMB	HML	RMW	CMA
Observations	1151	1151	1151	1151	1151
Minimum	0.2883	-0.9607	-1.5741	-3.3641	-1.7833
Maximum	1.8535	1.8893	2.7502	1.2130	1.0537
Median	1.0455	0.5944	0.1140	0.2345	0.0859
Arithmetic Mean	1.0425	0.5690	0.1891	0.0516	0.0070
Variance	0.0565	0.2444	0.3459	0.3813	0.1680
Stdev	0.2377	0.4944	0.5881	0.6175	0.4099
Skewness	0.0195	-0.0241	0.6907	-1.8403	-1.1156
Kurtosis	-0.0012	-0.7736	1.5592	3.7195	1.6887



6.3 Exhibit 3

Summary performance statistics as well as a risk return scatter plot for all decile portfolios and Asset Universe are shown below. The returns are calculated arithmetically and the risk is calculated as the simple standard deviation. The Asset Universe portfolio represents an equal weighted average of all securities in this study. Tabulated results are shown for both sets of quantile levels at which regression coefficients were estimated. A linear relationship between risk and return becomes evident on the scatter plot however this does not hold for the top two portfolios 100% - 80%. Similarly the tables show decreasing returns and increasing standard deviations in the deciles. The additional returns however make up for an increasing risk and this translates into an increasing Sharpe Ratio as we approach decile 10% - 0%. One notable thing is that the middle portfolios seem to be the worst performers.



Performance Summary Statistics Heat - Map [5% - 95%]

	100% - 90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%	AU
Annualized Return	0.16	0.08	0.08	0.09	0.10	0.12	0.14	0.17	0.23	0.41	0.16
Annualized Std Dev	0.26	0.21	0.20	0.21	0.22	0.22	0.24	0.25	0.27	0.31	0.23
Annualized Sharpe (Rf=0%)	0.63	0.39	0.38	0.44	0.46	0.54	0.59	0.67	0.84	1.31	0.69

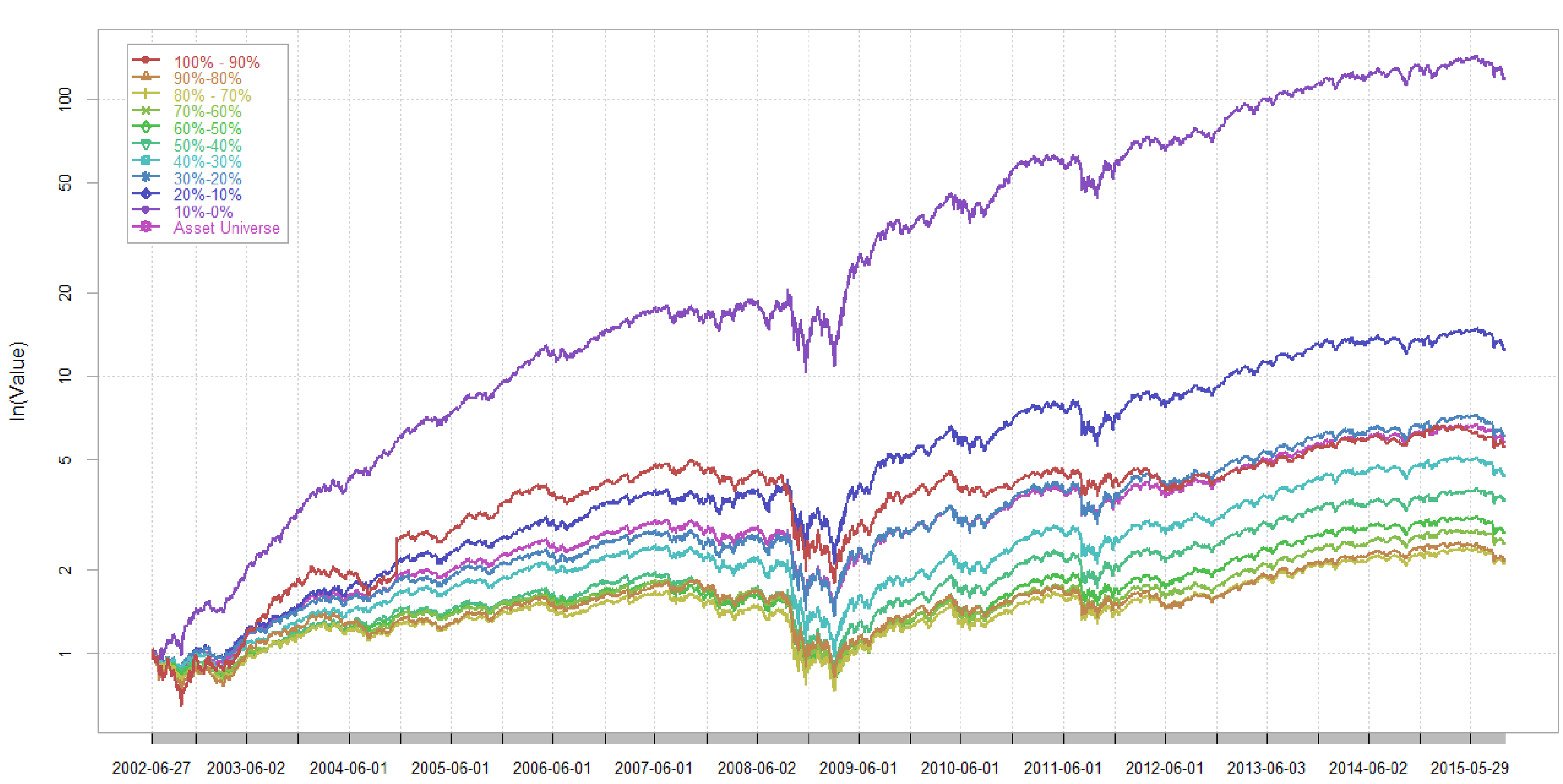
Performance Summary Statistics Heat - Map [10% - 90%]

	100% - 90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%	AU
Annualized Return	0.16	0.09	0.08	0.10	0.09	0.11	0.14	0.17	0.23	0.41	0.16
Annualized Std Dev	0.26	0.21	0.20	0.21	0.22	0.23	0.24	0.25	0.27	0.31	0.23
Annualized Sharpe (Rf=0%)	0.62	0.46	0.40	0.46	0.43	0.50	0.59	0.67	0.84	1.31	0.69

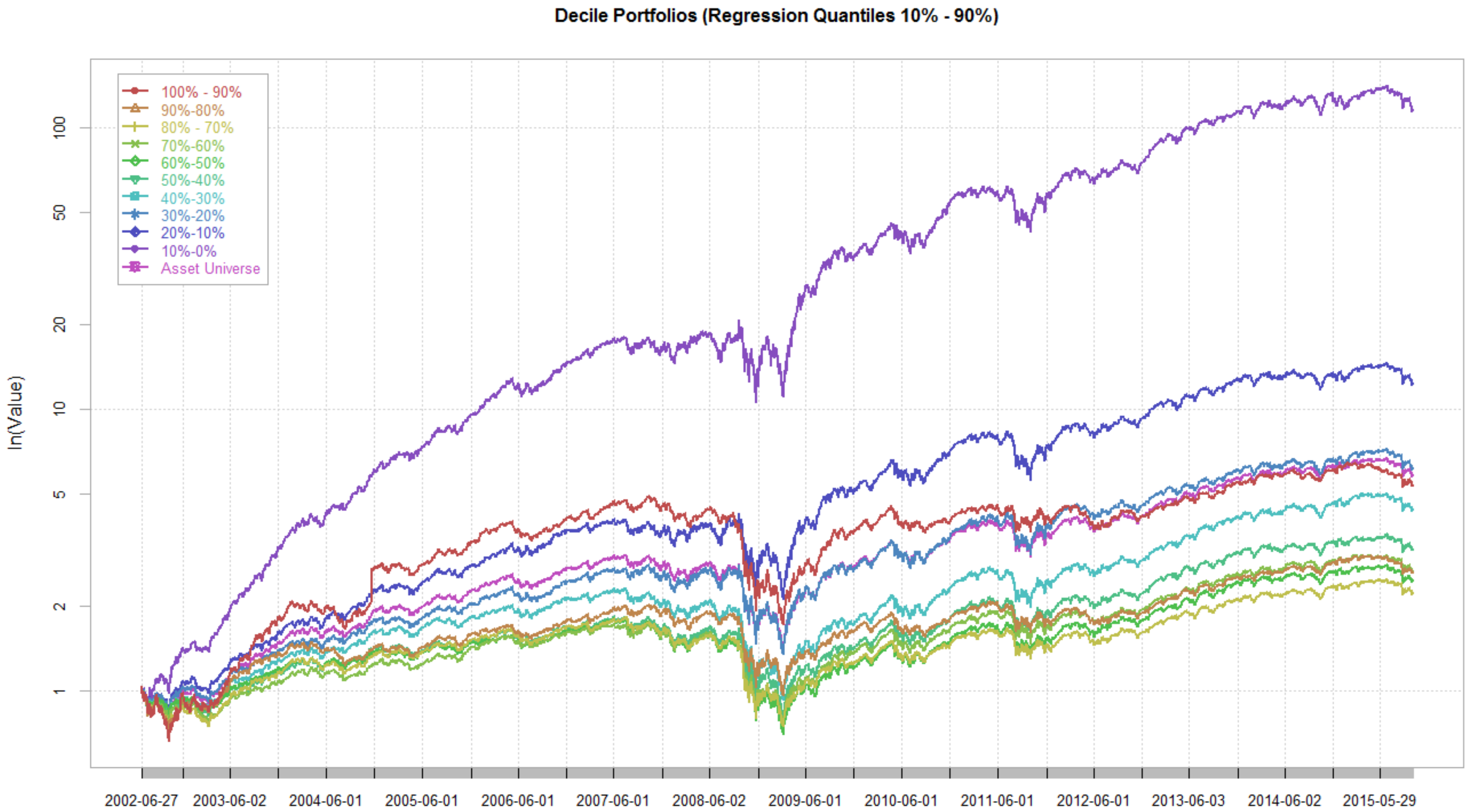
6.4 Exhibit 4

Charted performance for decile portfolios formed using slope coefficients estimated at the quantile levels 5% - 95%. The returns are calculated geometrically assuming re-investment of profits throughout the investment period. The Y axis is expressed in log scale for a more appealing visualization of the chart. The top performing portfolios are the 10% - 0% and the 20% - 10%. The outperformance of the bottom deciles is striking. If $1 was invested in the 10% -0% portfolio in 2002 when the investment simulation begins, and compounded continuously, that investment would be worth just under $120 at the end of the study. I assumes 0 transaction costs. The second best portfolio is also the second last decile producing $12.6 over the investment period.

Decile Portfolios (Regression Quantiles 5% - 95%)



Charted performance for decile portfolios formed using slope coefficients estimated at the quantile levels 10% - 90%. The returns are calculated geometrically assuming re-investment of profits throughout the investment period. The Y axis is expressed in log scale for a more appealing visualization of the chart. The top performing portfolios are the 10% - 0% and the 20% - 10%. The outperformance of the bottom deciles is striking. If $1 was invested in the 10% -0% portfolio in 2002 when the investment simulation begins, and compounded continuously, that investment would be worth just under $120 at the end of the study. I assumes 0 transaction costs. The second best portfolio is also the second last decile producing $12.6 over the investment period.



6.5 Exhibit 5

The portfolio returns over the entire analysis period of each bucket are decomposed using OLS regressions as shown in Eq. 2. Heat maps are applied to the table to better visualize the effect of ranking stocks on the DEA efficiency score e_i. Stocks with the highest coefficients are included on the second day post measurement in bucket 10% - 0%, conversely the lowest are included in the top decile. The final portfolios are shown to reflect this arrangement of factors: the risk factor coefficients are decreasing the deciles. Each factor is succeeded by a column representing the p-values of their T statistic. Most factors are significant at the 1% level. According the DEA analysis applied here, the CMA – Profitability factor does not look very robust, being insignificant across all ten portfolios. Alphas are shown to be decreasing in the deciles however the top two deciles exhibit positive Alphas expressed in percentage per day.

OLS Regressions Heat - Map [Quantile Levels 5% - 95%]

<table>
<tr><th>Decile</th><th>Alpha</th><th>Pr(>|t|)</th><th>RMRF</th><th>Pr(>|t|)</th><th>SMB</th><th>Pr(>|t|)</th><th>HML</th><th>Pr(>|t|)</th><th>RMW</th><th>Pr(>|t|)</th><th>CMA</th><th>Pr(>|t|)</th><th>R - Squared</th></tr>
<tr><td>100% - 90%</td><td>0.034%</td><td>0.0152</td><td>1.040</td><td>0.0000</td><td>0.200</td><td>0.0000</td><td>-0.161</td><td>0.0000</td><td>-0.629</td><td>0.0000</td><td>-0.067</td><td>0.1749</td><td>76.14%</td></tr>
<tr><td>90% - 80%</td><td>-0.001%</td><td>0.8052</td><td>0.955</td><td>0.0000</td><td>0.274</td><td>0.0000</td><td>-0.038</td><td>0.0002</td><td>-0.117</td><td>0.0000</td><td>-0.004</td><td>0.8595</td><td>93.34%</td></tr>
<tr><td>80% - 70%</td><td>-0.005%</td><td>0.3165</td><td>0.942</td><td>0.0000</td><td>0.373</td><td>0.0000</td><td>0.027</td><td>0.0007</td><td>0.034</td><td>0.0085</td><td>0.008</td><td>0.6149</td><td>95.64%</td></tr>
<tr><td>70% - 60%</td><td>-0.001%</td><td>0.8327</td><td>0.958</td><td>0.0000</td><td>0.450</td><td>0.0000</td><td>0.078</td><td>0.0000</td><td>0.085</td><td>0.0000</td><td>-0.016</td><td>0.2806</td><td>96.57%</td></tr>
<tr><td>60% - 50%</td><td>0.000%</td><td>0.9975</td><td>0.995</td><td>0.0000</td><td>0.505</td><td>0.0000</td><td>0.123</td><td>0.0000</td><td>0.158</td><td>0.0000</td><td>0.024</td><td>0.1114</td><td>96.88%</td></tr>
<tr><td>50% - 40%</td><td>0.007%</td><td>0.0754</td><td>1.020</td><td>0.0000</td><td>0.570</td><td>0.0000</td><td>0.183</td><td>0.0000</td><td>0.168</td><td>0.0000</td><td>0.021</td><td>0.1479</td><td>97.23%</td></tr>
<tr><td>40% - 30%</td><td>0.012%</td><td>0.0065</td><td>1.059</td><td>0.0000</td><td>0.656</td><td>0.0000</td><td>0.238</td><td>0.0000</td><td>0.193</td><td>0.0000</td><td>-0.010</td><td>0.5159</td><td>97.00%</td></tr>
<tr><td>30% - 20%</td><td>0.021%</td><td>0.0001</td><td>1.099</td><td>0.0000</td><td>0.743</td><td>0.0000</td><td>0.326</td><td>0.0000</td><td>0.222</td><td>0.0000</td><td>0.003</td><td>0.8633</td><td>96.45%</td></tr>
<tr><td>20% - 10%</td><td>0.041%</td><td>0.0000</td><td>1.152</td><td>0.0000</td><td>0.845</td><td>0.0000</td><td>0.422</td><td>0.0000</td><td>0.231</td><td>0.0000</td><td>0.011</td><td>0.6145</td><td>95.73%</td></tr>
<tr><td>10% - 0%</td><td>0.109%</td><td>0.0000</td><td>1.255</td><td>0.0000</td><td>1.011</td><td>0.0000</td><td>0.578</td><td>0.0000</td><td>0.180</td><td>0.0000</td><td>-0.049</td><td>0.1428</td><td>92.43%</td></tr>
</table>

Significance Legend | < 5 % | < 10 % | > 10%

The portfolio returns over the entire analysis period of each bucket are decomposed using OLS regressions as shown in Eq. 2. Heat maps are applied to the table to better visualize the effect of ranking stocks on the DEA efficiency score e_i. Stocks with the highest coefficients are included on the second day post measurement in bucket 10% - 0%, conversely the lowest are included in the top decile. The final portfolios are shown to reflect this arrangement of factors: the risk factor coefficients are decreasing the deciles. Each factor is succeeded by a column representing the p-values of their T statistic. Most factors are significant at the 1% level. According the DEA analysis applied here, the CMA – Profitability factor does not look very robust, being insignificant across all ten portfolios. Alphas are shown to be decreasing in the deciles however the top two deciles exhibit positive Alphas expressed in percentage per day.

OLS Regressions Heat - Map [Quantile Levels 10% - 90%]

Decile	Alpha	Pr(>|t|)	RMRF	Pr(>|t|)	SMB	Pr(>|t|)	HML	Pr(>|t|)	RMW	Pr(>|t|)	CMA	Pr(>|t|)	R - Squared
100% - 90%	0.032%	0.0196	1.048	0.0000	0.195	0.0000	-0.165	0.0000	-0.613	0.0000	-0.059	0.2335	76.37%
90% - 80%	0.004%	0.4737	0.951	0.0000	0.280	0.0000	-0.030	0.0032	-0.107	0.0000	0.019	0.3736	93.24%
80% - 70%	-0.003%	0.5234	0.942	0.0000	0.368	0.0000	0.034	0.0000	0.015	0.2486	0.003	0.8588	95.65%
70% - 60%	0.001%	0.8042	0.959	0.0000	0.452	0.0000	0.076	0.0000	0.086	0.0000	-0.003	0.8351	96.48%
60% - 50%	-0.003%	0.4478	0.994	0.0000	0.510	0.0000	0.130	0.0000	0.166	0.0000	0.013	0.3769	96.83%
50% - 40%	0.004%	0.3525	1.022	0.0000	0.580	0.0000	0.189	0.0000	0.159	0.0000	-0.002	0.8994	97.30%
40% - 30%	0.012%	0.0076	1.059	0.0000	0.648	0.0000	0.228	0.0000	0.197	0.0000	-0.003	0.8317	96.89%
30% - 20%	0.021%	0.0000	1.092	0.0000	0.743	0.0000	0.333	0.0000	0.206	0.0000	-0.004	0.8150	96.55%
20% - 10%	0.041%	0.0000	1.153	0.0000	0.842	0.0000	0.411	0.0000	0.232	0.0000	0.009	0.6860	95.82%
10% - 0%	0.108%	0.0000	1.255	0.0000	1.009	0.0000	0.571	0.0000	0.185	0.0000	-0.051	0.1248	92.45%

Significance Legend	< 5 %	< 10 %	> 10%

6.6 Exhibit 6

Two correlation analysis test are applied between the decile portfolios and also the asset universe in order to show the potential diversification effects of investing in this strategy. The first table represents a non-parametric measure of dependence that is the Kendall rank correlation coefficients. The second displays Pearson's correlation coefficients which is a measure of linear dependence between variables. The latter however is not considered to be a robust measure of association and can be very sensitive to outliers and non-normal distributions. Both take values between 1 and -1 where 1 represents identical total positive correlation and -1 the opposite. In both cases correlations with the Asset Universe are increasing in the deciles and both display similar rankings between the portfolios. Kendall however shows a much larger dispersion between the 11 portfolios analysed implying increasing differences in returns as we approach decile 10% - 0%.

Kendall Rank Correlation Tau Coefficients Heat - Map [5% - 95%]

	100%-90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%	AU
100%-90%	1.0000										
90%-80%	0.7728	1.0000									
80% - 70%	0.7360	0.8304	1.0000								
70%-60%	0.7105	0.8121	0.8463	1.0000							
60%-50%	0.6951	0.7995	0.8420	0.8582	1.0000						
50%-40%	0.6809	0.7830	0.8257	0.8503	0.8640	1.0000					
40%-30%	0.6694	0.7664	0.8103	0.8368	0.8581	0.8708	1.0000				
30%-20%	0.6585	0.7509	0.7945	0.8227	0.8415	0.8594	0.8745	1.0000			
20%-10%	0.6498	0.7358	0.7780	0.8056	0.8243	0.8420	0.8614	0.8708	1.0000		
10%-0%	0.6250	0.6996	0.7353	0.7609	0.7761	0.7934	0.8120	0.8304	0.8505	1.0000	
AU	0.7359	0.8294	0.8654	0.8820	0.8914	0.8953	0.8941	0.8859	0.8737	0.8302	1.0000

Pearson Correlation Coefficients Heat - Map [5% - 95%]

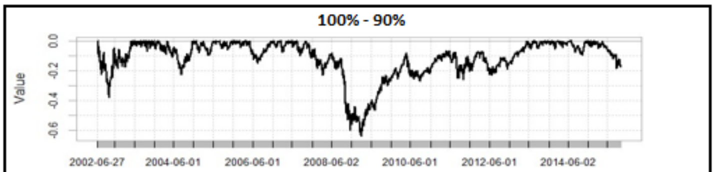
	100%-90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%	AU
100%-90%	1.0000										
90%-80%	0.8866	1.0000									
80% - 70%	0.8650	0.9702	1.0000								
70%-60%	0.8480	0.9611	0.9772	1.0000							
60%-50%	0.8373	0.9522	0.9735	0.9806	1.0000						
50%-40%	0.8269	0.9433	0.9661	0.9776	0.9819	1.0000					
40%-30%	0.8156	0.9322	0.9583	0.9714	0.9796	0.9838	1.0000				
30%-20%	0.8044	0.9194	0.9482	0.9630	0.9719	0.9801	0.9850	1.0000			
20%-10%	0.7910	0.9051	0.9344	0.9511	0.9609	0.9715	0.9789	0.9836	1.0000		
10%-0%	0.7652	0.8735	0.9038	0.9217	0.9331	0.9475	0.9566	0.9666	0.9756	1.0000	
AU	0.8703	0.9615	0.9781	0.9845	0.9872	0.9890	0.9880	0.9849	0.9787	0.9591	1.0000

6.7 Exhibit 7

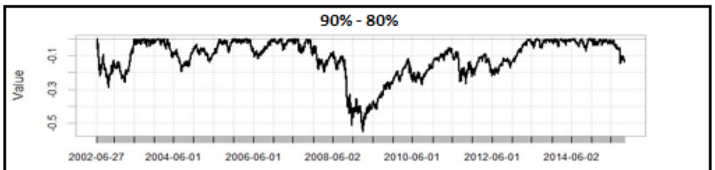
Drawdown analysis is performed on each decile portfolio as well as the Asset Universe. Drawdown is expressed as the maximum loss from a peak to a trough of a portfolio, before a new peak is attained. Each table displays individual portfolio Drawdown performance listing the date when it started, the date it reached the bottom and the date it reached a new high. This is followed by the depth expressed in return percentage points lost, the total length in days, length to the trough and the length of the recovery period. Top five drawdown are displayed for each portfolio and then the values are averaged and displayed in the Drawdown Summary Table. Heat maps are applied in the latter to better visualize the effect of ranking stocks on the efficiency score e_i. Interestingly even though the middle portfolios were the worst performers in terms of risk and return they display a lower average depth of Drawdown. Additionally the top decile has the worst drawdown. In terms of recovery the average length, average drop and average recovery are all decreasing in the deciles.

Drawdown Summary Table

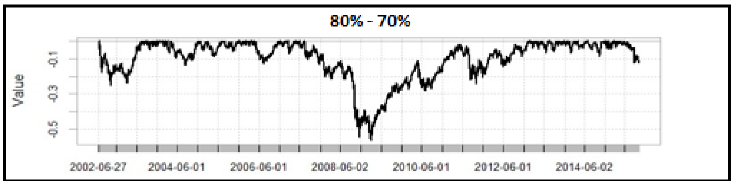
	Average Depth	Average Length	Avg. to trough	Avg. Recovery
100% - 90%	-31.51%	407.4	132	337.75
90%-80%	-25.97%	442	130.4	383
80% - 70%	-23.99%	414.4	141.2	335
70%-60%	-23.08%	398.6	146	309.25
60%-50%	-26.19%	328.6	138	231.75
50%-40%	-26.24%	298.8	144	187
40%-30%	-26.81%	305	165	174.25
30%-20%	-27.67%	188.8	121	84.25
20%-10%	-28.60%	159.8	85	74.8
10%-0%	-28.86%	132	67.4	64.6
Asset Universe	-27.19%	254	137.8	138.75



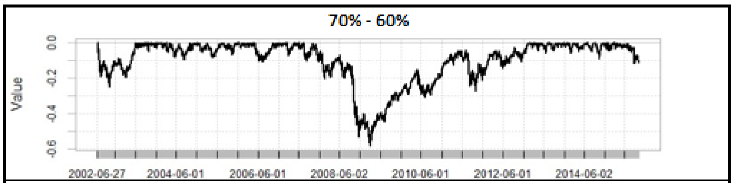
From	Trough	To	Depth	Length	To Trough	Recovery
11-Oct-07	9-Mar-09	21-May-13	-0.6423	1412	354	1058
1-Jul-02	9-Oct-02	6-May-03	-0.3773	214	71	143
6-Apr-04	13-Aug-04	18-Nov-04	-0.2229	158	90	68
16-Apr-15	25-Aug-15	End Of Data	-0.1832	118	92	NA
8-May-06	21-Jul-06	15-Nov-06	-0.15	135	53	82
Average			-31.51%	407.40	132.00	337.75



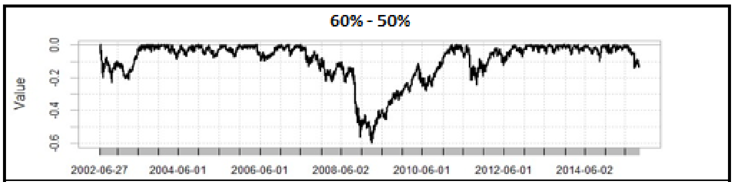
From	Trough	To	Depth	Length	To Trough	Recovery
1-Nov-07	9-Mar-09	25-Apr-13	-0.5556	1379	339	1040
1-Jul-02	9-Oct-02	5-Jun-03	-0.2887	235	71	164
6-Apr-04	12-Aug-04	28-Jul-05	-0.193	331	89	242
16-Apr-15	25-Aug-15	End Of Data	-0.1442	118	92	NA
20-Apr-06	17-Jul-06	15-Nov-06	-0.1168	147	61	86
Average			-25.97%	442.00	130.40	383.00



From	Trough	To	Depth	Length	To Trough	Recovery
20-Jul-07	9-Mar-09	15-Jan-13	-0.5653	1383	412	971
1-Jul-02	9-Oct-02	23-Jul-03	-0.2486	268	71	197
6-Apr-04	13-Aug-04	14-Dec-04	-0.1374	175	90	85
21-Apr-06	21-Jul-06	22-Nov-06	-0.1245	151	64	87
19-May-15	25-Aug-15	End Of Data	-0.1237	95	69	NA
Average			-23.99%	414.40	141.20	335.00

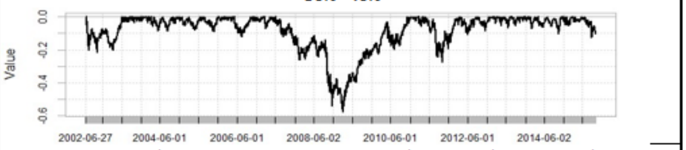


From	Trough	To	Depth	Length	To Trough	Recovery
20-Jul-07	9-Mar-09	17-Jan-13	-0.5858	1385	412	973
1-Jul-02	9-Oct-02	4-Jun-03	-0.2489	234	71	163
23-Mar-15	25-Aug-15	End Of Data	-0.1185	135	109	NA
9-May-06	17-Jul-06	16-Oct-06	-0.108	112	48	64
6-Apr-04	13-Aug-04	6-Oct-04	-0.093	127	90	37
Average			-23.08%	398.60	146.00	309.25



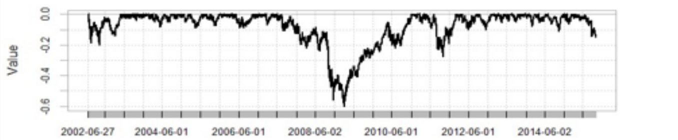
From	Trough	To	Depth	Length	To Trough	Recovery
5-Jun-07	9-Mar-09	11-Feb-11	-0.6018	932	444	488
8-Jul-11	3-Oct-11	6-Sep-12	-0.2406	295	61	234
1-Jul-02	9-Oct-02	7-Jul-03	-0.2272	256	71	185
24-Jun-15	25-Aug-15	End Of Data	-0.1398	70	44	NA
7-Jul-14	13-Oct-14	10-Nov-14	-0.1	90	70	20
Average			-26.19%	328.60	138.00	231.75

50% - 40%



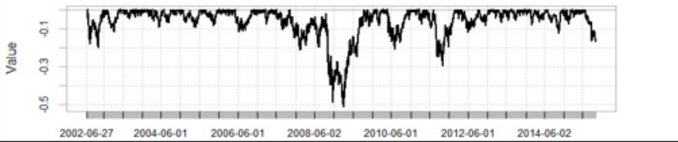
From	Trough	To	Depth	Length	To Trough	Recovery
5-Jun-07	9-Mar-09	5-Nov-10	-0.579	865	444	421
2-May-11	3-Oct-11	13-Mar-12	-0.2729	219	108	111
1-Jul-02	9-Oct-02	4-Jun-03	-0.2167	234	71	163
24-Jun-15	25-Aug-15	End Of Data	-0.124	70	44	NA
8-May-06	21-Jul-06	5-Oct-06	-0.1195	106	53	53
Average			-26.24%	298.80	144.00	187.00

40% - 30%



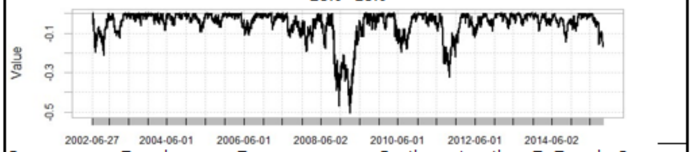
From	Trough	To	Depth	Length	To Trough	Recovery
5-Jun-07	9-Mar-09	3-Dec-10	-0.6038	884	444	440
2-May-11	3-Oct-11	3-Feb-12	-0.2744	193	108	85
1-Jul-02	9-Oct-02	2-May-03	-0.2011	212	71	141
23-Mar-15	28-Sep-15	End Of Data	-0.1494	135	132	NA
7-Jul-14	13-Oct-14	25-Nov-14	-0.1117	101	70	31
Average			-26.81%	305.00	165.00	174.25

30% - 20%



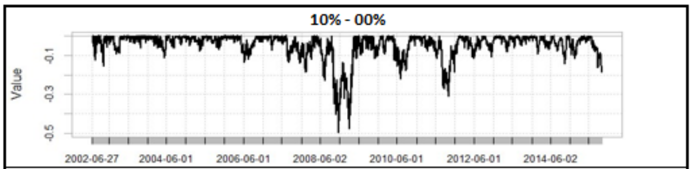
From	Trough	To	Depth	Length	To Trough	Recovery
10-Oct-07	9-Mar-09	16-Sep-09	-0.5137	488	355	133
8-Jul-11	3-Oct-11	1-Feb-12	-0.2969	144	61	83
26-Apr-10	6-Jul-10	4-Nov-10	-0.2067	136	50	86
1-Jul-02	9-Oct-02	27-Nov-02	-0.1973	106	71	35
24-Jun-15	29-Sep-15	End Of Data	-0.1687	70	68	NA
Average			-27.67%	188.80	121.00	84.25

20% - 10%



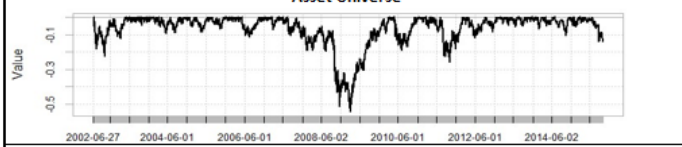
From	Trough	To	Depth	Length	To Trough	Recovery
22-Sep-08	9-Mar-09	22-Jul-09	-0.5066	210	116	94
8-Jul-11	3-Oct-11	1-Feb-12	-0.3227	144	61	83
1-Jul-02	9-Oct-02	22-Nov-02	-0.213	103	71	32
30-Apr-10	6-Jul-10	2-Nov-10	-0.1958	130	46	84
16-Jul-07	18-Jan-08	15-May-08	-0.1919	212	131	81
Average			-28.60%	159.80	85.00	74.80

10% - 00%



From	Trough	To	Depth	Length	To Trough	Recovery
22-Sep-08	20-Nov-08	21-Apr-09	-0.4995	146	44	102
8-Jul-11	3-Oct-11	10-Jan-12	-0.3096	129	61	68
2-May-08	15-Jul-08	18-Sep-08	-0.2276	97	51	46
26-Apr-10	6-Jul-10	5-Oct-10	-0.2217	114	50	64
16-Jul-07	18-Jan-08	24-Mar-08	-0.1847	174	131	43
Average			-28.86%	132.00	67.40	64.60

Asset Universe



From	Trough	To	Depth	Length	To Trough	Recovery
16-Jul-07	9-Mar-09	5-Mar-10	-0.5466	666	416	250
2-May-11	3-Oct-11	3-Feb-12	-0.2601	193	108	85
1-Jul-02	9-Oct-02	22-Apr-03	-0.224	204	71	133
26-Apr-10	6-Jul-10	5-Nov-10	-0.1891	137	50	87
24-Jun-15	25-Aug-15	End Of Data	-0.1399	70	44	NA
Average			-27.19%	254.00	137.80	138.75

6.8 Exhibit 8

The table below displays robustness test results, that is testing wither the DEA efficiency score rankings are robust to changes in the quantiles of the inputs provided. Two sets of inputs were estimated and calculated using Eq. 11 – 15. The table displays t-test for differences between means between each portfolio pair. The null hypothesis is that the difference between means is zero. This is not rejected in any of the portfolio pairs implying that they are almost one and the same.

	100% - 90%	90%-80%	80% - 70%	70%-60%	60%-50%	50%-40%	40%-30%	30%-20%	20%-10%	10%-0%
Degrees of Freedom	6676	6675.9	6676	6676	6676	6675.8	6676	6676	6676	6676
T - Statistic	0.0260	-0.1791	-0.0467	-0.0634	0.0941	0.0989	0.0022	-0.0047	0.0171	0.0197
P - Value	0.9793	0.8578	0.9627	0.9494	0.9251	0.9212	0.9983	0.9963	0.9863	0.9843
Confidence Interval	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%
Upper Bound	0.0008	0.0006	0.0006	0.0006	0.0007	0.0007	0.0007	0.0008	0.0008	0.0010
Lower Bound	-0.0008	-0.0007	-0.0006	-0.0006	-0.0006	-0.0006	-0.0007	-0.0008	-0.0008	-0.0009
Mean x	0.0006	0.0003	0.0003	0.0004	0.0004	0.0005	0.0006	0.0007	0.0009	0.0016
Mean y	0.0006	0.0004	0.0003	0.0004	0.0004	0.0004	0.0006	0.0007	0.0009	0.0016
Reject Null?	No	No	No	No	No	No	No	No	No	No

APPENDIX

There are several software tools designed specifically for DEA analysis however they lack in terms of the flexibility of the application. This paper performs a DEA analysis across all 1150 stocks in our sample for each time period meaning that over 3000 calculations are performed on each of the 120 portfolios presented in this paper. Due to the large number of repetitive analyses required this paper uses the programing language R, a free open source software which offers a wide variety of functions and graphical capabilities. The code below is the source code for R which performs the CCR DEA analysis. It was published as part of the 11th international Conference on Data Envelopment Analysis, Samsun, Turkey 2013. (Pessanha et. al. 2013)

```
namesDMU <- data[1]
inputs <- data[2]
outputs <-  data[c(3,4,5)]
N <- dim(data)[1] # number of DMU
s <- dim(inputs)[2] # number of inputs
m <- dim(outputs)[2] # number of outputs
```

```
f.rhs <- c(rep(0,N),1)  # RHS constraints
f.dir <- c(rep("<=",N),"=") # directions of the constraints
aux <- cbind(-1*inputs,outputs) # matrix of constraint coefficients in (6)
for (i in 1:N) {
    f.obj <- c(0*rep(1,s),outputs[i,]) # objective function coefficients
    f.con <- rbind(aux ,c(inputs[i,], rep(0,m))) # add LHS of b^Tz=1
    results <-lp("max",f.obj,f.con,f.dir,f.rhs,scale=1,compute.sens=TRUE) # solve LPP
    multipliers <- results$solution # input and output weights
    efficiency <- results$objval # efficiency score
    duals <- results$duals # shadow prices
    if (i==1) {
       weights <- multipliers
       effcrs <- efficiency
       lambdas <- duals [seq(1,N)]
    } else {
       weights <- rbind(weights,multipliers)
       effcrs <- rbind(effcrs , efficiency)
       lambdas <- rbind(lambdas,duals[seq(1,N)])
    }
}
```

7 REFERENCES

Asness, C. S., Frazzini, A., & Pedersen, L. H. (2014). Quality minus junk. Available at SSRN 2312432.

Barnes, M. L., & Hughes, A. T. W. (2002). A quantile regression analysis of the cross section of stock market returns.

Bhat, R., B. B. Verma, and E. Reuben (2001) “A note on data envelopment analysis [DEA]” Journal of Health Management 3, 309-328.

Black F, Jensen M, Scholes M (1972): The Capital Asset Pricing Model: Some Empirical Tests. In: Jensen M (Ed.): Studies in the Theory of Capital Markets. New York, Praeger Publishers

Buchinsky, Moshe. "Recent advances in quantile regression models: a practical guideline for empirical research." Journal of human resources(1998): 88-126.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429-444.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). A Data Envelopment Analysis Approach to Evaluation of the Program Follow through Experiment in US Public School Education (No. MSRR-432).

Cooper, W. W., Seiford, L. M. and Tone, K. (2007), Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA – Solver Software, 2nd Edition, Springer, New York.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. the Journal of Finance, 47(2), 427-465.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of financial economics, 33(1), 3-56.

Fama, E. F., & French, K. R. (2015). Incremental variables and the investment opportunity set. Journal of Financial Economics, 117(3), 470-488.

Fin, D. E. A. F., Gerrans, P., Singh, A. K., & Powell, R. (2009). Quantile regression: its application in investment analysis. Jassa, (4), 7.

Gardijan, M., & Kojić, V. (2012). DEA-based investment strategy and its application in the Croatian stock market. Croatian Operational Research Review, 3(1), 203-212.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. Econometrica: journal of the Econometric Society, 33-50.

Lintner J (1965): The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. Review of Economics & Statistics, 47(1):13–37

Lopes, A., Lanzer, E., Lima, M. and Newton da Costa, Jr. (2008), "DEA investment strategy in the Brazilian stock market" Economics Bulletin, Vol. 13, No. 2, pp. 1-10.

Mossin J (1966): Equilibrium in a Capital Asset Market. Econometrica, 34(4):768–783.

Pätäri, E. J., Leivo, T. H., & Samuli Honkapuro, J. V. (2010). Enhancement of value portfolio performance using data envelopment analysis. *Studies in Economics and Finance*, *27*(3), 223-246.

Pessanha, J. F. M., Marinho, A., LAURENCEL, L. D. C., & do Amaral, M. R. D. S. (2014). Implementing DEA models in the R program. *Data Envelopment Analysis and Performance Measurement*, *1*, 233.

Powers, J., & McMullen, P. (2000). Using data envelopment analysis to select efficient large market cap securities. Journal of Business and Management, 7(2), 31-42.

Sharpe WF (1964): Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. Journal of Finance, 19(3):425–442.

Sharpe, W. F. (1966). "Mutual Fund Performance". Journal of Business 39 (S1): 119–138

Sheridan Titman, K. C. John Wei, Feixue Xie Source: The Journal of Financial and Quantitative Analysis, Vol. 39, No. 4 (Dec., 2004), pp. 677-700

Singh, A. K., Allen, D. E., & Allen, D. E. (2010). Asset Selection Using Factor Model and Data Envelope Analysis-a Quantle Regression Approach. School of Accounting, Economics and Finance, Edith Cowan University.